# Analysis of the Risk Model of German Corona Warning App

Jens Braband

•

TU Braunschweig, Germany
j.braband@*tu-braunschweig*.de


Hendrik Schäbe

•

TÜV Rheinland
schaebe@de.tuv.com

## Abstract

*In Germany an App has been introduced to cope with the Corona epidemic. In this paper we describe, how the app works and analyze the semi-quantitative risk model that has been used in the app, because a large number of semi-quantitative risk models are known that are not consistent. Further we discuss in how far the Corona app in its current state can contribute to mitigate the pandemic. The risk model of the German Corona warning app has several interesting, somewhat puzzling properties. In this paper we describe the analysis and its results related to the underlying risk model.*

**Keywords:** risk model, Corona app, risk calculation

## I. Introduction

In Germany, the so-called Corona Warning App has been introduced to cope with the Corona epidemic. This app computes an individual infections risk based on contacts with infected persons. In this paper we describe, how the app version 1.7 works and analyzes the semi-quantitative risk model that has been used in the app, because a large number of semi-quantitative risk models are known that are not consistent. Further we discuss in how far the Corona app version 1.7 in its current state can contribute to mitigate the pandemic. Not that the current version of the app is 1.9 and changes are ongoing.

## II. How the app works

First a simplified overview is given in order to understand the risk model. More details are given in [1][2].
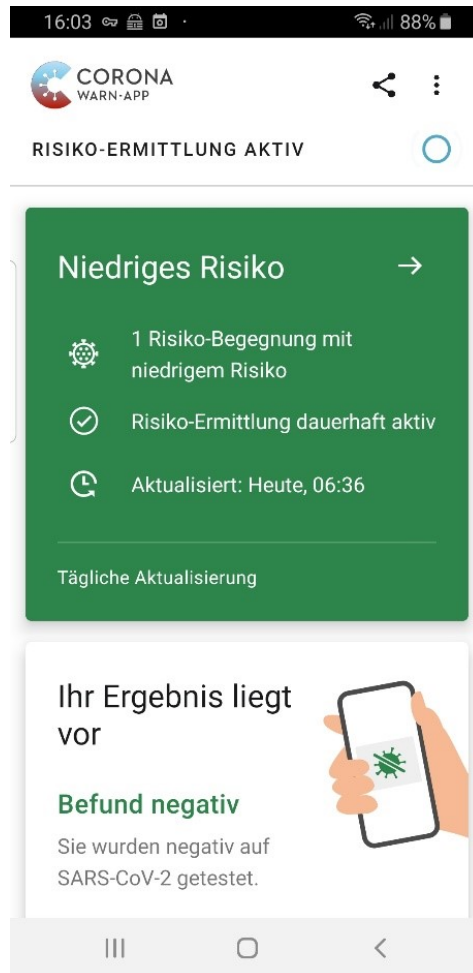


**Figure 1:** *Screenshot of the Corona Warning App indicating a low risk and a negative test result*

     After a user has installed the app, each day a new anonymous ID is created. Every few minutes the environment is scanned for Bluetooth signals emitted from other apps. Data like ID, signal attenuation, duration etc. are collected and aggregated for each day.

     If the user receives a positive test result and agrees to publish it, then the anonymous ID for the preceding 14 days are transmitted to the central server, from which it is transmitted to all subscribers. The actual risk evaluation is performed decentralized by each app.

     Figure 1 shows a screenshot of the app, with an indication of a low risk because of one encounter with an infected person. However the date of the encounter or a more detailed risk estimate is not given, although such data are known.

## III. The Basic Risk Model

The basic model is defined by four parameters [2], which in a first step are evaluated on a semi-quantitative scale each ranging from 0-8 for each day for each ID that reported a positive test result (see figure 2):

- The Days since Exposure (DE) is the time since exposure to the infected person, a value between 0 and 14, durations longer than 14 days are not considered.
- The Exposure Duration (ED) is the cumulative time of exposure on the day, takes values between 0 and 8.
- The Bluetooth Signal Attenuation (SA) is used as a measure of the distance to the infected person, takes values between 0 and 8.
- The Transmission Risk (TR) estimates the level of infectiosity of the person on that day, takes values between 0 and 8.

Then the Total Risk Score (TRS) is evaluated by multiplication of DE, ED, SA and TR, theoretically resulting in scores between 0 and 7168.

This resembles the approach known as Risk Priority Numbers (RPN) and suffers from the same limitations and flaws, which are known for about two decades [5][6]. For some application sectors the use of RPN is even deprecated [7].

The major problem is that the scores for the parameters are often only ordinal scale or rank numbers, for which operations like multiplication or division are not well-defined. As a consequence, the results may lead to under- or overestimation of the related risk [8].

However, in the practical implementation of the Corona warning app today the model is simplified and the coinciding ranges are limited [3] by

- Days since Exposure (DE) is set to 5 for values below 14 days, and 0 for above, leading to
  $\delta_{ED} = 5*I(DE \leq 14)$
- Exposure Duration (ED) is set to 0 for all values up to 10 minutes, and 1 for above, yielding
  $\delta_{DE} = I(ED > 10)$
- Signal Attenuation (SA) is set to 0 above 73 dB, and 2 for all values below, i.e.
  $\delta_{SA} = 2* I(SA \leq 73dB)$
- Transmission Risk (TR) is set to (6, 8, 8, 8, 5, 3, 1, 1, 1, 1, 1, 1, 1) [4], depending on DE, e. g. TR is 6 if DE=1, 8 if DE=2 etc., and 0, if DE=14 or above

Here, I(.) denotes the indicator function, which take value 1 if the expression in brackets holds true, and zero otherwise.

So, most of the parameters are only used as binary indicator variables and in the current configuration the Total Risk Score for a particular day and a particular ID is given by

$$TRS = 10 \cdot \delta_{ED} \cdot \delta_{SA} \cdot \delta_{DE} TR \qquad (1)$$

So, with the current implementation [3] there are only six possible scores: 0, 10, 30, 50, 60, 80. But also a Minimum Risk Score (MRS) of 11 is defined and all risks below are discarded. But the parametrization of the app may be changed.

So, as of today, we can conclude that the basic risk model as implemented is more a dosimetric model, depending on the estimated virus concentration, rather than on exposure and other parameters (but for some threshold values). It is not even a risk model as per the definition of many standards. Moreover, the risk model has been heavily discretized.
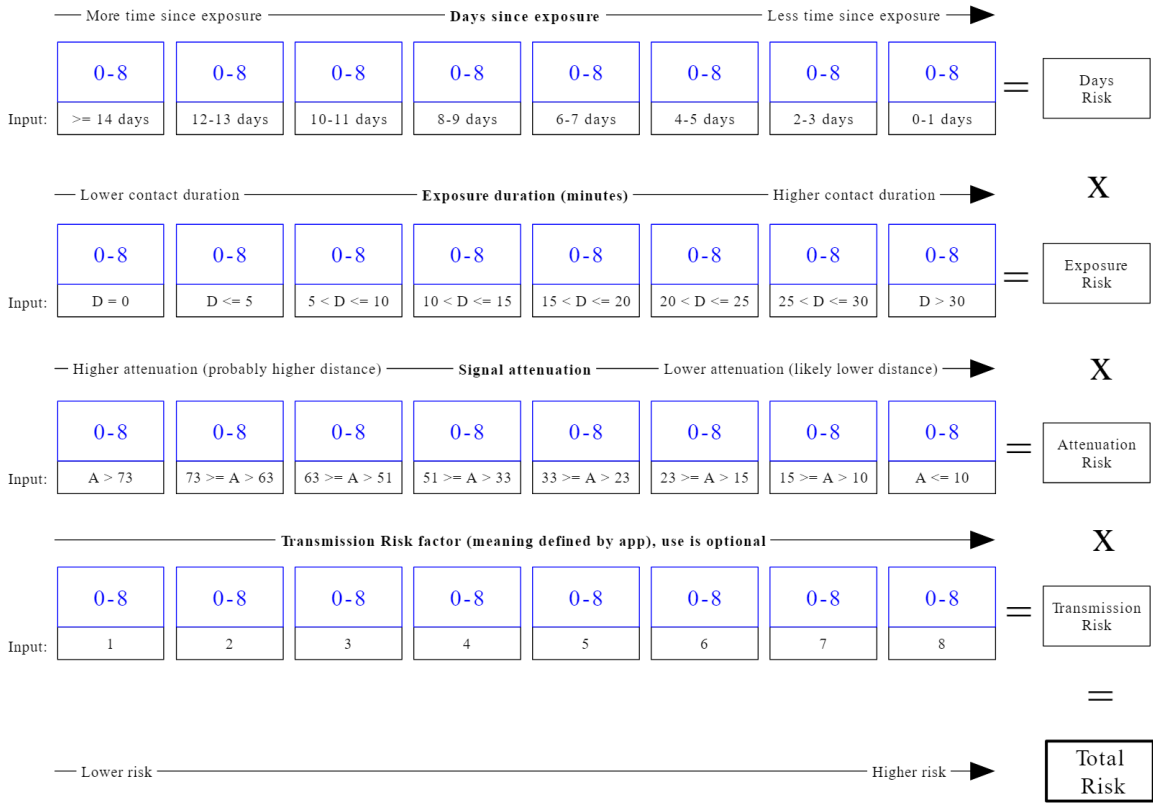
**Figure 2:** *Basic Risk Calculation [2]*

Example:

- Alice receives a positive test result on the 20th of the month, which she reports immediately.
- Bob is often taking the same bus as Alice. A ride takes 10 minutes and he has met her on the 16th (two rides) and the 9th (one ride). They have sat together with a distance of about 1m.
- For the 16th DE=4 and so TR=8. Both SA and ED are above the threshold and set to 2 and 1, respectively, so TRS=80.
- For the 9th DE=11 and so TR=1. However, ED is below the threshold and set to 0. So TRS=0. Otherwise the TRS would have been 5, which is below the MRS and would have been discarded anyhow.

## IV. Combined Risk Model

In a second step each app combines the scores for different encounters calculated by the basic risk model. Let $R_1$, $R_2$, …$R_n$ denote the individual TRS for different days and different IDs that are above the MRS.

In a first step the maximum value $R_{max}$ of the different TRS is determined. Then the ED of all the n encounters are summed up into three different classes: close, medium and far. Let their durations be $t_1$, $t_2$ and $t_3$, respectively. For each of the classes a weight is defined and additionally a weight offset, which are denoted by $w_1$, $w_2$, $w_3$ and $w_4$, respectively. Note that in practice the weights for the close and medium classes outweigh the others. Also, an Average Risk Score (ARS), currently 50 [3], is defined. Then the Total Combined Risk (TCR) is calculated as (see figure 3)

$$TCR = (t_1 w_1 + t_2 w_2 + t_3 w_3 + w_4) \frac{R_{max}}{ARS} \qquad (2)$$
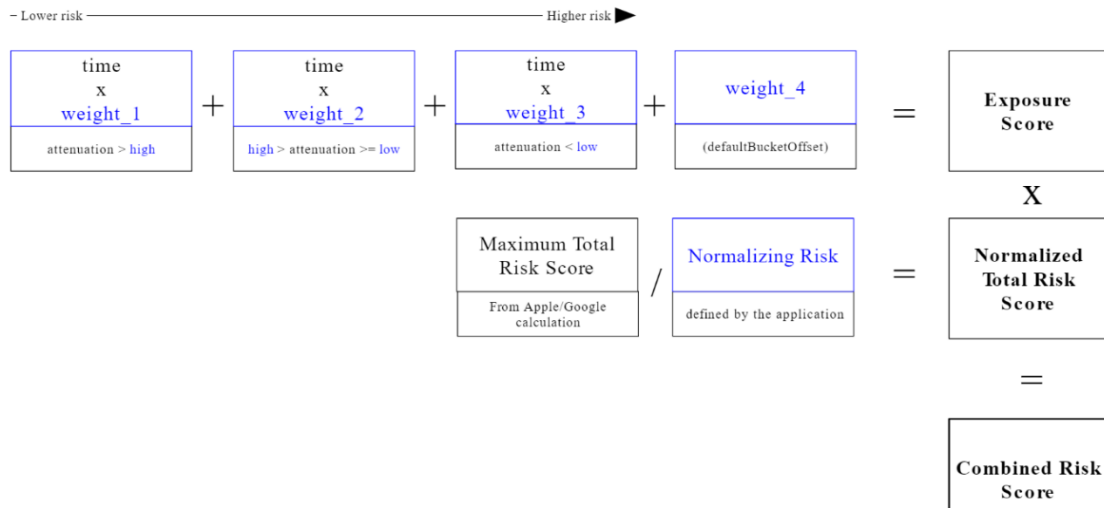
**Figure 3:** *Combined Risk Calculation [2]*

Surprisingly the TCR is in fact not a risk, but an exposure as the result is in minutes. The first term (in brackets) is a weighted exposure time which is adjusted by a relative factor (dimensionless) which depends on the maximum virus concentration compared with some average.

So overall, without full mathematical exactness, we can characterize the approach used today [3] by the German Corona warning app as

$$TCR = (t_1 w_1 + t_2 w_2 + t_3 w_3 + w_4) \frac{max\{10 \cdot \delta_{ED} \cdot \delta_{DE} \delta_{SA} \cdot TR\}}{ARS} \approx (t_1 + t_2/2) \frac{10 max\{TR\}}{ARS} \quad (3)$$

Basically this is not a risk in the narrow sense, it is a weighted exposure duration, the units of TCR are not expected damage per time or similar, but just minutes of exposure. And the weight applied is just a measure of relative infectiosity, which expresses the size of TR relative to a normalizing factor of 5, which is assumed as average infectiosity. And the impact of the factor is limited, the highest possible value being 1.6.

And, if we take additionally into account the uncertainty and spread of all the input parameters, e. g. noise in the signal attenuation, uncertainty in the exposure duration and infectiosity [4], or arbitrariness of the weights chosen, then the model boils down to a quite simple formula and decision procedure:

$$TCR \approx \frac{maxTR}{5} \sum ED \quad (4)$$

- Estimate the minutes that the person was exposed closely to infected persons ($\sum ED$)
- Weight the exposure ED by the infectiosity of the most infected person (max TR/5)
- Take action if the result is HIGH

Example (continued)
- Additionally, Charlie has received a positive test result on the 20th, but he reported it only on the 21st. But he has installed the app only a week ago.
- He has been on the same bus on the same days, but with some larger distance to Bob (2m)
- For the 16th, DE=5 (evaluated on the 21st) and so TR=5. Both SA and ED are above the threshold and set to 2 and 1, so TRS=50.
- For the 9th, Charlie had not installed the app yet, so there are no data.

- As Alice set close to Bob, and Charlie in medium distance, t1=t2=20 minutes. The weights are currently set [3] to w1=1, w2=0.5, w3=w4=0, and so the weighted sum results in 30 minutes
- So, the TCR amounts to 30 x 80 / 50, which gives 48 minutes.
- The warnings of the app are issued based on the TCR, which are configured [3] as LOW for values up to 15, and HIGH from 15 onwards. So finally, Bob would get a HIGH risk warning.

Note that the TCR is almost independent of the distance measured by Bluetooth Signal Attenuation. There is only a loose threshold defined and the exposure duration is weighted into two distance classes. However, it is known, see e. g. the FAQ by RKI [9], that transmission is mainly by aerosols, and the risk increases when the exposure distance decreases. Close contact to infected persons is also known to be a factor in superspreading events. The importance of the distance is also supported by the fact that it is a major factor in the German Corona rules AHA, where the first letter stands for Abstand (distance).

## V. Naive Risk Model

Unfortunately, the derivation of the Risk Model described above seems only partially published [4]. But we may formulate a general risk model stimulated by similar models from dosimetry.

The similarity that we exploit here is that we have sources of activity (the infected persons), here transmission of virus material with a certain intensity I (which is assumed to be constant at least temporarily e. g. over a day). The exposed persons are assumed to be exposed at a fixed (or average) distance D and time T.

The intensity I could be defined e. g. by virus material per volume or surface e. g. $[M/m^2]$, D would be measured by [m] and T by [min]. For the risk model we would have to take some assumptions how the intensity decreases with distance to the source.

Several assumptions on the dependence of concentration depending on distance can be made:

a. Thinning of the virus in a volume. The concentration of the virus is the amount of the virus per volume unit, assuming more or less equal distribution. Then the concentration would decrease with $R^3$, where R is the distance between the source (infected person) and the receiver. This assumption would only hold, if there is sufficient convection.

b. The virus is emitted as a spherical wave, e.g. as a result of playing a brass instrument, sneezing etc. Then the virus is approximately present on the surface of a sphere. So, the concentration would decrease with $R^2$.

c. The virus is emitted as a cylindrical wave. Then the concentration would decrease proportional to R. This can be the case if e.g. infected singers are standing on a stage, emitting their virus into a large room.

d. The virus is spread in the room, reflected from the walls, so that an equal concentration occurs in the entire room. Then, distancing would not decrease the virus concentration.

Therefore, we think that assuming a dependence of the intensity I on 1/R is a plausible assumption, which should cover typical cases or be conservative.

We may also define a nominal virus intensity I0 at a given distance, say 1 m. This would still depend on the infectiousness of the person.

As the exposure to virus material increases proportional to exposure time, we may define the dose for a particular encounter with an infectious person

$$R \approx \frac{TI_0}{D} \qquad (5)$$

Note that these parameters directly reflect parameters ED, SA and TR, while DE is not needed. Interestingly it is defined, but not used in the basic risk model, currently it is just a constant. And in the definition of many standards the dose R is not a risk.

Example (continued):
- Let's try to evaluate Alice's and Bob's encounter on the 16th by this approach. D=1 and T=20 mins. The absolute I0 value is not known, but it was the highest possible value. For the sake of simplicity, we take the same value of 8, which is also justified by the construction of the TR vector [4]. So, R=160 would result.
- For the 9th we have D=1 and T=10. I0 is at the lowest value, let's assume 1 and we get R=10.
- Now let's look at Charlie. He is at D=2m distance exposing Bob for T=20 min. His TR value was 5, thus resulting in R=50. The Combined Risk is just the sum of the partial risk and gives 220.

Note that in this model the influence of the distance is explicit and influential. E. g. if we assume that Charlie and Alice had close contact, e. g. at a party instead of on the bus ride, we might assume D=0.5, which would result then in R=200. Note that in the original model the TCR would only rise from 48 to 64 as Bob's 20 minutes of exposure to Charlie would now fall into the close distance category.

## VI. Other influences and limitations

Besides the risk model – or to be precise the dose model - inside the app we also would need to consider influences that can be identified outside the app.

Currently only 19.3 million persons in Germany have the app [10], out of 83.2 million inhabitants [11]. That means, that with a probability of 23% a person has the app installed. Now, this might have different reasons:

a. Some person do not own a smartphone. This holds for very young children or for elderly people.
b. Some persons have not yet installed the app, e. g. because their phone is not compatible or outdated
c. Some persons refuse to install the app.

Then, the probability that two persons meet that have both installed the app is 23%*23% =5.4%.

This computation is very rough, since there are correlations in different population groups with mobility and owning a smartphone. On the one hand side, elderly people who might not own a smartphone would on the other hand be less mobile or even be in an pensioner's home and would also be part of a group with high risk of fatality. On the other hand, small children do not own a smartphone, have many contacts but seem not to play a great role in transmitting infections.

In any case, the figure above gives a rough indication that the app overall would miss about 95% of all encounters between persons. However, for a person that has installed the app, it rises to 23%

Assume now that a person is infected but has no symptoms. There is a growing number of persons with no or light symptoms [12]. A precise number is hard to give. In [13] a number of 43% is given. Now, even if a person has no symptoms, Corona can be detected by a test. End of August 2020 a weekly number of tests of about 1 million has been carried out in Germany [12]. Assuming now a typical period of 14 days, i.e. two weeks that are relevant, the probability of being detected by mass tests as infected during the two weeks period is

$$1-(1-1/83.2)^2 = 2.4\%. \tag{6}$$

So, finally a person that is infected with Corona will only be detected by a test with probability

$$57\% + 43\%*2.4\% = 58\%. \tag{7}$$

Here we have even assumed that all persons with symptoms are tested and a Corona infection

is detected by test.

The infected and tested person has to register herself in the Corona app as infected. Assume now, that even everyone would do so.

We can now compute the probability that the Corona app would detect and handle such an encounter of an infected person with another. The result is

$$58\% * 5.4\% = 3.1\% \tag{8}$$

Only in this fraction of cases of encounters the risk or dose model described above would come to work and help to protect persons.

However, there is another serious influence when computing the overall dose or risk according to

$$TCR \approx \frac{maxTR}{5} \sum ED, \tag{9}$$

This value would be underestimated. The sum will run only over those encounters with another person also having the app, i.e. 23% and knowing that he / she is infected, i.e. 58%. That means, that in the sum above only

$$23\% * 58\% = 13\% \tag{10}$$

of all dangerous encounters will be considered so that the TCR value statistically heavily underestimate this value.

So we see that there are two influences on the risk or dose values currently computed by the Corona app:

1) Only a small number of encounters is handled in the app with a risk model that is more or less a dose model
2) A cumulative TCR value computed by the app is heavily biased and indicates too small values.

To make the use of the app more efficient it is important to increase the number of users. Since the fraction of users has a quadratic influence on the probability that an encounter of persons is analyzed, this is the most important possibility. For analyzing the accumulated risk, one needs to take into account the fraction of use of the app, too. The TCR value depends linearly on the fraction of persons that have installed the app. Possible target values on this value need to be adapted when the fraction of app users in the population grows.

Another approach to make the estimations more precise would be to allow the infected person to voluntarily release more or more precise data. E. g. the infected person might be asked to enter the date of the test instead of the day when she releases the test result. Also, the number of daily encounters with small distance could be shown by the app in order to increase the awareness for keeping distance. Additionally the user may voluntarily keep a contact diary in the app so that in case of infection the tracing of contacts may be easier.

## VII. Summary

The risk model of the German Corona warning app has several interesting, somewhat puzzling properties:

1. In the narrow definition of many standards it is not a complete risk model, as it estimates only single parameters of risk, but not a comprehensive risk. A partial explanation can be based on the decentralized architecture of the app and the incomplete and inaccurate information it uses, often explained by data privacy concerns
2. Out of four parameters defined in the basic risk model only one parameter, the transmission risk is fully evaluated, others are only used as binary parameters
3. While in the basic risk model the result could be expressed as an infectiosity or dose of infectious particles to which the individual is exposed, the resulting combined risk is expressed as an exposure time. This is very uncommon that the combination of

several similar risks is expressed in different measurement units.

4. The estimated distance to the infected person is only compared to a threshold and as a means to weight different exposure times, but it has a minor influence in the model compared to the other parameters, while distance keeping plays a major role in infection prevention.

Finally, the authors would not recommend using the full parameter set for the basic risk model as this would lead to the same problems encountered with risk priority numbers. It would be reasonable to develop a full risk model and not only a partial risk model based on a weighted exposure duration only. Some parameters could be collected voluntarily from the user, like the age as the predominant factor for serious consequences or the health status. Other important parameters could be estimated by data from other smartphone sensors like GPS, e. g. environmental parameters like indoor or outdoor.

Moreover, the authors have seen that the effectivity of the app is still small, since the action of persons using it is also limited.

# References

[1] RKI: So funktioniert die Corona-Warn-App im Detail, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Funktion_Detail.pdf, last retrieval 2020-10-06

[2] CWA Team: Corona-Warn-App Solution Architecture, https://github.com/corona-warn-app/cwa-documentation/blob/master/solution_architecture.md, last retrieval 2020-10-06

[3] CWA Team: Wie ermittelt die Corona-Warn-App ein erhöhtes Risiko ?, https://github.com/corona-warn-app/cwa-documentation/blob/master/translations/cwa-risk-assessment.de.md, last retrieval 2020-10-06

[4] CWA Team: Epidemiological Motivation of the Transmission Risk Level, 2020-06-15, https://github.com/corona-warn-app/cwa-documentation/blob/master/transmission_risk.pdf, last retrieval 2020-10-06

[5] J. Bowles: An Assessment of RPN Prioritization in a Failure Modes Effects and Criticality Analysis. In: Proc. RAMS2003, Tampa, January 2003

[6] J. Braband: Improving the Risk Priority Number Concept. In: Journal of System Safety. 3, 2003, S. 21–23

[7] Durivage, M.: Is It Time To Say Goodbye To FMEA Risk Priority Number (RPN) Scores?, in: Pharmaceutical Online, https://www.pharmaceuticalonline.com/doc/is-it-time-to-say-goodbye-to-fmea-risk-priority-number-rpn-scores-00012020-04-27

[8] J. Braband: Beschränktes Risiko. In: Qualität und Zuverlässigkeit. 53(2), 2008, S. 28–33.

[9] Robert-Koch-Institut: SARS-Cov-2 Steckbrief, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html, last retrieval 2020-10-06

[10] https://de.statista.com/statistik/daten/studie/1125951/umfrage/downloads-der-corona-warn-app/, accessed on 23.10.2020

[11] https://de.statista.com/statistik/daten/studie/1217/umfrage/entwicklung-der-gesamtbevoelkerung-seit-2002/, accessed on 23.10.2020

[12] https://www.aerzteblatt.de/nachrichten/116077/Wenig-Schwerkranke-trotz-gestiegener-Infektionszahlen

[13] https://www.faz.net/aktuell/wissen/wie-viele-corona-infizierte-frei-von-symptomen-bleiben-16816959.html