The significance of the detection ratio for predictions on the outcome of an epidemic - a message from mathematical modelers

Jan Fuhrmann[1,2] and Maria Vittoria Barbarossa[2,3]

[1]Jülich Supercomputing Center, Forschungszentrum Jülich, Jülich, Germany
[3]Institute for Applied Mathematics, Heidelberg University, Heidelberg, Germany
[3]Frankfurt Institute of Advanced Studies, Frankfurt, Germany
j.fuhrmann@fz-juelich.de, barbarossa@fias.uni-frankfurt.de

Abstract

In attempting to predict the further course of the novel coronavirus (COVID-19) pandemic caused by SARS-CoV-2, mathematical models of different types are frequently employed and calibrated to reported case numbers. Among the major challenges in interpreting these data is the uncertainty about the amount of undetected infections, or conversely: the detection ratio. As a result, some models include assumptions about the percentage of detected cases among total infections while others completely neglect undetected cases. Here, we illustrate how model projections about case and fatality numbers vary significantly under varying assumptions on the detection ratio. Uncertainties in model predictions can be significantly reduced by representative testing, both for antibodies and active virus RNA, to uncover past and current infections that have gone undetected thus far.

Background

The World Health Organization declared the outbreak of the respiratory disease COVID-19 (coronavirus disease 2019, caused by the virus SARS-CoV-2) a global pandemic on March 11, 2020. In the past months, numerous efforts have been undertaken to understand the properties of SARS-CoV-2 and control the spread of the disease. While it has been repeatedly observed that the disease occurs in different severities, from very mild to critical [7], it is yet to be clarified how pre-symptomatic or asymptomatic infections do contribute to the spread of the virus [27, 21, 8].

In the attempt to predict the course of the outbreak and to possibly achieve its mitigation, mathematical models have been devised to predict the course and possible outcomes for different countries have been presented [10, 19, 3, 24, 13]. It was noted by several authors [28, 5, 15] that an important parameter of the epidemic is the detection ratio, meaning the percentage of infections that are actually discovered. Most mathematical models are able to reproduce the chronology of case numbers under widely varying assumptions on the value of the detection ratio. In particular, as has been noted in [9], the dynamic parameters, like the reproduction number $\mathcal{R}$, derived from the case number data are completely independent of the detection ratio, assuming the latter is constant over time. Since detection ratios are also notoriously hard to obtain at early stages of an epidemic, many models use overly optimistic detection ratios of 90% or more [13], or do not even take undetected cases into account at all [6, 1, 20]. The differences in reported case numbers during the initial phase of an epidemic outbreak may not be noticeable. However, in the long run, the number of undetected infections, and in particular undetected recoveries, plays a major role in reducing the number of susceptible individuals within a population and thereby achieving the threshold of herd immunity. Moreover, as has been pointed out by, e.g., [25], the number of undiscovered infections is missing in the denominator when case fatality rates, i.e., the percentage of infected individuals dying from the disease, are being calculated. The uncertainty about detection ratios seems to be the major determinant of widely varying estimates for the mortality of COVID-19.

Unfortunately, different approaches to determining these ratios have delivered a huge range of estimates for different geographic regions, ranging from detection ratios as low as 2% [18] to approximately 35% [18, 17]. For Germany, preliminary unofficial results for Gangelt in the county of Heinsberg [26],

seem to yield about 10-20 % case detection. Consistently, screening for antibodies, and therefore counting recovered individuals as well as currently infected ones, seems to yield lower estimates than screening studies using PCR-tests for active virus RNA. Possible reasons may include that (i) mild infections could easily go unnoticed, or that (ii) false negative rates of PCR-tests being caused by less than perfect sample collection [29].

We have previously proposed mathematical models for the dynamics of COVID-19 infections in Germany, in particular taking into account the effect of current and possible non-pharmaceutical control measures [2, 3]. For the simulated scenarios in our most recent work [3], we assumed detection ratios closer to the upper end of the range detailed above and remarked that the predicted fatality numbers should be expected to look very different when a lower detection ratio is assumed.

To illustrate this effect, we present here the results of simulations assuming different detection ratios, while maintaining unchanged assumptions on the other basic model parameters. For the scenarios which we show below we do not take into account the limited capacity of the health care system (this factor would further aggravate the situation in scenarios with high numbers of active cases). Our simulations shall only show the time course of both *known* and *total* active cases, as well as the cumulative number of fatalities. The latter model output is not only of high importance but also particularly sensitive to assumptions on detection.

Results

In order to illustrate how different assumptions on the detection ratio (DR) affect predictions of the epidemic's course, we show here simulation results for a few scenarios under the assumptions of high detection ratio ($DR \approx 40\%$), medium detection ratio ($DR \approx 10\%$, and low detection ratio ($DR \approx 2.5\%$) each. The model used for simulation is an extended version of the classical susceptible - exposed - infectious - recovered (SEIR) system, with three age groups and different compartments of infectiuous individuals (based on our previous work [3]). Case and death counts reported in Germany by the Robert Koch Institute (RKI) [22] as of April 24, 2020 were used for model calibration. In essence, this means estimating the effective contact rates only up to the lock-down situation in force until April 19, 2020. Though lock-down measures were partially relaxed starting on April 20, 2020, this would not show in the most recent data due to both the latency time of infection and delays in reporting. We show model simulations of the following scenarios for Germany.

A Nonmedical interventions as of April 25, 2020, including the most recent relaxation of some lockdown measures (starting April 20) and partial reopening of schools.

B Interventions as in **A**, plus an additional fatigue effect leading to general awareness wearing off. People are assumed to become less careful in, e.g., sanitizing hands, keeping distance in public space, or in the coughing/sneezing protocol. This is assumed to gradually and partially reduce the effect of general awareness over the course of 8 weeks.

C Interventions as in **A**, assuming increased efforts isolating known and suspected infected individuals (called *strict case isolation*, sCI, in [3]) setting in on April 27. As long as the number of active cases remains relatively low, strict isolation could be practicable.

D The original baseline scenario from [3] with interventions in place as of April 14. This is a counterfactual scenario assuming the restrictions in place had not been relaxed starting on April 20.

Model simulations were run until the end of the epidemic, that is, until the number of active cases becomes insignificant due to the reproduction number $\mathcal{R}$ being persistently smaller than one. Note that this may be due to sufficiently low effective contact rates, or due to sufficiently many individuals having contracted the infection, hence having been removed from the pool of susceptibles. Needless to say, such long term projections are purely hypothetical since they neglect any possible reactions to the evolving situation. Specifically, it should be expected that significantly falling case numbers induce contact restrictions to be relaxed further, while rising case numbers might lead to new interventions. The precise numbers predicted by the simulations are not our main concern here. We rather want to emphasize the sensitivity of predictions to the detection ratio, that is, the different behavior exhibited by the system under the assumption of low, medium, or high detection ratios.

The effect of different detection ratios is most striking in scenarios **A** (Fig. 1a) and **B** (Fig. 1b). At the time of writing this, after having relaxed some of the measures, the reproduction number, $\mathcal{R}$, in
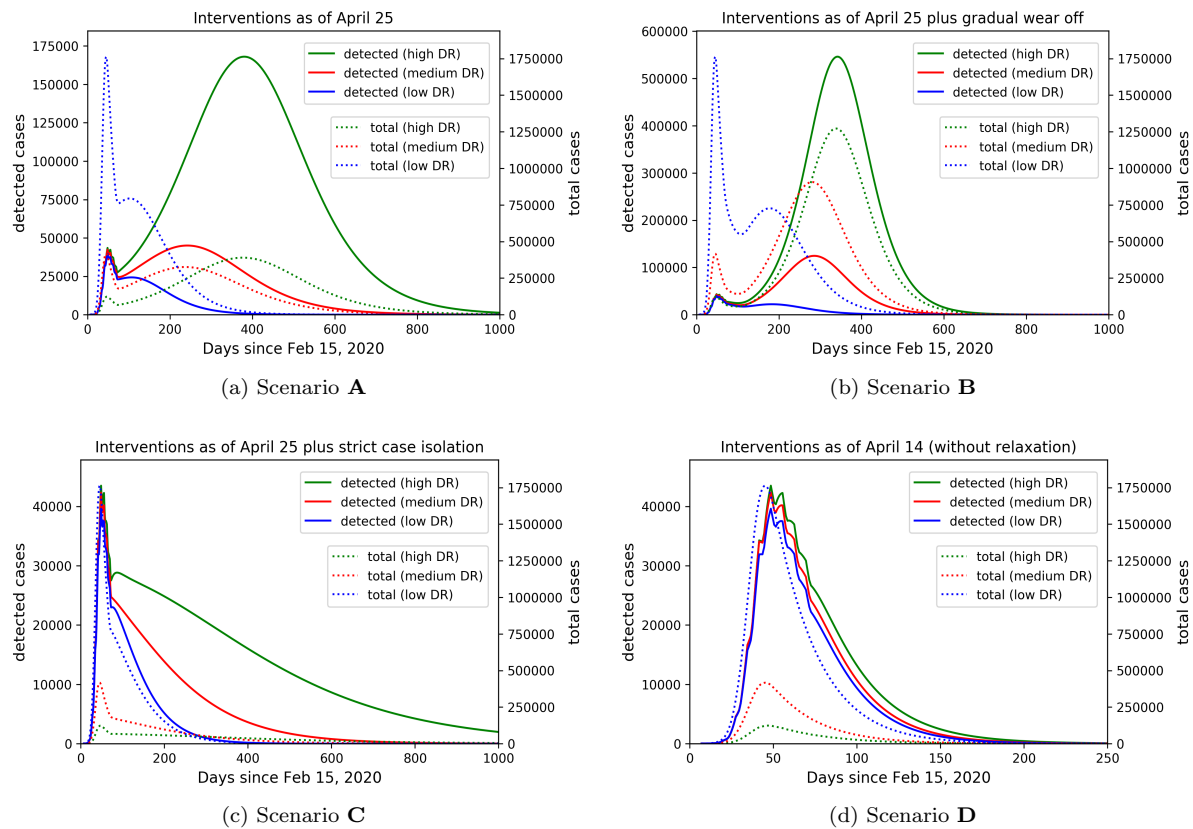
Figure 1: Model simulations for *detected* (solid) and *total* (dotted) active cases over time for different scenarios. Notice the different scaling for detected (left axis) vs. total (right axis) cases. For each scenario, the results for high ($DR \approx 40\%$), medium ($DR \approx 10\%$), and low ($DR \approx 2.5\%$) detection ratio are shown in green, red, and blue, respectively.

Germany is close to one [23], and the system is very sensitive to the proportion of susceptibles among the population. A high detection ratio implies that only a very limited number of infections remained undetected, and given the current number of detected cases, most individuals are still susceptible. In contrast, a low detection ratio suggests that a significant number of infections has been going on unobserved, and there would already be a significant number of recovered, hence immune, individuals. This can make the difference between $\mathcal{R} > 1$, leading to a second peak, or $\mathcal{R} < 1$ and the epidemic subsiding. In both other scenarios (**C** and **D**, shown in Fig. 1c and Fig. 1d, respectively) the reproduction number is overall smaller than in **A** and **B** for all assumed detection ratios. We therefore primarily observe a quantitative difference. Again, in these scenarios, assuming a low detection ratio means that more susceptibles could have already turned into recovered individuals, making $\mathcal{R}$ even smaller, and therefore leading to the epidemic subsiding faster. The smallest effect of variations in the detection ratio is observed in scenario **D**. In this scenario very low effective contact rates were maintained over time as of mid April. This fact reduces $\mathcal{R}$ to such small values that the lower number of susceptibles in the low DR-case does not make a significant difference.

Noteworthy is the effect of different detection ratios on the projected fatality number over the course of the epidemic. In scenarios **A** and **B**, the assumption of high or medium detection ratios lets us predict a second peak of the epidemic, over the course of which many more fatalities are to be expected. These would not be predicted if a low detection ratio is assumed (Fig. 2). The situation is aggravated by the fact that a low detection ratio means that the case fatality rate (percentage of fatalities among detected cases) is considerably higher than the actual lethality (or infection fatality rate, cf. [25]) of the disease. Projecting observed case fatality rates into the future therefore produces overestimations of total fatality numbers. The overestimation becomes more pronounced the lower the detection ratio.
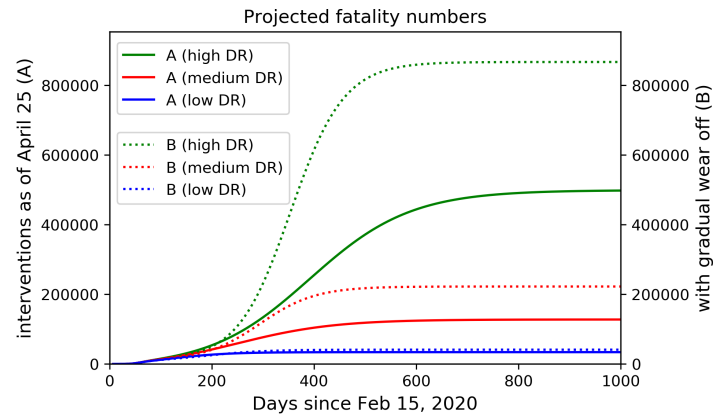
Figure 2: Projected fatality numbers under scenarios **A** (solid) and **B** (dotted) for different assumptions on the detection ratio.

Conclusion

While enormous efforts are undertaken all over the globe to stop the COVID-19 pandemic and limit its consequences, more and more studies indicate that a large portion of cases could have remained undocumented [14]. Here we showed how the knowledge of the detection ratio of COVID-19 infections is of crucial importance for model-based predictions on the further course of the outbreak and its control. This emphasizes the urgent need for screening representative samples of the population in order to determine the prevalence of antibodies against SARS-CoV-2. Studies like those reported in [26] for Germany, in [4] and in [12] for California, or in [16] for New York are promising first steps but more widespread screening with selection processes minimizing biases are necessary to obtain better estimates of past detection ratios. Continuous extensive testing of individuals for virus RNA may further help to uncover temporal changes of detection ratios. It remains to be determined whether wide range screening would also help limiting the spread of the disease [11]. We conclude by noting that the difficulties in predicting an outbreak outcome are not limited to COVID-19 but pertain to any novel infectious diseases, making it even more important to not forget this lesson, even after the COVID-19 pandemics will have been resolved.

References

[1] M. an der Heiden and O. Hamouda *Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland - Nowcasting* (2020) DOI 10.25646/6692.2

[2] M. V. Barbarossa et al. *A first study on the impact of current and future control measures on the spread of COVID-19 in Germany - March 28, 2020* medRxiv (2020) DOI: doi.org/10.1101/2020.04.08.20056630

[3] M. V. Barbarossa et al. *The impact of current and future control measures on the spread of COVID-19 in Germany - April 18, 2020* medRxiv (2020) DOI: doi.org/10.1101/2020.04.18.20069955

[4] E. Bendavid et al.*COVID-19 Antibody Seroprevalence in Santa Clara County, California. medRχiv* (2020) DOI: 10.1101/2020.04.14.20062463

[5] S. Chen *Mitigating COVID-19 outbreak via high testing capacity and strong transmission-intervention in the United States* (2020) *medRχiv* DOI: 10.1101/2020.04.03.20052720

[6] J. Dehning et al. *Inferring COVID-19 spreading rates and potential change points for case number forecasts* arXiv:2004.01105 (April 2020)

[7] EpiCentro *Epidemia COVID-19 - Aggiornamento nazionale del 22 Aprile 2020.* Istituto Superiore di Sanitá, Rom, Italy 2020. Available online: `https://www.epicentro.iss.it/coronavirus/bollettino/Infografica_22aprile%20ITA.pdf`

[8] X. He et al. *Temporal dynamics in viral shedding and transmissibility of COVID-19.* Nat Med (2020). DOI:10.1038/s41591-020-0869-5

[9] T. Hotz et al. *Monitoring the spread of COVID-19 by estimating reproduction numbers over time* 2020. Available online: `https://stochastik-tu-ilmenau.github.io/COVID-19/` (last accessed: April 24, 2020)

[10] N. Ferguson et al. *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand.* Imperial College London, March 16, 2020. DOI: 10.25561/77482

[11] N. C. Grassly et al. *Role of testing in COIVD-19 control.* Imperial College London, April 23, 2020 `https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-04-23-COVID19-Report-16.pdf` (accessed: April 23, 2020)

[12] Los Angeles County Public Health Communications *USC-LA County Study: Early Results of Antibody Testing Suggest Number of COVID-19 Infections Far Exceeds Number of Confirmed Cases in Los Angeles County.* `http://publichealth.lacounty.gov/phcommon/public/media/mediapubhpdetail.cfm?prid=2328` (accessed: April 25, 2020)

[13] S. Khailaie and T. Mitra et al. *Estimate of the development of the epidemic reproduction number Rt from Coronavirus SARS-CoV-2 case data and implications for political measures based on prognostics* medRχiv DOI: 10.1101/2020.04.04.20053637 (2020).

[14] R. Li et al. *Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)* Science (16 March 2020) DOI: 10.1126/science.abb3221

[15] Z. Liu et al. *Predicting the cumulative number of cases for theCOVID-19 epidemic in China from early data* arχiv 2002.12298 (2020)

[16] New York State, Official Press Release by the Governor from April 23, 2020. `https://www.governor.ny.gov/news/amid-ongoing-covid-19-pandemic-governor-cuomo-announces\-state-health-department-will-partner` (accessed: April 25, 2020)

[17] G. Ogris and C. Hofinger *COVID-19 Prevalence (Media information)* (April 10, 2020) `www.sora.at/uploads/media/Austria_COVID-19_Prevalence_BMBWF_SORA_20200410_EN_Version_fuer_HP.pdf`

[18] M. Park et al. *A systematic review of COVID-19 epidemiology based on current evidence* J. Clin. Med. 2020, 9(4), 967

[19] J. Riou and C. L. Althaus *Pattern of early human-to-human transmission of Wuhan 2019-nCoV.* bioRχiv **2020**, DOI:10.1101/2020.01.23.917351.

[20] E. Raimundez et al. *COVID-19 outbreak in Wuhan demonstratesthe limitations of publicly available casenumbers for epidemiological modelling* (2020) medRχiv, DOI:10.1101/2020.04.19.2007159

[21] C. Rothe et al. *Transmission of 2019-nCoV infection from an asymptomatic contact in Germany* N. Engl. J. Med. 2020 DOI:10.1056/NEJMc2001468

[22] RKI *Table with the current COVID-19 infections per day (time series) from the Robert Koch-Institute.* Robert Koch-Institute, Berlin, Germany 2020. Available online: `https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0/` (last accessed on April 25, 2020)

[23] RKI *Täglicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 (COVID-19). 25-04-2020 - Aktualisierter Stand für Deutschland.* Robert Koch-Institute, Berlin, Germany 2020. Available online: `https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-04-25-de.pdf?__blob=publicationFile`

[24] M. Shen et al. *Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China.* bioRχiv (2020) DOI: 10.1101/2020.01.23.916726v1

[25] P. Simon *Robust Estimation of Infection Fatality Rates during the Early Phase of a Pandemic.* medRχiv (2020) DOI: 10.1101/2020.04.08.20057729

[26] H. Streeck et al. *Vorläufiges Ergebnis und Schlussfolgerungen der COVID-19 Case-Cluster-Study (Gemeinde Gangelt)* Universitätsklinikum Bonn, April 9, 2020 `https://www.land.nrw/sites/default/files/asset/document/zwischenergebnis_covid19_case_study_gangelt_0.pdf`

[27] WHO Team *Modes of transmission of virus causing COVID-19:implications for IPC precaution recommendations* World Health Organisation - Department of Communication 2020. Available online: `https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc~\-precaution~-recommendations` (last accessed on April 25, 2020)

[28] J. Wu et al. *Ontarians need to rapidly increase their personal protection and testing to mitigate the COVID-19 spread in the province* (April 6, 2020). Available at SSRN `https://ssrn.com/abstract=3569731`

[29] Y. Yang et al. *Evaluating the accuracy of different respiratory specimens in thelaboratory diagnosis and monitoring the viral shedding of2019-nCoV infections. medRχiv* (2020) DOI: 10.1101/2020.02.11.20021493