Compliance with NPIs and possible deleterious effects on mitigation of an epidemic outbreak

Maria Vittoria Barbarossa [a], Jan Fuhrmann [b, *]

[a] Frankfurt Institute for Advanced Studies, Frankfurt, Germany
[b] Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany

ARTICLE INFO

ABSTRACT

The first attempt to control and mitigate an epidemic outbreak caused by a previously unknown virus occurs primarily via non-pharmaceutical interventions (NPIs). In case of the SARS-CoV-2 virus, which since the early days of 2020 caused the COVID-19 pandemic, NPIs aimed at reducing transmission-enabling contacts between individuals. The effectiveness of contact reduction measures directly correlates with the number of individuals adhering to such measures. Here, we illustrate by means of a very simple compartmental model how partial noncompliance with NPIs can prevent these from stopping the spread of an epidemic.

1. Introduction

Faced with an epidemic outbreak caused by a previously unknown virus, effective medication or vaccines are usually not available. Mitigation therefore tends to primarily rely on non-pharmaceutical interventions (NPIs). This was no different in case of the SARS-CoV-2 virus, which has caused the COVID-19 pandemic. Most countries applied NPIs aimed at reducing contacts between infectious and susceptible individuals. Such measures range from social distancing or school closure to most severe lockdown periods. Reducing most contacts between individuals, NPIs necessarily reduce epidemiologically relevant contacts, or *effective contacts*, viz., those between infectious and susceptible individuals during which the virus is successfully transmitted. The effectiveness of contact reduction measures directly correlates with the number of individuals actually adhering to the measures(Acuña-Zegarra et al., 2020). Is it better to have stricter measures followed by a small fraction of the population or almost universally adopted moderate measures? There is an obvious trade-off. Stricter measures are more effective in reducing contacts between compliant individuals but are also less likely to be realistically applicable to many people. Here, we illustrate by means of a very simple mathematical model how partial noncompliance with NPIs can prevent them from stopping the spread of an epidemic. Though the working example is the COVID-19 epidemic, specifically in Germany in late summer/early fall of 2020, the model is generally valid and flexible to be applied to other infectious diseases.

2. Methods

The core of the model used in this note extends the known *S-E-I-R* (susceptibles−exposed−infected−recovered) model for disease dynamics (Brauer et al., 2019). The ordinary differential equations (ODEs) approach that we use assumes that the population is homogeneous and well-mixed within a region. Individuals are classified according to their status with respect to the virus spread in the community. Susceptible individuals (*S*) can be infected. The time between exposure to the virus (becoming infected) and symptom onset, commonly known as "exposed phase" or incubation period is divided in three stages ($E_j, j = 1, 2, 3$), the last being presymptomatic and contagious. Besides allowing for the presymptomatic infectious period to be fixed as the last stage of the incubation period, splitting the exposed phase into three consecutive stages results in its duration being gamma distributed which is more realistic than the exponential distribution corresponding to a single compartment.

Infections might be reported (*I*) or remain undetected (*U*). The compartment *I* also accounts for severe infections, which might lead to death, the assumption being that all severe cases will be detected. Deceased (*D*) and recovered (*R*) individuals are removed from the chain of transmission, assuming long lasting immunity upon recovery. Susceptible individuals can be infected via contacts with presymptomatic (transmission rate β_E), undetected (transmission rate β_U), or detected (β_I) infectious individuals. We assume that presymptomatic and undetected infectious persons, lacking knowledge about being infectious, do not restrict their contacts to others, and therefore have higher transmission rates than detected infected individuals ($\beta_E, \beta_U > \beta_I$) who are expected to quarantine or isolate themselves at least to some degree. Further we include behavioral heterogeneity in the population. We assume that while everyone adheres to moderate restrictions being in place throughout the period under consideration, a certain fraction of the population might not comply with stricter measures as these are applied. Hence, we split the population into two groups, called compliant (subscript *c*) and noncompliant (subscript *n*), respectively. As a simplifying assumption we take the compliant group to perfectly adhere to prescribed contact reductions while the noncompliant group maintains its original contact level, regardless of imposed measures, be it because they do not accept such measures or because they are not able to implement them.

An overview of the model variables is given in Table 1. The dynamics of the model shown in Fig. 1 is given by the following system of differential equations:

$$
\begin{aligned}
\dot{S}_m &= -\lambda_m(t)S_m \\
\dot{E}_{1,m} &= \lambda_m(t)S_m - \gamma_E E_{1,m} \\
\dot{E}_{i,m} &= \gamma_E E_{i-1,m} - \gamma_E E_{i,m} \qquad , i \in \{2, 3\} \\
\dot{U}_m &= (1 - \tau_m)\gamma_E E_{3,m} - (\gamma_U + \eta_m)U_m \\
\dot{I}_m &= \tau_m \gamma_E E_{3,m} + \eta_m U_m - \gamma_I I_m \\
\dot{R}_m &= (1 - \delta)\gamma_I I_m + \gamma_U U_m \\
\dot{D}_m &= \delta \gamma_I I_m,
\end{aligned}
\tag{1}
$$

for $m, k \in \{c, n\}$, and with

$$
\lambda_m = \sum_{k=c,n} \left(\beta_{km,E}E_{3,k} + \beta_{km,U}U_k + \beta_{km,I}I_k \right)
\tag{2}
$$

The force of infection, λ_m, is determined by the specific transmission rates

$$
\beta_{km,X} = \beta_0 \mu_X a_k s_m
\tag{3}
$$

between the infectious compartment $X \in \{E_{3,k}, I_k, U_k | k \in \{n, c\}\}$, and the susceptible class $S_m, m \in \{c, n\}$. Here β_0 denotes a basic transmissibility rate specific to the virus, a_k is the specific infectiousness of population k (determined by their social behavior, in particular compliance with restriction rules), s_m the specific susceptibility of the susceptibles S_m (determined by their behavior), and μ_X is the specific weight of infectiousness for stage X of the infection. Disease progression through the different infectious stages is given by the rates γ_X, that is, $1/\gamma_X$ is the average duration of the stage X. The incubation time $3/\gamma_E$ is split equally among the three compartments E_j. Detection may occur with probability τ_m by the end of the incubation

Table 1
Model variables.

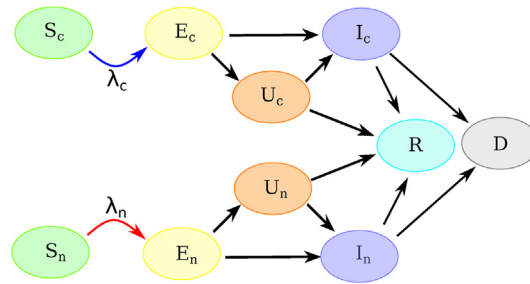| Notation	Description
$S_{c/n}$	compliant/noncompliant susceptible individuals
$E_{i,c/n}$	compliant/noncompliant exposed individuals in stage $i = 1, 2$ (not yet contagious)
$E_{3,c/n}$	compliant/noncompliant exposed individuals in stage 3 (already contagious)
$U_{c/n}$	compliant/noncompliant undetected infectious individuals
R	recovered individuals
D	deceased individuals

Fig. 1. Model structure for the transmission dynamics of an infectious disease with contact restrictions and partial compliance. Black arrows indicate transition from one compartment to another, red/blue arrows indicate new infections by virus transmission due to contact with infectious individuals. Upon infection, susceptible (S) individuals enter the exposed phase (E), divided into three consecutively passed stages, E_1, E_2, E_3, but represented here as a single stage for better clarity. After symptom onset, infections may be detected (I) or remain undetected (U). Severe cases potentially leading to death are assumed to be always detected. Infected individuals who recovered (R) or died (D) upon infections, are removed from the chain of transmission. All individuals who are relevant to the disease transmission dynamics are classified as compliant (S_c, E_c, U_c, I_c) or noncompliant (S_n, E_n, U_n, I_n), depending on their behavior response to imposed contact restrictions.

period. Later detection, when individuals are already in compartment U_m, may depend on the compliant/noncompliant status, m, of the infectious person and occurs with rate η_m. Detected individuals might die with probability δ.

For the simulations shown below we set the total population N to approximately 83 million, roughly Germany's population. A list of values for the other model parameters can be found in Appendix D.

2.1. The basic reproduction number $\mathcal{R}_0$ for the pre-intervention phase

As in the pre-intervention phase both the compliant and the noncompliant group are assumed to behave the same, for easiness of notation we omit the compliant/noncompliant index in the following computation. To employ the next generation matrix (NGM) approach for calculation of the basic reproduction number $\mathcal{R}_0$ (Diekmann et al., 1990), we split the compartments into infected ($x = (E_1, E_2, E_3, U, I)^T$) and non-infected ($y = (S,R,D)^T$) and write system (1) as

$$\begin{aligned} x' &= \mathcal{F}(x,y) - \mathcal{V}(x,y) \\ y' &= g(x,y) \end{aligned}$$

where $\mathcal{F}$ captures the inflow of new individuals into x from the non-infected compartments,

$$\mathcal{F}(x,y) = (\lambda S, 0, 0, 0, 0)^T,$$

while $\mathcal{V}$ collects the progression within the infected compartments as well as the outflow (recovery, deaths). Linearizing the equation for x about the disease free equilibrium (DFE) $\bar{x} = 0, \bar{y} = (N,0,0)$, we obtain

$$x' \approx Fx - Vx$$

where

$$V = \begin{pmatrix} \gamma_E & 0 & 0 & 0 & 0 \\ -\gamma_E & \gamma_E & 0 & 0 & 0 \\ 0 & -\gamma_E & \gamma_E & 0 & 0 \\ 0 & 0 & -(1-\tau)\gamma_E & \gamma_U + \eta & 0 \\ 0 & 0 & -\tau\gamma_E & -\eta & \gamma_I \end{pmatrix}, \quad \text{and} \quad F = \begin{pmatrix} 0 & 0 & \beta_E & \beta_U & \beta_I \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

are the Jacobians of $\mathcal{F}$ and $\mathcal{V}$, respectively, evaluated at the DFE. For the DFE to be locally asymptotically stable, all eigenvalues of $F - V$ must lie in the left half plane, or equivalently, the dominant eigenvalue of FV^{-1} must be smaller than 1. The dominant eigenvalue being the first entry of FV^{-1}, corresponds indeed to the basic reproduction number $\mathcal{R}_0$ (Diekmann et al., 1990). Short computation leads to

$$\mathcal{R}_0 = \mathcal{R}_I + \mathcal{R}_U + \mathcal{R}_E,$$

with

$$\mathcal{R}_I = \frac{\beta_I((\eta + \gamma_U)\tau + \eta(1-\tau))}{(\eta + \gamma_U)\gamma_I} = \frac{\beta_I(\eta + \tau\gamma_U)}{(\eta + \gamma_U)\gamma_I}, \quad \mathcal{R}_U = \frac{\beta_U(1-\tau)}{\eta + \gamma_U}, \quad \mathcal{R}_E = \frac{\beta_E}{\gamma_E}. \tag{4}$$

Let us now go back to the distinction into compliant and noncompliant groups. We shall assume that strict control measures are introduced in a very early phase of the outbreak (when we are very close to the DFE) and denote by ρ the fraction

of individuals that comply with the measures. That is we introduce one infectious individual in an entirely susceptible population split into compliant, $S_c(0) = \rho N$, and noncompliant, $S_n(0) = (1 - \rho)N$, group. Compliance corresponds to a reduction of effective contacts to a fraction $r \in [0, 1]$ of the original value. With these notations the initial controlled reproduction number would be

$$\mathcal{R}_c = (1 - (1 - r)\rho)\mathcal{R}_0. \tag{5}$$

The straightforward but somewhat lengthy derivation is given in the appendix. In Fig. 6b we show the ratio between $\mathcal{R}_c$ and $\mathcal{R}_0$ in dependence of ρ and r. The trivial limit cases are (i) $r = 0$ and $\rho = 1$, that is full compliance and reduction to zero contacts, yielding $\mathcal{R}_c = 0$, and (ii) $r = 1$, that is no intervention, or $\rho = 0$, no compliance, yielding $\mathcal{R}_c = \mathcal{R}_0$.

3. Results

The simulations that we show below are not calibrated on any specific time series but parametrized in a way to approximately reproduce the COVID-19 dynamics in different phases of the pandemic. We start the simulations with initial low incidence under moderate control measures, such that the resulting reproduction number is slightly larger than one. For a certain initial period, both the compliant and the noncompliant group behave the same, that is, the two subpopulations have the same transmission rates ($\beta_{kmJ} = \beta_{lpJ}$, for all k, l, m, $p \in \{c, n\}$, $J \in \{E, U, I\}$). After this initial phase we assume that transmission rates, hence the reproduction number, slightly increase over time. In the context of COVID-19, this setting might mimic the transition from the controlled situation in the summer 2020 to the fall 2020 in Germany and other European countries. We suppose that stricter intervention measures aiming at the reduction of transmission are introduced when a daily incidence of $\approx 20,000$ cases is reached. As an effect of these control measures, contacts in the population should significantly decrease and, if the whole population was behaving in compliance with the prescribed measures, contact rates would be reduced to a certain fraction $r < 1$ of their value before intervention. Upon the introduction of stricter control measures the compliant and noncompliant groups start behaving differently: noncompliants maintain their behavior (activity and susceptibility). In other words, if the whole population was noncompliant, the reproduction number would remain the same as before intervention. By a slight abuse of notation we shall denote by $\mathcal{R}_0$ the reproduction number before the modeled intervention. Here we set $\mathcal{R}_0 \approx 1.5$, approximately the value estimated for COVID-19 in late summer/early fall 2020 in Germany (Robert Koch Institute Coronavirus Disease 2019 (COVID-19), 2019). This value being already affected by some control measures, it is indeed significantly smaller than the uncontrolled reproduction number of SARS-CoV-2, mostly estimated above 2 (Kucharski et al., 2020; Zhao et al., 2020).

3.1. Scenarios for different contact reduction and compliance levels

In the following we show different scenarios for the dynamics of the outbreak under the variation of two major unknown factors:

1. **Reduction of contacts.** We assume that in accordance with control measures contacts would be reduced *by* a factor 75%, 50%, or 20% of the level previous interventions. This would lead, in case of perfect compliance, to a reduction *to* 25%, 50%, or 80% of the reproduction number $\mathcal{R}_0$ before interventions, corresponding to $r = 0.25$, 0.5, or 0.8, respectively.

(a) Incidence for different compliance levels

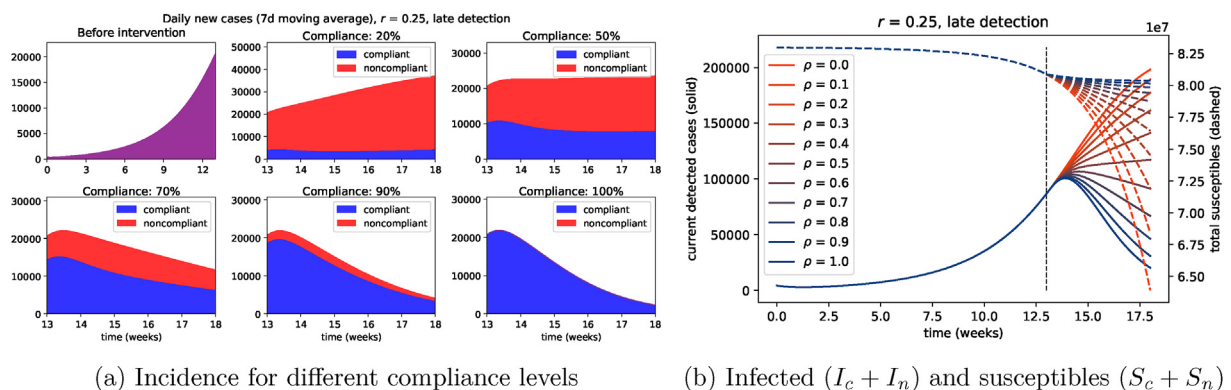(b) Infected ($I_c + I_n$) and susceptibles ($S_c + S_n$)

Fig. 2. Scenario 1: $\mathcal{R}_c = 0.25\mathcal{R}_0$. The upper left panel in (a) shows the incidence before the intervention. For different compliant fractions (ρ) of the total population we show (a) the evolution in time of daily cases reported among compliant (blue) and noncompliant (red) individuals, and (b) the currently known active cases ($I_c + I_n$, continuous curves) and susceptible individuals ($S_c + S_n$, dashed curves). The vertical dashed line shows the time of intervention. The effect of contact reduction is not immediately evident since new infections are not detected until several days later. Notice that the daily new cases in (a) are shown as 7-day moving average, meaning that weekly oscillations due to lower reporting rates on weekends are smoothed over and do not show in the plots.
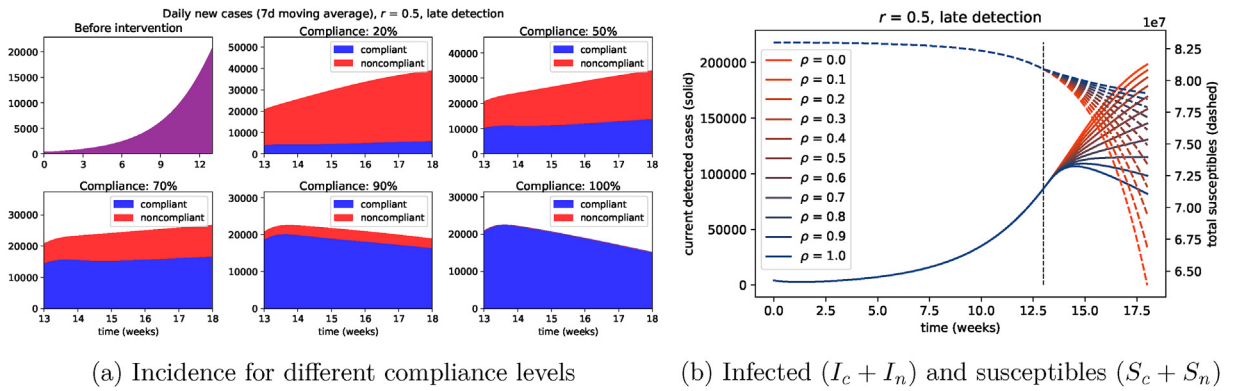
(a) Incidence for different compliance levels   (b) Infected $(I_c + I_n)$ and susceptibles $(S_c + S_n)$

Fig. 3. Scenario 2: $\mathcal{R}_c = 0.5\mathcal{R}_0$. The upper left panel in (a) shows the incidence before the intervention. For different compliant fractions (ρ) of the total population we show (a) the evolution in time of daily cases reported among compliant (blue) and noncompliant (red) individuals, and (b) the currently known active cases ($I_c + I_n$, continuous curves) and susceptible individuals ($S_c + S_n$, dashed curves). The vertical dashed line shows the time of intervention. The effect of contact reduction is not immediately evident since new infections are not detected until several days later.

2. **Compliant fraction of the population.** We vary the fraction $\rho \in [0, 1]$ of the population complying with restriction measures. We assume that the differentiation into compliant/noncompliant individuals occurs only once, namely at the time of intervention, and that individuals do not switch to the opposite behavior (noncompliant/compliant) for the entire course of the simulations.

It is yet unclear how many secondary cases of COVID-19 result from presymptomatic transmission, with estimates ranging from 6.4% (Wei et al., 2020) to 46% (Ferretti et al., 2020; He et al., 2020), or even above 50% (Ganyani et al., 2020). It is however widely accepted that almost half the infections are transmitted from pre- or asymptomatic infectious individuals (Buitrago-Garcia et al., 2020; Johansson et al., 2021). If we further assume that detected cases are well isolated and significantly reduce their contacts, these individuals will contribute only marginally to disease transmission. We therefore fix the relative infectiousness parameters (μ_X, $X \in \{E, U, I\}$) in the definition of $\beta_{km,X}$ such that

$$\mathcal{R}_U \gtrapprox \mathcal{R}_E \gg \mathcal{R}_I,$$

that is, individuals in U contribute slightly more to $\mathcal{R}_0$ than those in E_3, and the contribution of individuals in I is dwarfed by both of them. Given the other parameters listed in Appendix D, we achieve (i) $\mathcal{R}_I$ barely contributing by choosing $\mu_I = 0.1\mu_U$ and (ii) $\mathcal{R}_E$ accounting for a little less than 40% of $\mathcal{R}_0$ by fixing $\mu_E = 1.5\mu_U$.

For the following scenarios we furthermore consider two possible assumptions for the time of detection. We show as next the *late detection setting*, with $\tau = 0.1$ (corresponding to 10% of cases detected by symptom onset) and $\eta = 0.6\gamma_U$ ($\approx 38\%$ of U detected) at low prevalence. This leads to approximately 45% detection as long as the prevalence is sufficiently small. We

(a) Incidence for different compliance levels   (b) Infected $(I_c + I_n)$ and susceptibles $(S_c + S_n)$
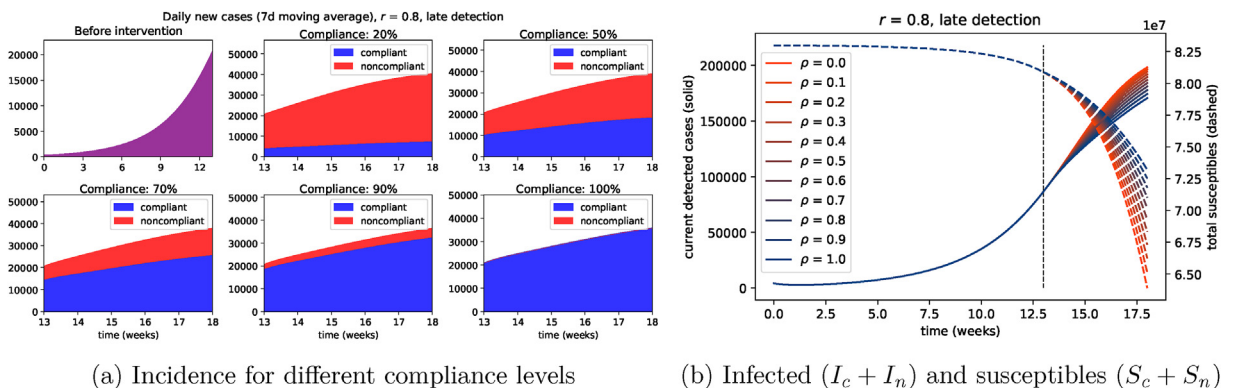
Fig. 4. Scenario 3: $\mathcal{R}_c = 0.8\mathcal{R}_0$. The upper left panel in (a) shows the incidence before the intervention. For different compliant fractions (ρ) of the total population we show (a) the evolution in time of daily cases reported among compliant (blue) and noncompliant (red) individuals, and (b) the currently known active cases ($I_c + I_n$, continuous curves) and susceptible individuals ($S_c + S_n$, dashed curves). The vertical dashed line shows the time of intervention. The limited effect of contact reduction is not stopping the increase in daily new cases and there is no qualitative difference between full compliance ($\rho = 1$) and full noncompliance ($\rho = 0$).
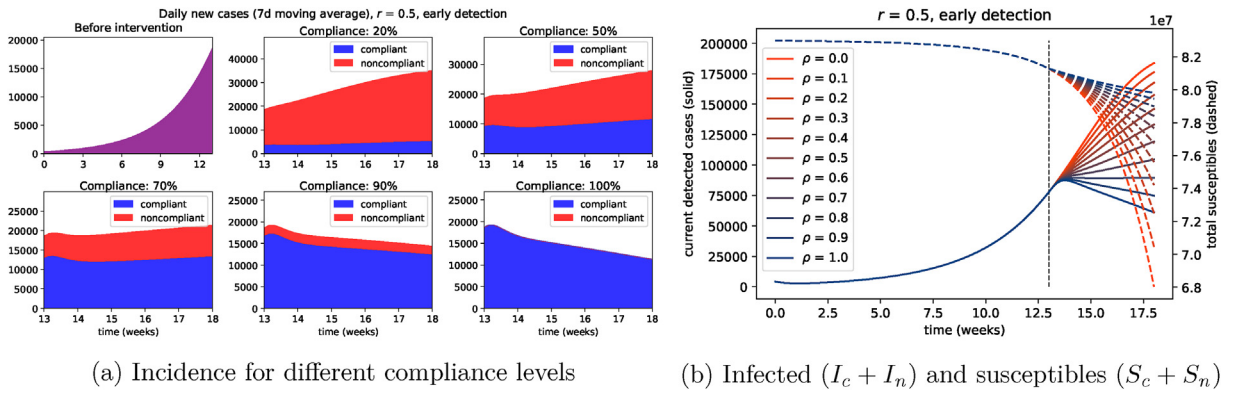
(a) Incidence for different compliance levels

(b) Infected $(I_c + I_n)$ and susceptibles $(S_c + S_n)$

Fig. 5. Scenario 2′: $\mathcal{R}_c = 0.5\mathcal{R}_0$. If infections are detected in an early stage, the case incidence follows closely the true incidence of new infections. The upper left panel in (a) shows the incidence before the intervention. For different compliant fractions (ρ) of the total population we show (a) the evolution in time of daily cases reported among compliant (blue) and noncompliant (red) individuals, and (b) the currently known active cases $(I_c + I_n$, continuous curves) and susceptible individuals $(S_c + S_n$, dashed curves). The vertical dashed line shows the time of intervention. Notice (panel (a)) the drop in incident cases at 70% compliance, followed by a slow increase. The reasons for this rebound will be discussed in Appendix C. Due to the shorter detection delay, the number of active detected cases drops (in case of high compliance) faster than in the corresponding settings in Fig. 3a.
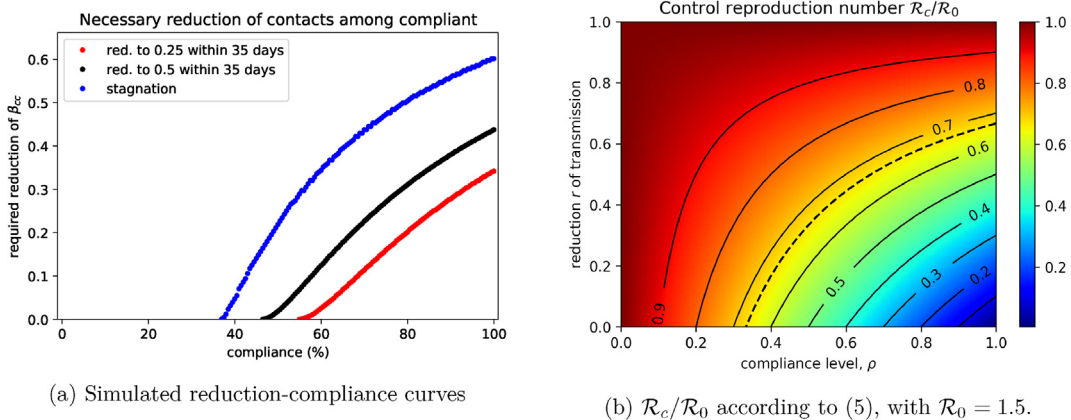

(a) Simulated reduction-compliance curves

(b) $\mathcal{R}_c/\mathcal{R}_0$ according to (5), with $\mathcal{R}_0 = 1.5$.

Fig. 6. (a) Required reduction of transmission among compliant individuals in order to achieve stagnation (blue) or a desired reduction within five weeks to 25% (red) or 50% (black) of case incidence compared to the value at the time of intervention. (b) Ratio between the control reproduction number, $\mathcal{R}_c$, and the pre-intervention reproduction number, $\mathcal{R}_0$, according to (5). Notice that reaching $\mathcal{R}_c = 1$ from $\mathcal{R}_0 = 1.5$ implies a ratio of 2/3 and that the corresponding level set (dashed line) is indeed very similar to the curve of required reductions for achieving stagnating incidence in (a). This level set being shifted to the left reflects the fact that the simulations do not operate at the DFE.

further assume that the detection rate decreases in the face of large numbers of (undetected) cases due to finite testing capacity, e.g.,

$$\eta = \bar{\eta} \frac{K}{U_c + U_n + \alpha E_{3,c} + \alpha E_{3,n} + K} \tag{6}$$

for some constant K and a maximal detection rate $\bar{\eta}$. As long as the total number of undetected infectious individuals is small as compared to K, the correction factor is close to one, $\eta \approx \bar{\eta}$, but as the prevalence approaches the order of K, an increasing proportion of infections goes undetected. A similar discussion applies to τ which is similarly composed of $\bar{\tau} = 0.1$ and a similar correction factor as derived for η. We remark that this nonlinearity does not affect the calculation of $\mathcal{R}_0$ or $\mathcal{R}_c$ but may well affect the reproduction number R_t far away from the disease free equilibrium. This makes intuitive sense: Since detected infectious individuals are assumed to (self-) isolate and produce few secondary infections after being detected, any reduction in detection rates will accelerate the spread of the epidemic while at the same time making reported case numbers appear smaller. A motivation for the precise shape of the correction factor is given in the appendix. An *early detection setting*, with $\tau = 0.3$ (corresponding to 30% of cases detected by symptom onset), $\bar{\eta} = 0.25\gamma_U$ (corresponding to 20% of U detected), yielding a comparable total detection ratio, was also considered, but simulations are shown or discussed only when differences with the late detection setting are significant.

3.1.1. Scenario 1: $\mathcal{R}_c = 0.25\mathcal{R}_0$ at full compliance

The first scenario considered is a prescribed contact reduction that would lead to a control reproduction number $\mathcal{R}_c$ being at 25% of $\mathcal{R}_0$. Assuming an initial $\mathcal{R}_0 \approx 1.5$, this would lead to $\mathcal{R}_c \approx 0.375$. The results are shown in Fig. 2. We observe that at full compliance (100% of the population, $\rho = 1$), the incidence would indeed quickly decrease. For lower levels of compliance, the decrease is expectedly slower. However, if only half the population adheres to the measures (50% compliance), this contact reduction is only sufficient to stall the rising incidence. This happens because the *effective reproduction number R_t*, resulting from transmission rates of compliant and noncompliant populations under the implemented control measures, approaches values close to 1 about two weeks after intervention. Fig. 2a evidences that the 50% noncompliant individuals make up way more than 50% of the cases. At even lower compliance, the measures may help to slow down the increase but are no longer sufficient to stop it.

In Fig. 2b, the red curve ($\rho = 0$) indicates the course of the epidemic with no compliance at all with the NPIs introduced in week 13. This corresponds to a situation without any new intervention measures, and five weeks after the (non-)intervention, already some 25% of the population would have been infected (cf. the dashed curve for the susceptible population). The number of detected cases does not rise as quickly as the loss of susceptibles would suggest. This is due to the limited test capacity, cf. (6), and the decreasing detection ratio as the prevalence becomes too large.

3.1.2. Scenario 2: $\mathcal{R}_c = 0.5\mathcal{R}_0$ at full compliance

The second scenario, shown in Fig. 3, assumes that at 100% compliance, the transmission rates and hence the control reproduction number would be cut in half, leading to $\mathcal{R}_c \approx 0.75$ at $\mathcal{R}_0 = 1.5$. Still, at perfect compliance, the incidence would start falling several days after the intervention but now even 70% compliance would not be sufficient to prevent the case numbers from rising.

3.1.3. Scenario 3: $\mathcal{R}_c = 0.8\mathcal{R}_0$ at full compliance

The implementation of moderate measures, reducing the transmission rate among compliant individuals by only 20%, leads to a reproduction number of $\mathcal{R}_c \approx 1.2$ at perfect compliance. As shown in Fig. 4, this is not sufficient to stop the increasing case numbers even if the whole population would adhere to the measures. The effect is due to the limited efficacy of the control measures (reducing the reproduction number from 1.5 to at best 1.2), rather than to the level of compliance.

3.1.4. Scenario 2′: $\mathcal{R}_c = 0.5\mathcal{R}_0$ at full compliance, with early detection

The assumptions on the reproduction numbers are the same as in Scenario 2, but here we consider the case of earlier detection ($\tau = 0.3$, $\eta = 0.25\gamma_U$). Qualitatively, the results shown in Fig. 5 are the same as in Scenario 2 (cf. Fig. 3), that is, for compliance levels below 50% the daily new cases keep increasing. Due to early case detection, however, the incidence of detected cases follows more closely the time course of "actual" new daily infections ($\gamma_E(E_{3,c} + E_{3,u})$). A consequence of this effect is that for 50% compliance, for which R_t is rather close to 1, after an initial drop the incidence increases again after a few days. The initial brief drop reflects transiently falling new infections. In case of late detection (cf. Fig. 2a, compliance 50%) this effect is smoothed out and not visible in the daily incidence time series.

3.2. Reduction levels necessary to either significantly reduce incidence or afford stagnation

So far, we have seen how the incidence of new detected cases progresses at different compliance levels if a given reduction of effective contacts among compliant individuals is prescribed. Let us now take a different point of view. For a given compliance level we ask how strong the reduction in transmission among compliant individuals needs to be in order to

(*i*) reduce the incidence to 25% or 50% of the value at intervention time within 5 weeks after intervention, or
(*ii*) reach permanent stagnation of incidence, meaning that after 3 weeks post intervention the incidence shall not increase anymore beyond a small tolerance of 0.25% over the last 10 days of the simulated period.

To this end, we performed simulations for 201 compliance levels (0% through 100% in steps of 0.5 percentage points) and screened for the reduction levels sufficient to achieve either of the goals (*i*) or (*ii*). Results are shown in Fig. 6. Since we start with a reproduction number $\mathcal{R}_0$ (pre-intervention) of about 1.5, hence rising incidence, it is not surprising that for low compliance even the complete elimination of contacts among compliant individuals ($\beta_{cc} \approx 0$) is not sufficient to achieve either one of the above goals. The population of susceptible noncompliant individuals is still sufficiently large and a significant reduction in the daily incidence is not feasible. At compliance levels close to 40% a very strong reduction in transmission among compliant individuals ($\beta_{cc} \approx 0$) allows to stall the rising incidence. In order to achieve the desired 50% reduction in daily cases, as stated in point (*i*), a compliance level of nearly 50% or more is necessary. Reducing the incidence to 25% within the same time requires even higher compliance levels. Conversely, even at full compliance, the contact reductions need to be sufficiently large in order to achieve stagnating or falling incidence. This is in agreement with what was already shown in Fig. 4, where a reduction of effective contacts by 20% was not sufficient to keep the incidence from rising further. Here, we see that reduction of effective contacts by a factor 0.4 to 0.6 of the pre-intervention level is required to either reach stagnation or

significant reduction of daily incidence. Qualitatively, the same conclusions are also obvious from consulting the reproduction numbers shown for comparison in Fig. 6b.

Simulated curves for the evolution of daily incidence in time corresponding to the limit cases (stagnation or reduction to a fixed percentage of pre-intervention value just being achieved) in Fig. 6a are shown in the appendix, Fig. A1a.

4. Conclusion and discussion

In order to illustrate possible effects of less than perfect compliance with non-pharmaceutical interventions (NPIs) on their effectiveness in curbing the spread of infectious diseases, we modeled and simulated a situation mimicking the status of the COVID-19 epidemic in Germany in the fall of 2020. The model captures both (i) reduced susceptibility of individuals adhering to the proposed NPIs (an effect similar to protection which could be achieved via vaccination) and (ii) reduced transmission from compliant individuals (acting similar to quarantine or treatment, cf. Chapter 9 in (Martcheva, 2015)). The simulations show that in implementing NPIs to rapidly reduce daily cases, a concurrence of a sufficient level of compliance (ρ) in the population and a significant reduction, r, of effective contacts among compliant individuals is required. For example, let the pre-intervention reproduction number be about 1.5, and let both the effective infectiousness, a_c, and the susceptibility, s_c, of half the population (compliance level $\rho = 0.5$) be reduced by about 30% each. This leads to a 51% reduction of transmission among compliant individuals, cf. (3), $r = 0.49$, which is by far not enough to stop the rise in new cases, as can be seen in the upper right panel in Fig. 3a. Only if the measures are sufficiently effective in reducing transmission *and* a large proportion of the population implements such measures, a stagnation or even reduction of case numbers can be achieved in a reasonably short time (in the above simulations: 5 weeks).

Moreover, if infections are detected and reported rather quickly (Scenario 2′, Fig. 5), reduced transmission among compliant individuals can lead to a brief reduction in newly reported cases before these start rising again (cf. Fig. 5 compliance 50% or 70%). If reported cases are used as a daily proxy for evaluating the effectiveness of control measures, such a short-time decline could be misleading.

The system (1) for transmission dynamics used for the above simulations was developed as a simplification of our previous models for COVID-19 in Germany (Barbarossa et al., 2020; Barbarossa & Fuhrmann, 2021). For the sake of simplicity we decided not to include age groups, stages of infection, hospitalizations or cases requiring intensive care, nor considered any reduction of contact rates due to self-control of individuals to high incidence values (Capasso & Serio, 1978; Fenichel et al., 2011). Further we assume that individuals are either compliant or noncompliant for the whole duration of control measures, and that there is no behavioral switching between the two groups. The model could be extended to include such a switch, as was done in the past by other authors (Poletti et al., 2009; D'Onofrio & Manfredi, 2020; Iboi et al., 2021).

We should note that all the simulations discussed above presume that most of the population is still susceptible at the time of intervention. This leads to the effective reproduction number R_t being only slightly smaller than the basic reproduction number $\mathcal{R}_0$ (before intervention) or the control reproduction number $\mathcal{R}_c$ (after intervention). This is one reason why the theoretical threshold curve shown in Fig. 6b is rather close to those found in Fig. 6a showing the required reduction of transmission among compliant individuals for given compliance levels. Clearly, the smaller the susceptible fraction of the population is at the beginning, the faster the relative change in the number of susceptibles over time and the more pronounced will the effect of depleting the pool of susceptibles be.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Calculation of the control reproduction number

For the calculation of the control reproduction number $\mathcal{R}_c$ we apply the next generation matrix (NGM) approach from (Diekmann et al., 1990). The population can be split into not infected compartments, $y = (S_c, S_n, R, D)^T$, and infected compartments $x = (E_{1,c}, E_{1,n}, E_{2,c}, E_{2,n}, E_{3,c}, E_{3,n}, U_c, U_n, I_c, I_n)^T$. The disease free equilibrium is

$$y = (\rho N, (1 - \rho)N, 0, 0)^T, \qquad x = 0$$

where ρ denotes the compliance level, i.e., the fraction of the population that is compliant. By the NGM approach we construct the 10×10 matrices

$$F = \beta_0 \begin{pmatrix} 0 & 0 & 0 & 0 & \mu_E a_c s_c \rho & \mu_E a_n s_c \rho & a_c s_c \rho & a_n s_c \rho & \mu_I a_c s_c \rho & \mu_I a_n s_c \rho \\ 0 & 0 & 0 & 0 & \mu_E a_c s_n \hat\rho & \mu_E a_n s_n \hat\rho & a_c s_n \hat\rho & a_n s_n \hat\rho & \mu_I a_c s_n \hat\rho & \mu_I a_n s_n \hat\rho \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where we wrote the transmission rates as $\beta_{mk,X} = \beta_0 \mu_X a_m s_k$ and abbreviated $\hat\rho = (1 - \rho)$, and

$$V = \begin{pmatrix} \gamma_E & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_E & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\gamma_E & 0 & \gamma_E & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\gamma_E & 0 & \gamma_E & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\gamma_E & 0 & \gamma_E & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\gamma_E & 0 & \gamma_E & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -(1-\tau_c)\gamma_E & 0 & \gamma_U + \eta_c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -(1-\tau_n)\gamma_E & 0 & \gamma_U + \eta_n & 0 & 0 \\ 0 & 0 & 0 & 0 & -\tau_c \gamma_E & 0 & -\eta_c & 0 & \gamma_I & 0 \\ 0 & 0 & 0 & 0 & 0 & -\tau_n \gamma_E & 0 & -\eta_n & 0 & \gamma_I \end{pmatrix}$$

In what follows, we assume $\tau_c = \tau_n = : \tau$ and $\eta_c = \eta_n = : \eta$, meaning that the chance of detection is independent of the behavior.[1] Then, the inverse of V is

$$V^{-1} = \begin{pmatrix} \frac{1}{\gamma_E} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\gamma_E} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{\gamma_E} & 0 & \frac{1}{\gamma_E} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\gamma_E} & 0 & \frac{1}{\gamma_E} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{\gamma_E} & 0 & \frac{1}{\gamma_E} & 0 & \frac{1}{\gamma_E} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\gamma_E} & 0 & \frac{1}{\gamma_E} & 0 & \frac{1}{\gamma_E} & 0 & 0 & 0 & 0 \\ \frac{1-\tau}{\gamma_U+\eta} & 0 & \frac{1-\tau}{\gamma_U+\eta} & 0 & \frac{1-\tau}{\gamma_U+\eta} & 0 & \frac{1}{\gamma_U+\eta} & 0 & 0 & 0 \\ 0 & \frac{1-\tau}{\gamma_U+\eta} & 0 & \frac{1-\tau}{\gamma_U+\eta} & 0 & \frac{1-\tau}{\gamma_U+\eta} & 0 & \frac{1}{\gamma_U+\eta} & 0 & 0 \\ \frac{\tau\gamma_U+\eta}{(\gamma_U+\eta)\gamma_I} & 0 & \frac{\tau\gamma_U+\eta}{(\gamma_U+\eta)\gamma_I} & 0 & \frac{\tau\gamma_U+\eta}{(\gamma_U+\eta)\gamma_I} & 0 & \frac{\eta}{(\gamma_U+\eta)\gamma_I} & 0 & \frac{1}{\gamma_I} & 0 \\ 0 & \frac{\tau\gamma_U+\eta}{(\gamma_U+\eta)\gamma_I} & 0 & \frac{\tau\gamma_U+\eta}{(\gamma_U+\eta)\gamma_I} & 0 & \frac{\tau\gamma_U+\eta}{(\gamma_U+\eta)\gamma_I} & 0 & \frac{\eta}{(\gamma_U+\eta)\gamma_I} & 0 & \frac{1}{\gamma_I} \end{pmatrix}.$$

The next generation matrix is hence

---

[1] This assumption may be motivated as follows. Although it may be expected that individuals who deliberately ignore the proposed contact reduction measures will be less likely to have themselves tested upon minor suspicions of being infected, the opposite may be true for those who cannot easily reduce their contacts due to their profession (e.g. medical doctors or caregivers). Individuals in this category could be tested periodically, making detection of asymptomatic or pauci-symptomatic infections more likely.

$$FV^{-1} = \beta_0 \begin{pmatrix} \left(\dfrac{\mu_I(\tau\gamma_U + \eta)}{(\gamma_U + \eta)\gamma_I} + \dfrac{1 - \tau}{\gamma_U + \eta} + \dfrac{\mu_E}{\gamma_E}\right)a_c s_c \rho & \left(\dfrac{\mu_I(\tau\gamma_U + \eta)}{(\gamma_U + \eta)\gamma_I} + \dfrac{1 - \tau}{\gamma_U + \eta} + \dfrac{\mu_E}{\gamma_E}\right)a_n s_c \rho & *...* \\ \left(\dfrac{\mu_I(\tau\gamma_U + \eta)}{(\gamma_U + \eta)\gamma_I} + \dfrac{1 - \tau}{\gamma_U + \eta} + \dfrac{\mu_E}{\gamma_E}\right)a_c s_n (1 - \rho) & \left(\dfrac{\mu_I(\tau\gamma_U + \eta)}{(\gamma_U + \eta)\gamma_I} + \dfrac{1 - \tau}{\gamma_U + \eta} + \dfrac{\mu_E}{\gamma_E}\right)a_n s_n (1 - \rho) & *...* \\ 0 & 0 & 0...0 \\ \vdots & \vdots & \vdots \end{pmatrix}$$

$$= \underbrace{\beta_0 \left(\dfrac{\mu_I(\tau\gamma_U + \eta)}{(\gamma_U + \eta)\gamma_I} + \dfrac{1 - \tau}{\gamma_U + \eta} + \dfrac{\mu_E}{\gamma_E}\right)}_{\mathcal{R}_0} \begin{pmatrix} a_c s_c \rho & a_n s_c \rho & *...* \\ a_c s_n (1 - \rho) & a_n s_n (1 - \rho) & *...* \\ 0 & 0 & 0...0 \\ \vdots & \vdots & \vdots \end{pmatrix}$$

That the common factor is $\mathcal{R}_0$ can be noted by recalling that $\beta_X = \mu_X \beta_0 as$ in the pre-intervention setting. Since we are interested in the dominant eigenvalue of this matrix and all but the first two rows are filled with zeros, we only need to calculate the eigenvalues of the upper 2×2 minor

$$\begin{pmatrix} a_c s_c \rho & a_n s_c \rho \\ a_c s_n (1 - \rho) & a_n s_n (1 - \rho) \end{pmatrix}$$

This leads to

$$\mathcal{R}_c = \mathcal{R}_0 (a_c s_c \rho + a_n s_n (1 - \rho)),$$

which is remarkably independent of the detection parameters, meaning that the threshold for achieving $\mathcal{R}_c = 1$ given any $\mathcal{R}_0 > 1$ should be the same for early and late detection, respectively. The ratio between $\mathcal{R}_c$ and $\mathcal{R}_0$ in dependence of the level of compliance ρ and the fraction $r = a_c s_c$ to which transmission rates among compliant individuals are reduced after intervention is shown in Fig. 6b.

Notice that this approach allows calculating the control reproduction number for the case that noncompliant individuals slightly change their behavior but to a lesser degree than compliant individuals ($a_c < a_n < 1$ and/or $s_c < s_n < 1$). Having assumed, however, that noncompliant individuals keep their pre-intervention behavior, we require that $a_n s_n = 1$. Moreover, since we assumed the reproduction number to be reduced by a factor $r < 1$ if the compliance were 100% (meaning $\rho = 1$),[2] we also require $r = a_c s_c$.

The simple formula for the control reproduction number is therefore as given in (5),

$$\mathcal{R}_c = (r\rho + 1 - \rho)\mathcal{R}_0 = (1 - (1 - r)\rho)\mathcal{R}_0.$$

The controlled reproduction number can be hence obtained reducing $\mathcal{R}_0$ by a fraction $(1 - r)\rho$. The latter corresponds to the practical efficacy of control measures, being the product of the fraction effective contacts to be reduced by the measures and the compliance level.

If we drop the assumption on homogeneous detection (that is we allow $\tau_c \neq \tau_n$ and $\eta_c \neq \eta_n$), then the next generation matrix is calculated in a completely similar fashion:

$$FV^{-1} = \begin{pmatrix} A_c a_c s_c \rho & A_n a_n s_c \rho & *...* \\ A_c a_c s_n (1 - \rho) & A_n a_n s_n (1 - \rho) & *...* \\ 0 & 0 & 0...0 \\ \vdots & \vdots & \vdots \end{pmatrix},$$

with the abbreviation

$$A_m = \frac{\mu_I(\tau_m \gamma_U + \eta_m)}{(\gamma_U + \eta_m)\gamma_I} + \frac{1 - \tau_m}{\gamma_U + \eta_m} + \frac{\mu_E}{\gamma_E}, \qquad m \in \{c, n\}.$$

The eigenvalues are again 0 and

---

[2] More precisely, we assume the reduction of transmission to result from (i) infectious compliant individuals spreading the infection to a lesser degree, encoded in $a_c < 1$, and (ii) susceptible infectious individuals being more cautious, hence less susceptible, encoded by $s_c < 1$. Both effects compound to reduced transmission between susceptible individuals by a factor $r = a_c s_c$.

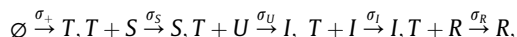$$\mathcal{R}_c = \beta_0(A_c a_c s_c \rho - A_n a_n s_n(1-\rho)).$$

Assuming again the behavior of the noncompliant population to be unaffected by the intervention, we obtain $\beta_0 A_n = \mathcal{R}_0$ and $a_n s_n = 1$, leading to

$$\mathcal{R}_c = \left(\frac{A_c}{A_n} r\rho - (1-\rho)\right)\mathcal{R}_0.$$

The ratio $\frac{A_c}{A_n}$ depends on the particular values of the parameters involved, but the important message is that $\mathcal{R}_c$ does not depend on whether the reduction in transmission is mediated by infectious individuals being cautious to not spread the virus or by susceptible individuals being cautious not to catch the virus. This will clearly be different once we are sufficiently far into the epidemic, i.e., sufficiently far away from the DFE. Since it should be expected that noncompliant individuals make up a larger share of the infectious population than their share in the total population is, while compliant individuals will more likely remain susceptible, the reduction a_c in spreading activity will play a less significant role compared to the reduction s_c in susceptibility.

Appendix B. Derivation of the correction factor of detection rates

To motivate the shape of the correction factor for the detection rate, η, given in (6) we consider a minimal model for an infectious disease with underascertainment. To this end, we track *S*usceptible, *U*ndetected infectious, detected *I*nfectious, and *R*emoved/recovered individuals. Moreover, we consider the population of available *T*ests.[3] These are supplied at a given rate σ_+ and are depleted by being applied to individuals who are tested. Clearly, not all tested individuals are infectious, hence tests are governed by the following quasi reactions:

$$\varnothing \xrightarrow{\sigma_+} T, T + S \xrightarrow{\sigma_S} S, T + U \xrightarrow{\sigma_U} I, \ T + I \xrightarrow{\sigma_I} I, T + R \xrightarrow{\sigma_R} R,$$

the first reaction describing the constant supply and the other reactions describing the pairing of a test with an individual (the person being tested). The test results allows only undetected individuals to change their state (from U to I), whereas leaves all other compartments unchanged. The resulting system of equations reads

$$
\begin{aligned}
\dot{T} &= \sigma_+ - (\sigma_S S + \sigma_U U + \sigma_I I + \sigma_R R)T \\
\dot{S} &= -(\beta_U U + \beta_I I)S \\
\dot{U} &= (\beta_U U + \beta_I I)S - \gamma_U U - \sigma_U UT \\
\dot{I} &= \sigma_U UT - \gamma_I I \\
\dot{R} &= \gamma_U U + \gamma_I I.
\end{aligned}
$$

It now makes sense to choose σ_+ and the σ_X sufficiently large to ensure that T is small. This reflects the fact that testing capacity cannot be stored since the limiting factor is laboratory time. Increasing testing capacity would result in larger σ_+ and would allow for larger σ_X, that is, a higher coverage, but should not result in large residual values of T.

This motivates the idea of viewing the test population as a fast variable, setting it to a quasi steady state

$$\dot{T} \approx 0 \Rightarrow T \approx \frac{\sigma_+}{\sigma_S S + \sigma_U U + \sigma_I I + \sigma_R R}$$

and writing the only relevant testing term as

$$\sigma_U UT \approx \sigma_U \frac{\sigma_+}{\sigma_S S + \sigma_U U + \sigma_I I + \sigma_R R} U = \frac{\sigma_+}{K} \frac{K}{K + U} U$$

where we put

$$K = \frac{\sigma_S}{\sigma_U} S + \frac{\sigma_I}{\sigma_U} I + \frac{\sigma_R}{\sigma_U} R.$$

With the notation

$$\bar{\eta} := \frac{\sigma_+}{K}$$

---

[3] More precisely: *T* corresponds to the capacity for administering and evaluating tests.

we recover an analogue of relation (6). There are several assumptions at play in this discussion. First, we assume that σ_U is significantly larger than σ_S, σ_I, and σ_R, meaning that for any given individual the chance of being tested is much larger if this individual is indeed infected. This chance may be influenced by the individual showing symptoms or having been in close contact to a confirmed case. While the derivation above remains valid for σ_U being similar to σ_S, σ_I, and σ_R and even if all σ_X were the same, which would be the case for random screening, the saturation effect would become less relevant. Indeed, if $\sigma_U = \sigma_S = \sigma_I = \sigma_R$ then $K = S + I + R$ and $\sigma_U U T \approx \sigma_+ U/N$ with the constant population size N. Moreover, for small σ_U, an unrealistically high number of tests would be required to reach a meaningful detection ratio at small prevalence. Considering again the case of random screening, to achieve a given detection ratio, this very fraction of individuals would have to be tested over the average infectious period which is about one week. For the sake of the argument, we may even assume $\sigma_I = 0$ since an already detected individual need not be detected again. We also assume the term K to be approximately constant. This is true if either the susceptible population is much larger than all other populations and σ_S is not much smaller than σ_R, or if the rates σ_S and σ_R are approximately the same and $S + R \gg I$. Both cases are plausible: If the detection ratio is high, most cases are detected and there are few recovered individuals who were not detected. Given the official COVID-19 data for Germany in the fall of 2020 (Robert Koch Institute Coronavirus Disease 2019 (COVID-19), 2019), this would mean the large majority of the population is still susceptible. If on the other hand, the detection ratio is low, there may be many recovered individuals who were never detected, but from the point of view of a test, these individuals are in no way different from susceptibles, and it should be expected that $\sigma_S \approx \sigma_R$.

In fact, simulating the above system (results not shown here) for $\sigma_S = \sigma_I = \sigma_R$ and, say, $\sigma_U = 20\sigma_S$, reveals that for most parameters, there is barely a difference between this system and the slow system

$$
\begin{aligned}
\dot{S} &= -(\beta_U U + \beta_I I)S \\
\dot{U} &= (\beta_U U + \beta_I I)S - \gamma_U U - \eta U \\
\dot{I} &= \eta U - \gamma_I I \\
\dot{R} &= \gamma_U U + \gamma_I I,
\end{aligned}
$$

with

$$
\eta = \bar{\eta} \, \frac{K}{K + U},
$$

and K assumed to be constant. In contrast, the same system with constant $\eta = \bar{\eta}$ shows significantly higher reported case numbers as soon as the prevalence increases. Moreover, if β_I is significantly smaller than β_U (as one would expect if detected cases are quarantined) the model with constant η predicts a slower progression of the disease than both the model with test population and the one with U-dependent detection rate.

Appendix C. Daily incidence in limit cases (cf. Fig. 6)

In Fig. 6a, we show the reductions of transmission required to achieve the goals of

(*i*) reducing within five weeks the incidence to 25% or 50% of the value at intervention, or
(*ii*) preventing the incidence from significantly rising at any time within weeks 2 through 5 after intervention,

introduced in subsection 3.1. Here we show the actual course of the incidence if contacts are reduced by these exact threshold levels. The corresponding curves are shown in Fig. A1 for selected prescribed compliance levels and under the assumptions of both late and early detection.

As already obvious from Fig. 6a for late detection (the curves for early detection being very similar), neither goal (*i*) or (*ii*) can be achieved at 30% compliance, and even without any transmission from or to compliant individuals (dashed curves, upper left panels in Fig. A1(a)-(b)) the incidence will rise after possibly briefly falling. At 40% compliance, stagnating incidence levels can be achieved but transmission among compliant individuals must be very small (blue curves, upper right panels in Fig. A1(a)–(b)). As compliance rises to 50%, a reduction of the incidence to half the value at intervention is possible within five weeks, and finally, if at least 60% of the population adhere to the contact reductions, either goal can be achieved, though clearly at different contact reductions. For compliance levels higher than 60% the incidence curves achieved by the threshold transmission reductions do not change significantly. This means that there is no big difference in the outcome upon very high compliance with moderately effective reduction measures, and moderate compliance with very strong reductions in transmission.
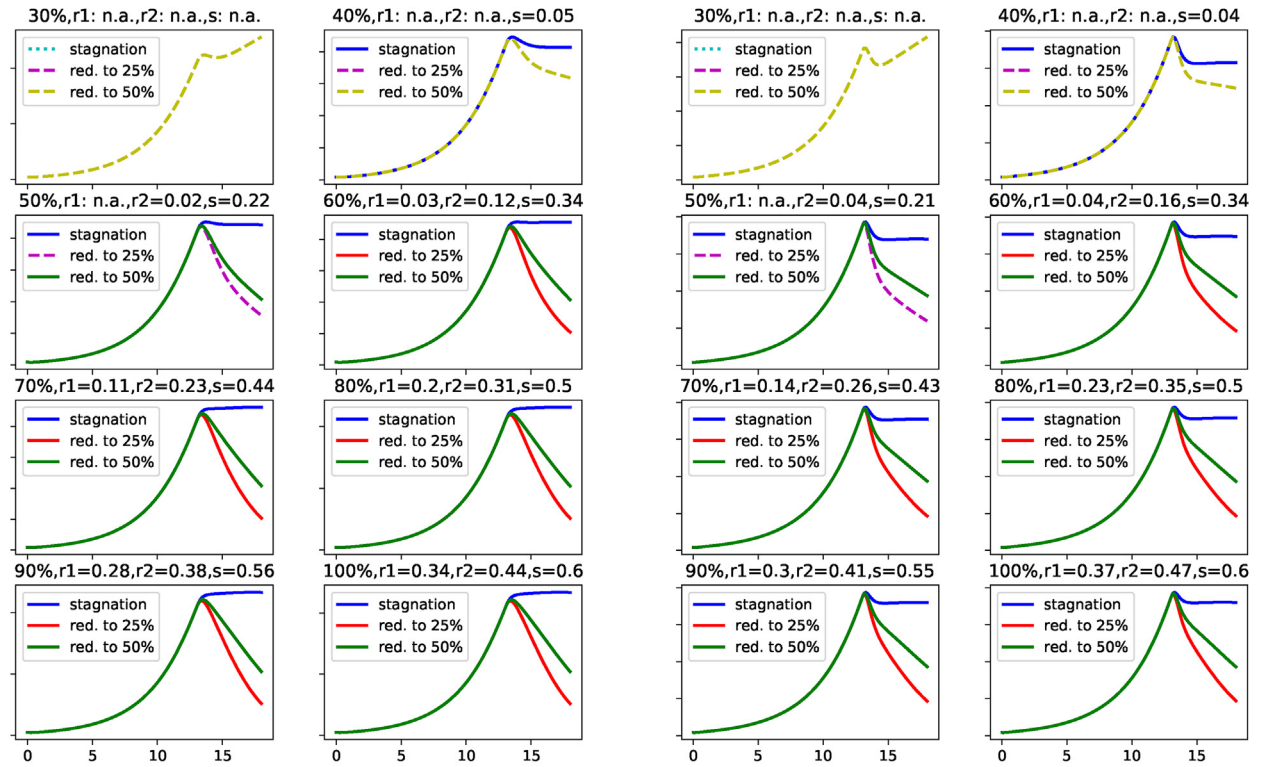
(a) Late detection

(b) Early detection

Fig. A1. Daily incidence for threshold cases in Fig. 6a. For different compliance levels (indicated as percentage on the top of each panel), the incidence over time is shown for the critical reductions according to Fig. 6a. The values $r1$ and $r2$ indicate the reduction of the transmission rate among compliant individuals required for lowering the incidence to 25% or 50%, respectively, of the value at the time of intervention, whereas s is the reduction factor required for stagnation as explained in Subsection 3.2. If a certain goal cannot be attained (denoted by "n.a.") at the given compliance level, the incidence for the $r = 0$, meaning that compliant individuals do not participate in the transmission at all, is shown as a dashed line.

Another effect, already alluded to in the discussion of scenario $2'$, becomes apparent from Fig. A1. In the case of late detection, the threshold contact reduction that is necessary for stagnation results in the incidence leveling out close to the maximum value. The latter is reached when the full effect of the intervention starts showing in the reported case numbers. Quite differently, for the case of early detection, the same threshold value leads to a significant decline in newly detected cases before a plateau well below the maximum incidence is reached. This can potentially lead to a deceptive situation where the rapidly falling incidence may make the contact reductions to appear way more effective than they are, and consequently reduce the sense of urgency among the population. If this in turn leads to a lower level of compliance, the incidence instead of falling further or stabilizing at a level slightly below the maximum may soon start rising again. However, as will become apparent from the following discussion, the incidence may start rising after a brief decrease even in the absence of any further changes of the transmission rates.

This effect is not caused by reporting delays or underascertainment but is inherent to the transmission dynamics as is apparent from the fact that the true incidence of new infections shows the same behavior. For the sake of simplicity, we shall illustrate here the argument on the basis of the SIR model, neglecting both the latent period and possible under ascertainment,

$$
\begin{aligned}
\dot{S}_c &= -(\beta_{cc}I_c + \beta_{nc}I_n)S_c \\
\dot{S}_n &= -(\beta_{cn}I_c + \beta_{nn}I_n)S_n \\
\dot{I}_c &= (\beta_{cc}I_c + \beta_{nc}I_n)S_c - \gamma I_c \\
\dot{I}_n &= (\beta_{cn}I_c + \beta_{nn}I_n)S_n - \gamma I_n,
\end{aligned}
\tag{A1}
$$

with the removed compartments, $R_m = N_m - S_m - I_m$, $m \in \{c, n\}$.

In keeping with our previous notation, we assume a fixed transmission parameter β_0 that is constant among all pairs (I_m, S_k), and describes transmission before the intervention. After interventions, the transmission parameters are again given by

$\beta_{mk} = a_m s_k \beta_0$, and we recall that $a_n = s_n = 1$. Instead of calculating next generation matrices we now linearize system (A1) about the disease free equilibrium $(\rho N, (1 - \rho)N, 0, 0)$, ρ again denoting the compliant fraction of the population. This leads to

$$\frac{d}{dt}\begin{pmatrix} S_c \\ S_n \\ I_c \\ I_n \end{pmatrix} \approx \begin{pmatrix} 0 & 0 & -\beta_0 a_c s_c \rho N & -\beta_0 s_c \rho N \\ 0 & 0 & -\beta_0 a_c (1 - \rho)N & -\beta_0 (1 - \rho)N \\ 0 & 0 & \beta_0 a_c s_c \rho N - \gamma & \beta_0 s_c \rho N \\ 0 & 0 & \beta_0 a_c (1 - \rho)N & \beta_0 (1 - \rho)N - \gamma \end{pmatrix} \begin{pmatrix} S_c \\ S_n \\ I_c \\ I_n \end{pmatrix}$$

The constant population size allows us to abbreviate $\beta = \beta_0 N$, and we only have to investigate the lower right submatrix

$$J = \begin{pmatrix} a_c s_c \rho \beta - \gamma & s_c \rho \beta \\ a_c (1 - \rho)\beta & (1 - \rho)\beta - \gamma \end{pmatrix}$$

Its eigenvalues are

$$\lambda_1 = \beta(a_c s_c \rho + (1 - \rho)) - \gamma \qquad \text{and} \qquad \lambda_2 = -\gamma < \lambda_1,$$

and obviously the sign of λ_1 determines the stability of the DFE. There is nothing new here, the same information was obtained by calculating the dominant eigenvalue of the next generation matrix that would have been

$$\mathcal{R}_c = \frac{\beta(a_c s_c \rho + (1 - \rho))}{\gamma}.$$

However, linearizing about the DFE allows us to investigate the corresponding eigenvectors

$$\xi_1 = (s_c \rho, 1 - \rho)^T \qquad \text{and} \qquad \xi_2 = (1, -a_c)^T$$

of J, which tell us how the infections have to be distributed among compliant and noncompliant individuals in order to create the maximal number of secondary infections.

Before intervention, there is no difference between compliant and noncompliant individuals and $a_c = s_c = 1$. Having assumed $\mathcal{R}_0 > 1$ means that the leading eigenvalue before intervention is $\lambda_1^0 > 0$, hence the system will have evolved to a situation where $(I_c, I_n)^T$ is proportional to $\xi_1^0 = (\rho, 1 - \rho)^T$. At the time of intervention, the system is still close to the DFE, $S_c + S_n$ still making up the vast majority of the population. Let us now assume that the control reproduction number after intervention is $\mathcal{R}_c \gtrsim 1$, meaning that the controlled eigenvalue $\lambda_1^c \gtrsim 0$ is still positive but small. However, the distribution of infectious individuals is no longer aligned with the corresponding eigenvector ξ_1^c but still with ξ_1^0. Hence, the component of the solution aligned with ξ_1^c will slowly increase at rate λ_1^c while the component aligned with ξ_2^c will decrease at rate $\lambda_2 = -\gamma$. If λ_1^c is sufficiently small, this leads to an overall decrease of the number of infectious individuals until the solution is sufficiently aligned to ξ_1^c and the component proportional to ξ_2^c becomes negligible.
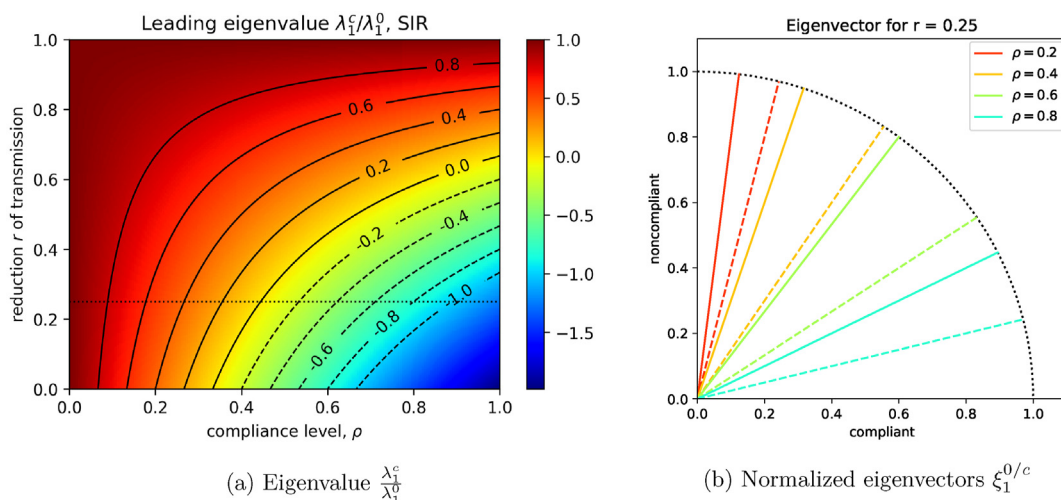

(a) Eigenvalue $\frac{\lambda_1^c}{\lambda_1^0}$ (b) Normalized eigenvectors $\xi_1^{0/c}$

Fig. A2. Illustration of the leading eigenvalue λ_1^c (left) of J as multiple of λ_1^0, depending on the transmission reduction factor, r, and the compliance level, ρ. Notice that the level $\lambda_1^c = 0$ corresponds to $\mathcal{R}_c = 1$ for system (A1). On the right, the corresponding eigenvector, ξ_1^c, in the I_c–I_n plane is shown for $r = 0.25$ and $a_c = s_c = \sqrt{r} = 0.5$, corresponding to dotted line in (a), and different compliance levels. For comparison, the dashed lines depict the eigenvectors ξ_1^0 at the same compliance levels. It is obvious that the eigenvectors ξ_1^c are shifted towards I_n. Notice that $\rho = 0.4$ approximately corresponds to $\lambda_1^c = 0$ at $r = 0.25$. The colors of the eigenvectors correspond to those of the point in (a) where the pair $(\rho, r = 0.25)$ is located.

In order to illustrate this effect in a numerical example, let us work with our previous assumption that $\mathcal{R}_0 = 1.5$, meaning in this simplified setting that $\beta = 1.5\gamma$ or $\lambda_1^0 = \beta - \gamma = 0.5\gamma$. This allows us to write

$$\lambda_1 = \gamma\left(\frac{3}{2}(a_c s_c \rho + (1 - \rho)) - 1\right) = \frac{\gamma}{2}(1 - 3(1 - a_c s_c)\rho) = \lambda_1^0(1 - 3(1 - r)\rho),$$

where $r = a_c s_c$. Just as we calculated the control reproduction number as a multiple of $\mathcal{R}_0$, we can now calculate λ_1^c as a multiple of λ_1^0 by a factor depending solely on the transmission reduction among compliant individuals, r, and the compliant fraction ρ of the population. It should not come as a surprise that Fig. A2a looks rather similar to Fig. 6b.

As the eigenvector ξ_1^c depends on s_c rather than r as a whole, we need to make a choice on how to distribute the contribution to r among reduced effective infectiousness, a_c, and reduced effective susceptibility, s_c, of compliant individuals. For the sake of showing the eigenvector in Fig. A2b, we choose $a_c = s_c = \sqrt{r}$, meaning that both effects contribute equally. Doing so allows us to illustrate how strongly ξ_1^c is shifted towards I_n as compared to ξ_1^0. We have chosen to fix $r = 0.25$ for figure A2b, and it should be obvious that the shift is stronger the smaller r, and hence s_c, is. Indeed, Fig. A3 illustrates how the solution vector shifts toward the new eigenvector after the intervention.

In fact, the observed phenomenon is not exclusively due to partial noncompliance. We may consider the classical SIR model ($\dot{S} = -\beta IS$, $\dot{I} = \beta IS - \gamma I$) with reproduction number $\mathcal{R}_0 = \frac{\beta}{\gamma} > 1$, and assume that at some point in time the contact rate β is reduced to $r\beta$ with $0 < r < 1$. At the time of intervention, the incidence of new infections will instantaneously drop by the reduction factor r. However, let if the contact reduction is so weak that $\mathcal{R}_c = r\frac{\beta}{\gamma} > 1$, then the incidence will immediately start rising again. Adding a latency period and explicitly modeling detection will smooth this process but an intermittently decreasing incidence followed by another rise may still be visible. This reasoning shows that there is no need to assume a reaction of the population to falling incidence values in order to observe rebounding incidence numbers. Conversely, these arguments should also caution against relaxing contact restrictions too quickly as this could exacerbate the rebound effect by increasing $\mathcal{R}_c$.
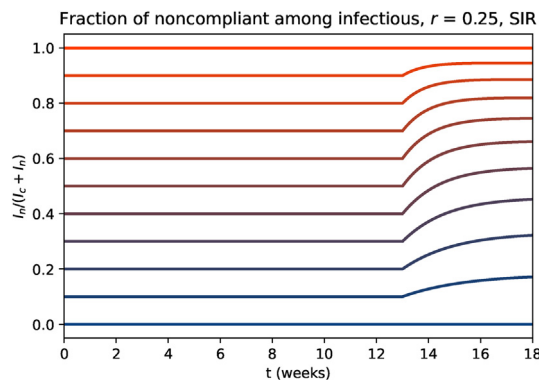

Fig. A3. Before the intervention (week 13), the fraction of noncompliant among all currently infectious individuals is precisely the same as their fraction among the whole population. After intervention, the share of noncompliant infectious individuals rises in accordance with the shifted eigenvector illustrated in Fig. A2b.

Appendix D. Model parameters

In the following list we summarize the model parameters relevant for the transmission dynamics of (1) and their values used for the simulations.

N: total population (assumed to be constant in good approximation)83, 000, 000approximately the population of Germany as of 2020

β_0: global transmission rate0.225 d^{-1}/N directly before the intervention; slightly smaller at the start of the simulated period as indicated at the beginning of Section 3fixed to achieve $\mathcal{R}_0 \approx 1.5$ given the other parameters

μ_U: relative transmission rate for undetected infectious individuals 1 can always be fixed as 1 by appropriately choosing the common factor β_0 which is exactly what we did

μ_E: relative transmission rate for presymptomatic infectious individuals1.5 assures that $\mathcal{R}_E$ contributes 35%–40% to $\mathcal{R}_0$

μ_I: relative transmission rate for detected infectious individuals 0.1 represents significantly decreased transmission due to (self) isolation

γ_E: three times inverse of mean duration of incubation period 3/(5 d)corresponds to mean incubation period of five days

γ_U: inverse of mean duration of symptomatic period1/(7 d)

γ_I: inverse of mean duration of infectious period after detection 1/(7 d) same as γ_U in order to ensure that detection does not affect mean infectious period

$\bar{\tau}$: maximal fraction of infectious individuals detected at symptom onset0.1

$\bar{\eta}$: maximal detection rate from symptomatic stage U $0.6\gamma_U$ chosen such that approximately 45% of infected individuals will be detected at low prevalence

K: prevalence of *undetected* infectious individuals at which the detection rates decrease to half their maximal value 500,000

References

Acuña-Zegarra, M. A., Santana-Cibrian, M., & Velasco-Hernandez, J. X. (2020). Modeling behavioral change and COVID-19 containment in Mexico: A trade-off between lockdown and compliance. *Mathematical Biosciences, 325*, 108370.

Barbarossa, M. V., & Fuhrmann, J. (2021). Germany's next shutdown - possible scenarios and outcomes. *Influenza Other Respi Viruses, 15*, 326—330. doi.org/10.1111/irv.12827.

Barbarossa, M. V., Fuhrmann, J., Meinke, J. H., Krieg, S., Varma, H. V., Castelletti, N., & Lippert, Th. (2020). Modeling the spread of COVID-19 in Germany: Early assessment and possible scenarios. *PLoS One, 15*(9), Article e0238559.

Brauer, F., Castillo-Chavez, C., & Feng, Z. (2019). *Mathematical models in epidemiology*. Springer.

Buitrago-Garcia, D., Egli-Gany, D., Counotte, M. J., Hossmann, S., Imeri, H., Ipekci, A. M., … Low, N. (2020). Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS Medicine, 17*(9), Article e1003346.

Capasso, V., & Serio, G. (1978). A generalization of the Kermack-McKendrick deterministic epidemic model. *Mathematical Biosciences, 42*, 43—61.

Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J. (1990). On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology, 28*(4), 365—382.

D'Onofrio, A., & Manfredi, P. (2020). The interplay between voluntary vaccination and reduction of risky behavior: A general behavior-implicit SIR model for vaccine preventable infections. In *Current trends in dynamical systems in biology and natural sciences* (pp. 185—203). Springer.

Fenichel, E. P., Castillo-Chavez, C., Ceddia, M. G., Chowell, G., Gonzalez Parra, P. A., Hickling, G. J., … Villalobos, C. (2011). Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences of the United States of America, 108*(15), 6306—6311.

Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, K., … Fraser, C. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science, 368*(6491), Article eabb6936.

Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J., & Hens, N. (2020). Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Euro Surveillance, 25*(17), Article pii=2000257.

He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., … Leung, G. M. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine, 26*(5), 672—675.

Iboi, E., Richardson, A., Ruffin, R., Ingram, D., Clark, J., Hawkins, J., … Akinyemi, L. (2021). Impact of public health education program on the novel coronavirus outbreak in the United States. *Frontiers of Public Health, 9*, 630974. https://doi.org/10.3389/fpubh.2021.630974

Johansson, M. A., Quandelacy, M. A., Kada, S., Prasad, P. V., Steele, M., Brooks, J. T., & Butler, J. C. (2021). SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA Network Open, 4*(1), Article e2035057.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., & Eggo, R. M. (2020). Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *The Lancet Infectious Diseases, 20*(5), 553—558.

Martcheva, M. (2015). *An introduction to mathematical epidemiology*. Springer.

Poletti, P., Caprile, B., Ajelli, M., Pugliese, A., & Merler, S. (2009). Spontaneous behavioural changes in response to epidemics. *Journal of Theoretical Biology, 260*(1), 31—40.

Robert Koch Institute Coronavirus Disease 2019 (COVID-19). Daily situation report of the robert Koch Institute. Updated status for Germany 10/31/2020. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Okt_2020/2020-10-31-en.pdf?__blob=publicationFile.

Robert Koch Institute Coronavirus Disease 2019 (COVID-19). Daily situation report of the robert Koch Institute. Updated status for Germany 11/10/2020. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Nov_2020/2020-11-10-en.pdf?__blob=publicationFile.

Wei, W. E., Li, Z., Chiew, C. J., Yong, S. E., Toh, M. P., & Lee, V. J. (2020). Presymptomatic transmission of SARS-CoV-2—Singapore, january 23—March 16, 2020. *MMWR Morb Mortal Wkly Rep, 69*(14), 411—415. https://doi.org/10.15585/mmwr.mm6914e1

Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., … Wang, M. H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International of Journal of Infectious Diseases, 92*, 214—217.