



Article

# A Statistical Model to Assess Risk for Supporting COVID-19 Quarantine Decisions

Sonja Jäckle <sup>1,\*</sup> , Elias Röger <sup>2</sup>, Volker Dicken <sup>3</sup> , Benjamin Geisler <sup>3</sup>, Jakob Schumacher <sup>4</sup>   
and Max Westphal <sup>3</sup>

<sup>1</sup> Fraunhofer Institute for Digital Medicine MEVIS, 23562 Lübeck, Germany

<sup>2</sup> Fraunhofer Institute for Industrial Mathematics ITWM, 67663 Kaiserslautern, Germany; yukio.elias.roeger@itwm.fraunhofer.de

<sup>3</sup> Fraunhofer Institute for Digital Medicine MEVIS, 28359 Bremen, Germany; volker.dicken@mevis.fraunhofer.de (V.D.); benjamin.geisler@mevis.fraunhofer.de (B.G.); max.westphal@mevis.fraunhofer.de (M.W.)

<sup>4</sup> Health Department Berlin-Reinickendorf, 13407 Berlin, Germany; jakob.schumacher@reinickendorf.berlin.de

\* Correspondence: sonja.jaekle@mevis.fraunhofer.de; Tel.: +49-3101-6106

**Abstract:** In Germany, local health departments are responsible for surveillance of the current pandemic situation. One of their major tasks is to monitor infected persons. For instance, the direct contacts of infectious persons at group meetings have to be traced and potentially quarantined. Such quarantine requirements may be revoked, when all contact persons obtain a negative polymerase chain reaction (PCR) test result. However, contact tracing and testing is time-consuming, costly and not always feasible. In this work, we present a statistical model for the probability that no transmission of COVID-19 occurred given an arbitrary number of negative test results among contact persons. Hereby, the time-dependent sensitivity and specificity of the PCR test are taken into account. We employ a parametric Bayesian model which combines an adaptable Beta-Binomial prior and two likelihood components in a novel fashion. This is illustrated for group events in German school classes. The first evaluation on a real-world dataset showed that our approach can support important quarantine decisions with the goal to achieve a better balance between necessary containment of the pandemic and preservation of social and economic life. Future work will focus on further refinement and evaluation of quarantine decisions based on our statistical model.

**Keywords:** COVID-19; decision support; bayesian statistics; quarantine; risk assessment



**Citation:** Jäckle, S.; Röger, E.; Dicken, V.; Geisler B.; Schumacher J.; Westphal M. A Statistical Model for Supporting COVID-19 Quarantine Decisions. *Int. J. Environ. Res. Public Health* **2021**, *18*, 9166. <https://doi.org/10.3390/ijerph18179166>

Academic Editor: Paul B. Tchounwou

Received: 30 June 2021

Accepted: 26 August 2021

Published: 31 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. COVID-19 Pandemic and Situation in Germany

Since spring 2021, COVID-19 spread to almost all countries around the world [1]. Many countries legislated strict lock-downs in order to decrease infections numbers. When the infection numbers have reached an acceptable low level, more liberal rules applied, such as social distancing and limits on larger gatherings. In addition, many countries have issued regulations requiring infected persons to be isolated and contact persons to be quarantined. In Germany, local health departments report information about their local, new cases daily to the Robert Koch Institute (RKI), the central German institution for infectious disease study and control. The RKI reports on a regular basis about the COVID-19 situation in Germany; see, for example, Reference [2]. Based on those reported numbers, the RKI recommends an implementation, adaption, or cancellation of procedural rules for political decision-makers in Germany.

Since August 2020, a second significant increase of new infections per week has been reported in Germany, as well as in many other European countries. Many of these cases were associated with vacation travels. In Germany, it resulted in several times higher numbers than reported in June, where the number of new infections per day had reached

a minimum so far [3]. This triggered strong societal discussions about risks involving the reopening of schools and re-allowing events, such as mass gatherings and shortening of the quarantine duration time for contact persons. This infection increase resulted in the so-called second wave in end of 2020, which continued until the start of 2021. Due to the spread of the alpha variant, the number of infections increased again in February and triggered a third wave of infections in spring of 2021. A similar situation occurred in summer 2021. Since June 2021, the delta variant, which already dominated the infections in July, has caused a further increase in infections.

### 1.2. Quarantine Decisions in Germany

The German Protection against Infection Act (“Infektionsschutzgesetz”) [4] defines the general rules for dealing with infectious diseases. It specifies which pathogens must be monitored and allows German health departments to quarantine people at home. In the current COVID-19 pandemic, all contacts with close contact to the infected person and, therefore, at high risk of infection have to stay at home. Since a large number of COVID-19 cases was reported by the RKI, a large number of contact persons probably had to be quarantined in Germany [2]. The RKI provides the procedural guidelines for these cases. In the end, however, the local health departments have to decide autonomously whether and for how long a person or a group of persons have to be quarantined.

This is the first time in many decades that health departments have had to quarantine such a large number of infected individuals and contacts. In recent decades, quarantines have been imposed only a few times. For example, in 2002/2003, the first severe acute respiratory syndrome (SARS-CoV) started to spread over several countries. In Germany, contact persons were quarantined at home, but the number of quarantined was substantially smaller than in the current COVID-19 pandemic [5]. A default workflow for contact tracing of COVID-19 cases is provided by the RKI [6]. All contact persons are classified into one of the three categories, that are listed in Table 1.

**Table 1.** Contact categories defined by the RKI.

Category	Infection Risk
1.	high, close contact to infectious person
2.	low, sparse contact to infectious person
3.	medical staff with adequate protection

The RKI defines category 1 as persons who

- had face-to-face contact for at least 15 min or
- had direct contact to bodily fluids or
- were exposed to a relevant aerosol concentration.

Other contact persons are classified as category 2. Medical staff with appropriate personal protective equipment is classified as category 3. The RKI recommends that contact persons of category 1 and 2 normally should reduce their contacts to other persons, and contact persons of category 1 additionally should stay quarantined at home until 14 days after exposure. Moreover, contact persons of category 1 should be tested, ideally one day after case documentation and a second time 5–7 days after contact with the COVID-19 case [6].

Since February 2020, a high number of contact persons of COVID-19 had to stay in home quarantine, especially during the infection waves, when the highest number of active COVID-19 cases occurred. The RKI collected data about all COVID-19 spreads until 11 August 2020 and reported the frequency and size for different event types [7]. The probability of a transmission and the number of infected people depends noticeably on the event type. As of 11 August 2020, 3902 spreads occurred in households and 709 spreads in retirements homes, whereas only 31 spreads were reported at schools and 33 at kindergartens [7].

This indicates that the risk of a COVID-19 spread in a group depends strongly on the scenario. This fact might inform quarantine decisions and might justify more liberal quarantine decisions in certain situations. However, the event type is currently not taken into account systematically according to the RKI recommendations. Furthermore, it is often not possible to get test results from all participants of a contact event. Reasons for this might be that testing capacity is limited, the persons could not be contacted at all or they did not go to a test center.

Moreover, quarantine can indirectly cause economic damage. Since the persons have to stay at home or employees have to take care of their quarantined children, they are unable to work during the quarantine. In addition, quarantines of persons, which work in system relevant positions, can lead to an undersupply of critical infrastructure. This includes medical care and food supply, but also other sectors, such as government administration and supply of water and energy.

Against this background, it is desirable to support these quarantine decisions with objective data and statistical models because many people are involved in the process. A software implementing such a model can streamline and support the decision making process and make it less prone to human mistakes. In literature, different models have been proposed for explaining COVID-19 transmission dynamics [8,9] or predicting the number of COVID-19 infections [10,11]. Those models allow the governments to adapt their COVID-19 regulations in order to control the COVID-19 pandemic. However, it is also important to directly support the institutions which are responsible for monitoring the COVID-19 infected persons in their district. In literature, a statistical model for describing COVID-19 transmissions of an infected person [12] and a model for estimating the infection risk in meetings [13] has been introduced. However, no model for decision support of group quarantines has been reported, to the best of our knowledge. Thus, this is the first work which describes a statistical model as decision support for group quarantines.

### 1.3. Structure of the Paper

In this work, we will provide a statistical model for the risk assessment of canceling group quarantines. It allows to estimate the risk of canceling a group quarantine, where a fraction of the group has been tested and all test results are negative. In Section 2, the Bayesian model including the prior choice is described. The proposed model can be used for any group event situation in any country. In this work, it is applied for a school class scenario and analyzed in Section 3. Here, the model specified for school classes was tested with a real-world data-set of a German school class. Finally, the model and the results are discussed in Section 4.

## 2. Materials and Methods

### 2.1. Use Case

In the following, we describe the idealized use case for our model. We assume that a group contact event took place, e.g., a family party, a sports event, or a teaching unit in school. A certain time after the event, one or multiple participants are tested positive for COVID-19, possibly only after the onset of symptoms. More generally, a series of contact events with the same group (e.g., in school or at work) might have taken place. We presume that there is only a single person known to be infected, the so-called *primary* or *index case*. Adaptations for multiple primary cases are discussed in Section 4. The remaining group, without counting the primary case, is assumed to consist of  $M$  individuals with unknown infection status. As a precautionary measure, this group of potential *secondary cases* is quarantined by health authorities to contain further infection dynamics. In addition, all contact persons are invited to take a polymerase chain reaction (PCR) test. Currently, all contacts should stay in quarantine for 14 days [14].

We are now interested whether there occurred any transmissions of COVID-19 from the primary case to any of the contact persons during the group event. Any other transmission paths outside of the group event are not considered. If there is at least one positive

test result in the remaining group, this indicates that a COVID-19 transmission occurred. In this case, the health offices in Germany are not allowed to cancel the group quarantine because other persons might be also infected, and this has to be checked. Thus, the case of interest is that all test results are negative. Figure 1 (left) illustrates the timeline of these events described above. The main question which guided the methodological developments in this work is:

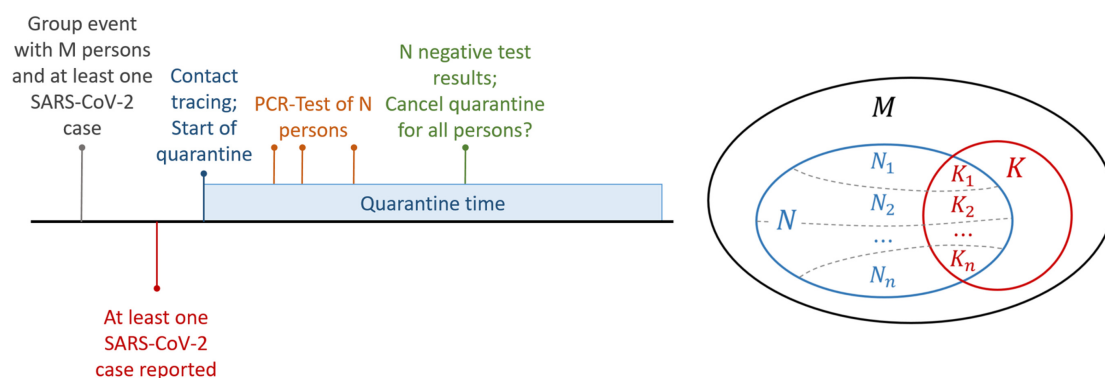
“Under which conditions can a group quarantine be released (earlier) such that the probability of overlooking a (secondary) infection is low?”

In an ideal world, all potential cases would be tested by a highly accurate test. However, we rather suppose that not all  $M$  individuals are tested due to resource constraints or other reasons. Instead, out of the  $M$  person group,  $N \leq M$  are successfully tested negative.

In our situation, we consider reverse transcription PCR tests with (time-dependent) imperfect accuracy. An extension to other tests, as, for example, the rapid antigen test, is possible as long as the specificity and sensitivity can be characterized. Thus, the probability that all test results are negative, even though there are some infected persons in the group, is not zero. As a further complication, we allow that not all  $N$  tests are conducted on the same day (after the group contact event). Rather, usually, test time points vary due to logistic or resource limitations. In summary, we have only partial (probabilistic) knowledge regarding the disease status of the remaining group. The above question, thus, can be formulated more precisely as:

“How likely is it that there is no secondary infection, given all  $N$  out of  $M$  people are tested negative?”

The model described in the following assumes that the people to be tested are chosen randomly. This is not likely to hold in the real world as most authorities will give priority to testing people that were close spatially or socially to the primary case(s). As a result, it is less likely that an individual with a high risk gets released from quarantine without a test. If that is the case, the model will slightly overestimate the risk of an infection. In consultation with the health office Berlin-Reinickendorf, we concluded that this calculation is appropriate.



**Figure 1.** Left: Timeline of the events starting from the group event until the quarantine decision. Right: Overview of the group with  $M$ , the possible number of infected  $K$ , and the negative tested persons  $N$ . For  $N$  and  $K$ , the contact persons are divided into subgroups  $N_d$  and  $K_d$ , which are tested on day  $D_d, d \in \{1, \dots, n\}$ , are also illustrated.

## 2.2. Statistical Model

The goal of the model is to estimate the probability, that no COVID-19 transmission occurred. As outlined above, we consider  $M$  contact persons,  $N \leq M$  of them having been tested and all tests results are negative. Moreover, we have an unknown number of infected persons  $K; K \leq M$ . We consider only the secondary transmissions caused by the group event, but no further potential infections outside of the participant group  $M$ . Figure 1 (right) gives

an overview over these (sub)groups. According to Bayes rule, the posterior distribution for a given number of  $K = k$  is then described as

$$\begin{aligned} P(K = k | N \text{ out of } M \text{ tested negative at specific days}) \\ = \frac{P(N \text{ out of } M \text{ tested negative at specific days} | K = k) \cdot P(K = k)}{P(N \text{ out of } M \text{ tested negative at specific days})}. \end{aligned} \quad (1)$$

For our use case, we are interested in the probability, that no one is actually infected ( $K = 0$ ) given that  $N$  out of  $M$  have been tested, and all tests result are negative. This probability is then given as

$$\begin{aligned} p_0 &:= P(K = 0 | N \text{ out of } M \text{ tested negative at specific days}) \\ &= \frac{P(N \text{ out of } M \text{ tested negative at specific days} | K = 0) \cdot P(K = 0)}{P(N \text{ out of } M \text{ tested negative at specific days})} \\ &= \frac{P(K = 0)}{\sum_{i=0}^M P(N \text{ out of } M \text{ tested negative at specific days} | K = i) P(K = i)}, \end{aligned} \quad (2)$$

since it is assumed, that healthy persons are always tested negative, which is described in more detail in Section 2.4. Thus,  $P(N \text{ out of } M \text{ tested negative} | K = 0) = 1$  holds.

This requires an a-priori probability distribution  $P(K)$  for the number of infected persons  $K$ . This prior describes which values of  $K$  are considered probable before considering the data. Furthermore, a likelihood function  $P(N \text{ out of } M \text{ tested negative} | K)$  is needed, which describes the probability distribution that  $N$  of  $M$  are tested negative under the condition that  $K$  persons are infected in the group. The modeling of the prior and the likelihood function are described in the following sections.

### 2.3. Prior Distribution

Transmissions of COVID-19 have a high heterogeneity, which means that most infected do not infect anybody else whereas some infected persons are causing super-spreading events. Lelieveld et al. [13] introduced a model based on the binomial distribution for SARS-CoV-19 aerosol transmission risk calculation. The negative binomial distribution was employed in the literature to model heterogeneity in secondary infections [12]. However, the negative binomial distribution is defined for an unrestricted support  $\{0, 1, 2, \dots\}$ , which is not fitting for a finite (and usually small) group size as considered in this work. We employ a Beta-binomial prior instead, which is a direct extension to the binomial prior. The uncertainty regarding the transmission probability  $p$  can be modeled using the Beta-binomial as prior distribution. The Beta-binomial distribution is defined as

$$K | \alpha, \beta, M \sim \text{BB}(M, \alpha, \beta), \quad (3)$$

with shape parameters  $\alpha$  and  $\beta$ . It is a discrete probability distribution on a finite support of non-negative integers, and it is frequently used when the success probability in each of the known number of trials is either unknown or random. Its probability mass function is defined as

$$P(K = k) = \binom{M}{k} \frac{B(k + \alpha, M - k + \beta)}{B(\alpha, \beta)}, \quad (4)$$

where  $B(\alpha, \beta)$  is the beta function. The probability that the contact persons become infected and the number of infected persons depends on the type of event, e.g., if it takes place outside or inside and how much the persons interact and talk. Thus, the parameters of the Beta-binomial distribution should be adjusted to the considered use case.

In order to define  $\alpha$  and  $\beta$ , we use two other parameters for the scenario specific modeling which allows for easier interpretation. We consider the probability that a transmission of COVID-19 to at least one person takes place

$$P(K > 0), \quad (5)$$

and the conditional expected value

$$E(K|K > 0) = \frac{\sum_{i=1}^M P(K = i) \cdot i}{P(K > 0)}. \quad (6)$$

With these two known values, the needed parameters  $\alpha, \beta$  can be determined by solving an optimization problem, for example, with the L-BFGS method [15].

#### 2.4. Likelihood

The accuracy of a test is described by the specificity (Sp, also called *True Negative Rate*) and the sensitivity (Se, also called *True Positive Rate*). In this paper, PCR tests are considered for testing. The U.S. Food and Drug Administration analyzed PCR tests and determined a specificity of nearly 100% [16] (the very few false positives are usually caused by laboratory handling errors). Thus, we assume a specificity of 1. However, the sensitivity is less than 1, and the probability that a person infected with COVID-19 is tested positive depends on the time point of testing; see Reference [17], as well. For our model, these probabilities were read and copied manually from the first graph of Figure 2 of publication [17]. Thus, the test of an infected person will be negative with almost 100% probability on the first and second day after the infection; between days six to ten after infection, this probability of a false negative test result drops below 25%, reaching a minimum of about 20% on day eight.

This day specific sensitivity is used for our realistic modeling of the likelihood function: The  $N$  tested persons are split in  $n$  day specific groups.  $N_1$  persons were tested on day  $D_1$  after the group event,  $N_2$  persons were tested on day  $D_2, D_2 \neq D_1, \dots$ , and  $N_n$  persons were tested on day  $D_n, D_n \neq D_d$  for  $d \in \{1, \dots, n-1\}$ . In the same way, the  $K$  infected persons are divided into days:  $K_1, \dots, K_n$ . For both the tested persons  $N_d$  and the infected tested persons  $K_d$ , we assume that they are only tested on a single day. Thus, every person  $x$  is tested on day  $D_i$  and is not tested on day  $D_j$  and  $i \neq j \forall i, j \in \{1, \dots, n\}$ . The remaining untested and infected persons  $N_0 = M - N, K_0 = K - \sum_{d=1}^n K_d$  are considered as an additional group. These subsets of  $N$  and  $K$  are also illustrated in the right image of Figure 1. Moreover, the day dependent sensitivity  $s_{D_d}$  is known for all days  $D_d, d \in \{1, \dots, n\}$ .

The distribution of COVID-19 infected persons  $K$  can now be considered as a random drawing from an urn containing  $M$  persons divided into the subsets  $N_0, N_1, \dots, N_n$ . Then, the probability of one specific draw  $\tilde{K} = (K_0, \dots, K_n), \sum_{d=0}^n K_d = K$  and  $K_d \leq N_d \forall d \in \{0, \dots, n\}$  can be calculated with the formula of the multivariate hyper-geometric distribution:

$$P(\tilde{K} = (K_0, \dots, K_n)) = \frac{\prod_{d=0}^n \binom{N_d}{K_d}}{\binom{M}{K}}. \quad (7)$$

Furthermore, the probability for a negative test on day  $d$  is needed. For healthy persons, this probability is 1, since we assume a specificity of 100%. However, for infected persons, this probability depends on the test date:

$$P(\text{Negative test on day } D_d | \text{person is healthy}) = 1; \quad (8)$$

$$P(\text{Negative test on day } D_d | \text{person is infected}) = 1 - s_{D_d}. \quad (9)$$

On each day  $D_d, N_d$  persons with an unknown number of infected persons  $K_d \leq N_d$  are tested and the likelihood of testing them all negative is given by

$$\begin{aligned} P(N_d \text{ negative tests on day } D_d | K_d \text{ persons are infected}) &= (1 - s_{D_d})^{K_d} \cdot 1^{N_d - K_d} \\ &= (1 - s_{D_d})^{K_d}. \end{aligned} \quad (10)$$

The probability for testing all infected persons on all days negative can be determined by multiplying all day specific probabilities, since the number of negative tests is given for each day; thus, the probabilities for each day are independent:

$$\begin{aligned}
 P(N \text{ total negative tests with } N_d \text{ at each day } D_d | K = \sum_{d=0}^n K_d \text{ of } N_d \text{ persons are infected}) \\
 &= \prod_{d=1}^n P(N_d \text{ negative tests on day } D_d | K_d \text{ of } N_d \text{ persons are infected}) \\
 &= \prod_{d=1}^n (1 - s_{D_d})^{K_d}.
 \end{aligned} \tag{11}$$

For the next step, we are considering a simple example, where all persons are tested on the same day, and it holds that  $N = N_1, N_0 = M - N$ . In this case, the following draws  $\tilde{K}$  can occur:

$$X_K = \{(\min(N_0, K), K - \min(N_0, K)), \dots, (K - \min(N_1, K), \min(N_1, K))\}. \tag{12}$$

Thus, the probabilities for all those possible draws defined in Equation (7) multiplied with the negative testing probabilities defined in Equation (11) must be summed up to obtain

$$P(N \text{ out of } M \text{ tested negative at specific days } | K) = \sum_{(K_0, K_1) \in X_K} \frac{\binom{N}{K_0} \binom{M-N}{K_1}}{\binom{M}{K}} (1 - s_{D_1})^{K_1}. \tag{13}$$

In general, the persons are tested on  $n$  different days, and we obtain:

$$X_K = \left\{ (K_0, \dots, K_n) \mid \sum_{d=0}^n K_d = K \text{ and } K_d \leq N_d \ \forall d \in \{0, \dots, n\} \right\} \tag{14}$$

as set of all possible draws. Again, the probabilities have to be summed up to obtain the probability

$$\begin{aligned}
 P(N \text{ out of } M \text{ tested negative at specific days } | K) \\
 &= \sum_{\tilde{K} \in X_K} P(\tilde{K} = (K_0, \dots, K_n) \cap N \text{ total negative tests with } N_d \text{ at each day } D_d),
 \end{aligned} \tag{15}$$

where  $P(\tilde{K} = (K_0, \dots, K_n) \cap N \text{ total negative tests with } N_d \text{ at each day } D_d)$  is the probability of testing all persons in one specific realization  $\tilde{K} = (K_0, \dots, K_n)$ . As in the simple example, the probability for the draw has to be multiplied with the probability, that all infected people are tested negative. Using Equations (7) and (11), we obtain

$$\begin{aligned}
 P(\tilde{K} = (K_0, \dots, K_n) \cap N \text{ total negative tests with } N_d \text{ at each day } D_d) \\
 &= P(\tilde{K} = (K_0, \dots, K_n)) \\
 &\quad \cdot P(N \text{ total negative tests with } N_d \text{ at each day } D_d | K = \sum_{d=0}^n K_d \text{ persons are infected}) \\
 &= \frac{\prod_{d=0}^n \binom{N_d}{K_d}}{\binom{M}{K}} \cdot \prod_{d=1}^n (1 - s_{D_d})^{K_d} \\
 &= \frac{\binom{N_0}{K_0}}{\binom{M}{K}} \cdot \prod_{d=1}^n \binom{N_d}{K_d} (1 - s_{D_d})^{K_d}.
 \end{aligned} \tag{16}$$

Thus, the likelihood function is modeled with the above equations, and the posterior probability can be calculated via Bayes Theorem, as outlined in Section 2.2.

### 3. Result

#### 3.1. Prior Distribution for School Classes

We conducted a literature search about the probability of a COVID-19 outbreak in schools. This search was done with Google Scholar with different combinations of the keywords SARS-CoV-2, COVID-19, transmission, school, pupils, education. As a result, we found four studies that analyzed the transmission probability in school classes: a French study of one school [18], an Irish study of all COVID-19 cases in schools [19], an Australian study [20] analyzing different educational scenarios, and a study in Luxembourg [21].

In the Australian study, 15 school classes with at least one COVID-19 infected person were analyzed and for 3 school classes new COVID-19 infections were detected [20]. No transmission of COVID-19 in 6 school classes was detected in the Irish study [19]. In the French study, one outbreak was analyzed in detail, especially the transmissions of COVID-19 in one high school [18] that took place at a time prior to enforcement of strict pandemic hygiene rules. This study analyzed an outbreak at a whole school and not in one single school class. The study in Luxembourg analyzed the primary and secondary transmissions in school; however, no information about transmission probability in school classes was reported [21]. Therefore, we decided not to take the last two studies into account. Based on the two remaining sources, we obtain a first crude estimation of  $P(K > 0)$ . Here, we considered all school classes equally. Then, the sum of all school classes with transmissions was calculated and divided by the sum of all considered school classes:

$$P(K > 0) = \frac{0 + 3}{6 + 15} = \frac{3}{21} \approx 0.14. \quad (17)$$

Moreover, the RKI recently published a report about COVID-19 outbreaks in different environments [7]. In this document, an analysis about the average number of infected persons per outbreak was made. For school classes, an average number of 4.8 infected persons was reported, which equals the conditional mean value:

$$E(K|K > 0) = 4.8. \quad (18)$$

With these two values, a unique Beta-binomial distribution can be determined as outlined in Section 2.3. The distributions for school classes with  $M = 15, 25, 35$  are plotted in Figure 2. For all cases, a suitable Beta-binomial distribution could be calculated. Small differences between the distributions can be only observed for the probability values with  $K > 0$ .

#### 3.2. Posterior Probabilities and Decision Support for School Classes

In this section, the school class scenario is further considered and the posterior probability is analyzed that no transmission occurred given that  $N$  out of  $M$  persons are tested negative. For this analysis, we used the prior specified for school classes which is determined in Section 3.1. In Germany, school classes have usually at least 13 and not more than 35 pupils [22]. Thus, we calculated this probability for different school class sizes  $M \in \{13, \dots, 35\}$  and different numbers of tests  $N$ , which were all done 4 days after the group event. The results are shown in the left panel of Figure 3. This probability ranges from 0.85 up to 1.0. One reason for this high minimum is the specific prior distribution, which should have the probability 0.86 for no transmission. Thus, the posterior probability cannot be smaller than 0.86 due to formula (2). Smaller values are only observed for  $M < 15$ , where no optimal parameters could be found with the optimization. Moreover, all determined posterior probabilities never reach 100% because there is always some small remaining risk, that there could be some infected persons in the group, who were tested negative.

With this calculated probability, one can make a decision whether the quarantine can be canceled despite the fact that not all persons have been tested. Using, for example, the rule  $p_0 > 0.9$ , the quarantine will be canceled in most cases, as shown in the right panel of



Figure 3. Only if less than around 30% of the group size have been tested is the result  $p_0$  less than 90%, and the quarantine is continued.

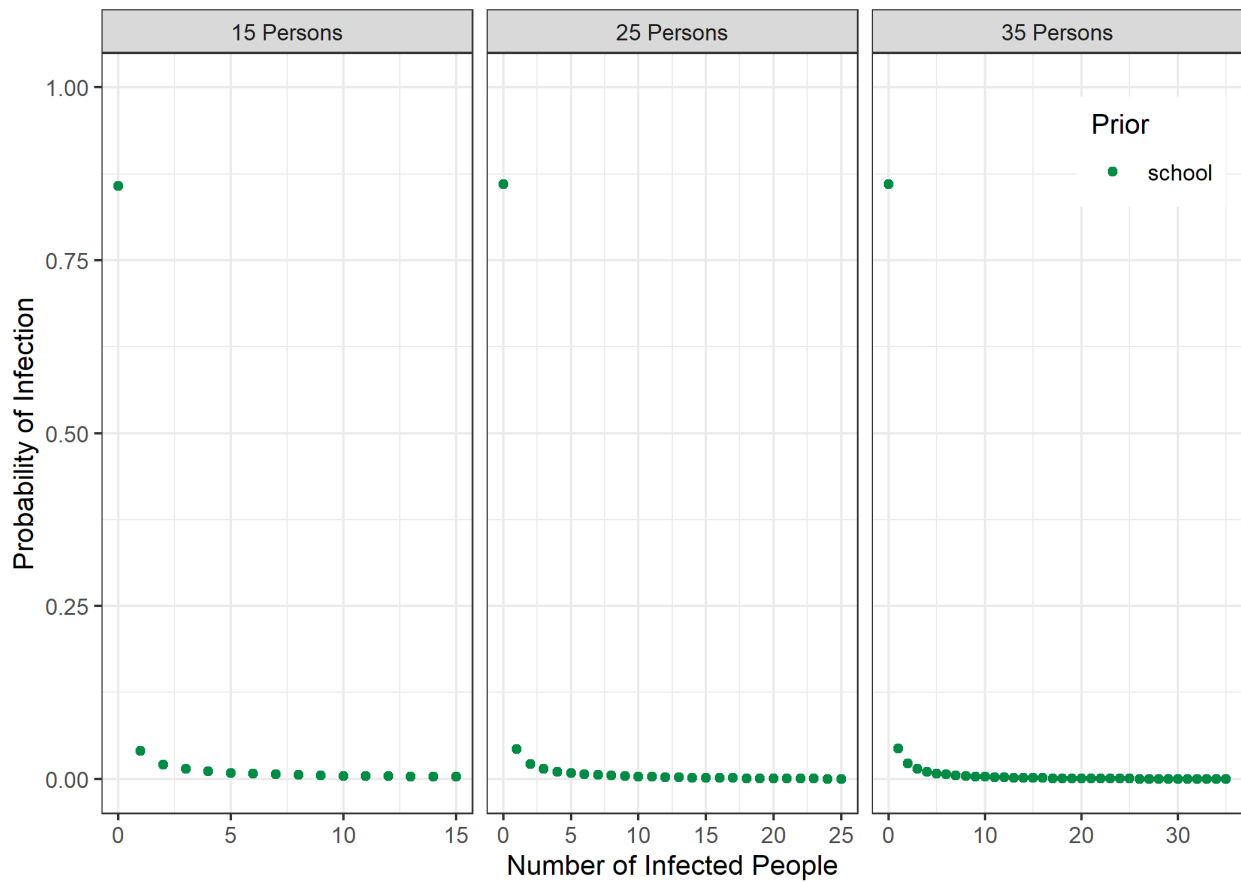


Figure 2. Plots of the Beta-binomial distribution as school class prior for  $M = 15, 25, 35$ .

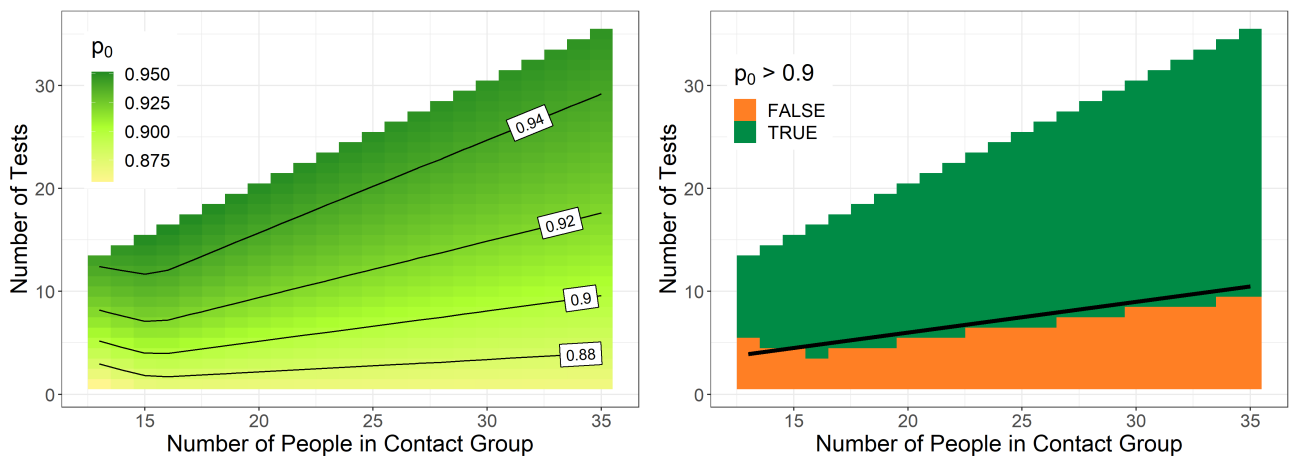


Figure 3. **Left:** Posterior probability that no transmission occurred given that  $N$  out of  $M$  persons are tested negative using the specific school prior for different group sizes and number of tests made 4 days after the group event. **Right:** Decision of a quarantine cancellation based the condition that  $p_0 > 0.9$ . The black line shows the situation that 30% are tested.

### 3.3. Decision about Quarantine Cancellation for School Classes

The statistical model with the prior specified for school classes (determined in Section 3.1) is tested and evaluated with one dedicated dataset provided by the health department Berlin-Reinickendorf. The dataset is shown in Table 2. We have 17 persons which were in contact with one infected person. All of them had contact on 10 August,

for most of them it was the last contact. However, two of them have met the person later on until 16 August and 18 August again. These two persons do not fit into our group event setting and hence were excluded. Employees at the German health departments consider and decide about the quarantine of these persons separately on an individual basis.

As a result, we now have 15 persons with last contact on August 10. Twelve persons were tested, and all test results are negative. One person was tested on August 18, ten persons on August 19, and one person on August 20. The timeline for this sequence of events is given in Figure 4. With this information, we can calculate the posterior probability, that nobody at all was infected according to our model. With the prior distribution determined for a school class with  $M = 15$  pupils, we obtain a probability of  $p_0 = 98\%$ , that no transmission of COVID-19 occurred in this setting. Based on the described decision rule of  $p_0 > 0.9$  suggested in Section 3.2, this result would support to cancel the quarantine for the whole group early. Indeed, no further COVID-19 cases were reported to the health department concerning this particular school class. Hence, the correct decision about cancellation of quarantine would have been made in this specific case.

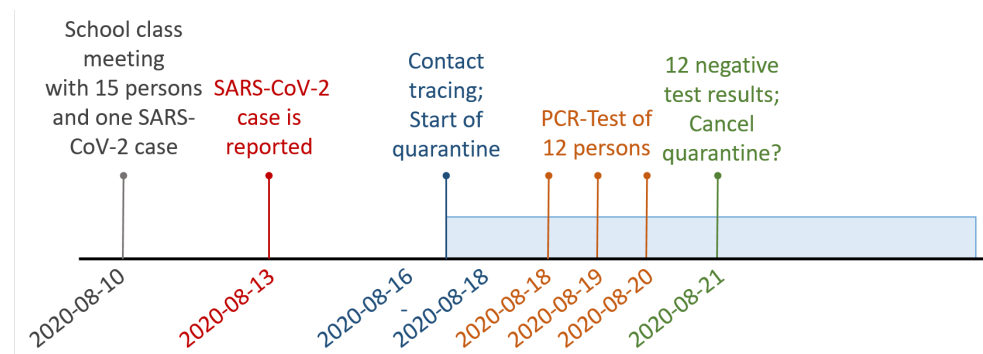


Figure 4. Timeline of the school class example.

Table 2. A dataset of a school class from the health department in Berlin-Reinickendorf, which contains the last contact date, the PCR test date, and the PCR results.

Case Number	Date of Last Contact	PCR Test Date	PCR Test Result
1	10 August 2020	—	—
2	10 August 2020	—	—
3	10 August 2020	—	—
4	10 August 2020	18 August 2020	negative
5	10 August 2020	19 August 2020	negative
6	10 August 2020	19 August 2020	negative
7	10 August 2020	19 August 2020	negative
8	10 August 2020	20 August 2020	negative
9	10 August 2020	19 August 2020	negative
10	10 August 2020	19 August 2020	negative
11	10 August 2020	19 August 2020	negative
12	10 August 2020	19 August 2020	negative
13	10 August 2020	19 August 2020	negative
14	10 August 2020	19 August 2020	negative
15	10 August 2020	19 August 2020	negative
16	16 August 2020	19 August 2020	negative
17	18 August 2020	24 August 2020	negative

#### 4. Discussion

The results show, that the proposed statistical model can support quarantine decisions by risk estimation. The generalizability of our approach to other countries and/or other types of group events is limited and depends on the availability of suitable evidence to inform the

prior distribution. For that reason, we would advise to conduct a literature research to derive meaningful prior parameters before adopting our model to other scenarios.

The results from the proposed model depend on the chosen prior distribution which is subject to uncertainty. In Bayesian analysis, a non-informative prior is sometimes phrased as a solution to this problem. We would, however, discourage using a “non-informative” uniform prior in the considered scenario because of epidemiological implausibility: for instance, under a uniform prior, the prior probability of no ( $K = 0$ ) or all ( $K = M$ ) contact persons being infected is equal, which we do not consider to be realistic.

Furthermore, general and quite simplified assumptions were made for the different probability distributions. Several aspects, such as the number of initially infected persons in the group or other aspects of the group event (inside/outside, ventilation, activity type, interactions, etc.), were not addressed. A paper was published, which describes a model for estimating the aerosol transmission risk depending on several parameters e.g., room properties, event duration or number of participants [13]. Based on this work, the probability that a transmission of COVID-19 occurred can be estimated. However, further information about the average number of infected persons is not available using this approach introduced in Reference [13]. Currently, there is still a lack of available data and published studies for extending our statistical model. In addition, for the likelihood, we assumed that every contact was tested once. However, normally, contact persons are tested two times.

Since the prior specification is always subjective and depends on the availability of adequate evidence for model parameters, the sensitivity of the results to the prior should be addressed in the future. Different frameworks to approach robust Bayesian analysis have been described in the literature [23]. Furthermore, we assumed a random choice of tested persons, but, in reality, the contact persons with the highest risk would be chosen for tests. The likelihood was specified for the PCR-tests, but this could be also adapted for other test types when the tests sensitivities and specificity are known. We also assumed that all tests results are negative because the health offices do not cancel the quarantine if at least one person is tested positive. However, this could change in future due to increasing vaccination numbers and one or two positive tested persons could be acceptable. In this case, a more precise estimation of tested specificity and, consequently, the false positive tested rate is needed to have a realistic model.

## 5. Conclusions

In this work, we introduced a statistical model for estimating the risk of an early group quarantine cancellation. Therefore, a day specific sensitivity and a specificity of 100% of the PCR-test were considered for the likelihood distribution. Moreover, the Beta-binomial distribution was chosen as prior for the transmission of the COVID-19, and a method for determining its parameters for specific situations was proposed. The whole model can be specified for any group event in any country. In this work, it was used for obtaining a prior distribution for school classes based on data obtained from literature. Based on the obtained prior distribution, we analyzed the usage of our statistical model for a decision support for quarantine cancellation and tested it on a real-world dataset of a German school class. Offered as a web application or integrated in a software solution, our statistical model could be used by employees of German health departments and could facilitate their daily work. A prototypical version of the web application (in German) is available upon request to the corresponding author.

In addition, we illustrated a decision rule based on the arbitrary threshold of 90% for the posterior probability that no transmission occurred. However, it is not clear, if this is the best rule for making a decision. The threshold could be also set higher, e.g.,  $p_0 > 0.95$ , to reduce the remaining risk. Future work should address the choice of the threshold in dependence of factors, such as available test capacities. Moreover, we will implement other circumstances of the contact event (location, interactions, etc.) in the model and will apply it for other specific scenarios. In a next step, the benefit of using our tool in routine

cases has to be evaluated. For this purpose, a broad evaluation study with German health departments is planned.

**Author Contributions:** Conceptualization, J.S.; investigation, S.J., V.D., B.G. and M.W.; methodology, S.J., E.R., V.D. and M.W.; resources, J.S.; software, S.J. and E.R.; supervision, M.W.; visualization, S.J. and E.R.; writing—original draft preparation, S.J., V.D. and M.W.; writing—review and editing, S.J., E.R., V.D., B.G., J.S. and M.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Fraunhofer Internal Programs under Grant No. Anti-Corona 840242.

**Institutional Review Board Statement:** Not Applicable.

**Informed Consent Statement:** Not Applicable.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author, upon reasonable request.

**Acknowledgments:** We thank the team from the Fraunhofer ITWM, in particular Andreas Wagner and Stefanie Grimm, for the fruitful collaboration in the CorASiV project, especially for the collaborative development of the prototypical web application for German health departments. Furthermore, we thank Ann Christina Foldenauer (Fraunhofer IME—Institute for Molecular Biology and Applied Ecology, Frankfurt, Germany) for her valuable feedback on an earlier version of the manuscript.

**Conflicts of Interest:** The authors declare no potential conflict of interests.

## References

- World Health Organization. COVID-19 Weekly Epidemiological Update Edition 45, Published 22 June 2021. Available online: <https://apps.who.int/iris/bitstream/handle/10665/342009/CoV-weekly-sitrep22Jun21-eng.pdf?sequence=1&isAllowed=y> (accessed on 30 June 2021).
- Robert Koch Institut. COVID-19: Fallzahlen in Deutschland und Weltweit. Available online: [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Situationsberichte/Jun\\_2021/2021-06-24-en.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Jun_2021/2021-06-24-en.pdf?__blob=publicationFile) (accessed on 30 June 2021).
- Robert Koch Institut. Täglicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 (COVID-19) 27.08.2020. Available online: [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Situationsberichte/2020-08-27-de.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-08-27-de.pdf?__blob=publicationFile) (accessed on 22 September 2020).
- Bundesamt für Justiz. Gesetz zur Verhütung und Bekämpfung von Infektionskrankheiten beim Menschen (Infektionsschutzgesetz—IfSG). Available online: <https://www.gesetze-im-internet.de/ifsg/IfSG.pdf> (accessed on 15 September 2020).
- Glasmacher, S.; Kurth, R. Globaler Alarm. *Spektrum der Wissenschaft*. 2006; pp. 12–17. Available online: [https://www.rki.de/DE/Content/InfAZ/N/Neue\\_Infektionskrankheiten/Globaler-Alarm.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/InfAZ/N/Neue_Infektionskrankheiten/Globaler-Alarm.pdf?__blob=publicationFile) (accessed on 30 August 2021).
- Robert Koch Institut. Kontaktpersonennachverfolgung bei Respiratorischen Erkrankungen durch das Coronavirus SARS-CoV-2. Available online: [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Kontaktperson/Management.html](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Kontaktperson/Management.html) (accessed on 31 August 2020).
- Robert Koch Institut. Infektionsumfeld von erfassten COVID-19-Ausbrüchen in Deutschland. Available online: [https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2020/Ausgaben/38\\_20.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2020/Ausgaben/38_20.pdf?__blob=publicationFile) (accessed on 31 August 2020).
- Chen, J.; Guo, X.; Pan, H.; Zhong, S. What determines city's resilience against epidemic outbreak: evidence from China's COVID-19 experience. *Sustain. Cities Soc.* **2021**, *70*, 102892. [CrossRef] [PubMed]
- Samui, P.; Mondal, J.; Khajanchi, S. A mathematical model for COVID-19 transmission dynamics with a case study of India. *Chaos Solitons Fractals* **2020**, *140*, 110173. [CrossRef] [PubMed]
- Forecasting Team. Modeling COVID-19 scenarios for the United States. *Nat. Med.* **2021**, *27*, 94–105. [CrossRef] [PubMed]
- Mandal, M.; Jana, S.; Nandi, S.K.; Khatua, A.; Adak, S.; Kar, T. A model based study on the dynamics of COVID-19: Prediction and control. *Chaos Solitons Fractals* **2020**, *136*, 109889. [CrossRef] [PubMed]
- Althouse, B.M.; Wenger, E.A.; Miller, J.C.; Scarpino, S.V.; Allard, A.; Hébert-Dufresne, L.; Hu, H. Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. *arXiv* **2020**, arXiv:2005.13689.
- Lelieveld, J.; Helleis, F.; Borrmann, S.; Cheng, Y.; Drewnick, F.; Haug, G.; Klimach, T.; Sciare, J.; Su, H.; Pöschl, U. Model Calculations of Aerosol Transmission and Infection Risk of COVID-19 in Indoor Environments. *Int. J. Environ. Res. Public Health* **2020**, *17*, 8114. [CrossRef] [PubMed]
- Robert Koch Institut. Abwägung der Dauer von Quarantäne und Isolierung bei COVID-19; Toleranzen Gegenüber Bioziden Wirkstoffen. Available online: [https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2020/Ausgaben/39\\_20.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2020/Ausgaben/39_20.pdf?__blob=publicationFile) (accessed on 14 October 2020).

15. Byrd, R.H.; Lu, P.; Nocedal, J.; Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **1995**, *16*, 1190–1208. [[CrossRef](#)]
16. Quest Diagnostics. SARS-CoV-2 RNA, Qualitative Real-Time RT-PCR (Test Code 39433): Package Insert. Available online: <https://www.fda.gov/media/136231/download> (accessed on 12 November 2020).
17. Kucirka, L.M.; Lauer, S.A.; Laeyendecker, O.; Boon, D.; Lessler, J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based SARS-CoV-2 tests by time since exposure. *Ann. Intern. Med.* **2020**, *173*, 262–267. [[CrossRef](#)] [[PubMed](#)]
18. Fontanet, A.; Tondeur, L.; Madec, Y.; Grant, R.; Besombes, C.; Jolly, N.; Pellerin, S.F.; Ungeheuer, M.N.; Cailleau, I.; Kuhmel, L.; et al. Cluster of COVID-19 in northern France: A retrospective closed cohort study. *medRxiv* **2020**, 20071134. [[CrossRef](#)]
19. Heavey, L.; Casey, G.; Kelly, C.; Kelly, D.; McDarby, G. No evidence of secondary transmission of COVID-19 from children attending school in Ireland, 2020. *Eurosurveill* **2020**, *25*, 2000903. [[CrossRef](#)] [[PubMed](#)]
20. Macartney, K.; Quinn, H.E.; Pillsbury, A.J.; Koirala, A.; Deng, L.; Winkler, N.; Katelaris, A.L.; O’Sullivan, M.V.; Dalton, C.; Wood, N.; et al. Transmission of SARS-CoV-2 in Australian educational settings: A prospective cohort study. *Lancet Child Adolesc. Health* **2020**, *4*, 807–816. [[CrossRef](#)]
21. Mossong, J.; Mombaerts, L.; Veiber, L.; Pastore, J.; Le Coroller, G.; Schnell, M.; Masi, S.; Huiart, L.; Wilmes, P. SARS-CoV-2 transmission in educational settings during an early summer epidemic wave in Luxembourg, 2020. *BMC Infect. Dis.* **2021**, *21*, 1–8. [[CrossRef](#)] [[PubMed](#)]
22. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Vorgaben für die Klassenbildung, Schuljahr 2019/2020. Available online: [https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Dokumentationen/2019-09-16\\_Klassenbildung\\_2019.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Dokumentationen/2019-09-16_Klassenbildung_2019.pdf) (accessed on 25 June 2021).
23. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*; Springer: New York, NY, USA, 1985.