

Original Paper

# A Novel Tool for Real-time Estimation of Epidemiological Parameters of Communicable Diseases Using Contact-Tracing Data: Development and Deployment

Bernard C Silenou<sup>1,2</sup>, MSc; Carolin Verset<sup>3</sup>, MSc; Basil B Kaburi<sup>1,2</sup>, MPhil, MD; Olivier Leuci<sup>3</sup>, MSc; Stéphane Ghozzi<sup>1</sup>, PhD; Cédric Duboudin<sup>3</sup>, MSc; Gérard Krause<sup>1,2,4</sup>, MD, Prof Dr

<sup>1</sup>Department of Epidemiology, Helmholtz Centre for Infection Research, Braunschweig, Germany

<sup>2</sup>Hannover Medical School, Hannover, Germany

<sup>3</sup>Agence Régionale de Santé de Bourgogne Franche-Comté, Dijon, France

<sup>4</sup>German Center for Infection Research, Braunschweig, Germany

**Corresponding Author:**

Gérard Krause, MD, Prof Dr

Department of Epidemiology

Helmholtz Centre for Infection Research

Inhoffenstraße 7

Braunschweig, 38124

Germany

Phone: 49 5316181 3100

Email: [Gerard.Krause@helmholtz-hzi.de](mailto:Gerard.Krause@helmholtz-hzi.de)

## Abstract

**Background:** The Surveillance Outbreak Response Management and Analysis System (SORMAS) contains a management module to support countries in their epidemic response. It consists of the documentation, linkage, and follow-up of cases, contacts, and events. To allow SORMAS users to visualize data, compute essential surveillance indicators, and estimate epidemiological parameters from such network data in real-time, we developed the SORMAS Statistics (SORMAS-Stats) application.

**Objective:** This study aims to describe the essential visualizations, surveillance indicators, and epidemiological parameters implemented in the SORMAS-Stats application and illustrate the application of SORMAS-Stats in response to the COVID-19 outbreak.

**Methods:** Based on findings from a rapid review and SORMAS user requests, we included the following visualization and estimation of parameters in SORMAS-Stats: transmission network diagram, serial interval (SI), time-varying reproduction number  $R(t)$ , dispersion parameter  $k$ , and additional surveillance indicators presented in graphs and tables. We estimated SI by fitting lognormal, gamma, and Weibull distributions to the observed distribution of the number of days between symptom onset dates of infector-infectee pairs. We estimated  $k$  by fitting a negative binomial distribution to the observed number of infectees per infector. Furthermore, we applied the Markov Chain Monte Carlo approach and estimated  $R(t)$  using the incidence data and the observed SI computed from the transmission network data.

**Results:** Using COVID-19 contact-tracing data of confirmed cases reported between July 31 and October 29, 2021, in the Bourgogne-Franche-Comté region of France, we constructed a network diagram containing 63,570 nodes. The network comprises 1.75% (1115/63,570) events, 19.59% (12,452/63,570) case persons, and 78.66% (50,003/63,570) exposed persons, including 1238 infector-infectee pairs and 3860 transmission chains with 24.69% (953/3860) having events as the index infector. The distribution with the best fit to the observed SI data was a lognormal distribution with a mean of 4.30 (95% CI 4.09-4.51) days. We estimated a dispersion parameter  $k$  of 21.11 (95% CI 7.57-34.66) and an effective reproduction number  $R$  of 0.9 (95% CI 0.58-0.60). The weekly estimated  $R(t)$  values ranged from 0.80 to 1.61.

**Conclusions:** We provide an application for real-time estimation of epidemiological parameters, which is essential for informing outbreak response strategies. The estimates are commensurate with findings from previous studies. The SORMAS-Stats application could greatly assist public health authorities in the regions using SORMAS or similar tools by providing extensive visualizations and computation of surveillance indicators.

**KEYWORDS**

COVID-19; disease outbreak; contact tracing; serial interval; basic reproduction number; infectious disease incubation period; superspreading events; telemedicine; public health; epidemiology; surveillance tool; outbreak response; pandemic; digital health application; response strategy

## Introduction

### Background

During the course of 2020, there was a substantial increase in the number and use of eHealth applications, mainly in response to the COVID-19 outbreak [1-3]. These applications are being used in different areas of digital health intervention, such as disease surveillance, vaccine delivery, hospital management, laboratory management, symptom journals, and education [4].

The Surveillance Outbreak Response Management and Analysis System (SORMAS) is an open-source digital tool that supports disease control and outbreak management procedures [5-8]. The objective of SORMAS is to ensure the availability of real-time surveillance data for priority diseases at all administrative levels. SORMAS supports task management, complies with data protection and data security standards, and enhances interoperability with other applications.

Essential epidemiological parameters governing COVID-19 transmissions such as serial interval (SI), instantaneous reproduction number  $R(t)$ , effective reproduction number  $R$ , and individual-level variation in transmission are context-specific, and thus often difficult to estimate precisely with publicly available data. Studies have been conducted to estimate these epidemiological parameters during the early phase of the outbreak, but most focused on coarse, aggregated, and publicly reported data sources that likely mask local specificities or are biased toward more severe cases [9-11]. The lack of easy access to outbreak data containing more epidemiological and clinical information has been reported as a limiting factor in improving the performance of epidemiological models [12]. Since then, most public health stakeholders used eHealth applications to document rich and large contact-tracing data in response to the COVID-19 outbreak. Nonetheless, such informative surveillance data are often not hosted on a centralized hub but scattered on different databases in the corresponding countries. In addition, the estimation of epidemiological parameters is frequently not done in real time and does not account for spatial and temporal variation, and thus, fails to provide comprehensive and timely information on outbreak evolution to best inform decision-making [13]. Further, public health stakeholders faced challenges in generating reproducible analyses for their routine situation reports because data are often manually exported from eHealth applications and analyzed with stand-alone statistical applications such as R, SAS, SPSS or STATA.

To address these challenges, we envisaged an approach whereby standardized statistical analysis methods are brought to where the rich and detailed surveillance data are hosted. To assist stakeholders (SORMAS users) with a reactive analytic platform that leverages the rich and detailed data documented in

SORMAS, we developed the SORMAS Statistics (SORMAS-Stats) application. SORMAS-Stats is a user-friendly R Shiny web application to estimate epidemiological parameters, compute country or region-specific indicators, and provide visualizations in real time.

### Objective

This study aims to describe the essential visualizations, epidemiological parameters, and surveillance indicators implemented in the SORMAS-Stats application. We illustrated its application in response to the COVID-19 outbreak using surveillance data captured with SORMAS in the Bourgogne-Franche-Comté region of France.

## Methods

### Design Process of SORMAS-Stats

We gathered the requirements essential to supporting outbreak response through SORMAS user requests (from Nigeria, Ghana, France, and Germany), which thus guided the design of SORMAS-Stats. SORMAS users are public health personnel of a country, such as field investigators and epidemiologists. We identified user requests through GitHub issues created by SORMAS users, the SORMAS user support platform, and the minutes of sprint planning meetings. The sprint planning meetings took place every 3 weeks, matching the software release cycle of SORMAS. In parallel, we conducted a rapid review of epidemiological publications and situation reports. Subsequently, we combined the essential requirements obtained from the review and the users' requests, implemented them, and released a beta version of the SORMAS-Stats application. Further, we conducted a field test of the beta version, got users' feedback, implemented them, and deployed a stable version. The time from requirement gathering to deployment was 12 months (July 2020 to June 2021).

### Overview of SORMAS-Stats

SORMAS-Stats is a web application that can be installed locally and uses advanced visualization and statistical analysis methods to analyze surveillance data in real time. SORMAS-Stats assists public health officials in managing outbreaks and permits the execution of reproducible routine epidemiological analysis. The workflow of SORMAS-Stats consists of the preprocessing phase and the analytics phase.

In the preprocessing phase, SORMAS-Stats imports pseudonymized data from an external database. The default integration of SORMAS-Stats is with the SORMAS PostgreSQL database. Only the records and associated attributes reported within the time interval specified in the SORMAS-Stats configuration file are extracted from the external database. Further data processing steps are the deletion of error records, deduplication, categorization, and the computation of derived

variables. In the analytics phase, SORMAS-Stats analyzes the preprocessed data. We classified the analytics phase into two types: (1) data visualization and the computation of summary statistics and (2) the estimation of essential disease-specific epidemiological parameters through statistical modeling.

### **Epidemiological Data**

The SORMAS-Stats application analyzes the entity-based surveillance data routinely collected by public health workers. Generally, surveillance data include the following entities: case person (a person infected with a disease), contact person (a noninfected person exposed to a case person), and event (any exposure or gathering that poses a threat to human health or may lead to the spread of diseases). The collection of all probable transmission chains (pairs of entities: case person–contact person or event–persons exposed to an event [event participants], that can result in disease transmission) forms the network data of the disease. During contact follow-up, a contact person or event participant may become symptomatic and meet the case definition of the disease in question, thus being converted or reclassified as a case. In such instances, 2 types of infector (ie, index case) and infectee (ie, secondary case) pairs are formed (infector-infectee pairs): first, between the case person and contact person, and second, between the event and the event participant.

The data used to illustrate the application of SORMAS-Stats consisted of confirmed cases of COVID-19 and their contacts documented using SORMAS between July 31 and October 29, 2021, in the Bourgogne-Franche-Comté region of France.

### **Estimation of Epidemiological Parameters and Surveillance Indicators**

#### ***Serial Interval***

We computed the observed SI as the difference in the number of days between symptom onsets of infector-infectee pairs. We excluded infector-infectee pairs for which the infector was an event or if one of a pair had missing data for symptom onset date. However, we only included symptomatic transmissions with available data for the date of symptom onset for both infector and infectee in a pair, as transmission data are often generated during contact-tracing under symptomatic settings [14]. We estimated the SI distribution by fitting lognormal, gamma, Weibull, and normal distributions to the observed SI data. The choice of these distributions was based on previous studies [15,16]. For all 4 types of distributions, we excluded observed SI greater than 30 days. For lognormal, gamma, and Weibull distributions, which do not take negative values, we dropped negative values before fitting the distributions. For each fitted distribution, we computed the goodness-of-fit criteria

(Akaike information criterion and Bayesian information criterion) and goodness-of-fit statistics (Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling) [17]. We chose the distribution with the best fit by several approaches: the smallest Akaike information criterion, a plot of the density function of the fitted distribution with the histogram of observed data, and a plot of the empirical and theoretical cumulative distribution functions of each fitted model. We calculated the 95% CI of the mean SI using the formula  $\mu \pm 1.96 \times (\sigma / \sqrt{n})$ , where  $\mu$  is the estimated sample mean,  $\sigma$  the estimated sample SD, and  $n$  the sample size. The analysis was performed using the R statistical software (R Foundation for Statistical Computing) package `fitdistrplus` [17].

#### ***Instantaneous Reproduction Number***

We estimated  $R(t)$ —on a weekly basis—using the approach proposed by Cori et al [18,19]. This method mainly requires the incidence and contact-tracing data, which are the types of data captured by SORMAS. We implemented 2 approaches to specify the SI distribution used to estimate  $R(t)$ . First, a parametric distribution for SI with values for the mean and SD, and second, a parametric distribution for SI and the observed data of infector-infectee pairs. For the second approach, we estimated the SI distribution by applying the Markov Chain Monte Carlo method. The possible choices of the parametric distribution for SI were gamma, Weibull, and lognormal. We computed the summary statistics of the posterior mean and plotted the posterior mean and 95% credible interval for  $R(t)$ . The analysis was performed using the R statistical software package `EpiEstim` [19].

#### ***Variation in Transmission Heterogeneity and Effective Reproduction Number***

Using the data for infector-infectee pairs of cases, we computed the observed offspring distribution as the number of infectees per infector. We applied the approach described by Lloyd-Smith et al [20] and fitted a negative binomial distribution to the observed offspring distribution. We estimated the effective reproduction number  $R$  and the variation in transmission heterogeneity as the mean and dispersion parameter  $k$  of the negative binomial distribution, respectively [20]. In addition, we computed the median and 95% percentile CI of both parameters using bootstrap. We performed the analysis using the R statistical software package `fitdistrplus` [17].

#### ***Visualizations and Surveillance Indicators***

We computed 6 surveillance indicators that may be informative in managing disease outbreaks using transmission network data. [Table 1](#) presents the definition of visualizations and surveillance indicators implemented in the SORMAS-Stats application.

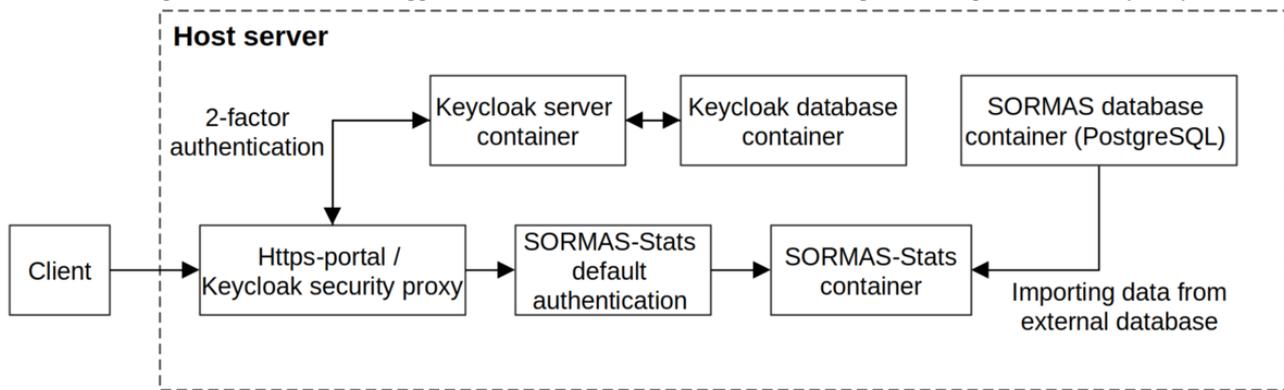
**Table 1.** Description and application of epidemiological parameters, surveillance indicators, and visualizations in SORMAS-Stats application.

Name of output or indicator	Description	Applications in disease surveillance
<b>Epidemiological parameters</b>		
Serial interval (SI)	The difference in the number of days between symptom onsets of infector-infectee pairs (see Methods).	To distinguish disease variants, design follow-up and quarantine duration, and determine time window for effective intervention strategies.
Instantaneous reproduction number $R(t)$	The average number of infectees per infector at a particular time $t$ (see Methods).	To assess the impact of intervention measures. $R(t) > 1$ signify an increase in infectiousness at time $t$ , $R(t) < 1$ signify a decrease in infectiousness [18].
Effective reproduction number ( $R$ )	The average number of infectees per infector (see Methods).	Similar to $R(t)$ .
Dispersion parameter ( $k$ )	A measure of how the number of infectees per infector (offspring distribution) is distributed around the mean value (see Methods).	To assess the evidence of superspreading events or formation of clusters. This can help to devise relevant control measures. Smaller values of $k$ indicate higher levels of dispersion, thus suggesting evidence of superspreading.
<b>Surveillance indicators</b>		
Proportion of exposed persons who became cases	The proportion of exposed persons—among all exposed persons—that converted to cases by exposure types.	To devise relevant control measures similar to the dispersion parameter $k$ . To assess the quality of contact tracing and better allocate resources.
Proportion of index infectors	The proportion of index infectors among all infector nodes (infector person, infectee person, or event).	To determine the quality of contact-tracing. A smaller proportion indicates a greater coverage of identified link-ages between infectors and infectees.
Variance-to-mean ratio (VMR)	The variance divided by the mean of the observed offspring distribution [21].	Similar to $k$ . $VMR > 1$ indicates higher levels of dispersion and thus signaling evidence of superspreading.
Edge density	The ratio of the number of edges (links between 2 nodes) and the maximum number of possible edges [22]. It represents how connected the nodes of the network diagram are to each other.	To assess the impact of control measures on overall societal behavior. Higher values may signify higher social interactions.
Number of individual transmission chains	The total number of transmission chains or infector nodes in the network diagram.	Similar to proportion of index infectors.
<b>Visualizations</b>		
Network diagram	A directed graph of all disease transmission chains consisting of the following types of nodes: case persons, contact persons, events, and event participants.	To prioritize investigation and follow-up of events with known confirmed cases.
Time series plots	A bar graph or line plot of entity counts over time (day, week, or month).	To gauge the efficacy of control measures in place and the need to implement new ones.
Tables	Tables of entity count, proportions, or incidence proportions by administrative area (regions, districts, community).	To target intervention measures to specific areas of the country such as hotspots.
Charts	Pie charts and bar graphs of entity counts or proportions by entity attributes (eg, age and sex).	To protect vulnerable groups.
Maps	Spatial-temporal display of entity counts, proportion, and incidence proportion on a map by administrative area (regions, districts, community).	Similar to tables.

## Architecture of SORMAS and SORMAS-Stats

The SORMAS application was developed on the VAADIN framework, JAVA EE, Payara server, and PostgreSQL database. SORMAS consists of 2 components: the mobile app and the web application. The mobile app communicates with the server via a REST-API and the VAADIN web client application. The SORMAS-Stats application analyzes the surveillance data documented in the PostgreSQL database of SORMAS. We developed SORMAS-Stats based on the R Shiny framework [23]. To secure the server hosting the application, we used an architecture with the following configuration: https-portal or

secure proxy, 2-factor authentication with Keycloak, and the default authentication of SORMAS-Stats (Figure 1). The default authentication of SORMAS-Stats used the shinyauthr R package authentication module to hash the user password [24,25]. SORMAS-Stats can be executed as a default R Shiny application or a Docker application. We deployed SORMAS, SORMAS PostgreSQL, and SORMAS-Stats applications as separate Docker containers and managed them with one Docker-compose file. The test version of SORMAS-Stats based on demonstration data is available online [26]. The code and description of deployment are hosted on GitHub [27].

**Figure 1.** Server setup for the SORMAS-Stats application. SORMAS: Surveillance Outbreak Response Management and Analysis System.

## Ethical Considerations

The Agence Régionale de Santé de Bourgogne Franche-Comté, France, as an administrative public health institution, granted permission for using the anonymized COVID-19 outbreak data in this study, under article 11 of Law No. 2020-546 of May 11, 2020. Under this law, the secondary data used in this study did not require individual informed consent.

## Results

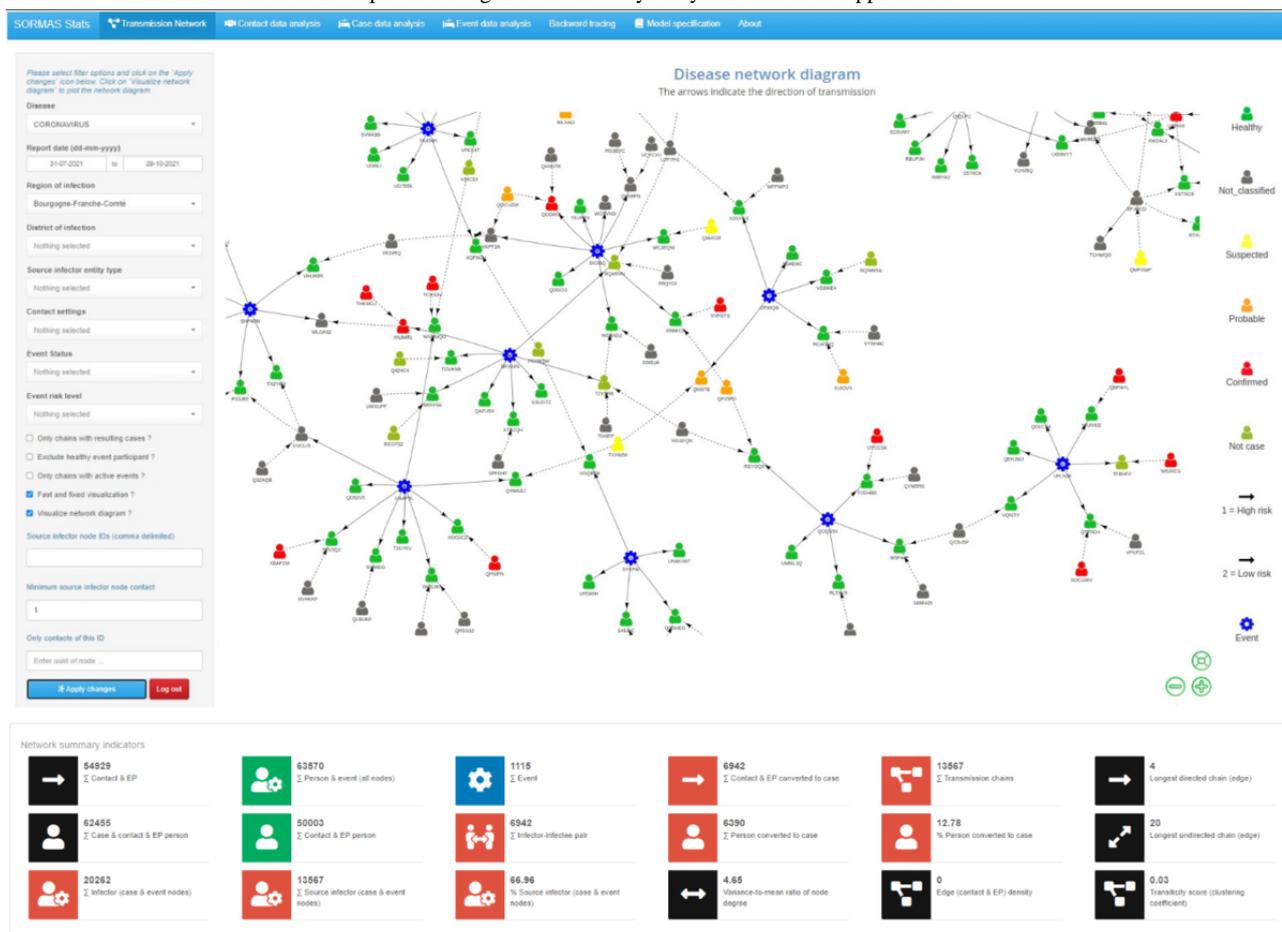
### Contents of the SORMAS-Stats User Interface

The SORMAS-Stats application has multiple dashboards with filters, permitting users to execute analysis and download the output. [Table 1](#) depicts the contents of the SORMAS-Stats user interface and their applications in disease surveillance or outbreak management. The contents of the interface are visualizations, epidemiological parameters, and surveillance indicators.

### Description of the Epidemiological Data Used to Illustrate SORMAS-Stats

We used SORMAS-Stats to estimate epidemiological parameters and surveillance indicators by analyzing the contact-tracing data documented in SORMAS between July 31 and October 29, 2021, in the Bourgogne-Franche-Comté region of France. [Figure 2](#) presents a network diagram constructed from the contact-tracing data consisting of 63,570 unique nodes, comprising 1.75% ( $n=1115$ ) events, 19.59% ( $n=12,452$ ) case persons, and 78.66% ( $n=50,003$ ) exposed persons. Of the 50,003 exposed persons, 6390 (12.78%) subsequently converted to cases. The network diagram consisted of 3860 transmission chains, with each chain comprising a minimum of 1 exposed person and a source infector node (case person or event). The length of the longest directed chain was 4 generations of infection. The average number of exposed persons per node (node degree) was 1.73 (IQR 1-228), whereas the variance-to-mean ratio was 4.65.

**Figure 2.** Screenshot of SORMAS-Stats showing the COVID-19 transmission network diagram and surveillance indicators for 63,570 entities reported between July 31 and October 29, 2021, in the Bourgogne-Franche-Comté region of France. The diagram comprises 1115 events (blue gear node), 12,452 case persons (nongreen person node), 50,003 exposed persons (green person node), and 54,929 exposures (directed arrow from infector node). SORMAS-Stats: Surveillance Outbreak Response Management and Analysis System Statistical application.



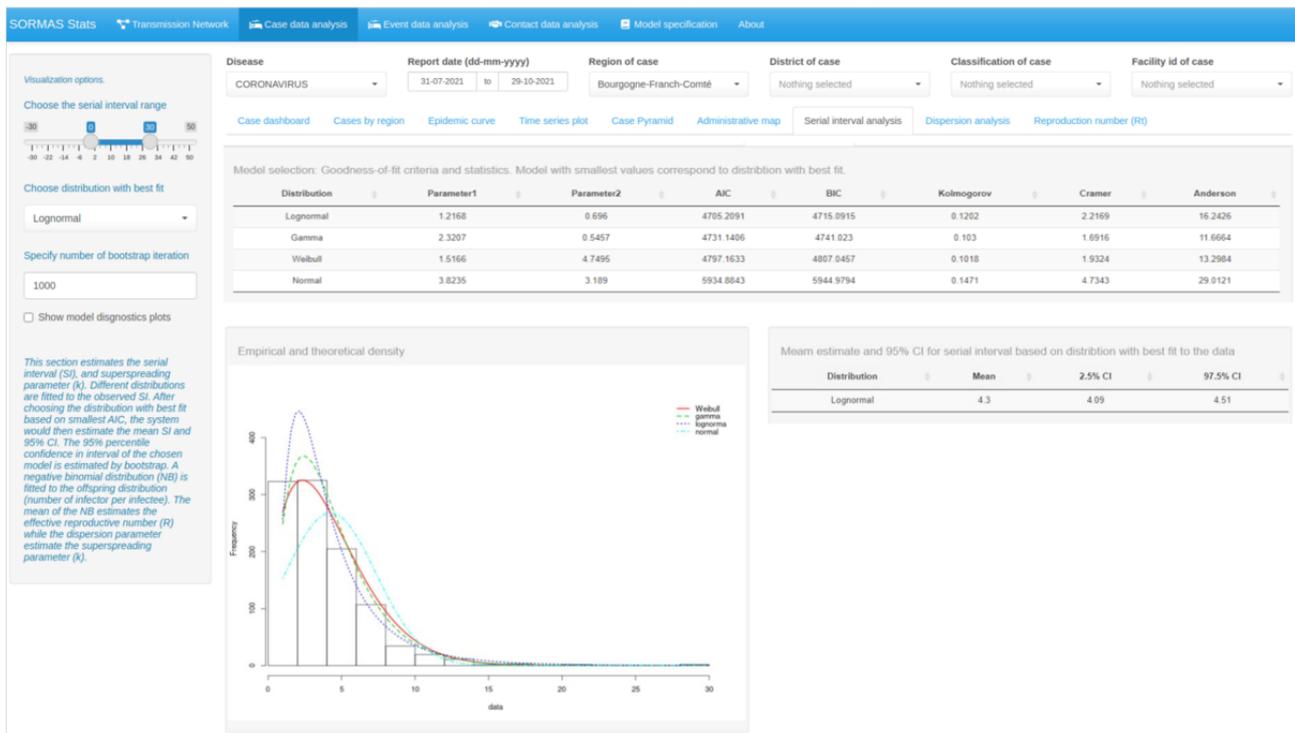
## Epidemiological Parameters and Surveillance Indicators

After subsetting the network diagram by considering chains with resulting cases only, 10,250 unique nodes remained, comprising 9.30% ( $n=953$ ) events and 90.70% ( $n=9297$ ) case persons. Of the 9297 case persons, 68.73% ( $n=6390$ ) were infectees. There were 3860 transmission chains, with 24.69% ( $n=953$ ) having an event as the index infector. The average node degree was 1.36 (IQR 1-36), whereas the variance-to-mean ratio was 1.05. The edge density of the complete or reduced network diagram considering chains with only resulting cases was  $<0.01$ .

Considering infector-infectee pairs consisting of person entities only (event nodes excluded) resulted in 1238 infector-infectee pairs, of which 31.26% ( $n=387$ ) had available data for symptom onset date. Of the 387 pairs with available data for symptom

onset date, 20.41% ( $n=79$ ) were pairs of asymptomatic transmission for which the onset date of the infectee preceded or was on the same date as that of the infector. After excluding negative SIs, the mean of the observed SI was 3.96 (IQR 0-27) days. The distribution with the best fit to the observed SI data was the lognormal distribution with a mean of 4.30 (95% CI 4.09-4.51) days (Figure 3). The mean of the observed offspring distribution was 1.36 (IQR 1-36). By fitting a negative binomial distribution to the observed offspring distribution, we estimated a dispersion parameter  $k$  of 21.11 (95% CI 7.57-34.66) and a reproduction number  $R$  of 0.9 (95% CI 0.58-0.60). Using the observed transmission data with a lognormal distribution for SI, the estimate of the average weekly posterior mean for  $R(t)$  was 0.98 (IQR 0.80-1.61) (Figure 4). The estimated range of  $R(t)$  was congruent with the values obtained when plugging in values for SI mean and SD obtained from the literature (5.19 and 4.23 days, respectively) [9].

**Figure 3.** Screenshot of SORMAS-Stats showing the COVID-19 serial interval distribution for 1238 infector-infectee pairs reported between July 31 and October 29, 2021, in the Bourgogne-Franche-Comté region of France. AIC: Akaike information criterion; BIC: Bayesian information criterion; SORMAS-Stats: Surveillance Outbreak Response Management and Analysis System Statistical application.



**Figure 4.** Screenshot of SORMAS-Stats showing an estimate of COVID-19 time-dependent reproduction number (line) with 95% credible interval (grey band) for 12,452 case persons reported between July 31 and October 29, 2021, in the Bourgogne-Franche-Comté region of France. SORMAS-Stats: Surveillance Outbreak Response Management and Analysis System Statistical application.



## Discussion

### Principal Findings

We developed and deployed SORMAS-Stats, an open-source web application for real-time visualization and the estimation of epidemiological parameters using contact-tracing data

captured with the SORMAS eHealth application. SORMAS-Stats is easy to deploy and requires no programming skills to perform analyses. Some epidemiological parameters included in SORMAS-Stats are SI, time-varying reproduction number  $R(t)$ , effective reproduction number  $R$ , and dispersion parameter  $k$ . We illustrated the use of SORMAS-Stats by

analyzing the contact-tracing data for COVID-19 captured between July 31 and October 29, 2021, in the Bourgogne-Franche-Comté region of France. The estimated mean SI was 4.30 days; this was commensurate with findings from previous studies [9,11]. The estimate for  $k$  was 21.11, whereas the variance-to-mean ratio of the offspring distribution was  $>1$  (1.05), signifying a probable clustering of infections compatible with the observation of superspreading events. However, the estimate was not of the same magnitude as findings from 2 previous studies that used contact-tracing data from other countries at the early phase of the pandemic and reported  $k$  of 0.58 and 0.43 [28,29]. This difference in estimates could be related to various factors that may have prevented superspreading events, such as (1) intervention measures (such as the closure of bars, schools, and gatherings) that stakeholders may have enforced during the study period or (2) differences in the geographical area. The possibility of estimating  $k$  in real time may assist stakeholders in understanding the current local infection dynamics and thus inform which, if any, control measure would be most appropriate. The weekly estimated  $R(t)$  values ranged from 0.80 to 1.61. The maximum value of 1.61 was at the start of the study period; this value was not well estimated since there were no data for the preceding weeks. However, the subsequent values were well estimated and predominantly fluctuated slightly below and above 1, signifying that the infection dynamics were stable within the study period. The computed density of the transmission network diagram was small ( $<0.01$ ), suggesting low social interactions among the persons in the population, thus resulting in a low transmission rate.

The visualizations included in SORMAS-Stats are maps, charts, tables, time series plots, and network diagrams. Further, SORMAS-Stats contains several filters to explore the network diagram by clusters, transmission chains, and superspreading events. The possibility to filter the network diagram was helpful to stakeholders, not only to assist the exploration of transmission chains but to also detect and correct errors created during the data collection phase. SORMAS-Stats is a stand-alone application, easily deployable owing to Docker technology and does not depend on the principal application or method used for data collection. Thus, users with programming skills in R statistical software only can easily configure and extend SORMAS-Stats to cover other types of statistical analyses.

SORMAS-Stats has a public GitHub repository that permits stakeholders from interested countries to make additional

requests and contributions. In this way, it stays an application developed by and for public health workers.

### Limitations

The current integration of SORMAS-Stats is with the SORMAS PostgreSQL database. However, with minor adjustment, SORMAS-Stats can be integrated with other databases or files that can be read by R statistical software as long as the relationship between the entities (such as case person, contact person, event, and event participant) are referenced across the tables in the database [30]. The range of the length of the directed transmission chain was 1 to 4; the majority of the chains had 2 or fewer generations of infection. This might be due to the short study period since the data for analysis were available for only 3 months. In France, data older than 3 months are deleted from the SORMAS database as demanded by the law. Some transmission chains may have continued to spread after the study period.

### Recommendations for Future Research

Further development of the SORMAS-Stats application can focus on the following features: (1) implement more epidemiological indicators that can inform the management of outbreaks such as statistics on hospitalization, immunization, or symptoms; (2) integrate data from other eHealth applications other than SORMAS; and (3) include outbreak detection, change point detection, or prediction models (eg, compartment models). In addition, as more public health workers use SORMAS-Stats, further research can investigate users' experience, such as desirability, usefulness, and performance, through incorporating the concepts of human-computer interaction [31].

### Conclusions

We have provided an application for the real-time estimation of communicable disease parameters, which are essential for outbreak response. The use of the application requires only basic statistical analysis skills. SORMAS-Stats may greatly assist public health authorities in countries using SORMAS or similar tools by providing extensive visualizations, the computation of surveillance indicators, the estimation of epidemiological parameters, and the facilitation of the generation of routine epidemiological reports. This study also showcases how epidemiologists with skills in R statistical software programming only can build a web application, integrate it with a database of another application, and deploy it in the field for outbreak response.

---

### Acknowledgments

This research was supported by the Helmholtz Association of German Research Centres (grant number SO-094).

---

### Authors' Contributions

GK conceived and designed the study. BCS was responsible for feature specification, implementation, and writing the original draft. SG contributed to supervision, code review, and testing. BBK, CV, OL, and CD contributed to feature specification and testing. All authors contributed to the writing of the final version of the manuscript.

---

### Conflicts of Interest

None declared.

---

## References

1. Ming LC, Untong N, Aliudin NA, Osili N, Kifli N, Tan CS, et al. Mobile health apps on COVID-19 launched in the early days of the pandemic: content analysis and review. *JMIR mHealth uHealth* 2020 Sep 16;8(9):e19796 [FREE Full text] [doi: [10.2196/19796](https://doi.org/10.2196/19796)] [Medline: [32609622](https://pubmed.ncbi.nlm.nih.gov/32609622/)]
2. Silenou BC, Nyirenda JL, Zaghoulou A, Lange B, Doerrbecker J, Schenkel K, et al. Availability and suitability of digital health tools in Africa for pandemic control: scoping review and cluster analysis. *JMIR Public Health Surveill* 2021 Dec 23;7(12):e30106 [FREE Full text] [doi: [10.2196/30106](https://doi.org/10.2196/30106)] [Medline: [34941551](https://pubmed.ncbi.nlm.nih.gov/34941551/)]
3. Almalki M, Giannicchi A. Health apps for combating COVID-19: descriptive review and taxonomy. *JMIR mHealth uHealth* 2021 Mar 02;9(3):e24322 [FREE Full text] [doi: [10.2196/24322](https://doi.org/10.2196/24322)] [Medline: [33626017](https://pubmed.ncbi.nlm.nih.gov/33626017/)]
4. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. World Health Organization. 2018. URL: <https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf?sequence=1&isAllowed=y> [accessed 2021-12-05]
5. Fähnrich C, Denecke K, Adeoye O, Benzler J, Claus H, Kirchner G, et al. Surveillance and Outbreak Response Management System (SORMAS) to support the control of the Ebola virus disease outbreak in West Africa. *Euro Surveill* 2015 Mar 26;20(12):21071 [FREE Full text] [doi: [10.2807/1560-7917.es2015.20.12.21071](https://doi.org/10.2807/1560-7917.es2015.20.12.21071)] [Medline: [25846493](https://pubmed.ncbi.nlm.nih.gov/25846493/)]
6. Perscheid C, Benzler J, Hermann C, Janke M, Moyer D, Laedtke T, et al. Ebola outbreak containment: real-time task and resource coordination with SORMAS. *Front. ICT* 2018 Apr 10;5:7. [doi: [10.3389/fict.2018.00007](https://doi.org/10.3389/fict.2018.00007)]
7. Silenou BC, Tom-Aba D, Adeoye O, Arinze CC, Oyiri F, Suleman AK, et al. Use of Surveillance Outbreak Response Management and Analysis System for human monkeypox outbreak, Nigeria, 2017-2019. *Emerg Infect Dis* 2020 Feb;26(2):345-349 [FREE Full text] [doi: [10.3201/eid2602.191139](https://doi.org/10.3201/eid2602.191139)] [Medline: [31961314](https://pubmed.ncbi.nlm.nih.gov/31961314/)]
8. Tom-Aba D, Silenou BC, Doerrbecker J, Fourie C, Leitner C, Wahnschaffe M, et al. The Surveillance Outbreak Response Management and Analysis System (SORMAS): digital health global goods maturity assessment. *JMIR Public Health Surveill* 2020 Apr 29;6(2):e15860 [FREE Full text] [doi: [10.2196/15860](https://doi.org/10.2196/15860)] [Medline: [32347809](https://pubmed.ncbi.nlm.nih.gov/32347809/)]
9. Rai B, Shukla A, Dwivedi LK. Estimates of serial interval for COVID-19: a systematic review and meta-analysis. *Clin Epidemiol Glob Health* 2021 Jan;9:157-161 [FREE Full text] [doi: [10.1016/j.cegh.2020.08.007](https://doi.org/10.1016/j.cegh.2020.08.007)] [Medline: [32869006](https://pubmed.ncbi.nlm.nih.gov/32869006/)]
10. Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat Med* 2020 Nov 17;26(11):1714-1719. [doi: [10.1038/s41591-020-1092-0](https://doi.org/10.1038/s41591-020-1092-0)] [Medline: [32943787](https://pubmed.ncbi.nlm.nih.gov/32943787/)]
11. Lu QB, Zhang Y, Liu MJ, Zhang HY, Jalali N, Zhang AR, et al. Epidemiological parameters of COVID-19 and its implication for infectivity among patients in China, 1 January to 11 February 2020. *Euro Surveill* 2020 Oct;25(40):2000250 [FREE Full text] [doi: [10.2807/1560-7917.ES.2020.25.40.2000250](https://doi.org/10.2807/1560-7917.ES.2020.25.40.2000250)] [Medline: [33034281](https://pubmed.ncbi.nlm.nih.gov/33034281/)]
12. Mena RH, Velasco-Hernandez JX, Mantilla-Beniers NB, Carranco-Sapiéns GA, Benet L, Boyer D, et al. Using posterior predictive distributions to analyse epidemic models: COVID-19 in Mexico City. *Phys Biol* 2020 Sep 22;17(6):065001. [doi: [10.1088/1478-3975/abb115](https://doi.org/10.1088/1478-3975/abb115)] [Medline: [32959788](https://pubmed.ncbi.nlm.nih.gov/32959788/)]
13. Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs HP, Bosse NI, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Res* 2020 Dec 8;5:112. [doi: [10.12688/wellcomeopenres.16006.2](https://doi.org/10.12688/wellcomeopenres.16006.2)]
14. Nakajo K, Nishiura H. Exploring secondary SARS-CoV-2 transmission from asymptomatic cases using contact tracing data. *Theor Biol Med Model* 2021 Jul 16;18(1):12 [FREE Full text] [doi: [10.1186/s12976-021-00144-z](https://doi.org/10.1186/s12976-021-00144-z)] [Medline: [34271962](https://pubmed.ncbi.nlm.nih.gov/34271962/)]
15. Griffin J, Casey M, Collins A, Hunt K, McEvoy D, Byrne A, et al. Rapid review of available evidence on the serial interval and generation time of covid-19. *BMJ open* 2020;10(11):e040263. [doi: [10.1136/bmjopen-2020-040263](https://doi.org/10.1136/bmjopen-2020-040263)]
16. Lehtinen S, Ashcroft P, Bonhoeffer S. On the relationship between serial interval, infectiousness profile and generation time. *J R Soc Interface* 2021 Jan 06;18(174):20200756 [FREE Full text] [doi: [10.1098/rsif.2020.0756](https://doi.org/10.1098/rsif.2020.0756)] [Medline: [33402022](https://pubmed.ncbi.nlm.nih.gov/33402022/)]
17. Delignette-Muller ML, Dutang C. fitdistrplus: an R package for fitting distributions. *Journal of Statistical Software* 2015;64(4):1-34 [FREE Full text] [doi: [10.18637/jss.v064.i04](https://doi.org/10.18637/jss.v064.i04)]
18. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol* 2013 Nov 01;178(9):1505-1512 [FREE Full text] [doi: [10.1093/aje/kwt133](https://doi.org/10.1093/aje/kwt133)] [Medline: [24043437](https://pubmed.ncbi.nlm.nih.gov/24043437/)]
19. Thompson R, Stockwin J, van Gaalen R, Polonsky J, Kamvar Z, Demarsh P, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* 2019 Dec;29:100356 [FREE Full text] [doi: [10.1016/j.epidem.2019.100356](https://doi.org/10.1016/j.epidem.2019.100356)] [Medline: [31624039](https://pubmed.ncbi.nlm.nih.gov/31624039/)]
20. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature* 2005 Nov 17;438(7066):355-359 [FREE Full text] [doi: [10.1038/nature04153](https://doi.org/10.1038/nature04153)] [Medline: [16292310](https://pubmed.ncbi.nlm.nih.gov/16292310/)]
21. Snijders TA. The degree variance: an index of graph heterogeneity. *Social Networks* 1981;3(3):163-174. [doi: [10.1016/0378-8733\(81\)90014-9](https://doi.org/10.1016/0378-8733(81)90014-9)]
22. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems* 2006 Jan 11;1695(5):1-9 [FREE Full text]

23. Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, et al. shiny: web application framework for R. The R Project for Statistical Computing. 2021 Oct 02. URL: <https://CRAN.R-project.org/package=shiny> [accessed 2021-12-05]
24. Campbell P, Dewar M. shinyauthr: 'Shiny' authentication modules. The R Project for Statistical Computing. 2021 Jul 20. URL: <https://CRAN.R-project.org/package=shinyauthr> [accessed 2021-12-05]
25. Ooms J. sodium: A Modern and Easy-to-Use Crypto Library. The R Project for Statistical Computing. 2021 Oct 21. URL: <https://CRAN.R-project.org/package=sodium> [accessed 2021-12-05]
26. SORMAS-Stats demonstration server (Username: stats Password: aeJahpho3shi). URL: <https://sormas.helmholtz-hzi.de/sormas-stats-shinyapp/> [accessed 2021-12-05]
27. SORMAS-Stats GitHub repository. URL: <https://github.com/hzi-braunschweig/SORMAS-Stats-next-gen/> [accessed 2021-12-05]
28. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect Dis* 2020 Aug;20(8):911-919 [FREE Full text] [doi: [10.1016/S1473-3099\(20\)30287-5](https://doi.org/10.1016/S1473-3099(20)30287-5)] [Medline: [32353347](https://pubmed.ncbi.nlm.nih.gov/32353347/)]
29. Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat Med* 2020 Nov 17;26(11):1714-1719. [doi: [10.1038/s41591-020-1092-0](https://doi.org/10.1038/s41591-020-1092-0)] [Medline: [32943787](https://pubmed.ncbi.nlm.nih.gov/32943787/)]
30. R data import/export. The R Project for Statistical Computing. 2021 Nov 01. URL: <https://cran.r-project.org/doc/manuals/r-release/R-data.html> [accessed 2021-12-05]
31. Stolterman E, Wiberg M. Concept-driven interaction design research. *Human-Computer Interaction* 2010 Jun 15;25(2):95-118. [doi: [10.1080/07370020903586696](https://doi.org/10.1080/07370020903586696)]

## Abbreviations

**k:** dispersion parameter

**SI:** serial interval

**SORMAS:** Surveillance Outbreak Response Management and Analysis System

**SORMAS-Stats:** Surveillance Outbreak Response Management and Analysis System Statistical application

*Edited by T Sanchez, A Mavragani; submitted 27.12.21; peer-reviewed by O Ngwenyama, G Yuan; comments to author 25.01.22; revised version received 06.03.22; accepted 26.04.22; published 31.05.22*

*Please cite as:*

*Silenou BC, Verset C, Kaburi BB, Leuci O, Ghozzi S, Duboudin C, Krause G*

*A Novel Tool for Real-time Estimation of Epidemiological Parameters of Communicable Diseases Using Contact-Tracing Data: Development and Deployment*

*JMIR Public Health Surveill* 2022;8(5):e34438

URL: <https://publichealth.jmir.org/2022/5/e34438>

doi: [10.2196/34438](https://doi.org/10.2196/34438)

PMID: [35486812](https://pubmed.ncbi.nlm.nih.gov/35486812/)

©Bernard C Silenou, Carolin Verset, Basil B Kaburi, Olivier Leuci, Stéphane Ghozzi, Cédric Duboudin, Gérard Krause. Originally published in *JMIR Public Health and Surveillance* (<https://publichealth.jmir.org>), 31.05.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <https://publichealth.jmir.org>, as well as this copyright and license information must be included.