medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Shut and re-open: the role of schools in the spread of COVID-19 in Europe

Helena B. Stage[1,2,†,*], Joseph Shingleton[3,†,*], Sanmitra Ghosh[4,†], Francesca Scarabel[1,2,5,6], Lorenzo Pellis[1,2,7], Thomas Finnie[3]

1 Department of Mathematics, The University of Manchester, UK
2 Joint UNIversities Pandemic and Epidemiological Research, `https://maths.org/juniper/`
3 Emergency Response Department, Public Health England, UK
4 MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
5 Laboratory of Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto Ontario, Canada
6 CDLab - Computational Dynamics Laboratory, Department of Mathematics, Computer Science and Physics, University of Udine, Italy
7 The Alan Turing Institute, London, UK

*** joseph.shingleton@phe.gov.uk, helena.stage@manchester.ac.uk**
† These authors contributed equally to the production of this manuscript.

Abstract

We investigate the effect of school closure and subsequent reopening on the transmission of COVID-19, by considering Denmark, Norway, Sweden, and German states as case studies. By comparing the growth rates in daily hospitalisations or confirmed cases under different interventions, we provide evidence that school closures contribute to a reduction in the growth rate approximately 7 days after implementation. Limited school attendance, such as older students sitting exams or the partial return of younger year groups, does not appear to significantly affect community transmission. In countries where community transmission is generally low, such as Denmark or Norway, a large-scale reopening of schools while controlling or suppressing the epidemic appears feasible. However, school reopening can contribute to statistically significant increases in the growth rate in countries like Germany, where community transmission is relatively high. In all regions, a combination of low classroom occupancy and robust test-and-trace measures were in place. Our findings underscore the need for a cautious evaluation of reopening strategies.

Keywords: COVID-19, school closure, school reopening, non-pharmaceutical interventions

Introduction

Throughout the course of the ongoing COVID-19 pandemic, the role of young people and children in transmission has been a question of particular concern [1, 2]. This question is not only motivated by the goal of protecting the younger generations; it is also known from other respiratory diseases that, because younger people tend to have more prolonged and physical contacts among themselves [3], they pose a greater risk of infection to each other as well as being likely to introduce the infection to their respective households and so can drive the epidemic [4, 5]. Consequently, school closure is often one of the first measures considered when non-pharmaceutical interventions are needed [6]. However, during the COVID-19 pandemic it has often been implemented concurrently with other measures, making it difficult to assess its individual impact [7, 8].
Many of the challenges inherent in quantifying the impact of closure remain when policy-makers subsequently turn to the reopening of schools. Reopening presents a myriad of further questions, such as the ages of those returning, the physical circumstances and timing of their return, and the necessary conditions which must be met on a community level before a return can be deemed safe enough. For new or emerging infections, answers to these questions require new efforts to establish an age-stratified understanding of the infection and transmission dynamics [9, 10].

Earlier work exists concerning the effectiveness of school closure as a means of controlling the spread of COVID-19, with mixed conclusions depending on the studied age group, country, and modelling assumptions [6–9, 11–13]. These sought to estimate the impact of school closure on nationwide transmission levels, be it for instance a reduction in the

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

peak number of cases or the timing thereof. The challenges and questions related to school reopening have also been addressed from a theoretical or modelling perspective of scenario planning [14–16], estimating the number of new or severely ill cases resulting from school reopening. These use varying assumptions of the underlying community transmission, and consider various scenarios of the ages and timing of students returning to school.
While such models are valuable means of quantifying the expected impact of measures without increasing the risk of exposing the wider community to infection, their inclusion of other interventions is necessarily limited. The times surrounding school closure and reopening have seen a myriad of other non-pharmaceutical interventions being implemented. Through social contact patterns, observed changes in community transmission are the result of non-trivial, underlying interactions between current interventions. We believe our work fills an important knowledge gap in the literature by addressing the context of school interventions alongside other measures, and the *de facto* impact of schools in a broader framework of non-pharmaceutical interventions.
School closure and reopening not only affect transmission occurring on school premises; they also affect (and are affected by) community transmission, transmission within households with young children, and wider measures taken to monitor and curb an outbreak. In addition to the well-being of children, school interventions also impact a nation's workforce via the time dedicated to childcare. It must be remembered that the observed effects of these interventions are a product of underlying testing, reporting, and isolation (or other physical or social distancing) measures.

The aim of this work is to carry out a comparative analysis of school interventions, making use of the diversity of available data streams, to serve as a complement to theoretical modelling efforts. Ours is a data-driven approach which seeks not to establish the individual role of schools interventions on outbreak management, but instead assesses their impact in the context of wider societal interventions. Specifically, we wish to examine roles in transmission played by a) the different age cohorts of students, b) the timing of the school interventions (closure and reopening), and c) the background or community incidence. We hope these results can serve as a series of lessons learned from nations which have already reopened schools.
For school closure, we address these questions by estimating the time between intervention and a response being observed in the recorded data, as well as the changes in the growth rate pre- and post-intervention. For school reopening, we track the growth rate in cases over the intervention timeline and search for correlations between these interventions and changes in the growth rate.

Methods

Data selection criteria

We consider four countries: Denmark, Germany, Norway, and Sweden due to their geographic proximity, demographic similarities, and the relative timing and scope of their interventions to allow for a better comparison. We distinguish between countries with medium-to-high (Germany and Sweden) and low (Denmark and Norway) levels of community transmission on the basis of daily COVID-19 cases, rather than cases per capita [17]. This is motivated by the feasibility of testing, tracing, and isolating cases, which need not scale with population size. At the time of school closure, we saw the following cumulative cases: 800 (Denmark), 6500 (Germany), 1100 (Norway), and 1400 (Sweden). By the time schools started reopening, the total cases had risen to: 7000 (Denmark), 158000 (Germany), 7200 (Norway), and 19400 (Sweden on the day German schools first reopened). Only German states with at least 50 cases at the point of school closure, and at least 10 days of non-zero daily cases prior to closures, have been selected for analysis.
We considered hospital admissions as the primary data source in our analyses, where the numbers were available and sufficiently large to do so. All studied countries expanded their hospital surge capacities to accommodate patients to a sufficient degree that we are not aware of instances of COVID-19 patients being turned away. Since clinically ill patients are unlikely not to present themselves for treatment at hospital, admissions data are a practical measure of community infection and, unlike confirmed cases, are not as susceptible to variable testing rates in the wider population. However, by studying only a subset of the entire population, the data will be biased toward older and sicker individuals which may for example lengthen the delay from an intervention to a visible signal in the data.
Confirmed cases were used in situations where hospitalisation data were not available or insufficient to reliably infer the effect of interventions - this was particularly relevant in the case of school reopening, which has predominantly been recommended in communities with significantly reduced daily incidence (and hence hospitalisation) counts. Although we do not correct for variable testing rates in the confirmed cases, we have sought out data sets where there was

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

evidence of consistent and thorough testing. However, we acknowledge this was a challenge for most countries in early March. We document the number of tests carried out and comment further on the reliability of confirmed cases as representative of the community epidemic in the Supplementary Material (S3, Testing data).
Consistent test numbers for both Germany and Norway around the time of school reopening (see Supplementary Material, S3) suggest that the confirmed number of cases is less prone to biases than earlier in the pandemic. We are therefore less concerned about using these data streams in a reopening context.

National data streams

The effect of school closure was estimated using hospitalisation data for Denmark and Norway, and daily confirmed cases for Germany and Sweden. Given other interventions were implemented after school closure and their effect is hard to disentangle, it is implicitly assumed that the inversion of epidemic trend from growing to declining is not solely a result of school closure. Therefore, the analysis of the impact of school closure is restricted to data before the peak in reported cases or hospital admissions.
Denmark reopened schools quickly enough following sweeping nationwide interventions that hospitalisation data could still be used, though we cross-checked these findings by analysing confirmed cases. Official estimates at the time suggest a delay from infection to hospitalisation of 10 to 14 days [18]. Official Norwegian estimates suggest this same delay to be 14 days [19]. As Norwegian hospitalisation data were too sparse to reliably infer the effect of school reopening, daily confirmed cases were analysed instead.
In Germany, daily confirmed cases are reported specifically for students under 18 in schools, kindergartens, holiday camps, after school clubs, etc. as well as for the staff working in these facilities. We used these numbers, and national hospital admissions, to analyse school reopening instead of population aggregates on the state or federal level.
References for each data stream, together with further discussion on their limitations, are provided in the Supplementary Material (S1, Data availability).

Estimating the effect of school closure by simulating the early epidemic

Our aim is to assess the impact of school closure on transmission dynamics. This includes any associated changes in behaviour (e.g. parents not accompanying children to school or working from home due to caring responsibilities), as well as all other interventions (if any) occurring on the same day, which cannot be disentangled from school closure. The impact is assessed by comparing differences in the growth rate of cases or hospitalisations before and after the intervention.
Rather than naïvely comparing the growth rate at different points in time, we follow a more sophisticated procedure with the aim of separating the decrease in growth rate due to interventions implemented before school closure, and the impact of school closure itself. This is achieved by generating a *counterfactual* projection of daily cases or hospital admissions, which accounts inasmuch as possible for events prior to, but excluding, school closure, and identifying when there is a clear deviation between the real data and such a projection. We then compare the growth rate observed in the real data and in the modelled counterfactual at the time when this deviation is detected, and interpret this difference as the likely impact of school closure.

To construct the counterfactual, we use a compartmental ODE model fitted to the pre-intervention data. However, fitting a model to data from the earlier part of the epidemic is extremely challenging since the observations are generally scarce, noisy and coloured through various reporting issues, in particular systematic ones, such as a strong weekend effect.
Fitting a simple compartmental model without accounting for these factors will result in parameter estimates that are systematically biased [20]. These inaccuracies in parameter estimates propagate to any projection drawn from the model. Mitigating this model discrepancy in the fitting process is an area of active research; see [21] for a recent review. Generating a counterfactual projection using a compartmental model, without compensating for such discrepancies, will erroneously understate or overstate the effect of the intervention.

To alleviate the challenges brought on by scarce and noisy data we argue that, given the 4.8 day mean incubation period for SARS-CoV-2 [22], we expect the impact of any intervention to be delayed by at least 5 days, and in particular we expect cases on the first 5 days following school closure to predominantly reflect only earlier interventions (whether imposed or not, e.g. spontaneous physical distancing). We then use a two-step approach for fitting an ODE

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

model and correcting for the discrepancy between model and data as follows. In the first step, a selection of sample trajectories are generated via Approximate Bayesian Computation (ABC) fitting of the ODE from the first day of data until 5 days after school closure. In a second post-processing step, a Bayesian regression model is then trained on the same data used to fit the ODE model, while using the sampled trajectories as the covariates (inputs). Essentially, the regression model attempts to capture the structural part of the discrepancy between ODE simulations and the observed data (predominantly, potential deviation from exponential growth and weekend effects). We formulate this regression model as a Gaussian process (GP) with a Negative Binomial likelihood. Once trained, the regression model is used to project the trajectory of cases for the time period following the 5 days after school closure. This projection is then treated as the desired counterfactual.

We identify the first day on which there is a clear and sustained deviation from the modelled data, hereafter referred to as the *response date*. Such a deviation must (a) occur more than 5 days after the date of school closure, (b) fall outside of the 75th percentile of the projected data, and (c) persist in doing so for at least 5 days. The time window from school closure to response date defines the *lag time* (Table 1, column 2), which runs from the date of closure (acting as the zeroth day) up to but not including the response date.
The growth rates are obtained as point estimates (see following description of the instantaneous growth rate) at the time of school closure, for the observed data (r_{pre}^{obs}, Table 1, column 3), and at the response date for both the modelled (r_{post}^{mod}, Table 1, column 4) and the observed data (r_{post}^{obs}, Table 1, column 5). The relative changes in the estimated growth rates between modelled and observed data at the response date can be used to assess the impact of school closure. The observed growth rate at the time of school closure can be used to cross-check the growth rate in the modelled data; these could be significantly different if the impact of interventions prior to school closure had a strong effect on transmission which the GP was able to capture in the counterfactual projection. However, a causal link cannot be established between the interventions and the growth rates, calling for a cautious interpretation of the specific values of the growth rates and the reductions therein.

The ABC fitting of the SEIR model was carried out using the PyGOM package for Python [23]. The GP regression method, devised as a Bayesian latent variable approach, was carried out using the PyMC3 probabilistic programming package for Python [24]. Further details about the introduced methods can be found in the Supplementary Material (S2, Numerical methods).

Estimating the effect of closure and reopening using the instantaneous growth rate

With the number of sequential changes in interventions and loosened restrictions on personal movement and the operation of businesses, it is misleading to estimate a constant growth rate in new cases before and after schools reopened. We therefore consider a method whereby the growth rate can be quantified following successive changed measures. A smoother $\rho(t)$ is applied to the data over time t, such that the instantaneous growth rate is $\rho'(t)$ (cf. a constant value in a phase of pure exponential growth or decline). It is assumed that the daily new confirmed cases (or daily new hospital admissions) $c(t)$ grow or decay exponentially, with noise added to account for small case numbers, i.e. $c(t) \propto e^{\rho(t)}$. To estimate $\rho'(t)$ we adapt a General Additive Model (GAM) from the R package *mgcv*, using a Negative Binomial family with canonical link [25]. Smoothing is achieved using default thin plate regression splines.
Where case numbers are sufficiently high, and there is a clear weekend effect in the reporting of data, a weekend effect has been accounted for in the GAM by the addition of a fixed effect on those days of the week. Specifically, we add a quantity ω_d for each day of the week $d \in [1, 7]$, such that cases follow $c(t) \propto e^{\omega_d + \rho(t)}$. Setting $\omega_d = 0$ for all but two days yields a weekend-specific description, and $\omega_d = 0 \; \forall \; d$ recovers the GAM with no day-of-week effects. This method has previously been used in [22].

In the case of school closure, the GAM approach with a weekend effect has been used to estimate the instantaneous growth rate at different points in time. In the case of school reopening, the instantaneous growth rate with a day-of-week effect has been used to identify trends in the data. Our aim here is to assess if there is any correlation between changes in the growth rate, and the timing of school reopening.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Results

As many of our findings are based on the premise of analysing interventions at different points in time, or in different geographical regions, all results are inherently conditional on the assumptions of *stability* and *homogeneity*. Firstly, in order to make comparisons throughout the same time series, we assume that the only changes to behaviour are those directed by changing public guidelines, and that adherence to these is constant throughout (stability). Secondly, in order to compare different regions of the same country, we must assume that there are no fundamental differences in adherence, testing, implementation of national policies, or similar such aspects (homogeneity). Deviations from these assumptions are taken to be too small to affect the data in a way to qualitatively alter our conclusions.

Closing of schools in Germany

We consider the date of school closure as the first day on which all schools in a state were closed as a response to state or national government intervention. In most cases, however, there were local school closures prior to enforced closures. Furthermore, most primary schools continued to be open to both vulnerable children and the children of key workers after national and state closures.
Table 1 provides an overview of the observed changes in the daily growth rates in the period during and after school closures. These growth rates are consistent with previous estimates [26].

State	Lag time (days)	r^{obs}_{pre} (day^{-1})	r^{mod}_{post} (day^{-1})	r^{obs}_{post} (day^{-1})	$1 - r^{obs}_{post}/r^{mod}_{post}$
Baden-Württemberg	8	0.143 (0.104 - 0.182)	0.167 (0.148 - 0.185)	0.051 (0.013 - 0.089)	69% (40 - 93)
Bavaria	8	0.216 (0.176 - 0.255)	0.214 (0.208 - 0.221)	0.109 (0.072 - 0.146)	49% (29 - 67)
Berlin	–*	0.145 (0.103 - 0.187)	–	–	–
Hesse	7	0.251 (0.195 - 0.308)	0.274 (0.265 - 0.283)	0.067 (0.017 - 0.117)	75% (56 - 94)
Lower Saxony	7	0.223 (0.179 - 0.267)	0.229 (0.213 - 0.244)	0.069 (0.032 - 0.107)	70% (50 - 87)
North Rhine-Westphalia	6	0.192 (0.156 - 0.228)	0.206 (0.200 - 0.213)	0.061 (0.026 - 0.096)	70% (52 - 88)
Rhineland-Palatinate	7	0.257 (0.205 - 0.310)	0.235 (0.211 - 0.259)	0.043 (0.001 - 0.086)	82% (59 - 100)

Table 1. Estimated lag time and pre- and post-intervention (and for the latter, modelled and observed) daily growth rates in different German states, and relative change between the modelled and observed growth rate. The 95% credible intervals (CrI) are given in brackets. Their equivalent formulation as doubling times can be found in the Supplementary Material (Table S3). Sensitivity analysis of the training period on the lag time suggests these can vary by up to two days (see Supplementary Material, S2.7). The overlapping CrIs between the pre-intervention and the post-intervention modelled growth rates suggest a limited deviation from exponential growth between the day of school closure and the end of the training window.
*The peak in daily incidence is reached before a response is seen in the data. A lag time which may be attributable to school closures therefore cannot be determined.

All states in Germany saw a reduction in growth rate after the closure of schools, typically after a delay of 7 days, or about 1.2 generations [27]. With the exception of Baden-Württemberg and Berlin, all German states closed schools on March 16th. As this was a Monday, we set the effective date of school closures as Saturday March 14th, under the assumption that school activity is significantly reduced on weekends. Schools in Baden-Württemberg and Berlin closed on Tuesday March 17th. It should be noted that all states experienced further interventions around the same time as school closures. The presence of concurrent interventions makes it difficult to attribute the fall in cases solely to the closure of schools, and it is likely that a combination of factors contributed to the observed decay in growth rate.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

However, comparison between Baden-Württemberg and North Rhine-Westphalia, which saw similar case numbers, yields comparable lags and overall trajectories of the epidemic curves when accounting for the three-day delay in school closures in Baden-Württenberg (see Supplementary Material, Figure S9). This is indicative of school closures being at least partially responsible for the reduction in growth rate.

The reduction between the modelled and observed post-response growth rates serves as a measure of the overall effectiveness of interventions (Table 1, column 6). Overall, lower relative reductions in the growth rate are weakly correlated with states which had higher (daily and cumulative) incidence counts at the time of intervention (Baden-Württemberg, Bavaria, and North Rhine-Westphalia). This supports the generally held expectation that non-pharmaceutical interventions are more effective when implemented early, capable of breaking transmission chains while community transmission is relatively low.

The states of Hesse and Rhineland-Palatinate allowed students aged 18-19 to sit in-school examinations in late March, under strict social distancing measures and other precautions. Neither of the states permitting examinations saw any less evident reduction in growth rates compared to states which had similar case numbers prior to school closure, but where exams did not take place during this time period (e.g. Lower Saxony). Further, the largest reduction in the growth rate was observed in Rhineland-Palatinate. Assuming stability and homogeneity, this suggests that under controlled conditions with limited social interaction, the beginning of the examination period for older students was likely not a significant driver of epidemic growth. We cannot comment on the full effect of the entire examination period. We include the detailed results from the highlighted German states in Figure 1 and a timeline of key interventions from [28] below, with the remaining states detailed in the Supplementary Material (S4, School closures analyses).

- 10/03 - Banned gatherings of more than 1000 people (DE-G1).
- 14/03 - **Hesse, Lower Saxony, and Rhineland-Palatinate closed schools** (effective date, DE-S1).
- 16/03 - Shut borders with France (FR), Switzerland (CH), Austria (AT), Denmark (DK) and Luxembourg (LU) (DE-B1); closure of non-essential business and public service (DE-P1).
- 17/03 - **Baden-Württemberg closed schools** (DE-S2); shut borders with EU (DE-B2).
- 22/03 - National stay at home orders, with exceptions for essential trips, and banned gatherings of more than 2 people (DE-G2); banned all social events and gatherings (DE-P2); closure of non-essential retail and leisure, with exceptions for restaurant takeout (DE-R1).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

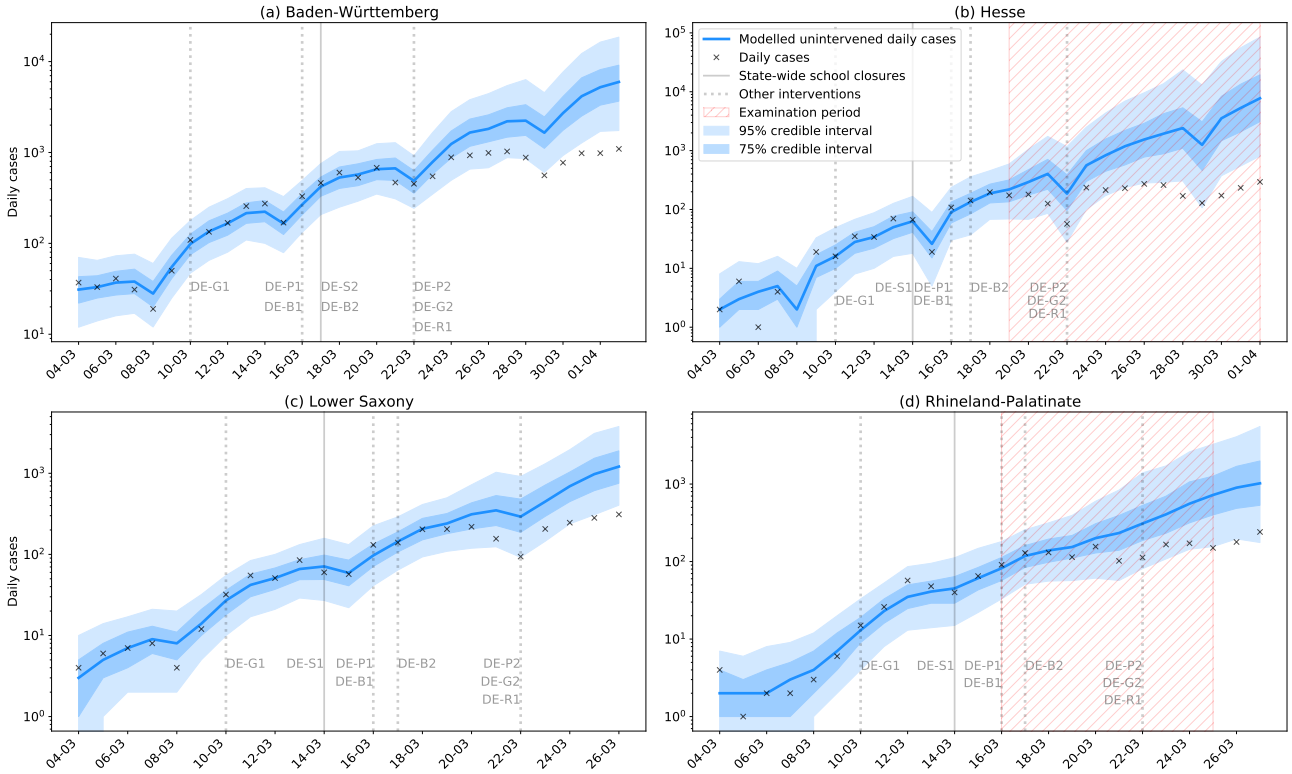


Figure 1. Modelled and observed cases in (a) Baden-Württemberg, (b) Hesse, (c) Lower Saxony, and (d) Rhineland-Palatinate. Hesse and Rhineland-Palatinate (Figure 1 (b), (d)), where final year high school exams took place in late March, saw a similar response to interventions as other German states with moderate incidence (Figure 1 (c)) where exams did not take place at that time. While there is insufficient scope in the data to assess the effect of the full examination period, we should in principle be able to detect a signal related to the beginning of the examination period. Assuming stability and homogeneity across states, and given the lack of such a signal, it is unlikely that these exams significantly contributed to the overall outbreak.

Closing of schools in Denmark, Norway, and Sweden

In all three countries there were provisions in place to allow key workers' children to continue attending school. Hospital admissions were analysed for Denmark and Norway, as testing was deemed too variable during this time period (see Supplementary Material, S3) to reliably use confirmed cases. However, the expected lag time from infection to hospital admission in Denmark and Norway is 10-14 days [18, 19], whereby any signal observed in the data is too early to be attributable to school closures. For completeness, we include the fits to daily hospital admissions in the Supplementary Material (S4, School closure analyses).

Sweden's school closures were less restrictive than other countries', with only educational establishments for students aged 16 or over being required to close. Despite no official nationwide closing of primary or secondary schools in Sweden, there were local closures in response to outbreaks within the community. There is no evidence of a sustained reduction in the growth rate within a time period attributable to school closures (Figure 2). It is notable, however, that the limited closures in Sweden were imposed in the absence of large-scale social restrictions. This indicates that school closures affecting older students without more widespread social interventions are unlikely to have significant national effects, and that school closures are most effective when implemented concurrently with other interventions.
It is notable that there was an increase in weekly testing between March 30th and April 6th, which may have contributed to the apparent limited reduction in growth rate during this time. However, this falls outside of the time window in which we would expect to see a response attributable to school closures.

Sweden saw the following interventions introduced around the same time as school closures:

- 11/03 - Banned gatherings of more than 500 people (SE-G1).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

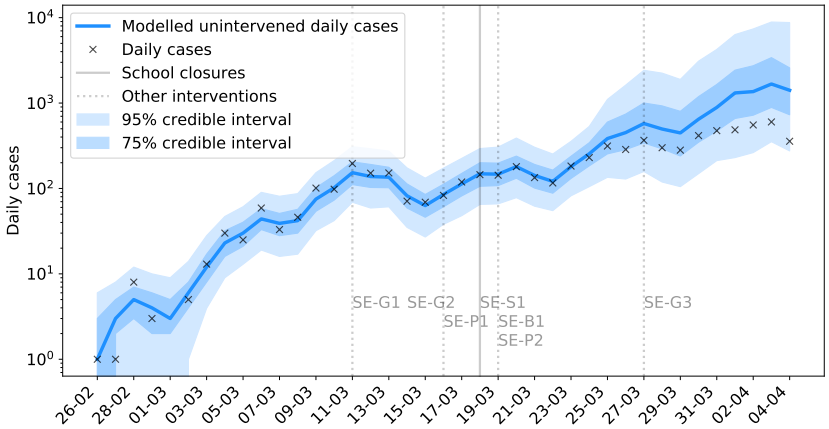


Figure 2. Modelled and observed daily cases in Sweden.

- 14/03 - Advice against non-essential travel (SE-G2).
- 16/03 - Social distancing advised but not enforced (SE-P1).
- 18/03 - **Closed all education for students aged 16 or over** (SE-S1).
- 19/03 - Restrictions on international travel (SE-B1); advice against national travel (SE-P2).
- 27/03 - Banned gatherings of more than 50 people (SE-G3).

Reopening of schools

Germany

The following key interventions, sourced from [28], are possible confounders for the effects of school reopening:

- 20/04 - Opening of some retail venues (DE-R2).
- 22/04 - Mandatory mask wearing in certain public spaces (DE-P3).
- 27/04 - **Return of Year 10, final year exam students (ages 15, 18-19)** (DE-S3).
- 29/04 - Extension of mask-wearing requirements (DE-P4).
- 03/05 - Expiry of stay-at-home order (DE-G3).
- 04/05 - **Return of Year 4 primary schools students (age 9)** (DE-S4); opening of retail (DE-R3) and public spaces (DE-P5).
- 11/05 - **Return of primary and secondary school students (ages 9, 15, 17-19)** (DE-S5).
- 15/05 - Relaxation of border controls (DE-B3).
- 18/05 - **Staggered return of primary and secondary school students (ages 9-11, 15-19)** (DE-S6); meeting of two households allowed (DE-G4).
- 29/05 - Gatherings of up to 10 people allowed (DE-G5).
- 02/06 - Pubs reopen (DE-R4).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Due to differing policies across German states, the dates of school reopening and the ages of students returning were variable. Where an age group appears over multiple dates, the return of students in this age group took place in different states. We present a summary of the overall national trend, as our data only distinguish between staff and students on the national scale. On three occasions the recorded cases were inconsistently reported, and values were imputed using cases reported on preceding and proceeding days. Our findings do not change significantly upon exclusion of these data points. We contrast these demographically specific findings by comparison with national hospital admissions.

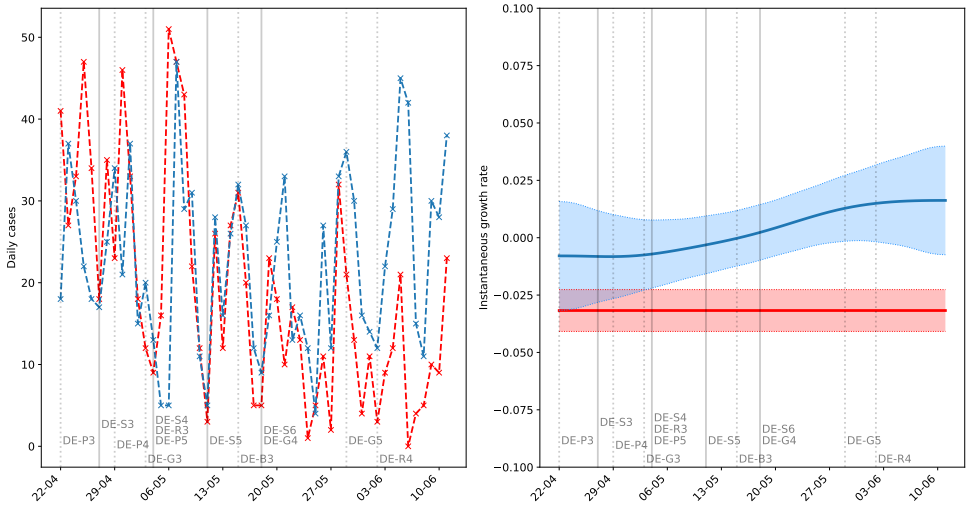


Figure 3. Confirmed cases in staff (red) and students (blue) in schools, kindergartens, holiday camps, and other educational facilities for under-18s (age distribution not known) in Germany. Left shows daily new confirmed cases, and right shows the instantaneous growth rate (shaded regions are 95% confidence intervals). Solid vertical lines indicate when students returned to school, and dashed lines indicate other changes to public measures. In April and early May with small numbers of primary school or exam students returning, there was no notable difference between the incidence among students and staff. Accounting for the lag time, the incidence among students was higher than that of staff following May 18th.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

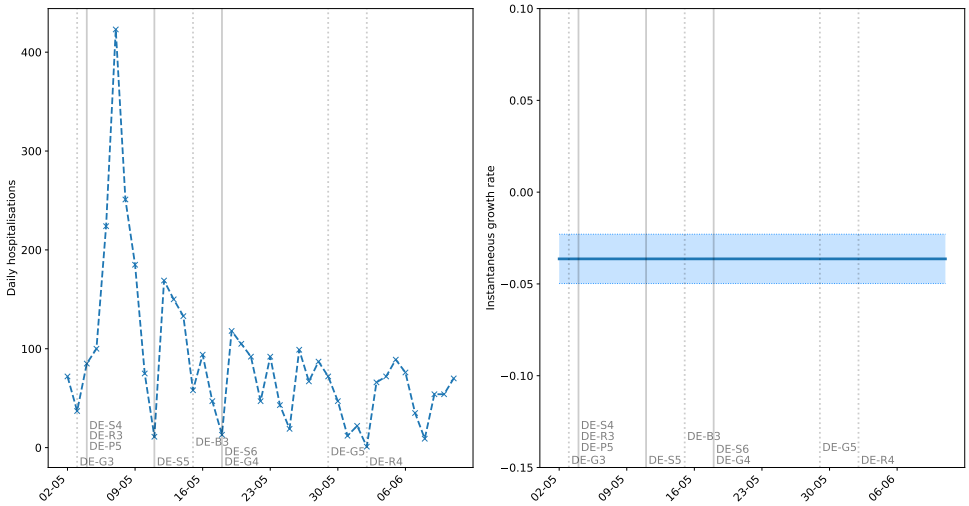


Figure 4. Daily hospital admissions with COVID-19 in Germany, excluding those working in education, front-line healthcare workers, carers, catering, and hospitality, thus representing transmission in the general, average-exposure population. Left shows daily admissions, and right shows the instantaneous growth rate (shaded regions are 95% confidence intervals). The continuing decline in admissions suggests that the return of younger (and exam) students did not present a statistically significant impact on the general hospitalised population.
It is worth bearing in mind that hospital admissions lag further behind than confirmed cases. Additionally, since very few young people have been hospitalised, an additional generation time of 6 days [27] may need to be added to this lag to account for students infecting older age groups.

The spike in daily cases observed around May 7-8th (Figure 3, left) may be a result of increased presentation for testing following a national announcement of school reopening on May 4th (allowing for testing delay), or increased community transmission following reopening of other parts of society which was subsequently contained. Overall the incidence among staff decreased, which is supported by the growth rate among staff being negative. The incidence among students first decreased, and subsequently increased with a predominantly positive growth rate from the end of May (Figure 3, right).

The stable, low, values of the incidence and growth rate until the middle of May indicate that the return of final year and year 4 students either a) did not significantly increase transmission in schools or the community, or b) did increase transmission, but this was mitigated due to safety protocols of prevention and monitoring. This observed effect is quite a strong signal as the daily case counts remain low even across a background of increased community transmission from late April onward with, for example, shops reopening. It is therefore reasonable to conclude that these age groups do not strongly increase transmission in a setting of effective social distancing.

However, the impact of most students returning to school from late May was different. In this time period, the incidence among staff qualitatively agreed with the national trend in hospitalisations (Figure 4), i.e. staff did not immediately appear to be at greater risk following the return of more students. In contrast, the growth rate in student cases increased following May 18th. The constant staff growth rate, and the small effect of the return of (mostly) younger years, suggests that the increased incidence may be due to a) increased transmission among older students, b) low feasibility of effective physical distancing in venues at full capacity, or c) a combination of these.

Denmark

Schools reopened alongside the following key interventions sourced from [28]:

- 08/04 - 7-day ban on gatherings of over 10 people (DK-G5).
- 14/04 - Partial return of employees to work (DK-P2).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

- 15/04 - **Return of nursery, kindergarten, Year 0-5 primary school, and final year exam students (ages 0-12, 18-19)** (DK-S3).
- 20/04 - Partial reopening of retail and small business (DK-R2).
- 21/04 - Assemblies limited to 500 people (DK-G6).
- 11/05 - Full reopening of shopping and retail (DK-R3).
- 18/05 - **Return of Year 6-10 secondary school students (ages 12-16), and examinations requiring physical attendance** (DK-S4); restaurants and cafés reopen (DK-R4); reopening of houses of worship (DK-P3).
- 21/05 - Reopening of leisure and cultural facilities (DK-P4).
- 25/05 - Relaxation of border restrictions with Nordic countries and Germany (DK-B2).
- 27/05 - **Return of secondary school students (ages 16-18) and adult education** (DK-S5).

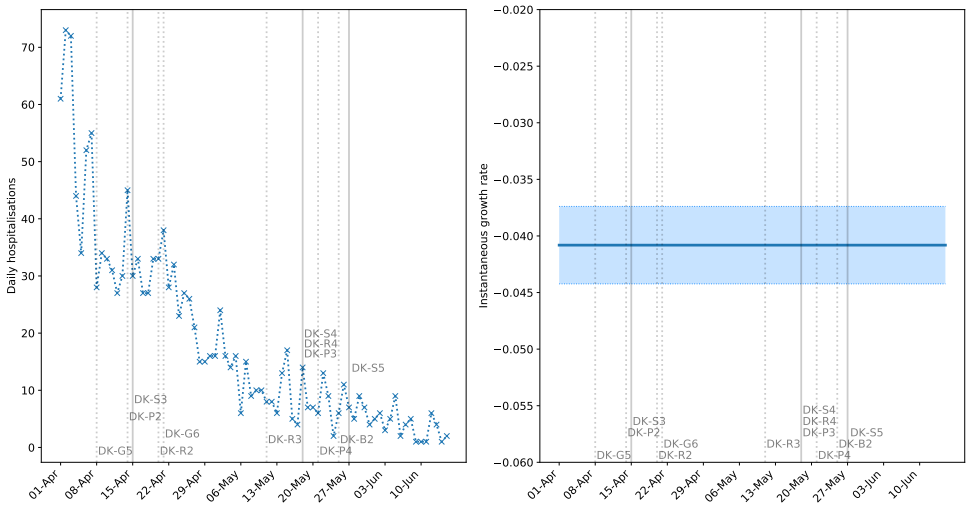


Figure 5. Daily hospitalisations with COVID-19 in Denmark. Admissions are shown left, and right shows the instantaneous growth rate (shaded regions are 95% confidence intervals). A longer lag time of 10-14 days is in effect from infection to hospitalisation [18], with a further 6 days' generation time [27] to account for subsequent infection generations due to the low hospitalisation rate among children. Solid vertical lines indicate when students returned to school, and dashed lines indicate other interventions.

There is no significant observable increase in the growth rate of hospital admissions following school reopening to younger years, even bearing in mind the subsequent reopening of some businesses (Figure 5). The low growth rate and small relative number of admissions suggests that the return of younger years to school with social distancing did not contribute significantly to community transmission. The subsequent reopening stage on May 18th also did not have a significant impact on hospital admissions, which we verify using confirmed cases (see Supplementary Material S5, School reopening analyses).
These findings are further supported by a lower proportion of adults testing positive for COVID-19 among those working with children aged 0-16 than those working with students aged 16 or over (see Supplementary Material, S5) [29]. However, these numbers alone do not distinguish between infection acquired from students and infection acquired elsewhere.

Norway

The following events are possible confounders in the data, and key dates for school reopening sourced from [28]:

- 01/04 - Exceptions made to entry restrictions (NO-B4).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

- 08/08 - Easing of entry restrictions from EEA workers (NO-B5).
- 20/04 - **Return of kindergarten students (ages 1-5)** (NO-S2); travel to cabins allowed (NO-G5).
- 27/04 - **Return of Year 1-4 (ages 6-10) and final year students (ages 18-19), vocational training, and higher education requiring physical attendance** (NO-S3); partial reopening of retail and small business (NO-R1).
- 07/05 - Events, and some public sports and leisure facilities open, but limited to 50 people (NO-P1); group size for social gatherings increased from 5 to 20 people (NO-G6).
- 11/05 - **Return of students aged 10-18 this week** (NO-S4); reopening of bingo halls and driving schools (NO-P2).
- 12/05 - Easing on entry restrictions (NO-B6).

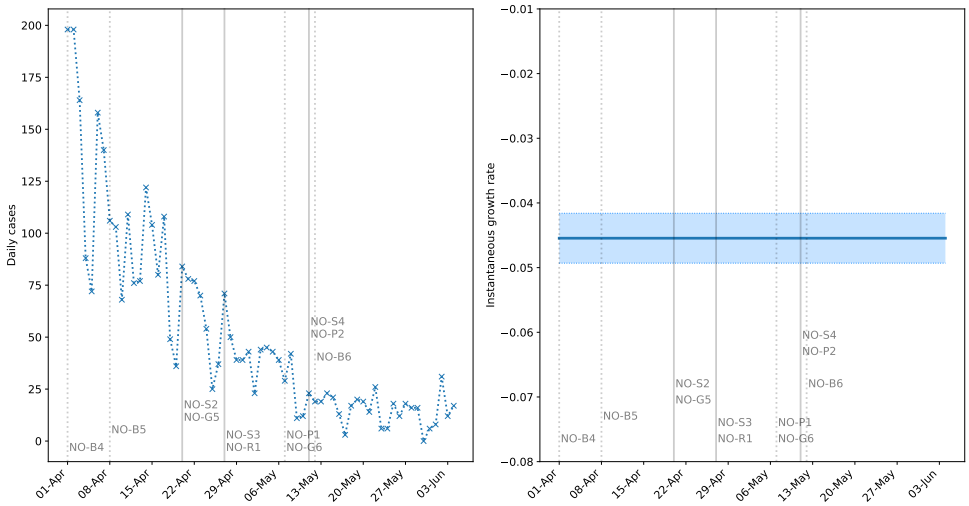


Figure 6. Daily confirmed cases in Norway. The left panel shows new cases, and the right panel shows the instantaneous growth rate (shaded regions are 95% confidence intervals). Solid vertical lines indicate when students returned to school, and dashed lines indicate other interventions.

There is no notable change in the growth rate, even following the return of students in May (Figure 6). The consistently negative growth rate and small number of cases indicate that the return of most students to school (with social distancing) did not contribute significantly to community transmission. However, this effect is likely conditional on high levels of testing, with very low community transmission.

Discussion

Decoupling the effect of school closure, and subsequent reopening, from other interventions is not straightforward. This work does not claim to have achieved this; however, the consistency of the signal across regions with different intervention timelines suggests a distinct effect of school closure on the subsequent growth in cases. The consistently lower post-intervention growth rates in German states when compared to the modelled scenarios with no interventions (see Table 1) suggests that school closures contributed to reducing the epidemic growth rate. School interventions were typically followed by a response in the data approximately 7 days later (corresponding to approximately 1.2 generations). Table 1 shows that this lag time was comparable across states which closed schools at different times. High school students sitting their final examinations under social distancing does not appear to have significantly impacted case numbers. Sweden implemented partial school closures which affected students aged 16 or above. However, there is no evidence to suggest that this intervention affects the later daily incidence within the expected time frame. These findings are consistent with earlier works suggesting that school closure in isolation is insufficient to prevent the

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

spread of COVID-19 [6,8,11]. The evidence for the impact of school closures on growth rates in Norway and Denmark is more limited. While there was a reduction in growth rate of hospitalised cases after school closures, it has not been possible to link this effect with school closures.

While school closures are often among the first implemented control measures, school reopenings are typically staggered with other eased restrictions, often with a small initial cohort of returning students. Since, to our knowledge, no stringent restrictions were introduced to compensate for the additional transmission risk due to school reopening, a lack of signal in the growth rate after reopening would be indicative that schools do not contribute substantially to community transmission. From our analysis, the reopening of schools to younger year groups and exam students in Germany, Denmark and Norway has not resulted in a significant increase in the growth rate. The return of all students (up to age 16 in Denmark) does not appear to have increased transmission in Denmark and Norway. However, the added return of most (primarily older) students in Germany has increased transmission among students, but not staff. It is unclear whether older students transmit more, or if physical distancing is practically unfeasible in classrooms at high capacity. The distinction between the impact of younger and older students is echoed in other modelling studies [14,15]. Although our findings cannot be interpreted as causal links between individual interventions and changes in national case numbers, they represent a realistic assessment of the effects of school reopening in their natural context of wider societal interventions.

Our findings are not without limitations. The presence (or lack) of signals in the data following school interventions are limited by the reliability of the available data. We have worked with hospital admissions data, as they are less affected by variable testing, while bearing in mind that hospitalisations only affect a subset of the infected population. Where these data were unavailable, we have considered confirmed cases while monitoring the degree of testing in place to ensure such numbers were indicative.

The GP regression method allows one to account for differences between the simulated epidemic trajectories from the ODE model and the observed cases. However, the fact that closures occurred very early on in the epidemic means that the GP method often had to be trained on a limited number of data points.
Since the instantaneous growth rate relies on the derivative of splines, it is subject to increased error at the boundaries of the data. However, the observed signals are qualitatively robust to this limitation. Due to the noisiness of some data streams from relatively low incidence following mass quarantine, the values of the instantaneous growth rate should be taken as a quantification of the trend in incidence rather than the true value on any given day.

The data have generally not made it possible to account for inevitable geographic variability, the age distribution of those studied, and their occupation (i.e. likelihood of exposure to infected individuals) in our analyses. The analysis of German school reopening, particularly the comparison of staff and student infections, warns against the reliability of using national-level data to understand the immediate effects or impact on a single population. Rather, such impact may only become visible in national data in subsequent generations. We must therefore be concerned not only with the lag time from intervention to a signal in the data, e.g. 7 days in Germany, but also with the following generation of infections.

Our analysis is restricted to countries with high monitoring and intervention efficacy (including but not limited to high testing, tracing, and adherence to isolation), hence great care should be taken when translating our findings on the impact of school reopening to other nations. For instance, continued low incidence following the return of younger students does not imply that such a measure is inherently safe, but may rather be a result of the successful implementation of a complete system of monitoring and interventions jointly with low daily incidence, as observed in Denmark and Norway. In many instances, the students were spread over more classrooms, with greater levels of physical distancing from each other and teachers, conditions which are not always practically feasible.
Caution is warranted surrounding the return of older students, in particular regarding the likely increased number of crowded classrooms, as well as their added impact to community transmission. The correlation with community transmission is particularly clear in Germany, with confirmed cases increasing among students, and the halted decay in hospital admissions on the national level. While all three countries seem to be effectively managing transmission, the volatility of new German hospital admissions warns that a failure in control, or a sudden spike in cases, will likely have a stronger effect in Germany than it would have in Denmark or Norway. Key to this observation is the aforementioned delay before which the ripple effects of school reopening will travel from students to the general population.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Furthermore, we highlight that the tenuous balance (net zero growth in June) in Germany exists despite a swift and robust test and trace infrastructure and school-level stratified monitoring. We question the possibility of an equally effective reopening in countries with a monitoring process reliant on national-level incidence data. The swiftness and effectiveness of targeted interventions become increasingly crucial as the daily incidence increases, due to the correspondingly greater challenges presented in managing the localised outbreaks across e.g. reopened schools.

Policy-makers should carefully consider their nations' respective capacities and associated effectiveness of infection management before considering a partial or full reopening of schools. Our findings suggest a small, strategically chosen, proportion of students to return in the first instance, with dedicated testing and monitoring of cases among staff and students (over time scales where the effect can be assessed). Any significant return of students to schools, particularly in countries with a high incidence, should not be considered unless an infrastructure is in place which would be able to swiftly identify and isolate most new cases as they appear [16]. Such a strategy may not be feasible if the community incidence is too high.
When used in conjunction with household transmission models, and knowledge of the length of immunity associated with SARS-CoV-2, our findings may be used to inform age-targeted vaccine allocation protocols.

Authors' contributions

JS modelled school closure from early data. HS analysed school reopening from late data. SG, LP, JS and HS drafted the manuscript. SG implemented the Gaussian process model. All authors contributed to the writing of the manuscript.

Funding

HS and LP are funded by the Wellcome Trust and the Royal Society (grant 202562/Z/16/Z). FS is funded by the CIHR 2019 Novel Coronavirus (COVID-19) rapid research program. LP and FS are also supported by the UKRI through the JUNIPER modelling consortium (grant number MR/V038613/1). SG was supported by the Medical Research Council (Unit programme number MC UU 00002/11). JS and TF are funded through the Department for Health and Social Care grant in aid funding to Public Health England.

Acknowledgements

We gratefully acknowledge Thomas House, Ian Hall and Pantelis Samartsidis for useful discussions and contribution to the methodology. In particular we acknowledge the work of Ian Hall in developing the GAM approach with a day-of-week effect.

Competing interests

The authors declare no competing interests.

Data accessibility

All data used in the production of this work, including the relevant scripts, can be found in the electronic supplementary material and online at `https://github.com/HelenaStage/COVIDschools`.

Ethical approval

This study does not require ethical approval, as the only data used are publicly available.

References

1. Ping Ing Lee, Ya Li Hu, Po Yen Chen, Yhu Chering Huang, and Po Ren Hsueh. Are children less susceptible to COVID-19? *Journal of Microbiology, Immunology and Infection*, 53(3):371, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2. Jonas F. Ludvigsson. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. *Acta Paediatrica, International Journal of Paediatrics*, 109(6):1088–1095, June 2020.

3. Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, 5(3):0381–0391, March 2008.

4. Thomas House, Marc Baguelin, Albert Jan van Hoek, Peter J White, Zia Sadique, Ken Eames, Jonathan M Read, Niel Hens, Alessia Melegaro, W John Edmunds, and Matt J Keeling. Modelling the impact of local reactive school closures on critical care provision during an influenza pandemic. *Proceedings of the Royal Society B: Biological Sciences*, 278(1719):2753–2760, 2011.

5. Simon Cauchemez, Alain Jacques Valleron, Pierre Yves Boëlle, Antoine Flahault, and Neil M Ferguson. Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature*, 452(7188):750–754, 2008.

6. Russell M Viner, Simon J Russell, Helen Croker, Jessica Packer, Joseph Ward, Claire Stansfield, Oliver Mytton, Chris Bonell, and Robert Booy. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *The Lancet Child & Adolescent Health*, 4(5):397–404, May 2020.

7. Seth Flaxman, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W. Eaton, Mélodie Monod, Pablo N. Perez-Guzman, Nora Schmit, Lucia Cilloni, Kylie E.C. Ainslie, Marc Baguelin, Adhiratha Boonyasiri, Olivia Boyd, Lorenzo Cattarino, Laura V. Cooper, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra Djaafara, Ilaria Dorigatti, Sabine L. van Elsland, Richard G. FitzJohn, Katy A.M. Gaythorpe, Lily Geidelberg, Nicholas C. Grassly, William D. Green, Timothy Hallett, Arran Hamlet, Wes Hinsley, Ben Jeffrey, Edward Knock, Daniel J. Laydon, Gemma Nedjati-Gilani, Pierre Nouvellet, Kris V. Parag, Igor Siveroni, Hayley A. Thompson, Robert Verity, Erik Volz, Caroline E. Walters, Haowei Wang, Yuanrong Wang, Oliver J. Watson, Peter Winskill, Xiaoyue Xi, Patrick Gt Walker, Azra C. Ghani, Christl A. Donnelly, Steven M. Riley, Michaela A.C. Vollmer, Neil M. Ferguson, Lucy C. Okell, and Samir Bhatt. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 2020.

8. Xiaohui Chen and Ziyi Qiu. Scenario analysis of non-pharmaceutical interventions on global COVID-19 transmissions. *arXiv*, 2020.

9. Nicholas G Davies, Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, CMMID COVID-19 working group, and Rosalind M Eggo. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature medicine*, 2020.

10. D. F. Gudbjartsson, A. Helgason, H. Jonsson, O. T. Magnusson, P. Melsted, G. L. Norddahl, J. Saemundsdottir, A. Sigurdsson, P. Sulem, A. B. Agustsdottir, B. Eiriksdottir, R. Fridriksdottir, E. E. Gardarsdottir, G. Georgsson, O. S. Gretarsdottir, K. R. Gudmundsson, T. R. Gunnarsdottir, A. Gylfason, H. Holm, B. O. Jensson, A. Jonasdottir, F. Jonsson, K. S. Josefsdottir, T. Kristjansson, D. N. Magnusdottir, L. le Roux, G. Sigmundsdottir, G. Sveinbjornsson, K. E. Sveinsdottir, M. Sveinsdottir, E. A. Thorarensen, B. Thorbjornsson, A. Löve, G. Masson, I. Jonsdottir, A. D. Möller, T. Gudnason, K. G. Kristinsson, U. Thorsteinsdottir, and K. Stefansson. Spread of SARS-CoV-2 in the Icelandic population. *New England Journal of Medicine*, 382(24):2302–2315, 2020.

11. Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Yan Wang, Wei Wang, Shanlu Zhao, Qianhui Wu, Stefano Merler, Cécile Viboud, Alessandro Vespignani, Marco Ajelli, and Hongjie Yu. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*, 368(6498):1481–1486, 2020.

12. Jan Markus Brauner, Sören Mindermann, Mrinank Sharma, Anna B Stephenson, Tomáš Gavenčiak, David Johnston, John Salvatier, Gavin Leech, Tamay Besiroglu, George Altman, Hong Ge, Vladimir Mikulik, Meghan Hartwick, Yee Whye Teh, Leonid Chindelevitch, Yarin Gal, and Jan Kulveit. The effectiveness and perceived burden of nonpharmaceutical interventions against COVID-19 transmission: a modelling study with 41 countries. *medRxiv*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

13. Nicolas Banholzer, Eva van Weenen, Bernhard Kratzwald, Arne Seeliger, Daniel Tschernutter, Pierluigi Bottrighi, Alberto Cenedese, Joan Puig Salles, Werner Vach, and Stefan Feuerriegel. Impact of non-pharmaceutical interventions on documented cases of COVID-19. *medRxiv*, 2020.

14. Laura Di Domenico, Giulia Pullano, Chiara E. Sabbatini, Pierre-Yves Boëlle, and Colizza Vittoria. Expected impact of reopening schools after lockdown on COVID-19 epidemic in Ile-de-France. *medRxiv*, 2020.

15. Matt J Keeling, Michael J Tildesley, Benjamin D Atkins, Bridget Penman, Emma Southall, Glen Guyver-Fletcher, Alex Holmes, Hector McKimm, Erin E Gorsich, Edward M Hill, and Louise Dyson. The impact of school reopening on the spread of COVID-19 in England. *medRxiv*, 2020.

16. Jasmina Panovska-Griffiths, Cliff C. Kerr, Robyn M. Stuart, Dina Mistry, Daniel J. Klein, Russell M. Viner, and Chris Bonell. Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study. *The Lancet Child and Adolescent Health*, 4(11):817–827, November 2020.

17. European Centre for Disease Prevention and Control. COVID-19 situation update for the EU/EEA and the UK, as of 20 June 2020. `https://www.ecdc.europa.eu/en/cases-2019-ncov-eueea`, June 2020.

18. Statens Serum Institut. Ekspertrapport af den 6. maj 2020. Technical report, Statens Serum Institut, 2020.

19. Norwegian Institute of Public Health. Koronavirus-modellering ved FHI. `https://www.fhi.no/sv/smittsomme-sykdommer/corona/koronavirus-modellering`, June 2020.

20. Jenni Brynjarsdóttir and Anthony Ohagan. Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30(11), 2014.

21. Chon Lok Lei, Sanmitra Ghosh, Dominic G. Whittaker, Yasser Aboelkassem, Kylie A. Beattie, Chris D. Cantwell, Tammo Delhaas, Charles Houston, Gustavo Montes Novaes, Alexander V. Panfilov, Pras Pathmanathan, Marina Riabiz, Rodrigo Weber Dos Santos, John Walmsley, Keith Worden, Gary R. Mirams, and Richard D. Wilkinson. Considering discrepancy when calibrating a mechanistic electrophysiology model. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 378(2173):20190349, 2020.

22. Lorenzo Pellis, Francesca Scarabel, Helena B. Stage, Christopher E. Overton, Lauren H. K. Chappell, Katrina A. Lythgoe, Elizabeth Fearon, Emma Bennett, Jacob Curran-Sebastian, Rajenki Das, Martyn Fyles, Hugo Lewkowicz, Xiaoxi Pang, Bindu Vekaria, Luke Webb, Thomas House, and Ian Hall. Challenges in control of Covid-19: short doubling time and long delay to effect of interventions. *medRxiv*, page 2020.04.12.20059972, March 2020.

23. Public Health England. pygom 0.1.6. `https://github.com/PublicHealthEngland/pygom`, 2020.

24. John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2016(4):e55, April 2016.

25. Simon N. Wood. Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. `https://cran.r-project.org/web/packages/mgcv/`, 2019.

26. Jonas Dehning, Johannes Zierenberg, F. Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*, 369(6500), 2020.

27. Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491), 2020.

28. World Health Organization. Tracking Public Health and Social Measures. `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/phsm`, 2020.

29. Danish National Health Data Authority. COVID-19 og ansatte indenfor institutioner og undervisning. `https://files.ssi.dk/COVID-19-ansatte-institutioner-og-undervisning-uge23-kks8`, June 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

30. Statens Serum Institut. Overvågning af COVID-19. https://www.ssi.dk/sygdomme-beredskab-og-forskning/sygdomsovervaagning/c/covid19-overvaagning, June 2020.

31. Robert Koch Institute. Daily Situation Reports on COVID-19. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Gesamt.html?nn=13490888, June 2020.

32. Robert Koch Institute. COVID-19: Fallzahlen in Deutschland und weltweit. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html, April 2020.

33. Norwegian Institute of Public Health. Dagsrapport og statistikk om koronavirus og covid-19. https://www.fhi.no/sv/smittsomme-sykdommer/corona/dags--og-ukerapporter/dags--og-ukerapporter-om-koronavirus/#mer-covid19-statistikk, June 2020.

34. The Public Health Agency of Sweden. Bekräftade fall i Sverige – daglig uppdatering. https://www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/bekraftade-fall-i-sverige, June 2020.

35. C.E. Rasmussen and C.K.I. Williams. *Gaussian processes for machine learning.* MIT Press, 2006.

36. Federal Statistical Office of Germany. Population by federal states. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/bevoelkerung-nichtdeutsch-laender.html, June 2019.

37. Statistics Denmark. Population in Denmark. https://www.dst.dk/en/Statistik/emner/befolkning-og-valg/befolkning-og-befolkningsfremskrivning/folketal, March 2020.

38. Statistics Norway. Key figures for the Population. https://www.ssb.no/en/befolkning/nokkeltall/population, January 2018.

39. Statistics Sweden. Population Statistics. https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-composition/population-statistics/, May 2020.

40. Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P.H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, 6(31):187–202, 2009.

41. Matthew D Hoffman and Andrew Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

42. M. Filippone and M. Girolami. Pseudo-marginal bayesian inference for gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2214–2226, 2014.

43. Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, Yu-Sung Su, et al. A weakly informative default prior distribution for logistic and other regression models. *The annals of applied statistics*, 2(4):1360–1383, 2008.

44. Andrew Gelman et al. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, 1(3):515–534, 2006.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Material

S1 Data availability

The data streams used to carry out the analysis of this work for each country can be found in the following sources:

- Denmark: daily tests, hospital admissions (used for school closure and reopening), and confirmed cases (used for school reopening) from the National Serum Institute [30]. Sentinel survey among educational staff is from [29].
- Germany: daily (and weekly) tests, hospital admissions (used for school reopening), and confirmed cases (state-level used for school closure, and separated by staff and students used for school reopening) from the Robert Koch Institute [31, 32].
- Norway: daily tests, hospital admissions (used for school closure), and confirmed cases (used for school reopening) from the Norwegian Institute of Public Health [33].
- Sweden: weekly tests, and confirmed cases (used for school closure) from the Public Health Agency of Sweden [34].

Note that some of these sources are updated over time, so that the numbers used in these analyses may not correspond exactly to those reported at the time of reading. Where we suspect this to be the case, our analysed data are enclosed with the relevant scripts.
In the main manuscript and the analyses below, we highlight key interventions and measures concurrent with school closures and reopening. A comprehensive overview of these can be found in [28] and sources therein.

All data sources present their own biases, even beyond whether someone is suspected or confirmed as being infected with COVID-19. The available tools, procedures, and protocols for identifying and reporting cases have all changed over the course of this pandemic.
Hospitalisations only capture those cases sick enough that they require medical attention, and is thus a biased representation of transmission towards the elderly and those with underlying health conditions. Furthermore, there are differences in reporting due to personal judgements from individual doctors and nurses. For example, the correct classification of nosocomial cases has been unclear in many cases. Similar challenges arise for patients admitted to hospital for a different treatment, but who also happened to be infected with SARS-CoV-2 with mild symptoms.
Confirmed cases are strongly biased by testing; both the availability of the tests, their logistical deployment, and the laboratory capacity to analyse them. Further biases are present in the populations these are offered to: front-line healthcare staff, essential workers, care home residents, etc. This is further complicated if the allocation to various groups changes over time. Another bias in the data is what makes the general public eligible for a test: the number and severity of symptoms, their age, known recent contacts, and so on. If tests are only offered to symptomatic cases, the results will be biased away from younger people, as we know they are disproportionately likely to be asymptomatic or only mildly symptomatic.

S2 Numerical methods

S2.1 Overview of the method for counterfactual projection

We adopt a two-step approach for generating the counterfactual projection. The first step consists in fitting a simple SEIR-type ODE-based compartmental model, using Approximate Bayesian Computation (ABC). Because we assume a constant transmission parameter, with the possible exception of a transient initial phase when the initial numbers propagate through the compartments, the model trajectories are characterised by exponential growth and cannot account for the impact of interventions prior to school closure (potentially deviating from exponential growth), nor for weekend effects. Therefore, in a second post-processing step we use a posterior sample of these trajectories as inputs to a Bayesian regression model that is in turn fitted to the same data. This post-processing step captures at least part of the discrepancy between the data and the smooth simulated ODE trajectories.

For this regression model, we use a latent Gaussian process (GP) with a Negative Binomial likelihood. The choice of the Negative Binomial likelihood is dictated by the overdispersion inherent in the noisy incidence data, and hence

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

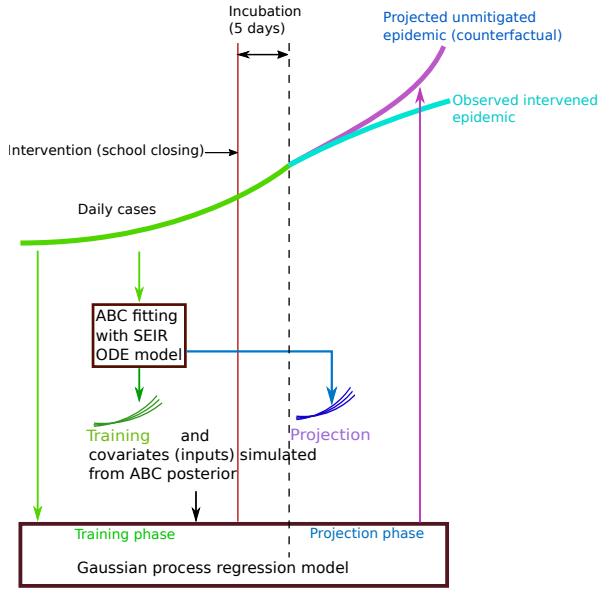


Figure S1. A cartoon illustration of the various steps that constitute our method for drawing counterfactual projections of an unmitigated epidemic.

captures the aleatoric uncertainty in the observational noise. The use of Gaussian processes allows us to specify priors on function spaces [35] for the Bayesian regression. The use of a Bayesian framework guarantees that not only a best fitting map linking the ODE trajectories and the data is used, but the entire posterior distribution of possible maps is explored. This properly quantifies and propagates both the uncertainty in the random observational noise, captured by the Negative Binomial likelihood, and the uncertainty in how well we are able to learn the structural discrepancy between the ODE trajectories and the data, which in this work is mainly due to weekend effects and possible systematic deviation from pure exponential growth before school closure can have any effect (most visible in the case of Rhineland-Palatinate, Figures 1 and S3).

Once the GP is trained from the ODE trajectories and the data on the fitting interval (in the main results, up to 5 days after school closure, but sensitivity to this choice is explored in Table S2), the same ODE trajectories past the end of the fitting interval are used by the GP to generate the counterfactual. In doing so, we assume that both the structural discrepancy between the ODE trajectories learnt by the GP and the overdispersion in the Negative Binomial observational noise remain unchanged within the counterfactual horizon.

Figure S1 presents an illustration of the various steps that constitute our method. Further details of the method are described in detail in the following sections.

S2.2 The compartmental ODE model

The compartmental model used to generate sample trajectories is outlined in Figure S2. Multiple compartments have been used for the exposed (E) populations to model an Erlang-distributed incubation period compatible with available estimates of the mean and standard deviation of its duration [22]. The early infection compartment I_0 splits into detected (I_d) and undetected (I_u) infectious populations. The same model is used for hospitalisations, with hospitalisations taking the place of I_d, and non-hospitalised cases taking I_u. A higher variability, possibly country-dependent, in the time from onset of symptoms to detection/hospitalisation [22], and limited knowledge on the duration of the infectious period and non-modelled pathways of hospitalised cases suggest a single compartment (i.e. exponential holding time) for these states is a reasonable and parsimonious choice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Parameter	Interpretation	Value	Priors
N_0	Population	Various	N/A
β	Transmission rate	Inferred in ABC fitting	$0.25 \leq \beta \leq 4$
α	Rate of transition through latent period	$1/4.8$ days^{-1} [22]	N/A
k	Proportion of infectious individuals who are detected	Inferred in ABC fitting	$0 \leq k \leq 1$
γ	Rate of transition through infectious period	Inferred in ABC fitting	$0.05 \leq \gamma \leq 1$
δ_d	Rate of removal of detected infectious individuals	Inferred in ABC fitting	$0.03 \leq \delta_d \leq 1$
δ_u	Rate of removal of undetected infectious individuals	Inferred in ABC fitting	$0.03 \leq \delta_u \leq 1$

Table S1. A summary of the parameters used in the ODE model. The only parameters not inferred are the rate of transition through the latent period, α, and the population, N_0. Population data are taken from various official sources [36–39]. The priors for the respective parameters are chosen as uniform distribution with ranges furnished in the last column.

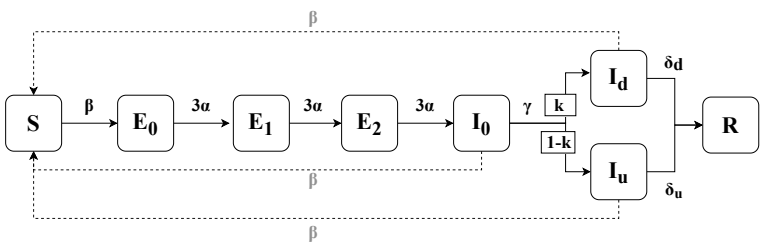


Figure S2. The epidemic model used to simulate cases. The model uses multiple exposed compartments (E_0, E_1, E_2) to account for an Erlang-distributed incubation period.

Denoting by $I = I_0 + I_d + I_u$, the model equations read

$$\frac{dS}{dt} = -\beta S \frac{I}{N_0} \tag{S1}$$

$$\frac{dE_0}{dt} = \beta S \frac{I}{N_0} - 3\alpha E_0 \tag{S2}$$

$$\frac{dE_1}{dt} = 3\alpha(E_0 - E_1) \tag{S3}$$

$$\frac{dE_2}{dt} = 3\alpha(E_1 - E_2) \tag{S4}$$

$$\frac{dI_0}{dt} = 3\alpha E_2 - \gamma I_0 \tag{S5}$$

$$\frac{dI_d}{dt} = k\gamma I_0 - \delta_d I_d \tag{S6}$$

$$\frac{dI_u}{dt} = (1-k)\gamma I_0 - \delta_u I_u \tag{S7}$$

$$\frac{dR}{dt} = \delta_d I_d + \delta_u I_u \tag{S8}$$

The data for Germany are only available starting from the 4th of March, by which time, most states had already seen a number of cases. As such, we have inferred the initial values for states $\Phi = \{E_0(0), E_1(0), E_2(0), I_0(0), I_d(0), I_u(0)\}$. For consistency the same initial conditions have been fitted for Sweden, Denmark and Norway.

Denoting the daily symptomatic case count as $D(t)$, we set the priors on the initial states $\Phi_i \in \Phi$, $i = 1, .., 6$, as $D(0)/4 \leq \Phi_i \leq 4 \cdot D(0)$ for all i.

We denote the unknown parameter vector of the model as: $\boldsymbol{\theta} = (\beta, \gamma, k, \delta_d, \delta_u; \Phi)$. The parameter α is not estimated in this method, and is instead taken from the mean incubation period [22] to be $\alpha = 1/4.8$ days^{-1}. The interpretations of these parameters are provided in Table S1.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

S2.3 Step 1: Approximate Bayesian Computation for fitting the ODE

We used the Approximate Bayesian Computation (ABC) method to carry out Bayesian inference for estimating the ODE parameters, which we describe next. We denote by $y(t)$ the daily cases observed on day t and we regard them as the noisy observation corresponding to the unobserved true number of symptomatic cases $D(t)$ on the same day. Having put a prior distribution $\pi(\boldsymbol{\theta})$ (see Table S1 for a description of the priors) on the parameter vector we can obtain the posterior distribution, using the Bayes' theorem as:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta}) p(\boldsymbol{y}|\boldsymbol{D}^{\boldsymbol{\theta}}, \boldsymbol{\theta}), \tag{S9}$$

where $p(\boldsymbol{y}|\boldsymbol{D}^{\boldsymbol{\theta}}, \boldsymbol{\theta})$ is the likelihood term. The vectors $\boldsymbol{y} = [y(1), \ldots, y(T)]$ and $\boldsymbol{D}^{\boldsymbol{\theta}} = [D(1), \ldots, D(T)]$ denote the observed data and corresponding simulation from the model for a period of $t = 1, \ldots, T$ days, given the parameters $\boldsymbol{\theta}$.

In ABC the likelihood is replaced by a distance function $d(\boldsymbol{y}, \boldsymbol{D}^{\boldsymbol{\theta}})$, between the data, $\boldsymbol{y}$, and the model simulation, $\boldsymbol{D}^{\boldsymbol{\theta}}$. The ABC posterior given by

$$p_\epsilon(\boldsymbol{\theta}|\boldsymbol{y}) := p(\boldsymbol{\theta}|\mathbb{1}\{d(\boldsymbol{y}, \boldsymbol{D}^{\boldsymbol{\theta}}) \leq \epsilon\}), \tag{S10}$$

which depends upon a user defined tolerance threshold ϵ. We can obtain the exact posterior when $\epsilon \to 0$. $\mathbb{1}(\cdot)$ denotes the indicator function. Rather than using a single tolerance threshold, we used the ABC-SMC [40] algorithm that generates a sequence of intermediate distributions, each corresponding to a decreasing tolerance threshold. The final distribution in the sequence is the desired ABC posterior.

We employ a Negative Binomial probability mass function to calculate the probability of observing the data when the unobserved true values are given by the simulated trajectories. That is, for each day t in a set of T data points used for the fit, given the observed value $y(t) \in \boldsymbol{y}$ and the modelled value $\hat{y}(t) \in \boldsymbol{D}^{\boldsymbol{\theta}}$, the probability mass function is given by:

$$P(y(t); \hat{y}(t), k) = \frac{\Gamma(k + y(t))}{\Gamma(k) y(t)!} \left(\frac{k}{k + \hat{y}(t)}\right)^k \left(\frac{\hat{y}(t)}{k + \hat{y}(t)}\right)^{y(t)}, \tag{S11}$$

where k is the over-dispersion parameter, which is also estimated from the data set. The distance function is then defined as:

$$d(\boldsymbol{y}, \boldsymbol{C}^{\boldsymbol{\theta}}) = -\sum_{t=1}^{T} \ln(P(y(t); \hat{y}(t), k)). \tag{S12}$$

S2.3.1 Generating covariates for the GP regression

After obtaining the posterior distribution we generate simulated epidemic curves $\boldsymbol{C_{pre}} := \left[D^{\boldsymbol{\theta}^*}(1), \ldots, D^{\boldsymbol{\theta}^*}(T)\right]$, by solving the compartmental model using samples from the ABC posterior $\boldsymbol{\theta}^* \sim p_\epsilon(\boldsymbol{\theta}|\boldsymbol{y})$ for the period up to 5 days after school closures. Similarly, we generate posterior predictive trajectories $\boldsymbol{C_{post}}$ for the post-intervention period of T^* days. This completes the first step of our method. We then select $M = 15$ simulated epidemic curves $(\boldsymbol{C_{pre}}, \boldsymbol{C_{post}})$ evenly distributed over the Bayesian credible interval of the posterior predictive distribution of $(\boldsymbol{C_{pre}}, \boldsymbol{C_{post}})$, to be used as the respective training and projection covariates for the GP regression which constitutes the second step of our method. Essentially, with these M covariates we are simply summarising the posterior predictive distribution at each time point. Note that these M summaries are not treated as random variables but as fixed covariates. Testing of the method with different numbers of covariates ($M = 5$, $M = 40$ and $M = 150$) has lead to qualitatively and quantitatively similar results, suggesting the method is robust to the choice of M.

S2.4 Step 2: Learning discrepancy using GP regression

In order to generate the unmitigated epidemic, the counterfactual, we first learn a map $\boldsymbol{C_{pre}} \mapsto \boldsymbol{y}$, that represents the discrepancy as a nonlinear function. To achieve this we model the case data $\boldsymbol{y}$ prior to the intervention using a hierarchical latent variable formulation of GP regression with discrete outcomes. At the first level of this hierarchy we model each data-point $y(t)$, independently, using a Negative Binomial distribution given by

$$y(t) \sim \text{NegBin}(\mu(t), \eta), \tag{S13}$$

where η is a dispersion parameter such that $\mathbb{E}[y(t)] = \mu(t)$ and $\text{Var}[\boldsymbol{y}(t)] = \mu(t)(1 + \mu(t)/\eta)$. The mean $\boldsymbol{\mu} = [\mu(1), \ldots, \mu(T)]$ is in turn driven by a latent Gaussian process as:

$$\log \boldsymbol{\mu} \sim \mathcal{GP}\big(\boldsymbol{C_{pre}}\boldsymbol{\beta} + \boldsymbol{b}, \boldsymbol{\kappa}(\boldsymbol{C_{pre}}, \boldsymbol{C_{pre}}'; \boldsymbol{\phi})\big), \tag{S14}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

where we have used a linear regression form to describe the mean function of the GP with coefficients $\boldsymbol{\beta}$, an M-dimensional vector, and intercept term $\boldsymbol{b}$, a T-dimensional vector with elements set to b. This ensures that the GP, while extrapolating further away from the training data, follows the ODE predictions. $\boldsymbol{\kappa}(\boldsymbol{C_{pre}}, \boldsymbol{C_{pre}}'; \boldsymbol{\phi})$ is the covariance function (also known as covariance *kernel*) parameterised by $\boldsymbol{\phi}$, the details of which are described in the next section.

S2.4.1 Capturing the weekend effect

Gaussian processes are considered to be a prior on function spaces [35]. The prior knowledge about a function modelled by the GP is generally introduced through the covariance function. For our problem we expect the function $\log \boldsymbol{\mu}$ to be a smooth yet periodic function. The periodicity assumption is introduced to cater for the weekend effect. To design such a prior we first introduce a new categorical *day-of-week* covariate $\boldsymbol{W_{pre}} = (w(1), \ldots, w(T)), \quad w(t) \in \{1, \ldots, 7\}$, that indicates the day of a week for the T consecutive training days, and $\boldsymbol{W_{post}}$ for the forecasting days. We then replace the covariance of the GP, in Equation (S14), with the sum of two separate covariances:

$$\boldsymbol{\kappa}(\boldsymbol{C_{pre}}, \boldsymbol{C_{pre}}', \boldsymbol{W_{pre}}, \boldsymbol{W_{pre}}'; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) = \boldsymbol{\kappa}_1(\boldsymbol{C_{pre}}, \boldsymbol{C_{pre}}'; \boldsymbol{\phi}_1) + \boldsymbol{\kappa}_2(\boldsymbol{W_{pre}}, \boldsymbol{W_{pre}}'; \boldsymbol{\phi}_2), \tag{S15}$$

based on these different covariates, and having individual parameters $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2$. By choosing valid covariance kernels components $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$, it is ensured that we produce a valid overall covariance kernel $\boldsymbol{\kappa}$.

For the first component we chose the *Matern* kernel [35] given by

$$\boldsymbol{\kappa}_1(\boldsymbol{C_{pre}}, \boldsymbol{C_{pre}}'; \boldsymbol{\phi}_1) = \alpha_1^2 \left(1 + \sum_{j=1}^{M} \frac{\sqrt{3(C_{pre_j} - C'_{pre_j})^2}}{\rho_j}\right) \left(\sum_{j=1}^{M} \exp\left[-\frac{\sqrt{3(C_{pre_j} - C'_{pre_j})^2}}{\rho_j}\right]\right), \tag{S16}$$

where M is the number of covariates, that is the number of simulated trajectories $\boldsymbol{C_{pre}}$ generated using the ABC posterior as described previously. The parameter ρ_j quantifies the characteristic length scale along the jth covariate and α_1 denotes the marginal variance of the GP prior. Together they constitute the parameter vector $\boldsymbol{\phi}_1 = [\alpha_1, \rho_1, \ldots, \rho_M]$ (note that ρ_j is distinct from the quantity $\rho(t)$ used to estimate the instantaneous growth rate for school reopening).

For the second component, we chose the *exponentiated quadratic* kernel:

$$\boldsymbol{\kappa}_2(\boldsymbol{W_{pre}}, \boldsymbol{W_{pre}}'; \boldsymbol{\phi}_2) = \alpha_2^2 \exp\left(-\frac{\|\boldsymbol{W_{pre}} - \boldsymbol{W_{pre}}'\|^2}{2\rho^2}\right), \tag{S17}$$

where $\boldsymbol{\phi}_2 = (\alpha_2, \rho)$ are the covariance parameters.

S2.4.2 Estimation of the GP parameters using MCMC

The latent GP represents the second level in the hierarchy and finally we place priors $p(\boldsymbol{\beta}, b, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ on all the parameters to completely specify the generative model. Thus, the likelihood corresponding to all T observations is given by

$$p(\boldsymbol{y}|\boldsymbol{C_{pre}}, \boldsymbol{W_{pre}}, \boldsymbol{\beta}, b, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \eta) = \Big(\prod_{t=1}^{T} \text{NegBin}(e^{\mu(t)}, \eta)\Big)\mathcal{N}(\boldsymbol{\mu}; \boldsymbol{C_{pre}}\boldsymbol{\beta} + \boldsymbol{b}, \boldsymbol{\Sigma}_{TT}), \tag{S18}$$

where $\boldsymbol{\Sigma}_{TT}$ represents the covariance function (Eq. (S16)) evaluated on all $T \times T$ pairs of the covariates $\boldsymbol{C_{pre}}$ and $\boldsymbol{W_{pre}}$:

$$\boldsymbol{\Sigma}_{TT} = \begin{pmatrix} \boldsymbol{\kappa}(\boldsymbol{C_{pre}}^1, \boldsymbol{C_{pre}}^1, \boldsymbol{W_{pre}}^1, \boldsymbol{W_{pre}}^1; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) & \cdots & \boldsymbol{\kappa}(\boldsymbol{C_{pre}}^1, \boldsymbol{C_{pre}}^T, \boldsymbol{W_{pre}}^1, \boldsymbol{W_{pre}}^T; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\kappa}(\boldsymbol{C_{pre}}^T, \boldsymbol{C_{pre}}^1, \boldsymbol{W_{pre}}^T, \boldsymbol{W_{pre}}^1; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) & \cdots & \boldsymbol{\kappa}(\boldsymbol{C_{pre}}^T, \boldsymbol{C_{pre}}^T, \boldsymbol{W_{pre}}^T, \boldsymbol{W_{pre}}^T; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) \end{pmatrix}. \tag{S19}$$

We then combine the likelihood with the priors to obtain the posterior distribution of the parameters given by

$$p(\boldsymbol{\beta}, b, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \eta|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{C_{pre}}, \boldsymbol{W_{pre}}\boldsymbol{\beta}, b, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2)p(\boldsymbol{\beta}, b, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \eta). \tag{S20}$$

We used MCMC, applying the No-U-Turn (NUTS) [41] algorithm provided by the PyMC3 package, to sample from this posterior. Note that to apply MCMC we need to sample the latent Gaussian process for the mean $\boldsymbol{\mu}$. It is well known

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(see [42]) that sampling both the GP parameters $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2$ as well as the latent GP itself drastically reduces the efficiency of the MCMC algorithm and poor mixing is observed. For this reason a non-centred reparameterisation is used which gives the GP sample as:

$$\boldsymbol{\mu} = \boldsymbol{C_{pre}}\boldsymbol{\beta} + \boldsymbol{b} + \boldsymbol{L}_{TT}\boldsymbol{\epsilon}, \tag{S21}$$

where, $\boldsymbol{L}_{TT}$ is obtained through the Cholesky decomposition of $\boldsymbol{\Sigma}_{TT}$ and $\boldsymbol{\epsilon} \in \mathbb{R}^T$ is a sample from the standard Normal density $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ ($\boldsymbol{I}$ being the identity matrix).

S2.4.3 Projecting the counterfactual

Having estimated the posterior $p(\boldsymbol{\beta}, b, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2|\boldsymbol{y})$ we can obtain the posterior predictive distribution of the daily cases, in the case of no intervention, $\boldsymbol{y^*} = [y^*(1), \ldots, y^*(T^*)]$, where T^* is the number of days we want to project after the intervention date. This distribution is given by

$$p(\boldsymbol{y^*}|\boldsymbol{C_{pre}}, \boldsymbol{C_{post}}, \boldsymbol{W_{pre}}, \boldsymbol{W_{post}}, \boldsymbol{y}) = \int \Big(\prod_{t=1}^{T^*} \text{NegBin}(e^{\mu(t)}, \eta) \Big) \mathcal{N}(\boldsymbol{\mu}; \boldsymbol{m}, \boldsymbol{V}) p(\boldsymbol{\beta}, b, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2|\boldsymbol{y}) d\boldsymbol{\mu} d\boldsymbol{\beta} db d\boldsymbol{\phi}_1 d\boldsymbol{\phi}_2 d\eta, \tag{S22}$$

where the posterior mean $\boldsymbol{m}$ and variance $\boldsymbol{V}$ of the GP is given by [35]:

$$\begin{aligned} \boldsymbol{m} &= (\boldsymbol{C_{post}}\boldsymbol{\beta} + \boldsymbol{b}) + \boldsymbol{\Sigma}_{T^*T}\boldsymbol{\Sigma}_{TT}^{-1}(\boldsymbol{y} - \boldsymbol{C_{pre}}\boldsymbol{\beta} - \boldsymbol{b}) \\ \boldsymbol{V} &= \boldsymbol{\Sigma}_{T^*T^*} - \boldsymbol{\Sigma}_{T^*T}\boldsymbol{\Sigma}_{TT}^{-1}\boldsymbol{\Sigma}_{TT^*}. \end{aligned} \tag{S23}$$

$\boldsymbol{\Sigma}_{T^*T}$ and $\boldsymbol{\Sigma}_{TT^*}$ denote the $T^* \times T$ and $T \times T^*$ matrices of covariance function evaluations between the training (pre-intervention) $\boldsymbol{C_{pre}}, \boldsymbol{W_{pre}}$ and projection (post-intervention) $\boldsymbol{C_{post}}, \boldsymbol{W_{post}}$ covariates:

$$\boldsymbol{\Sigma}_{T^*T} = \begin{pmatrix} \boldsymbol{\kappa}(\boldsymbol{C_{post}}^1, \boldsymbol{C_{pre}}^1, \boldsymbol{W_{post}}^1, \boldsymbol{W_{pre}}^1; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) & \cdots & \boldsymbol{\kappa}(\boldsymbol{C_{post}}^1, \boldsymbol{C_{pre}}^T, \boldsymbol{W_{post}}^1, \boldsymbol{W_{pre}}^T; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\kappa}(\boldsymbol{C_{post}}^{T^*}, \boldsymbol{C_{pre}}^1, \boldsymbol{W_{post}}^{T^*}, \boldsymbol{W_{pre}}^1; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) & \cdots & \boldsymbol{\kappa}(\boldsymbol{C_{post}}^{T^*}, \boldsymbol{C_{pre}}^T, \boldsymbol{W_{post}}^{T^*}, \boldsymbol{W_{pre}}^T; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) \end{pmatrix}, \tag{S24}$$

$$\boldsymbol{\Sigma}_{TT^*} = \begin{pmatrix} \boldsymbol{\kappa}(\boldsymbol{C_{pre}}^1, \boldsymbol{C_{post}}^1, \boldsymbol{W_{pre}}^1, \boldsymbol{W_{post}}^1; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) & \cdots & \boldsymbol{\kappa}(\boldsymbol{C_{pre}}^1, \boldsymbol{C_{post}}^{T^*}, \boldsymbol{W_{pre}}^1, \boldsymbol{W_{post}}^{T^*}; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\kappa}(\boldsymbol{C_{pre}}^T, \boldsymbol{C_{post}}^1, \boldsymbol{W_{pre}}^T, \boldsymbol{W_{post}}^1; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) & \cdots & \boldsymbol{\kappa}(\boldsymbol{C_{pre}}^T, \boldsymbol{C_{post}}^{T^*}, \boldsymbol{W_{pre}}^T, \boldsymbol{W_{post}}^{T^*}; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) \end{pmatrix}, \tag{S25}$$

and $\boldsymbol{\Sigma}_{T^*T^*}$ is the covariance evaluated at the projection inputs $\boldsymbol{C_{post}}$ only:

$$\boldsymbol{\Sigma}_{T^*T^*} = \begin{pmatrix} \boldsymbol{\kappa}(\boldsymbol{C_{post}}^1, \boldsymbol{C_{post}}^1, \boldsymbol{W_{post}}^1, \boldsymbol{W_{post}}^1; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) & \cdots & \boldsymbol{\kappa}(\boldsymbol{C_{post}}^1, \boldsymbol{C_{post}}^{T^*}, \boldsymbol{W_{post}}^1, \boldsymbol{W_{post}}^{T^*}; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\kappa}(\boldsymbol{C_{post}}^{T^*}, \boldsymbol{C_{post}}^1, \boldsymbol{W_{post}}^{T^*}, \boldsymbol{W_{post}}^1; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) & \cdots & \boldsymbol{\kappa}(\boldsymbol{C_{post}}^{T^*}, \boldsymbol{C_{post}}^{T^*}, \boldsymbol{W_{post}}^{T^*}, \boldsymbol{W_{post}}^{T^*}; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2) \end{pmatrix}. \tag{S26}$$

We use Monte Carlo integration, using the samples of $\boldsymbol{\beta}, b, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \eta$ obtained through MCMC and samples of $\boldsymbol{\mu}$ generated using the aforementioned reparameterisation, to implicitly evaluate the desired posterior predictive distribution, given by Eq. (S22), and subsequently the counterfactual epidemic projections.

S2.5 Priors for the GP and MCMC details

For each of the coefficients $\boldsymbol{\beta}$ and the intercept b parameters we chose a Student's t-distribution with 5 degrees of freedom. The mean is set to 0 and the standard deviation is set to 1. This was done following the recommendations in [43] for choosing priors for regression (since we modelled the GP's mean function using a linear regression form).

We used a half-Cauchy prior with scale set to 5 for the marginal variances α_1, α_2, following recommendations in [44]. For each characteristic length scale parameter we have placed a Gamma$(2, 1)$ prior.
For the dispersion parameter, η, of the negative binomial we placed a Gamma$(2, 0.5)$ prior.

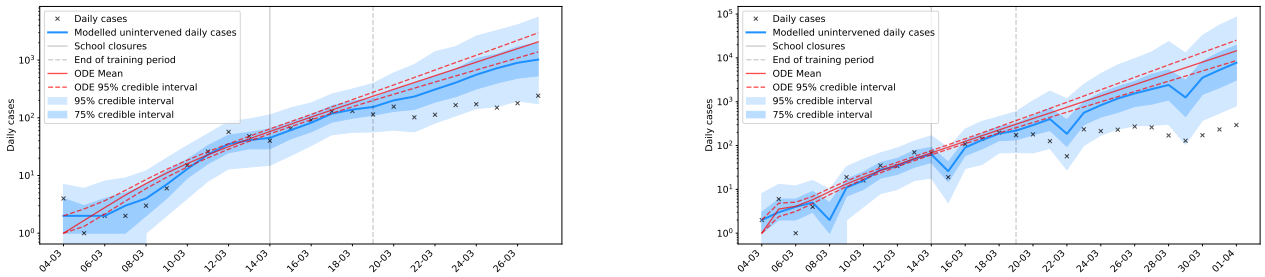
We have also used a logarithmic transform of $\boldsymbol{C_{pre}}, \boldsymbol{C_{post}}$ before passing them to the GP.

We ran NUTS with 2 chains for a total 1000 iterations, where we discarded the first 500 iterations as burnin (warmup). We set a target acceptance rate of 0.98 for NUTS during the warmup phase iterations. As with the SEIR model fit, the GP model is trained with data up to 5 days after school closure.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

S2.6 Projection comparison: the benefit of using the GP

Figure S3 shows the sample regression outputs in two German states, plotted alongside the mean and 95% credible intervals of the ABC posterior projections (simulations from the fitted ODE model) used as covariates in the GP model. In both cases, the ODE model fails to capture the change in growth rate, as well as the weekend effect (which is visibly more pronounced in one state than the other). The post-processing with the GP regression compensates for this failure by learning the discrepancy as a function mapping the ODE simulation to the observations.



(a) Gaussian Process and ABC fit for Rhineland-Palatinate. **(b)** Gaussian process and ABC fit for Hesse.

Figure S3. In (a) the ABC fitting process failed to account for a change in growth rate in the observed data directly before intervention. The GP method, however, is able to capture this change in the training period and adapt the projection accordingly. Similarly, in (b) the GP model is able to pick up both the slight change in growth rate and the strong weekend effect in the data, factors which are not addressed by the ABC fitting process.

S2.7 Sensitivity of the lag time to the training period

Due to the noisiness of the data, our findings will inevitably be sensitive to the number of data points which we include in any analysis we do. This is most apparent in the fitting and training periods of the ABC and GP, which in this work extend to 5 days past school closure. This duration is chosen as it is the mean incubation period for SARS-CoV-2 [22], and thus we would not expect any signal during those days to be strongly attributable to school closures. Five days is thus our *minimal lag time*, which acts as an input to our models. While a full sensitivity analysis of our modelling choices is beyond the scope of our investigation, we include some choice illustrations of the effects of changing this minimal lag time in Table S2:

State	Minimal lag time of 4 days	Minimal lag time of 5 days (in manuscript)	Minimal lag time of 6 days
Baden-Württemberg	6	8	8
Bavaria	12	8	12
Berlin	5	-	-
Hesse	6	7	7
Lower Saxony	6	7	7
North Rhine-Westphalia	6	6	7
Rhineland-Palatinate	5	7	7

Table S2. Lag times (in days) for different German states using various minimal lag times for the ABC fit and GP training period. The standard choice employed in the findings of the manuscript is 5 days.

In general, we note that a shorter minimal lag time gives a faster response in the data, and consequently a shorter lag time. This is not surprising, as we allow for a greater attribution of effects to school closures which may be the result of earlier interventions. Reassuringly, we find that extending the minimal lag time by a day, thereby further limiting the

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

attribution of any response to school closures, does not significantly change the lag time. This alleviates some concern that 5 days, being solely the mean of a distribution of incubation times, is insufficient for the method to adequately capture the pre-intervention trajectory. Expanding the sensitivity analysis to even longer minimal lag times is not meaningful: most lag times in the manuscript are 7 days, and by training the GP up until this point, any resulting lag times would likely be forced to be even longer. This amounts to forcing a signal, rather than searching for possible variability in the signal within a reasonably expected range. The effect of this can be seen in North Rhine-Westphalia.

In Berlin, we find that a shorter minimal lag time of 4 days does allow for a deviation to be seen already on the fifth day when comparing the projected and observed trajectories. Inspection of Figure S11 shows that there is a very gradual approach to the peak, which explains why more data points' signal needs to be attributed to school closures before a lag time can be determined.
We encounter a similar limitation with Bavaria. In Figure S10, the data points from the 22nd to at least the 25th of March are very close to the boundary of the 75% credible intervals of the GP. Although a visual inspection confirms that an 8 day lag time is consistent with a change in the data, small changes to the minimal lag can bring those observed points to fall just inside the boundary. When a 12 day lag time is analysed for Bavaria the relative reduction in the growth rate is 72%, which is consistent with the other German states. The effect of interventions is consistently impacted in the state: either a smaller reduction the growth rate is seen compared to other states, or a similar reduction is obtained, but after a longer duration of time.

S3 Testing data

S3.1 Testing data for Germany

Figure S4 shows the number of tests carried out per day in German medical laboratories, along with the positive test ratio over the same period. This is not equivalent to the total number of tests carried out in Germany, as not all laboratories provided this type of data; however, it can be used as an indication of general testing trends. There is a weekend effect occurring in the testing data for Germany, with lower relative testing occurring on March 7th-9th, 14th-16th and 21st-23rd. No corresponding change is seen in the positive test ratio, indicating that case numbers were likely consistent across these periods. As such, any fall in confirmed cases over these periods can likely be attributed to reduced testing, rather than a response to intervention. Ignoring the weekend effect, the number of tests carried out across the period between March 17th and 27th was fairly stable. As most school closures in Germany occurred on March 16th, we can expect the confirmed cases over this period to provide a reasonable representation of the underlying epidemic.

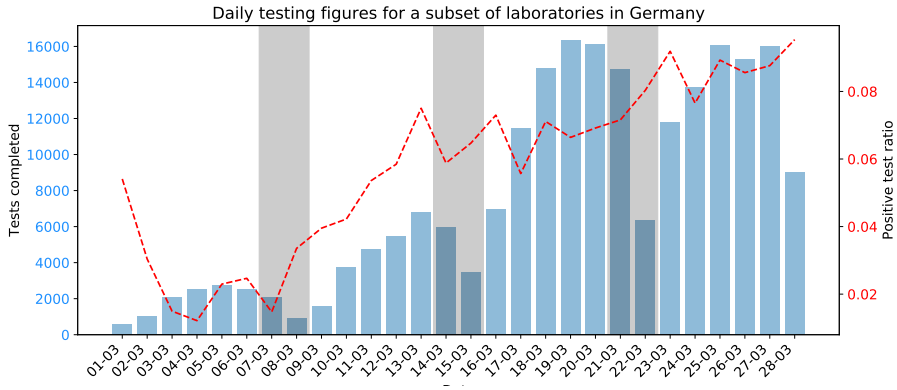


Figure S4. Daily testing from a subset of German testing laboratories during March. Weekends are highlighted in grey. There is a periodic drop in testing occurring on weekends, particularly evident on Sundays. These drops do not coincide with any changes to the positive test ratio.

Daily testing data for Germany are not available after March. As such, it will be necessary to consider the weekly testing totals, which are made available through the RKI. Figure S5 shows the weekly testing numbers for Germany, along with the weekly positive test ratio. Note that weekly testing data for Germany are released every Wednesday.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

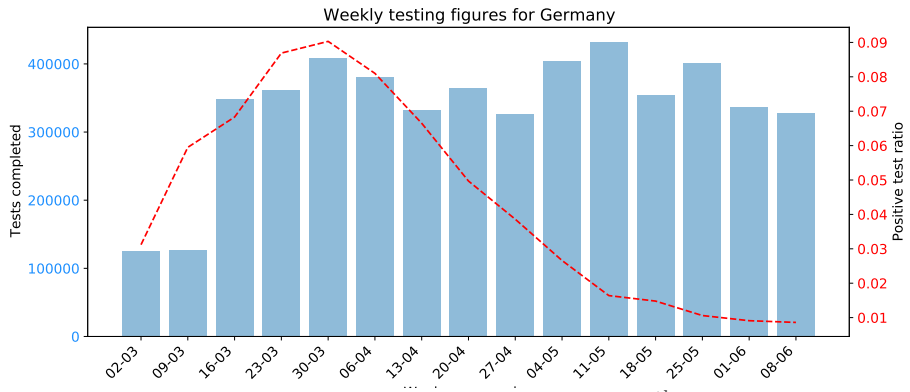


Figure S5. Weekly testing in Germany remained consistent from March 18th, however the weekend effect (see Figure S4) was likely present across the entire period. There were no abrupt changes in the positive test ratio.

S3.2 Testing data for Scandinavia

Both Denmark and Norway saw a similar weekend effect in testing numbers, with midweek testing figures roughly 50% higher than weekend figures in Denmark, and almost three times higher in Norway. The weekly testing figures for Denmark and Norway are shown in Figure S6 and S7 respectively.

Denmark displays two clear increases in testing capacity between March 23rd and 30th and again between April 13th and 20th. The increase in late March, combined with a relatively high positive test ratio, indicates that confirmed cases during this period might not be a suitable metric.

Similarly in Norway there was a large increase in testing in the week commencing March 16th, very close to the date of school closures. As such, for both Norway and Denmark it will be necessary to consider hospitalisations as a metric for assessing the dynamics of the epidemic.

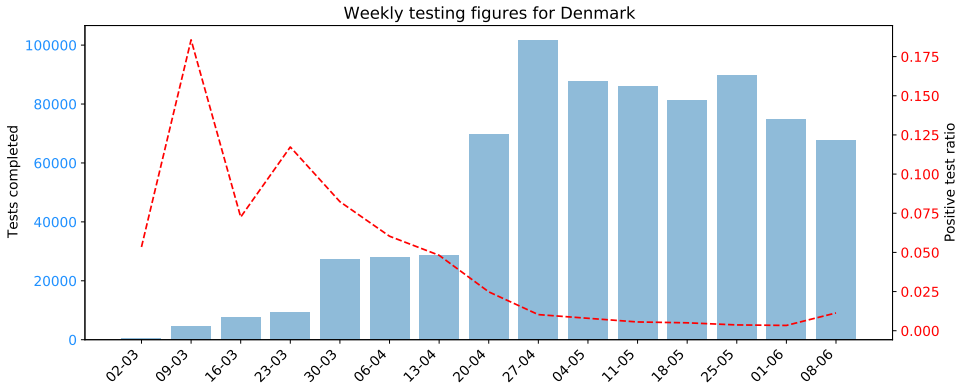


Figure S6. Weekly testing in Denmark was not consistent across the period of this investigation, and so confirmed cases up to April 20th cannot be relied upon to provide a reliable representation of the underlying epidemic.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

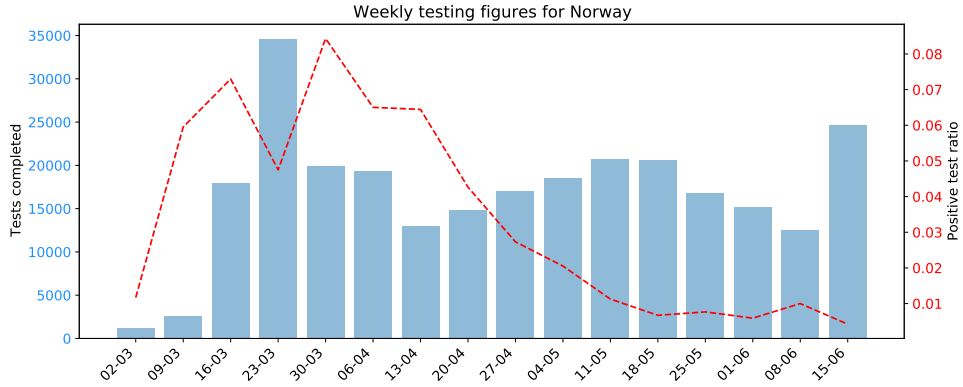


Figure S7. Norway saw inconsistent testing during March, making confirmed cases an inappropriate metric for assessing school closures. More consistent testing was apparent in April and May.

The weekly testing figures for Sweden are highlighted in Figure S8, along with the positive test ratio for the same period. Testing rates around the time of school closures (March 18th) were generally increasing, with a large increase occurring during the week beginning March 30th. This increase was accompanied by an increase in positive test ratio, indicating an increasing capability to identify and test infected individuals, or a change in testing policy. As a result, it will not be possible to attribute any change in the increase in case numbers after March 30th solely to the effect of interventions.

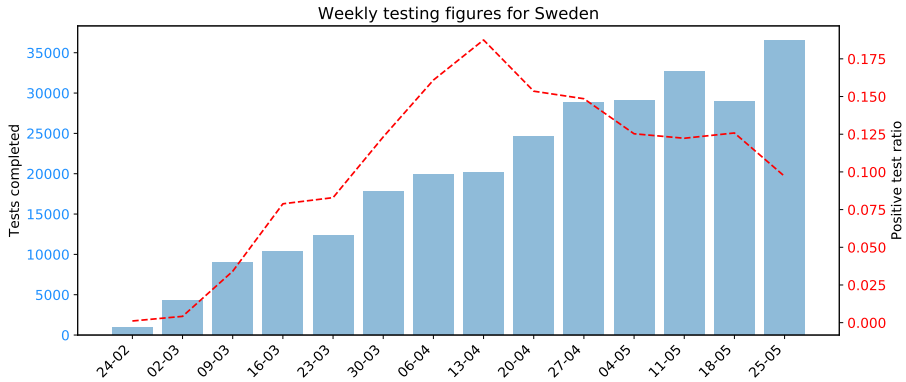


Figure S8. Reported weekly tests carried out in Sweden. These numbers rose throughout March and April.

Despite the changes in testing rates around the time of school closures, Sweden is still able to provide very useful insight. The decision by the Swedish government to (a) leave schools open to all students under the age of 16 and (b) do so with a background of limited social interventions is useful for partially decoupling the effect of school closures from other controls.

S4 School closure analyses

We present the equivalent of Table 1 in the manuscript, but expressed via the doubling time (ln(2)/growth rate).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

State	Lag time (days)	Pre-response doubling time (days)	Modelled post-response doubling time (days)	Observed post-response doubling time (days)
Baden-Württemberg	8	4.9 (3.8 - 6.6)	4.2 (3.7 - 4.7)	13.5 (7.7 - 51.9)
Bavaria	8	3.2 (2.7 - 3.9)	3.2 (3.1 - 3.3)	6.3 (4.7 - 9.6)
Berlin	–*	4.8 (3.7 - 6.7)	–	–
Hesse	7	2.8 (2.3 - 3.6)	2.5 (2.4 - 2.6)	10.3 (5.9 - 40.0)
Lower Saxony	7	3.1 (2.6 - 3.9)	3.0 (2.8 - 3.3)	10.0 (6.5 - 21.9)
North Rhine-Westphalia	6	3.6 (3.0 - 4.4)	3.4 (3.3 - 3.5)	11.3 (7.2 - 26.3)
Rhineland-Palatinate	7	2.7 (2.2 - 3.4)	3.0 (2.7 - 3.3)	16.0 (8.1 - 1188.4)

Table S3. Comparison of estimated lag time and pre- and post-intervention doubling times in different German states. *The peak in daily incidence is reached before a response is seen in the data. A lag time which may be attributable to school closures therefore cannot be determined.

In Figure S9 we compare states which closed schools at different times to examine the robustness of the response time being associated with school closures, and not a fixture particular to a certain day. The comparison between states with different dates of school closure is limited to Baden-Württemberg and North Rhine-Westphalia for a number of reasons. Firstly, under the hypothesis that the effectiveness of interventions, or the timing of a response, would depend on the community prevalence, we are limited to comparing Baden-Württemberg with Bavaria, and North Rhine-Westphalia. However, Bavaria imposed stay-at-home orders and held local elections on different dates than the rest of the analysed states. In order to fix as many other factors as possible to be constants, we are left only with the comparison between Baden-Württemberg and North Rhine-Westphalia.
A similar comparison using Berlin, which also closed schools on the 17th of March was not possible, due to the lack of a lag time resulting from the analysis. Furthermore, Berlin had a different timeline of interventions than other states, making for an unjust comparison.

We shift the cases in Baden-Württemberg back in time by three days, so as to coincide the dates of effective school closure (March 14th and 17th). Further, to aid comparison we rescale the cases in Baden-Württemberg by a multiplicative constant (approximately 0.76), so that the cases in both states are identical on the day of school closure. We note that the pure exponential growth rate is unchanged by these transformations. The profiles of the data can now be compared.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

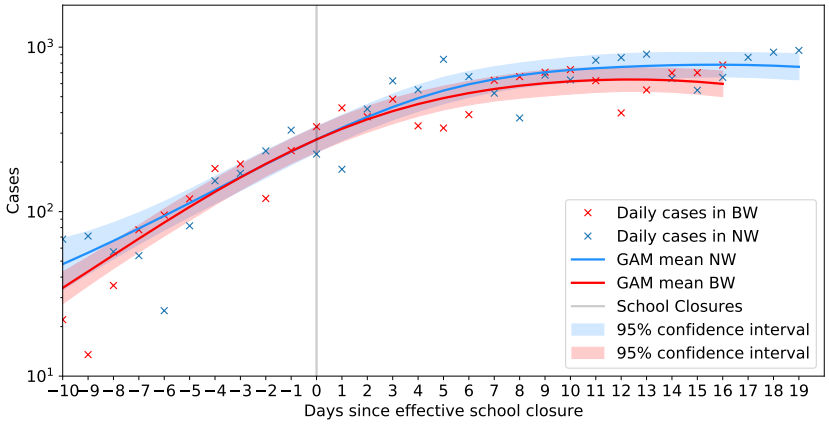


Figure S9. Daily cases for Baden-Württemberg and North Rhine-Westphalia when corrected for the three-day shift in school closure between the two. The effective day of school closure in both states is shown by the solid grey line. The smoothed trajectory, obtained via a GAM analysis, is shown by the solid lines with the shaded regions representing the 95% confidence intervals. The data for North Rhine-Westphalia are scaled so that the GAM smoothed incidence are identical on the day of closures.
There is very good agreement between the two data streams despite the time difference in the school closure, suggesting comparable underlying transmission in the two states following school closure. Additionally the lag times are comparable between the two states.

The two states appear to be comparable both in terms of the overall cases following school closure, as well as the time taken until a response from an intervention can be observed in the data. While the lag times are not identical, sensitivity analysis (see Table S2) suggests that these lag times overlap. Furthermore, it is reassuring that there does not seem to be a fixed date following school closures when a response is seen. Were this the case, we would expect Baden-Württemberg to have a lag time which was 3 days shorter than North Rhine-Westphalia. This suggests a signal in both data sets following school closure. Clearly it would be unrealistic to assume school closures to be wholly responsible for the observed fall in cases, but the above-detailed correlations suggest that they may have partially contributed to the overall effect.

Bavaria

Bavaria saw a sustained decrease in epidemic growth rate occurs 8 days after school closures (Figure S10). Bavaria saw the following interventions around the time of school closures:

- 10/03 - Banned gatherings of more than 1000 people (DE-G1).
- 14/03 - **School closures** (effective date, DE-S1).
- 15/03 - Local elections went ahead, with a high turnout (BY-P1). A large number of votes were submitted by post.
- 16/03 - State of emergency declared; shut borders with France (FR), Switzerland (CH), Austria (AT), Denmark (DK) and Luxembourg (LU) (DE-B1); closure of non-essential businesses and public services (DE-P1).
- 17/03 - Shut borders with EU (DE-B2).
- 21/03 - Stay at home orders, with exceptions for essential trips, and banned gatherings of more than 2 people (BY-G1);
- 22/03 - Banned all social events and gatherings (DE-P2); closure of non-essential retail and leisure, with exceptions for restaurant takeout (DE-R1).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

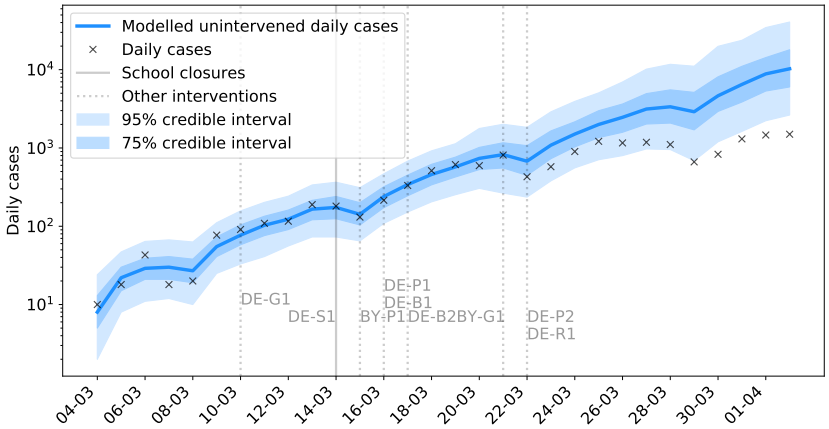


Figure S10. Modelled and observed cases in Bavaria.

Berlin

Berlin did not see a response which may be partially attributable to school closures, as a peak in daily cases was reached before a signal was observed in the data (Figure S11). This suggests that the subsequent decline in cases in the state was brought about by other interventions. Testing rates throughout this period were consistent, indicating that confirmed cases are likely a good indicator of general trends in the epidemic (Figure S11).

Berlin saw the following interventions around the time of school closures:

- 10/03 - Banned gatherings of more than 1000 people (DE-G1).
- 14/03 - Banned events with more than 50 people (BE-G1).
- 16/03 - Shut borders with France (FR), Switzerland (CH), Austria (AT), Denmark (DK) and Luxembourg (LU) (DE-B1); closure of non-essential business and public service (DE-P1).
- 17/03 - **School closures** (DE-S2); shut borders with EU (DE-B2).
- 22/03 - National stay at home orders, with exceptions for essential trips, and banned gatherings of more than 2 people (DE-G2); banned all social events and gatherings (DE-P2); closure of non-essential retail and leisure, with exceptions for restaurant takeout (DE-R1).

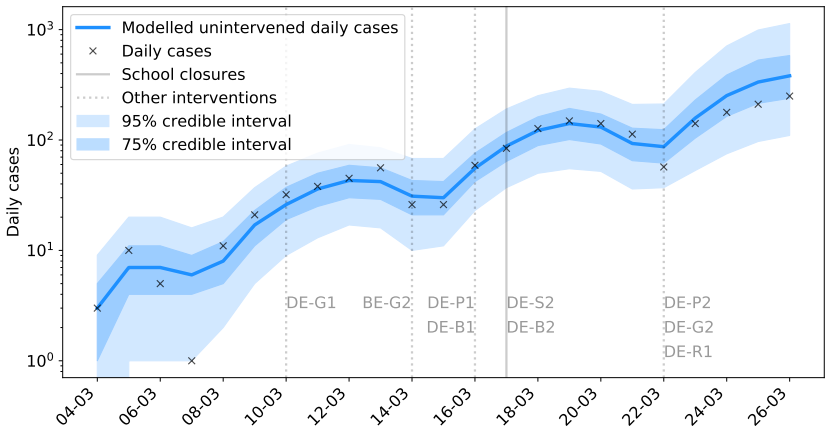


Figure S11. Modelled and observed cases in Berlin.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

North Rhine-Westphalia

North Rhine-Westphalia saw a sustained drop in epidemic growth rate occurs on the 20th, 6 days after school closures (Figure S12).

- 05/03 - **Local school closures in Heinsberg.**
- 10/03 - Banned gatherings of more than 1000 people (DE-G1).
- 14/03 - **School closures** (effective date, DE-S1).
- 16/03 - Shut borders with France (FR), Switzerland (CH), Austria (AT), Denmark (DK) and Luxembourg (LU) (DE-B1); closure of non-essential business and public service (DE-P1).
- 17/03 - Shut borders with EU (DE-B2).
- 22/03 - National stay at home orders, with exceptions for essential trips, and banned gatherings of more than 2 people (DE-G2); banned all social events and gatherings (DE-P2); closure of non-essential retail and leisure, with exceptions for restaurant takeout (DE-R1).

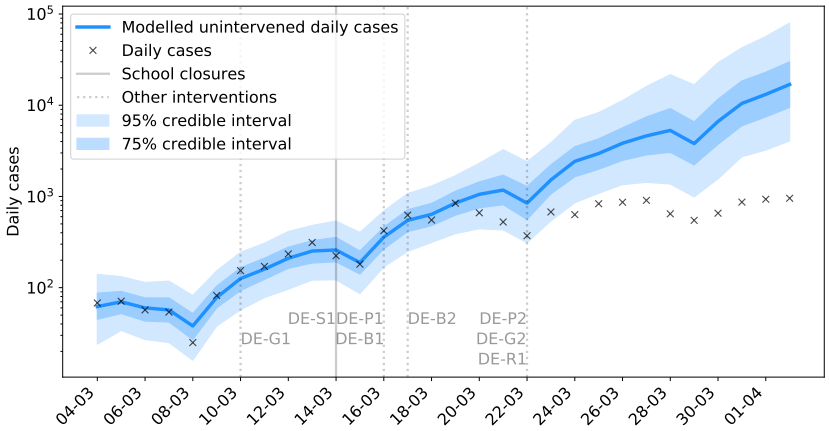


Figure S12. Modelled and observed cases in North Rhine-Westphalia.

Denmark

Denmark saw a staged closing of schools, with primary school closing on Friday 13th of March, and all other schools following on Monday the 16th of March. The effective date of secondary school closures is taken to be Saturday 14th of March.
While there is a decrease in growth rate in the period following school closures (Figure S13), it is not attributable to school closures, as the peak in hospitalisations occurs much sooner than the expected time from infection to hospitalisation (10-14 days).

Denmark saw the following interventions introduced around the same time as school closures:

- 06/03 - Recommendation against events with more than 1000 people (DK-G1).
- 11/03 - Recommendation against public gatherings of more than 100 people (DK-G2).
- 13/03 - **School closure for students aged 16 or over** (DK-S1); closure of non-essential businesses, public sports or leisure facilities, and cultural institutions (DK-P1); work from home order for non-essential workers, with exceptions for essential travel (DK-G3).
- 14/03 - **School closures** (effective date, DK-S2); borders closed (DK-B1).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

- 18/03 - Banned gatherings of more than 10 people (DK-G4); closed non-essential retail, with exceptions for takeaway and food delivery (DK-R1).

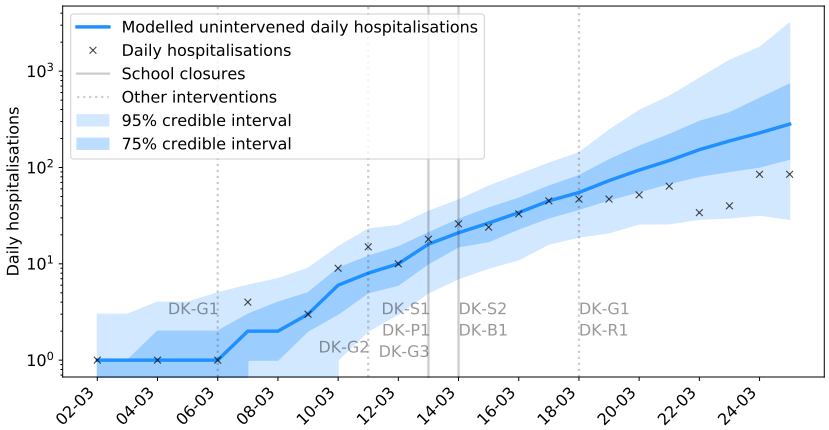


Figure S13. Modelled and observed daily hospitalisations in Denmark. The peak in daily hospitalisations occurs too soon after school closures to suggest that those interventions could have contributed to the observed reduction in the growth rate of cases.

Norway

Norway closed schools at the same time as introducing a range of other restrictions on social life. As such it is not possible to attribute the observed change in hospital admissions solely to school closures. It is notable, however, that the observed reduction in hospitalisations is comparable in Denmark and Norway; both countries which simultaneously targeted schools and non-essential businesses.

For completeness, we fit the GP model to daily hospitalisations (Figure S14). However the short training period for the model resulted in a wide credible interval for this projection. Furthermore, as with Denmark, the proximity of school closures to the point of peak incidence makes it difficult to adequately assess either the lag time or the change in growth rate occurring after closures. As such, it is difficult to draw any firm conclusions from the data.

Norway saw the following interventions introduced around the same time as school closures:

- 08/03 - Docking restrictions for large ships (NO-B1).
- 10/03 - Flight restrictions (NO-B2); working from home encouraged (NO-G1).
- 11/03 - Recommendation against gatherings with more than 500 people (NO-G2).
- 12/03 - **School closures** (NO-S1); closure of non-essential businesses, public leisure facilities, and non-essential retail; essential travel only (NO-G3). Exceptions were made for venues serving food, where social distancing could be observed.
- 14/03 - Recommendations against foreign travel (NO-B2).
- 16/03 - Border and entry restrictions (NO-B3).
- 24/03 - Banned outdoor gatherings of more than 5 people (NO-G4).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

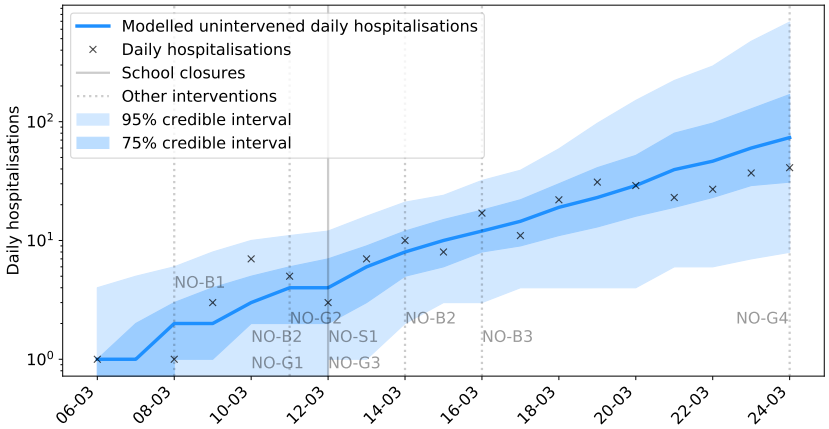


Figure S14. Modelled and observed hospitalisations in Norway.

S5 School reopening analyses

We here present results in support of the main results of the paper, but which are not essential to the exposition of our findings.

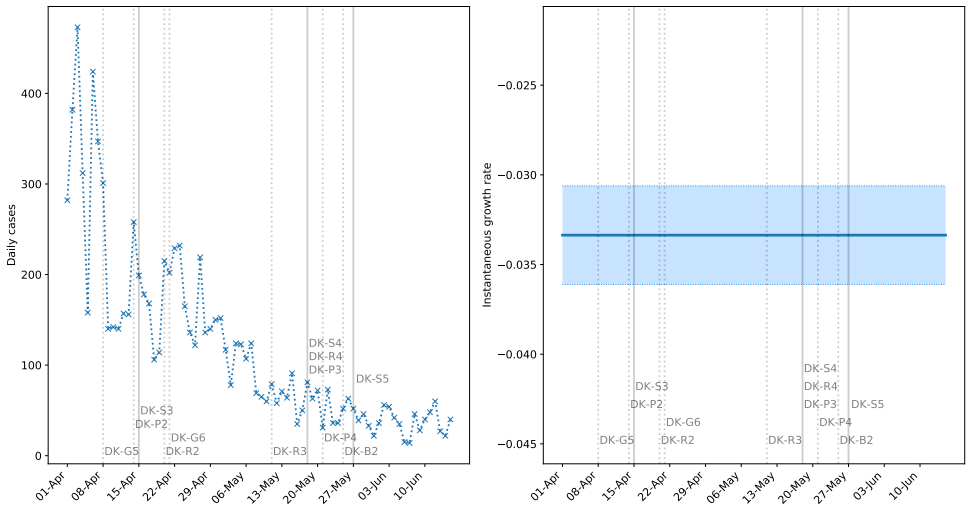


Figure S15. Reported daily confirmed cases in Denmark. The left panel shows new cases, and the right panel shows the instantaneous growth rate (shaded regions are 95% confidence intervals). Solid vertical lines indicate when students returned to school, and dashed lines indicate other loosened measures. We present these numbers in support of the observations made for daily hospital admissions due to the larger numbers of cases recorded here. These results are not qualitatively different from those obtained from hospitalisation data, but support the conclusions which are harder to draw from that data set due to the longer delay from infection to hospitalisation.
We refrain from interpreting positive cases before early May due the increase in testing from the 20th of April.

Sentinel survey information indicates the following for staff in different educational settings. A smaller proportion of staff working with young children have tested positive compared to staff working with older students. However, these numbers alone do not distinguish between infection acquired from the students, and infection acquired elsewhere (Table S4).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.24.20139634; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Educational Level	**Tested Population (%)**	**Positive Tests (%)**	**Tests**
Nursery	9.63	1.18	593
Kindergarten	12.85	0.90	2773
Primary school (ages 7 to 15/16)	13.36	1.23	14855
Secondary school (ages 16 to 19)	8.79	1.35	3343
Higher education	8.30	1.85	3354
Adult education	13.22	1.43	2875

Table S4. A comparison of tests carried out among staff working in different stages of the Danish educational and childcare sector dated June 2nd. We indicate the proportion of tested staff relative to estimated employee numbers in each group, and the percentage of those tested who test positive. For reference, the absolute numbers of tests are also shown.