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

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

1 Department of Mathematics, University of Manchester, UK
2 Emergency Response Department, Public Health England, UK
3 MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.
4 Laboratory of Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto Ontario, Canada.

*** joseph.shingleton@phe.gov, 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 the effect of school closure is visible as a reduction in the growth rate approximately 9 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. A large-scale reopening of schools while controlling or suppressing the epidemic appears feasible in countries such as Denmark or Norway, where community transmission is generally low. However, school reopening can contribute to significant increases in the growth rate in countries like Germany, where community transmission is relatively high. Our findings underscore the need for a cautious evaluation of reopening strategies that ensure low classroom occupancy and a solid infrastructure to quickly identify and isolate new infections.

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 to protect 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 [4] and can drive the epidemic [5, 6]. Consequently, school closure is often one of the first measures considered when non-pharmaceutical interventions are needed to curb the spread of an epidemic. However, it is often implemented concurrently with other measures, making it difficult to assess its individual impact. 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 condition which must be met on a community level before a return can be deemed safe. Some work has already addressed these questions from a theoretical perspective of scenario planning [7], a valuable means of quantifying the expected impact of various measures without unnecessarily exposing staff and students to an increased risk of infection. In this work we present a complementary data-based approach which focuses on the effects of school closure and reopening in Denmark, Germany, Norway, and Sweden. We hope that our contribution can serve as a series of lessons learned from the outcomes of nations who have already reopened schools. Clearly, modelling is essential in informing future decisions when faced with numerous possible actions. However, we believe our work fills a clear 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 epidemiological interventions. School closure and reopening not only affect transmission occurring on the premises; they also affect (and are affected by) the community transmission, the transmission within households with young children, and the measures taken to monitor and curb the outbreak. 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, the varying age groups of the affected school children, and the variability in the timings of these interventions. Where possible, we wish to examine roles in transmission played by a) different age cohorts of students, b) the timing of the school interventions (closure and reopening), and c) the background or community incidence of each country. We distinguish between countries with high and low levels of community transmission on the basis of daily COVID-19 cases, rather than cases per capita. This is motivated by the feasibility of testing, tracing, and isolating cases, which need not scale with population size.

Methods

Data selection criteria

Data have been selected with the intention of studying a broad range of interventions. However, care has been taken to ensure that the data are representative of the underlying epidemic, and that incidence numbers at the time of intervention are large enough so that proper conclusions can be drawn without any signal being dominated by stochastic effects.

Where possible, we consider hospital admissions as the primary data source. This is a reasonably practical incidence measure and, unlike confirmed cases, is not as susceptible to variable testing rates in the wider population. However, it should be noted that hospitalisation only tells us about the specific subset of the population which is hospitalised. There are likely biases in such data toward older and sicker individuals. This could lead to longer than expected delays from school closures to an observed reduction in new admissions, as hospitalisations are more likely caused by second or third generation infections from an initially infected student. Confirmed community cases have been used as a metric only in situations where hospitalisation data were not available or insufficient to reliably infer the effect of interventions - particularly relevant in the case of school reopening, which has predominantly been recommended in communities with significantly reduced daily incidence counts. In this case, care has been taken to ensure that there is evidence of consistent and thorough testing. We document the numbers of tests carried out and comment further on the reliability of confirmed cases as representative of the community epidemic in the Supplementary Material.

Only German states with at least 100 cases at the point of school closure, and at least 10 days of consecutively increasing cases prior to closures, have been selected for analysis. The selection was guided by the availability of data, the relative timings and scope of interventions, and the similar demographic profiles allowing for a better comparison.

National data streams

The effect of school closure is estimated using hospitalisation data for Denmark and Norway, and daily confirmed cases for Germany and Sweden. In all countries, we restrict the analysis of school closures to fall before the peak in reported cases or hospital admissions.

Hospital admissions data for Norway starts 4 days prior to school closure, totaling just 18 admissions. We expect the mean time until interventions would be observed in hospitalisation data to be 14 days after school closures [8]. As Norwegian hospitalisation data was too sparse to reliably infer the effect of school reopening, daily confirmed cases were analysed instead.

Swedish confirmed cases saw two distinct growth phases in the early part of the epidemic, likely caused by different testing regimes. For consistency with later data, we chose to work with the latter phase.
In Germany, daily confirmed cases are reported specifically for students under 18 in the schools, kindergartens, holiday camps, after school clubs, etc. as well as for the staff working in these facilities. We concern ourselves with these numbers for the analysis of school reopening, rather than the total population aggregates on the state or federal level.
Consistent test numbers for both Germany and Norway around the time of school reopening 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, though they are still likely to exhibit greater weekend effects than hospital admissions.
Denmark reopened schools quickly enough following mass quarantine that hospitalisation data could be still used to monitor its effect, though we also verify these findings by analysing confirmed cases. The expected delay to observation in the Danish hospitalisation data is 10 to 14 days [9].
The announced re-opening of Swedish upper secondary and higher education in mid-June will likely not be observed in the data until late June, and is thus not included in this analysis. Of the four countries selected for comparison, Denmark and Norway are considered countries with low incidence, whereas Germany and Sweden are considered countries of medium to high incidence [10].

Simulating the unmitigated epidemic

In order to assess the impact of school closures, we developed a method to project forward the trajectory of cumulative cases or hospital admissions, under the scenario of no intervention. The method couples an ODE epidemic model with a Poisson Gaussian process (GP) regression model. First, a selection of sample trajectories are generated via Approximate Bayesian Computation (ABC) fitting the ODE compartmental model from the first day of data until 5 days after school closures. This is based on the assumption that any change in growth rate within these time windows is unlikely to be attributable to the closures, given a 4.8 day mean incubation period for COVID-19 [11]. The GP regression model is then trained on the sampled trajectories and the data and is used to generate an estimate of the trajectory of cases. The GP model uses the same assumed lag period of 5 days as was used in the ABC fitting process, allowing more data points to be used in training the model. Motivated by recent approaches proposed in [12,13], the GP model helps to account for some of the structural discrepancies between the observed data and the simulated trajectories produced via the ABC fitting method. Additionally, the use of the GP model ensures that the modelled data points immediately after school closures closely follow the observed data, thereby making changes in the growth rate slightly easier to identify.

We identify the first day on which there is a clear and sustained deviation from the modelled data, referred to as the *response date*. More precisely, a change in growth rate is considered an effective response to school closures if (a) it occurs more than 5 days from the intervention date, (b) the deviation persists for at least 5 days, and (c) exceeds the 75^{th} percentile of the modelled data. 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 defined as the first day of deviation from the projection.

The growth rates are estimated through a weighted Negative Binomial regression applied to: the data during the lag time window since school closure (Table 1, column 3) and both the modelled (Table 1, column 4) and the observed data (Table 1, column 5) during equally long time windows after the response date (which marks the first included date in this window). In the case of Baden-Württemberg, Berlin, North Rhine-Westphalia, and Rhineland Palatinate, it was necessary to shorten the period over which the observed post-response growth rate was calculated to ensure that the window did not exceed the point of peak daily incidence.
The relative changes in the estimated growth rates can be used to assess the impact of interventions. However, we stress that the computed values should be viewed as representative of general trends in the epidemic, rather than definitive growth rates.

The ABC fitting of the SEIR model was achieved through the PyGom package for Python [14]. The Poisson Gaussian process regression method, carried using a Bayesian latent variable approach, uses the PyMC3 probabilistic programming package for Python [15]. Further details about the introduced methods can be found in the Supplementary Material.

Estimating the effect of 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 different method whereby the growth rate can be quantified following successive loosened measures. A smoother $\rho(t)$ is applied over time t, such that the instantaneous growth rate is $\rho'(t)$ (c.f. a constant value in a phase of pure exponential growth). 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 from the R package *mgcv*, using a quasi-Poisson family with canonical link [16]. Smoothing is achieved using default thin plate regression splines.

Results

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. With the exception of Baden-Württemberg, all German states closed schools on the 16^{th} of March 2020. As this was a Monday, we have assumed the effective date of school closures is Saturday March 14^{th}, under the assumption that school activity is significantly reduced on weekends. Schools in Baden-Württemberg closed on Tuesday March 17^{th}.
School closures in Hesse and Rhineland Palatinate were only partially observed: final year high school students were permitted to take their exams as planned in late March.

Summary of results

Table 1 provides an overview of the observed change in daily growth rate in the period after school closures. These growth rates are consistent with previous estimates [17]. We show the lag times until the effect of an intervention can be seen in the data, the growth rates pre- and post-response, and the relative change between the modelled and observed post-response growth rate. For ease of reading, a corresponding table in the Supplementary Materials displays these findings using the doubling time.

State	Lag time (days)	Pre-response growth rate (day^{-1})	Modelled post-response growth rate (day^{-1})	Observed post-response growth rate (day^{-1})	Relative growth rate reduction
Baden-Württemberg	8	0.183* (0.158 - 0.206)	0.228 (0.222 - 0.233)	0.115 (0.097 - 0.132)	50%
Bavaria	11	0.219 (0.210 - 0.227)	0.207 (0.201 - 0.213)	0.154 (0.141 - 0.166)	26%
Berlin	9	0.202 (0.188 - 0.215)	0.204 (0.188 - 0.219)	0.142 (0.135 - 0.149)	30%
Hesse	8	0.231 (0.213 - 0.249)	0.255 (0.241 - 0.369)	0.120 (0.106 - 0.135)	53%
Lower Saxony	10	0.224 (0.208 - 0.240)	0.258 (0.243 - 0.273)	0.132 (0.120 - 0.144)	49%
North Rhine-Westphalia	7	0.189 (0.178 - 0.200)	0.202 (0.198 - 0.206)	0.143* (0.130 - 0.155)	31%*
Rhineland Palatinate	6	0.326 (0.279 - 0.372)	0.353 (0.350 - 0.356)	0.124 (0.103 - 0.144)	65%

Table 1. Comparison of estimated lag time and pre- and post-intervention daily growth rates in different German states. Their equivalent formulation as doubling times can be found in the Supplementary Material. Note that the pre-response growth rate in Baden-Württemberg is influenced by a strong weekend effect. If the corresponding data points from the 22$^{\text{nd}}$ and 23$^{\text{rd}}$ of March are omitted from the fitting process, then the pre-response growth rate is 0.196 (0.179 - 0.213).
Similarly, a weekend effect is observed in North Rhine-Westphalia on the 21$^{\text{st}}$ and 22$^{\text{nd}}$ of March. If these data points are omitted from the fit, then the observed post-response growth rate is 0.115 (0.106 - 0.124). This yields a relative reduction in the post-intervention growth rate of 44%.

All states in Germany saw a reduction in growth rate after the closure of schools, typically after a delay of around 9 days. It should be noted, however, that all states experienced further interventions around the same time as school closures, namely the closure of national borders with France, Switzerland, Austria, Denmark and Luxembourg on March 16$^{\text{th}}$, and the declaration of a national state of emergency - leading to the closure of sports facilities and non-essential shops, and restrictions on restaurants - on March 17$^{\text{th}}$. In all states, except Baden-Württemberg, these interventions came 2 and 3 days after the effective date of school closures, respectively. In Baden-Württemberg school closures occurred one day after the closure of borders and on the same day as the declaration of a state of emergency.
These concurrent interventions make it difficult to attribute the fall in cases solely to the closure of schools, and it is likely that there is a combination of factors contributing to the observed decay in growth rate. However, attention should be paid to Baden-Württemberg, which saw a similar lag time as other states, despite having different school closure dates relative to other interventions. We compare this state to North Rhine-Westphalia which saw comparable case numbers, and account for the three day difference in the closing of schools. The two curves exhibit very similar growth following school closure, and yield similar lag times when accounting for the weekend effect (see Supplementary Material). This is indicative of school closures being at least partially responsible for the reduction in growth rate.

The states of Hesse and Rhineland Palatinate allowed students aged 18-19 to sit examinations in late March. In all cases the exams were taken in schools under strict isolation conditions. Neither of the states permitting examinations saw any significant detrimental effect on 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). This can be seen from comparable reductions in the growth rate in all three states (when accounting for the errors in the rates). Further, Rhineland Palatinate managed to reduce the post-intervention growth to a similar (observed post-response) value as the other two states while holding exams despite a higher pre-response rate. This result indicates that, under controlled conditions with limited social interaction, older students sitting exams in school were likely not a significant driver of epidemic growth. We include the detailed results from the highlighted German states below, with the remaining analysed states detailed in the Supplementary Material.

The difference between the modelled and observed post-response growth rates serves as an indicator regarding the effectiveness of interventions. Bavaria and Lower Saxony had comparable pre-response growth rates, but Lower Saxony achieved a significantly greater reduction in the rate post-response. We argue this difference might be due to Lower Saxony having, at the time of intervention, a lower daily (and cumulative) incidence. Additionally, North Rhine-Westphalia (when correcting for the strong weekend effect), Hesse, and Rhineland Palatinate reached similar post-response growth rates despite Rhineland Palatinate having a much greater pre-response growth rate. At the time of intervention, Rhineland Palatinate and Hesse also had lower daily (and cumulative) incidences. This underscores the increased effectiveness of earlier interventions, capable of breaking many transmission chains when community transmission remains low.
The state of Baden-Württemberg clearly saw a decrease in the growth rate following school interventions, but we do not directly compare the growth rates in this state to others. A necessary change in the training period for the GP on the data means there is an unequal basis of comparison relative to other states.

Baden-Württemberg

Baden-Würtemberg saw a reduction in growth rates very quickly after school closures - this is likely a result of reduced weekend testing rather than any response to school interventions. It is likely that the genuine response is seen on March 25^{th}, after which there is a clear and sustained reduction in the growth rate (Figure 1).

Baden-Württemberg saw the following interventions around the time of school closures:

- 10/03 - Banned gatherings of more than 1000 people.
- 16/03 - Shut borders with France (FR), Switzerland (CH), Austria (AT), Denmark (DK) and Luxembourg (LU).
- **17/03 - Closed schools.**
- 17/03 - State of emergency: closed sports and leisure facilities, closed non-essential shops, restrictions imposed on restaurants.
- 23/03 - Banned gatherings of more than 2 people. Closed all restaurants.

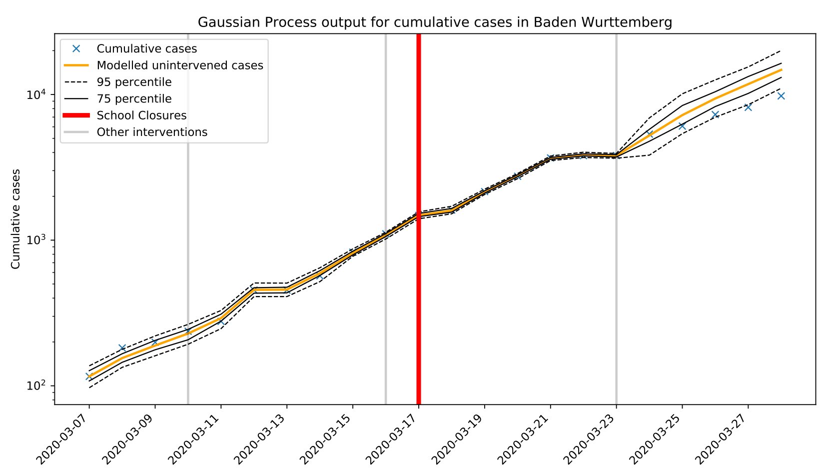

Figure 1. Modelled and observed cases in Baden-Württemberg. Note that the strong weekend effect 5 and 6 days after school closure lead to artificially deflated values. As a result, it was necessary to fit the GP model to 6 days after closure.

Hesse

Hesse saw a similar response to school closures as other German states with moderate incidence (Figure 2), despite schools in the state permitting older high school students to sit examinations towards the end of March.

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

- 10/03 - Banned gatherings of more than 1000 people.
- **14/03 - School closures (effective date).**
- 16/03 - Shut borders with FR, CH, AT, DK and LU.
- 17/03 - State of emergency: closed sports and leisure facilities, closed non-essential shops, restrictions imposed on restaurants.
- 23/03 - Banned gatherings of more than 2 people. Closed all restaurants.

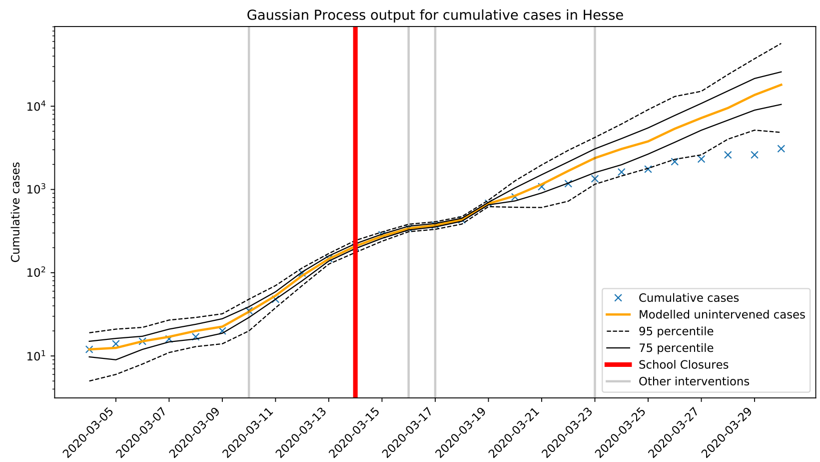

Figure 2. Modelled and observed cases in Hesse.

Lower Saxony

Lower Saxony had a similar number of cases to the two states which permitted students to sit examinations in March. Despite this, there was little difference between the reductions in growth rates - indicating that the examinations did not have a significant impact on the epidemic (Figure 3).

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

- 10/03 - Banned gatherings of more than 1000 people.
- **14/03 - School closures (effective date).**
- 16/03 - Shut borders with FR, CH, AT, DK and LU.
- 17/03 - State of emergency: closed sports and leisure facilities, closed non-essential shops, restrictions imposed on restaurants.
- 23/03 - Banned gatherings of more than 2 people. Closed all restaurants.

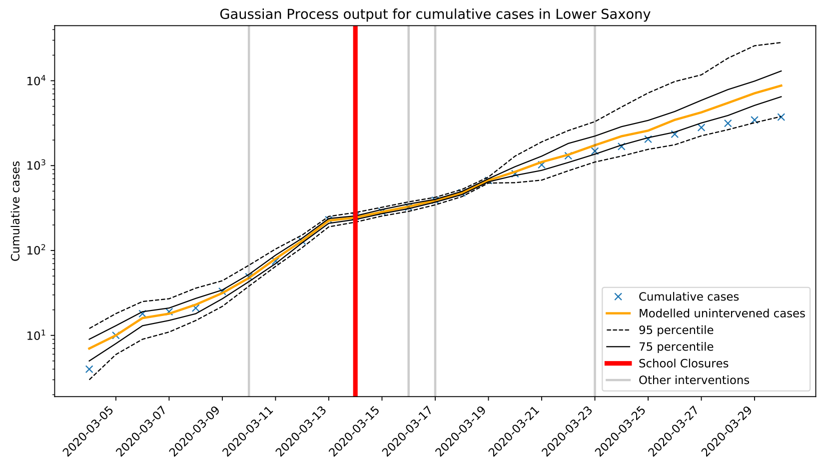

Figure 3. Modelled and observed cases in Lower Saxony.

Rhineland Palatinate

Final year high school students in Rhineland Palatinate were required to sit examinations in the last two weeks of March. This does not appear to have had any significant effect on case numbers. Note that the drop in growth rate on March 21st and 22nd is again likely to be a result of fewer tests reported during the weekend (Figure 4).

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

- 10/03 - Banned gatherings of more than 1000 people.
- **14/03 - School closures (effective date).**
- 16/03 - Shut borders with FR, CH, AT, DK and LU.
- 17/03 - State of emergency: closed sports and leisure facilities, closed non-essential shops, restrictions imposed on restaurants.
- 23/03 - Banned gatherings of more than 2 people. Closed all restaurants.

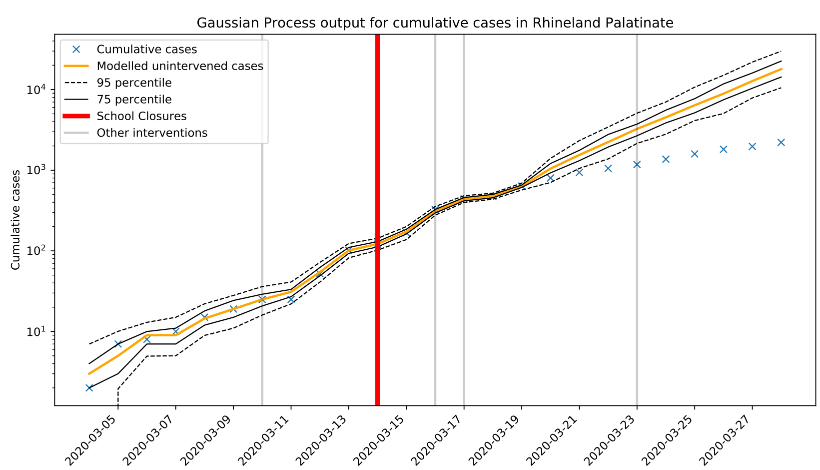

Figure 4. Modelled and observed cases in Rhineland Palatinate.

Closing of schools in Scandinavia

This section considers the effect of closing schools in Denmark, Norway and Sweden. For Norway and Denmark we consider hospital admissions, while for Sweden we consider confirmed cases. As with Germany, we use the officially announced date of school closures as a reference. The official date of closures in Denmark fell on Monday March 16$^{\text{th}}$, and as such we take the effective closure date as Saturday March 14$^{\text{th}}$.
In all three countries there were provisions in place to allow key workers' children to continue attending school. Despite no official nation-wide closing of primary or secondary schools in Sweden, there were local closures in response to outbreaks within the community.

Due to the low incidence in Denmark and Norway, the data for these two countries does not lend itself to reliably determining the response date of interventions, and estimates for the growth rates are therefore not reported. However, a visual inspection of the daily hospital admissions data in both countries (see the Supplementary Material) clearly finds a fall in cases following interventions. However, school closures in both Denmark and Norway occurred in conjunction with other interventions, and as such it is difficult to attribute any effect solely to the closures themselves. For completeness, we include the fits to daily and cumulative hospital admissions using the GP method in the Supplementary Material.
The response date in Sweden was 17 days after school closure, and 8 days after a ban on large gatherings. It is therefore unlikely that this signal can be attributed to the closures themselves. 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 are most effective when implemented concurrently with other interventions.

Sweden

Sweden's school closures were less restrictive than other countries, with only educational establishments for students aged 16 or over being required to close. The first sustained reduction growth rate occurs 17 days after school closures, and 8 days after the banning of mass gatherings (Figure 5).
The limited response to the closure of education establishments should be viewed in the context of looser social restrictions. The eventual response after the banning of large gatherings indicates that school closures affecting older students without more widespread social interventions are unlikely to be effective.
It is notable that there was an increase in weekly testing between March 30$^{\text{th}}$ and April 6$^{\text{th}}$, which may have contributed to the apparent limited reduction in growth rate during this time.

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

- 11/03 - Banned gatherings of more than 500 people.
- 16/03 - Social distancing advised but not enforced.
- **18/03 - Closed all education for students aged 16 or over.**
- 18/03 - Advised that travelling within the country should be reduced.
- 27/03 - Banned gatherings of more than 50 people.

Reopening of schools

Germany

The following events are possible confounders in the data (on a national level), and general indicators for the key dates of schools reopening:

- Shops were allowed to reopen on April 20$^{\text{th}}$; museums, gardens and zoos on April 30$^{\text{th}}$, and hairdressers on May 4$^{\text{th}}$.
- **Return of final year (exam) students on April 27$^{\text{th}}$.**

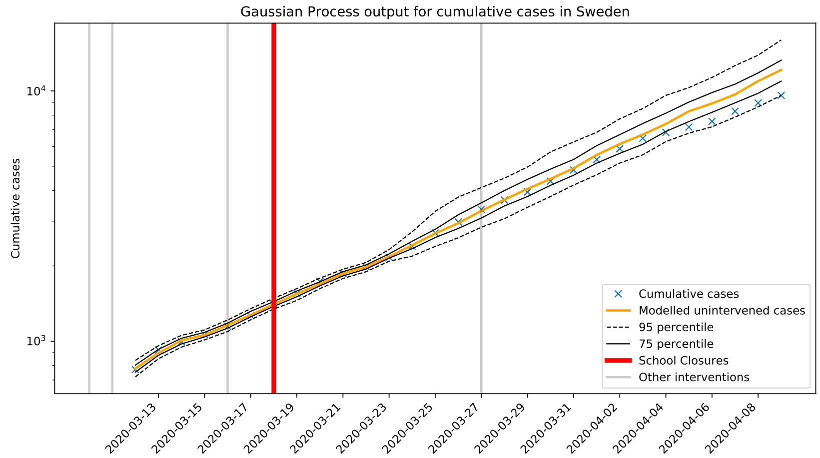

Figure 5. Modelled and observed cases in Sweden.

- **Return of primary school year 4 on May 4th or May 18th.**
- **Return of mixed years on May 11th or May 18th.**

Due to the variability in policies across German states, the dates of school reopening, and the ages of students returning were variable. The above is a summary of the general national trend.

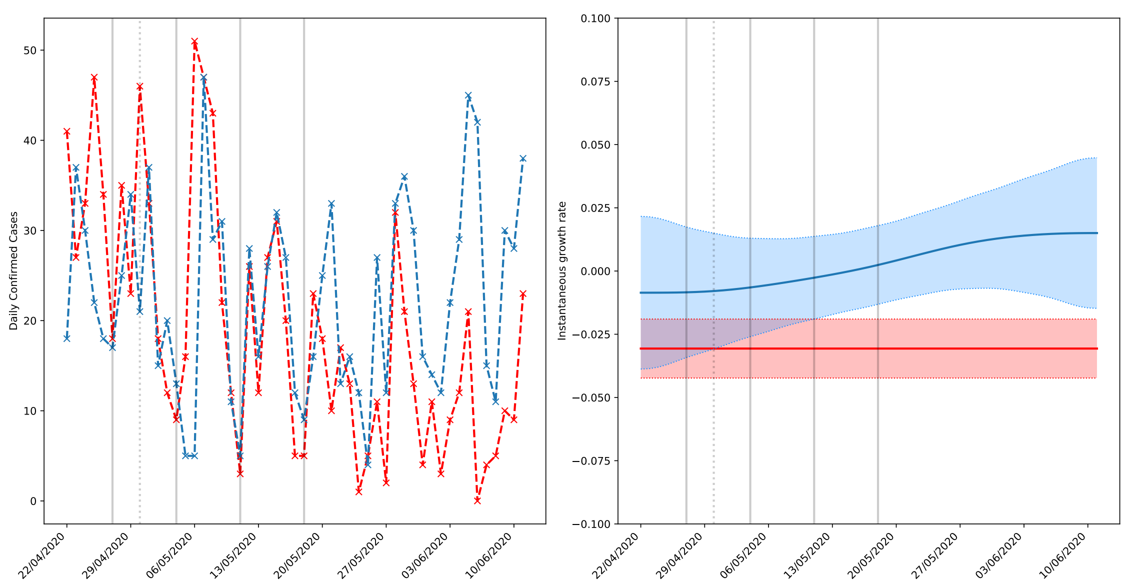

Figure 6. Confirmed cases of COVID-19 in staff (red) and students (blue) in schools, kindergartens, holiday camps, and other educational venues or institutions for under-18s. The exact age distribution of those tested is not known. Left shows daily new confirmed cases, and right shows the instantaneous growth rate (shaded regions are 95% confidence intervals). Solid vertical lines indicate when students returned to school, and dashed lines indicate other loosened 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 detection delay, the incidence among students was higher than that of staff following the return of more (and older) students on May 18th.

The spike in daily cases observed following May 4$^{\text{th}}$ (Figure 6, left) is likely a result of increased presentation for testing following a national announcement of school reopening, or increased community transmission following reopening of other parts of society which was subsequently contained[1].
Overall, there is a downwards trend in the incidence of staff, which is supported by the growth rate among staff being negative. The incidence among students decreases, and subsequently increases with a predominantly positive growth rate from the end of May (Figure 6, right).

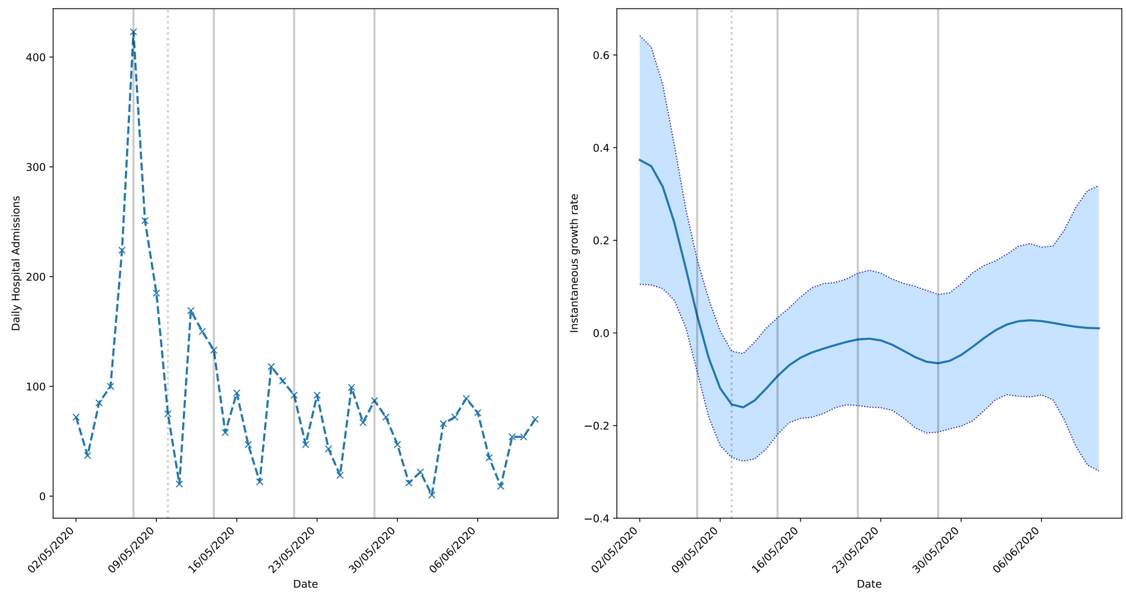

Figure 7. Reported daily hospital admissions in Germany, excluding those working in education, front-line healthcare workers, carers, catering, and hospitality. These numbers indicate transmission in a general, average-exposure population. Left shows daily admissions, and right shows the instantaneous growth rate (shaded regions are 95% confidence intervals). The large confidence intervals on the instantaneous growth rates do not allow one to conclude if, following the reopening of schools, the growth rate has continued to be negative, or whether it is approximately zero. This suggests that the return of younger (and exam) students did not significantly impact the general hospitalised population.

The stable, low, values of the incidence and growth rate until the middle of May indicate that the return of final year exam students and year 4 students (9 year-olds) either

a) does not significantly increase transmission in schools or the community, or

b) can increase transmission, but this is prevented due to the increased distancing in e.g. more spacious classrooms, and an effective testing and tracing system.

This observed effect is quite a strong signal as small case counts appear even across a background of increased community transmission from late April onward with the opening of shops. It is therefore reasonable to conclude that these age groups do not strongly increase transmission in a setting of effective social isolation.

However, the impact of most students returning to school from late May is different. In this time period, the incidence among staff qualitatively agrees with the national trend in hospitalisations (Figure 7), i.e. staff do not appear to become increasingly infected following the return of more students. In contrast, there is a clear increase in the growth rate of students following May 18$^{\text{th}}$. Given that staff incidence is unchanged, and there was little effect of the return of younger years on their own, this suggests that the increase is due to either

[1]At three points in this data set, the recorded cumulative cases (from which the above daily cases were obtained) were inconsistent. These values were imputed using cases reported on the days immediately before and after. The findings do not change significantly upon exclusion of these data points.

a) increased transmission among older students, or

b) the impossibility of adequately carrying out physical distancing in venues at full capacity.

Note that we examine the impact of reopening for a longer period in Germany than other countries, due to the possibility of specifically understanding the impact of reopening on the young student population. We have not done so for other countries because an equivalent data set was not available.

Denmark

The following events are an inexhaustive list of possible confounders in the data, and key dates for the return to school:

- Small businesses were allowed to reopen on April 20$^{\text{th}}$; further shops on May 11$^{\text{th}}$, and some shops, restaurants, and cafes on May 18$^{\text{th}}$; zoos, museums, cinemas and similar venues opened on May 27$^{\text{th}}$.
- **Return of years 0 to 5 on April 15$^{\text{th}}$.**
- **Return of years 6 to 10 and exam students on May 18$^{\text{th}}$.**
- **Return of secondary school students on May 27$^{\text{th}}$.**

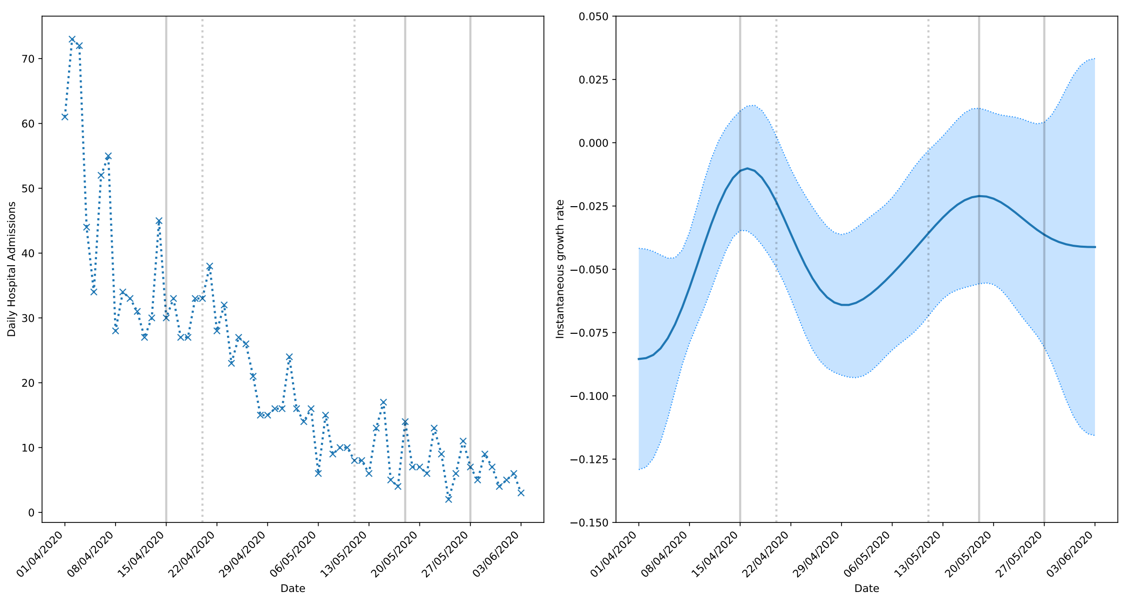

Figure 8. Reported daily hospitalisations in Denmark. New admissions are reported left, and right shows the instantaneous growth rate (shaded regions are 95% confidence intervals). Given the small numbers of children with COVID-19 admitted to hospital for treatment, the return to schools will be seen in the following generations, implying longer delays until an effect might be observed. Note that large confidence intervals on the growth rate are a result of the small number of hospitalisations at the end of May. Solid vertical lines indicate when students returned to school, and dashed lines indicate other loosened measures.

There is no significant observable increase in the growth rate of hospital admissions following school re-opening to younger years, even bearing in mind the subsequent reopening of some businesses (Figure 8). The low growth rate and small relative number of admissions indicate that the return of younger years to school settings did not contribute significantly to transmission, subject to adherence to social distancing. There are only very few data points following the reopening stage on May 18$^{\text{th}}$ when accounting for the expected delay. However, the present data does not suggest this event to have had a significant effect on national hospitalisations. We

verify this observation by comparison with confirmed cases (as they are subject to a shorter delay, and are recorded in greater numbers) in the Supplementary Material.
This finding is further supported by the observation that the proportion of adults testing positive for COVID-19 is lower among those working with children aged 0 to 16 than those working with students aged 16 or over (further details in the Supplementary Material) [18]. However, these numbers alone do not distinguish between infection acquired from the students, and infection acquired elsewhere.

Norway

The following events are possible confounders in the data, and key dates for the return to school:

- Small businesses were allowed to reopen on April 27$^{\text{th}}$; larger gatherings of up to 50 people were allowed from May 7$^{\text{th}}$.
- **Return of kindergarten students on April 20$^{\text{th}}$.**
- **Return of years 1 to 4, final year students, vocational training, and higher education which requires physical presence on April 27$^{\text{th}}$.**
- **Return of the remainder of primary and secondary school the week commencing May 11$^{\text{th}}$.**

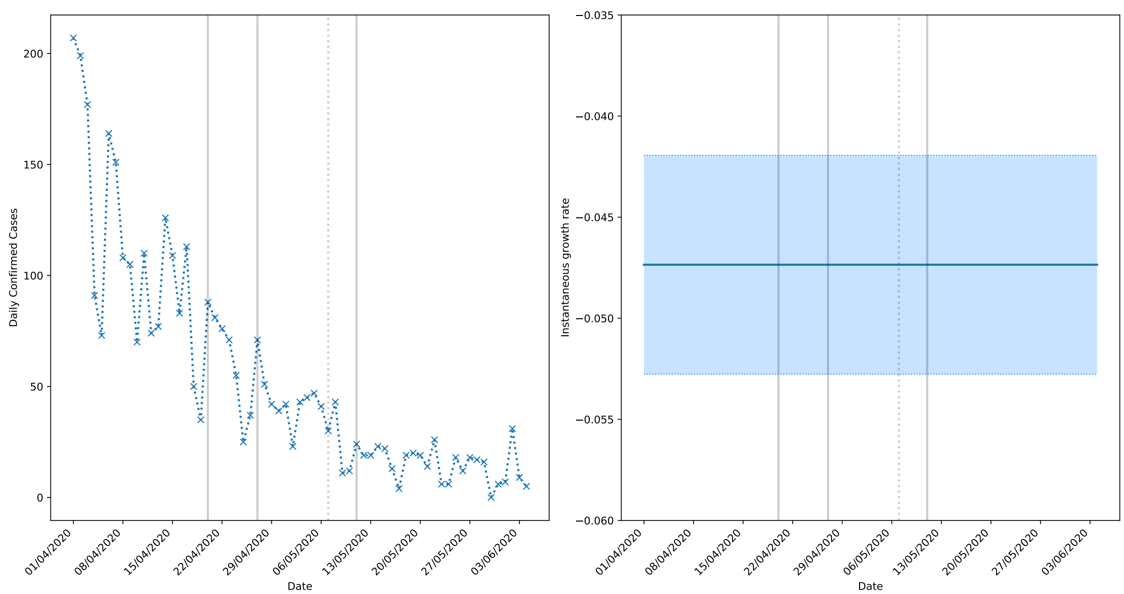

Figure 9. Reported daily confirmed cases in Norway. New cases are reported left, and right shows the instantaneous growth rate (shaded regions are 95% confidence intervals). Solid vertical lines indicate when students returned to school, and dashed lines indicate other loosened measures. These numbers are obtained over a time period of increased testing, so there is little reason to believe the case numbers to be higher than reported. Norway has a significantly large testing rate per capita.

There is no observed increase in the growth rate of confirmed cases following any school reopening, even bearing in mind the subsequent reopening of some businesses (Figure 9). The consistently low growth rate and small number of admissions indicate that the return of most students to school settings did not contribute significantly to transmission, subject to adherence to social distancing. However, we emphasise that this effect is subject to very high levels of testing, with very low community transmission.

Discussion

Our analysis suggests that school closures in Denmark, Norway and Germany had some impact on epidemic growth rates, compared to the modelled un-intervened data. Interventions always reduced the growth rate compared to our modelled scenario with no intervention; however the extent of this reduction strongly varied with the level of community transmission. School closures in Germany, when analysed on a state level, typically resulted in a reduction in epidemic growth rate 9 days after the intervention. We find comparable results in the state of Baden-Württemberg, which closed all schools later than other states. This indicates that school closures are at least partially responsible for the reduction in growth rate. However, the decision of two German states (Hesse and Rhineland Palatinate) to permit final year high school students to sit examinations in late March does not appear to have had a significant impact on state-level case numbers. Sweden implemented partial school closures which affected students aged 16 or above. However, there is no clear evidence to suggest that this intervention is the cause of a later decline in the growth rate.
The evidence for the impact of school closures on growth rates in Norway and Denmark is more limited. While there was a clear and significant reduction in growth rate of hospitalised cases after school closures, it has not been possible to disentangle this response from other interventions occurring at the same time.

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; countries with low community transmission. 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. We argue that the different impact of reopening schools in Germany may be due to higher levels of community transmission. Despite an established system of testing, tracing, and isolation, the growth rate of the national German epidemic has increased in the period following reopening and is now close to zero.

Decoupling the effect of school closures from other interventions is not straightforward. This work does not claim to have achieved this, however the consistency of the signal across regimes with different intervention timelines suggests a positive effect of school closings. The degree of this effect is correlated with other interventions introduced in parallel which aim to reduce community transmission by different means.
One cannot assume that school closures are a "mirror-image" of school reopenings. The two can provide context to one another, however a positive result for school closures is not a justification or repudiation of school openings. School closures are often one of the first measures introduced in populations with high incidence, whereas school reopening has been staggered with other gradually eased restrictions; often with only a small cohort of students returning initially. Consequently, due to the additional backdrop of other interventions, e.g. retail closure and reopening, our findings should not be interpreted as presenting a causal link between individual interventions and changes in national case numbers. However, as most countries leave mass quarantine, reopening schools will likely only be one of several relaxation measures. In light of this, we believe our findings are a realistic assessment of the effects of school reopening in their natural context of wider societal changes.

The presence (or lack) of signals in the data following school interventions are limited by the reliability of the available data. Efforts have been made to consider the most reliable data streams, which we argue is hospital admissions, while bearing in mind that hospitalisations only affect a subset of the infected population. Where this data was unavailable, we have considered confirmed cases while monitoring the degree of testing in place to ensure such numbers were indicative. Where the availability and nature of tests changed (e.g. nasal swabs or antibody blood tests), it is not automatically clear that data from these different regimes can be compared. This is also true as the availability of tests expands from front-line workers to the general population. Hospitalisations are less prone to some of these biases, but are still affected by protocols in reporting, e.g. the hospital admission of an individual with urgent medical needs who also tested positive for COVID-19, albeit with mild symptoms. School closures occurred early in the outbreak, with potential variations in diagnostic protocols, testing availability, and the nature of the data being collected. Later in the pandemic, the quality and diversity of the data is now increasing as we shift from emergency measures to an infrastructure of long-term monitoring.

However, as case numbers have fallen due to mass quarantines and other severe measures, over time some data streams simply have too few cases to be used for analyses. This has been seen in Norway with the steady decrease in hospital admissions requiring a shift to using testing data for analysis.
Efforts have been made to clearly present these biases, and to quantify their impacts. Where possible, we have cross-validated our observations using several data streams.

The simulation of an unmitigated epidemic used to analyse school closures has been built around an SEIR model. While the system of ODE's underpinning this model are well justified, they are inevitably unable to capture some of the processes present in the early stages of the epidemic. Early, highly localised interventions - such as the quarantining of the first early cases - are likely to have had some impact on the transmission rate which is not able to be captured in the model. This is particularly evident in states in which there was an abrupt change in growth rate in the period prior to school closures.
The Gaussian process regression method allows one to account for differences between the simulated epidemic trajectories and the observed cases, however this process is not without its limitations. The fact that closures occurred very early on in the epidemic means that the Gaussian process method often had to be trained on a limited number of data points (the short posterior predictive trajectories from the ODE before the intervention). Although a GP, itself being a probabilistic model, is tailored to handle small data with exact uncertainty quantification, the process of estimating its hyperparameters, using MCMC, becomes challenging due to less than optimal mixing of some of these parameters.
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 presence of weekend effects and the noisiness of some data streams due to the 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.

With some exceptions, we have considered the effects of school closure and reopening on the national level without accounting for inevitable geographic variability, the age distribution of those studied, and their profession (i.e. likelihood of exposure to infected individuals). The analysis of Germany, particularly the comparison of staff and student infections, warns against the reliability in using national-level data to understand the immediate effects or impact of a single population. However, this does not imply that national data cannot be used, much like the lack of a signal on the national level following school reopenings does not guarantee a limited national impact. Instead, this impact may only become first visible in subsequent generations. We must therefore be concerned not only with the delay until a signal might first be seen in the studied population, e.g. 9 days, but also with the following generation of infections. This puts the minimum time scale between intervention and impact (assuming a change can be inferred from consistent data over a week, and a constant 5 day incubation period) from 16 to 21 days if using confirmed cases, or 21 to 26 days if using hospital admissions (using an estimated 14 day delay from infection to admission). A statistically challenging yet positive result is that, in many countries, there is insufficient data to analyse the impact of various interventions on an age-stratified level due to low incidence numbers. This is similarly true if examining cases stratified by occupation.

Our analysis is constrained to settings with high monitoring and intervention efficacy (including but not limited to high testing, tracing, and adherence to isolation). Concerning e.g. the return of younger years, continued low incidence following their return to school does not imply that such a measure is inherently safe. 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.
Similar caution should be had regarding the small or manageable effect of the return of older students, in particular with regard to: the likely increased number of crowded classrooms, as well as their added impact to community transmission if the latter is already relatively high. While all three studied countries seem to be effectively managing transmission, the volatility of the German growth rate in 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 population at large. Furthermore, we highlight that the tenuous

balance (net zero growth) 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.

Our findings generally underscore the precarious nature of transmission control as it relates to the reopening of schools, particularly in numbers whereby physical distancing is unfeasible. The safe return of most (or even large proportions of) students to school is conditional on the successful implementation of a complete system of monitoring and interventions, jointly with low daily incidence, as observed in Denmark and Norway.
This correlation with community transmission can be seen particularly clearly in Germany, with confirmed cases increasing among students, and the halted decay in hospital admissions on the national level. The most severe impact appears to occur following the return of most older students to schools when community transmission remains significant (if managed).
Great care should be taken when attempting to infer the impacts of school reopening in other nations by comparison with the presented subset. It is insufficient to compare the number of daily tests carried out (or similar monitoring metrics), or the daily incidence (be it total or per capita). The speed of decline in daily cases is also a key quantity, as it informs us about the effectiveness of tracing, individual or household isolation, and the adherence thereto. 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 myriad 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 appropriately 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. Such a strategy may not be feasible if the incidence is too high.

Authors' contributions

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

Funding

LP and HS 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. 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 and Ian Hall for useful discussions and contribution to the methodology.

Data accessibility

All data used in the production of this work can be found in the electronic supplementary material.

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.
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, Jun 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, Mar 2008.
4. Nicolo Principi, S Esposito, R Gasparini, P Marchisio, and P Crovari. Burden of influenza in healthy children and their households (Retraction in: Archives of Disease in Childhood (2006) 91:9 (797)). *Archives of Disease in Childhood*, 89(11):1002–1007, 2004.
5. 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.
6. 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.
7. 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*.
8. Norwegian Institute of Public Health. Koronavirus-modellering ved fhi. `https://www.fhi.no/sv/smittsomme-sykdommer/corona/koronavirus-modellering/`, June 2020.
9. Statens Serum Institut. Ekspertrapport af den 6. maj 2020. Technical report, Statens Serum Institut, 2020.
10. 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.
11. 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. *arXiv*, 2020.
12. Jenni Brynjarsdóttir and Anthony Ohagan. Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30(11), 2014.
13. 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.
14. Public Health England. pygom 0.1.6. `https://github.com/PublicHealthEngland/pygom`, 2020.
15. John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2016(4):e55, Apr 2016.

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

17. 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*, page eabb9789, apr 2020.

18. 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.

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

20. 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.

21. 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.

22. 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.

Supplementary Material

Data availability

The data streams used to carry out the analysis of this work can be found at:

- Denmark: daily tests, hospital admissions, and confirmed cases from the National Serum Institute [19]. Sentinel survey among educational staff is from [18].
- Germany: daily (and weekly) tests, hospital admissions, and confirmed cases from the Robert Koch Institute [20].
- Norway: daily tests and hospital admissions from the Norwegian Institute of Public Health [21].
- Sweden: weekly tests, and confirmed cases from the Public Health Agency of Sweden [22].

Numerical methods

The compartmental model used to generate sample trajectories is outlined in Figure S1. Multiple compartments have been used for the exposed (E) populations to model an Erlang-distributed incubation period compatible with available estimates of mean and standard deviation of its duration [11]. The I_0 compartment splits into both 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 [11] and limited knowledge on duration of infection and non-modelled pathway of hospitalised cases suggest a single compartment (i.e. exponential holding time) for these states is a reasonable and parsimonious choice.

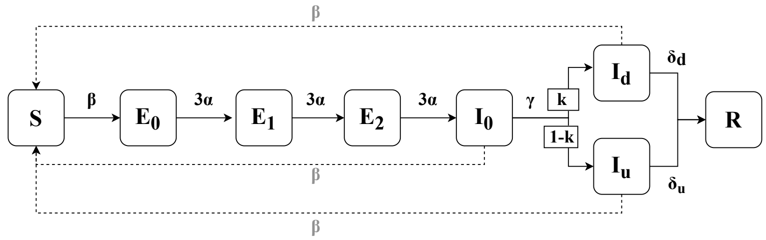

Figure S1. The epidemic model used to simulate cases. The model uses multiple exposed compartments to account for an Erlang-distributed incubation period.

In order to provide more data points for the fitting process, the model has been fitted to 5 days after school closure. This is based on the assumption that any change in growth rate within these time windows is unlikely to be attributable to the closures, given the 4.8 day mean incubation period for COVID-19 [11].
Simulation of the SEIR model (Figure S1) is achieved through Approximate Bayesian Computation (ABC). This fitting process is used to estimate the parameter vector:

$$\theta = (\beta, \gamma, k, \delta_d, \delta_u; E_0(0)). \tag{1}$$

The parameter α is not estimated in this method, and is instead taken from the mean incubation period [11] to be $\alpha = 1/4.8$ days^{-1}. We employ a weighted negative binomial loss function to measure the distance between the observed data and the posterior predictive trajectories generated by ABC fitting. Over-representation of earlier data points in the cumulative data is accounted for by introducing a weighting, w_i, such that for a data point, x_i, we have:

$$w_i = \frac{x_i - x_{i-1}}{x_i} \tag{2}$$

The ABC method generates 150 posterior predictive trajectories, from which we select 15 evenly distributed samples to be used as covariates with which to train the Poisson Gaussian Process (GP) machine learning model. This model attempts to use the trajectories obtained via ABC to replicate the observed data. As with the SEIR model fit, the GP model is trained with data 5 days after closure for confirmed cases (6 days, in the case of Baden-Württemberg). Care has been taken to ensure the reproducability and robustness of these results, including testing with different distributions of 30 and 50 covariates, and running the model with re-calculated ABC posteriors.

Testing data

Testing data for Germany

Figure S2 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 7$^{\text{th}}$-9$^{\text{th}}$, 14$^{\text{th}}$-16$^{\text{th}}$ and 21$^{\text{st}}$-23$^{\text{rd}}$. 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 17$^{\text{th}}$ and 27$^{\text{th}}$ was fairly stable. As most school closures in Germany occurred on March 16$^{\text{th}}$, we can expect the confirmed cases over this period to provide a reasonable representation of the underlying epidemic.

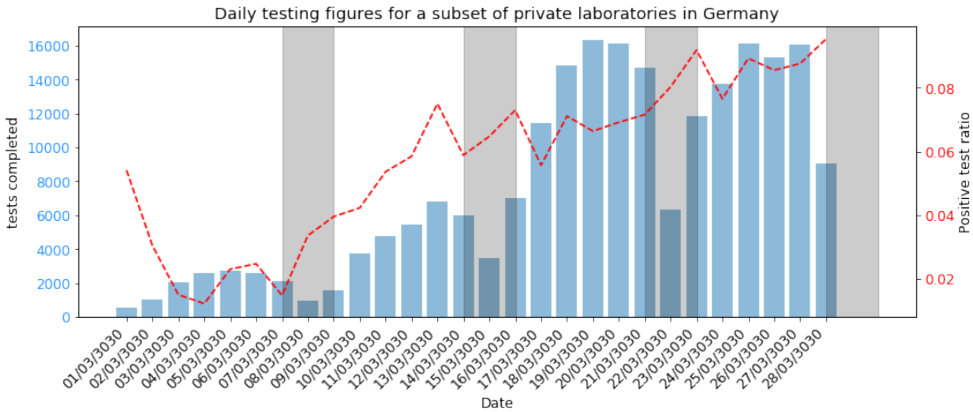

Figure S2. Daily testing from a subset of German testing laboratories during March. Weekends are highlighted in grey. There is a periodic drop in testing occurring on weekends, particularly evident on Sundays. These drops do not coincide with any changes to the positive test ratio.

Daily testing data for Germany is not available after March. As such, it will be necessary to consider the weekly testing totals, which are made available through the RKI. Figure S3 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.

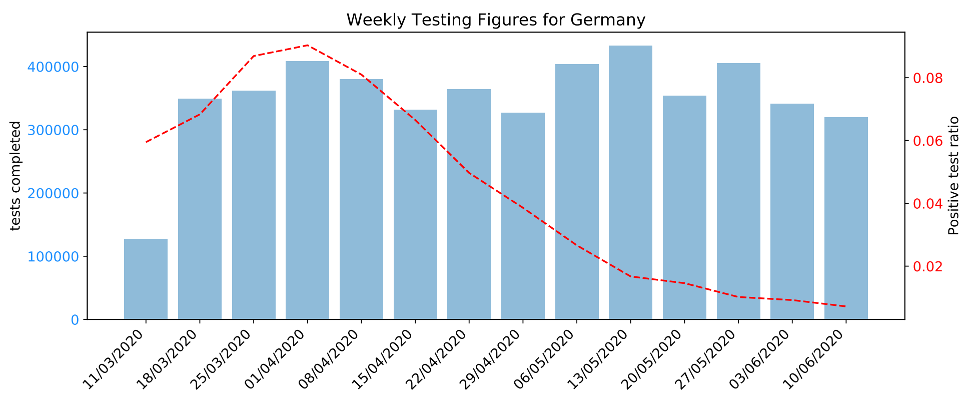

Figure S3. Weekly testing in Germany remained consistent from March 18$^{\text{th}}$, however the weekend effect (see Figure S2) was likely present across the entire period. There were no abrupt changes in the positive test ratio.

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 S4 and S5 respectively.
Denmark displays two clear increases in testing capacity between March 23$^{\text{rd}}$ and 30$^{\text{th}}$ and again between April 13$^{\text{th}}$ and 20$^{\text{th}}$. 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 16$^{\text{th}}$, 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.

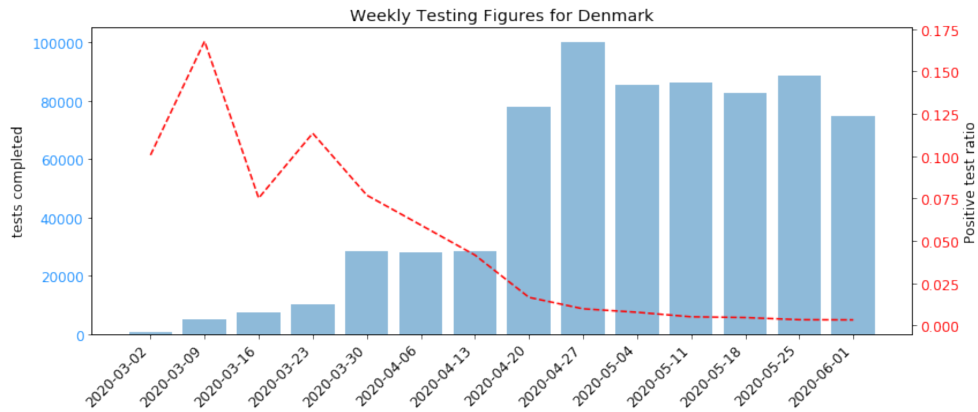

Figure S4. Weekly testing in Denmark was not consistent across the period of this investigation, and so confirmed cases up to April 20$^{\text{th}}$ cannot be relied upon to provide a reliable representation of the underlying epidemic.

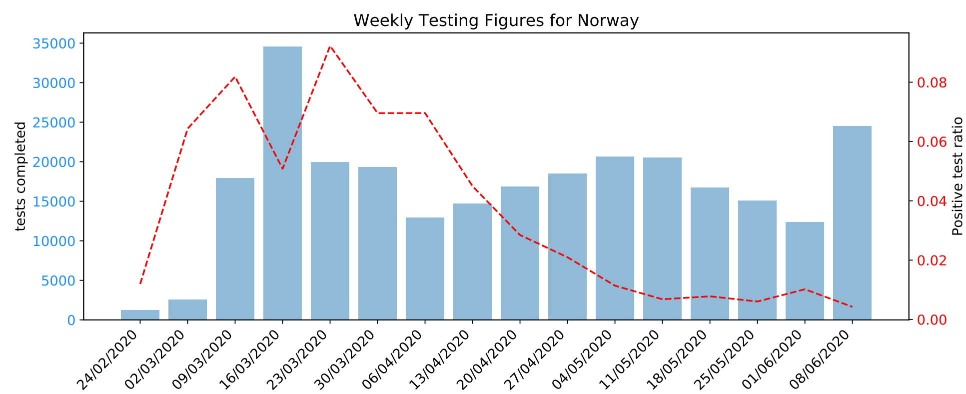

Figure S5. Norway saw inconsistent testing during March, making confirmed cases an inappropriate metric for assessing school closures. More consistent testing was apparent in April and May.

The weekly testing figures for Sweden are highlighted in Figure S6, along with the positive test ratio for the same period. Testing rates around the time of school closures (March 18$^{\text{th}}$) were generally increasing, with a large increase occurring during the week beginning March 30$^{\text{th}}$. This increase was accompanied by an increase in positive test ratio, indicating an increasing capability to identify and test infected individuals. As a result, it will not be possible to attribute any increase in case numbers after March 30$^{\text{th}}$ solely to the effect of interventions.

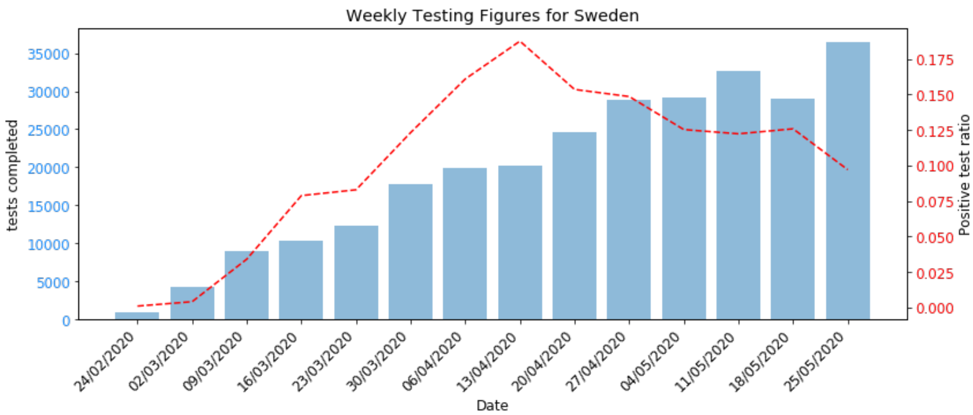

Figure S6. Reported weekly tests carried out in Sweden.

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) to do so with a background of limited social interventions is useful for partially decoupling the effect of school closures from other controls.

School closure analyses

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

State	Lag time (days)	Pre-response doubling time (days)	Modelled post-response doubling time (days)	Observed post-response doubling time (days)
Baden-Württemberg	8	3.8* (3.4 - 4.4)	3.0 (3.0 - 3.1)	6.0 (5.3 - 7.1)
Bavaria	11	3.2 (3.1 - 3.3)	3.3 (3.3 - 3.4)	4.5 (4.2 - 4.9)
Berlin	9	3.4 (3.2 - 3.7)	3.4 (3.2 - 3.7)	4.9 (4.6 - 5.1)
Hesse	8	2.8 (2.5 - 3.0)	2.6 (1.8 - 2.8)	5.9 (5.3 - 6.6)
Lower Saxony	10	3.1 (2.9 - 3.3)	2.7 (2.5 - 2.9)	5.3 (4.8 - 5.8)
North Rhine-Westphalia	7	3.7 (3.5 - 3.9)	3.4 (3.3 - 3.4)	4.8* (4.5 - 5.3)
Rhineland Palatinate	6	2.1 (1.9 - 2.5)	2.0 (1.9 - 2.0)	5.9 (5.6 - 6.2)

Table S1. Comparison of estimated lag time and pre- and post-intervention doubling times in different German states. Note that the pre-response doubling time in Baden-Württemberg is influenced by a strong weekend effect. If the corresponding data points from the 22nd and 23rd of March are omitted from the fitting process, then the pre-response doubling time is 2.9 (2.7 - 3.3) days.
Similarly, a weekend effect is observed in North Rhine-Westphalia on the 21st and 22nd of March. If these data points are omitted from the fit, then the observed post-response doubling time is 6.0 (5.6 - 6.5) days.

In Figure S7 we consider the cumulative cases in both Baden-Württemberg and North Rhine-Westphalia on account of the states being very similar. We shift the cases in Baden-Württemberg back in time by three days, so as to coincide the dates of effective school closure (March 14$^{\text{th}}$ and 17$^{\text{th}}$). Further, to aid comparison we rescale the cases in Baden-Württemberg by a multiplicative constant (approximately 0.76), so that the cumulative cases in both states are identical on the day of school reopening. We note that the pure exponential growth rate is unchanged by these transformations. The profiles of the data can now be compared, with the results from the GP fit to North Rhine-Westphalia providing a baseline from which the lag times can be found.

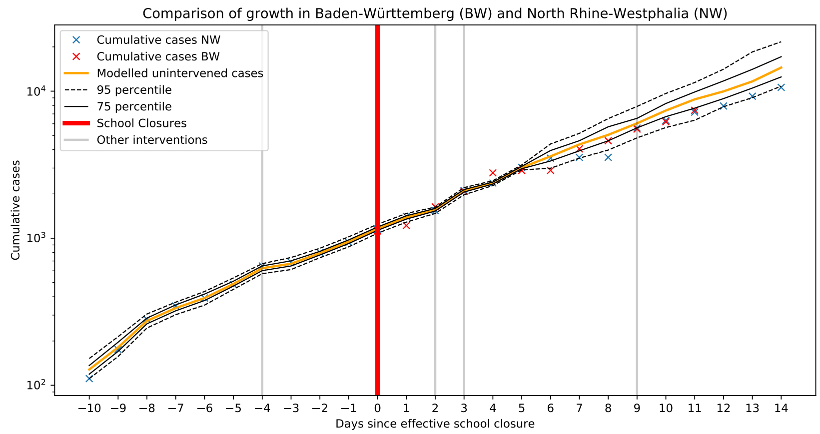

Figure S7. Cumulative cases for Baden-Württemberg and North Rhine-Westphalia when corrected for the three-day shift in school closure between the two. The effective day of school closure in both states is shown in red, with the timings of other interventions which took place in North Rhine-Westphalia included for reference. 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, when considering the weekend effect occurring in Baden-Württemberg, 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. Clearly it would be unrealistic to assume school closure to be wholly responsible for the observed fall in cases, but the above-detailed correlations are evidence to suggest that they may have partially contributed to the total effect.

Bavaria

Bavaria saw a small decrease in case numbers immediately after school closures, although this is likely a result of changes in testing rates over this period. The first sustained decrease in epidemic growth rate occurs 11 days after school closures (Figure S8).

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

- 10/03 - Banned gatherings of more than 1000 people.
- **14/03 - School closures (effective date).**
- 15/03 - Local elections went ahead, with a high turnout. A large number of votes were submitted by post.
- 16/03 - Shut borders with FR, CH, AT, DK and LU.
- 17/03 - State of emergency: closed sports and leisure facilities, closed non-essential shops, restrictions imposed on restaurants.
- 23/03 - Banned gatherings of more than 2 people. Closed all restaurants.

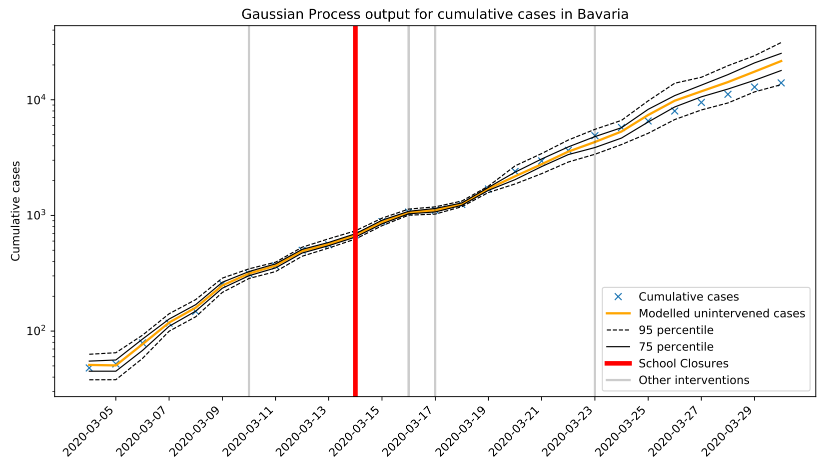

Figure S8. Modelled and observed cases in Bavaria.

Berlin

Berlin saw a limited response to school closures, with a small increase in growth rate occurring 9 days after intervention. Testing rates throughout this period were consistent, indicating that confirmed cases are likely a good indicator of general trends in the epidemic (Figure S9).

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

- 10/03 - Banned gatherings of more than 1000 people.
- **14/03 - School closures (effective date).**
- 16/03 - Shut borders with FR, CH, AT, DK and LU.
- 17/03 - State of emergency: closed sports and leisure facilities, closed non-essential shops, restrictions imposed on restaurants.
- 23/03 - Banned gatherings of more than 2 people. Closed all restaurants.

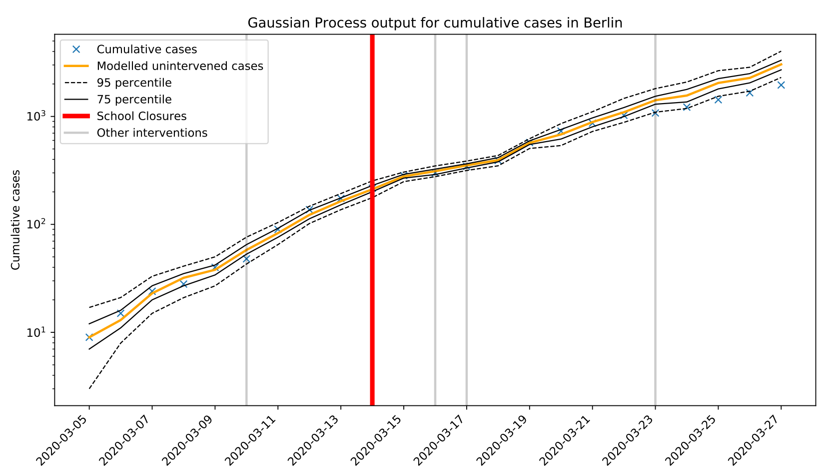

Figure S9. Modelled and observed cases in Berlin.

North Rhine-Westphalia

North Rhine-Westphalia saw a drop in cases between March 20th and 22nd. This was likely a result of decreased testing or reporting in this period. When accounting for this effect, the first sustained drop in epidemic growth rate occurs 7 days after school closures (Figure S10).

- **05/03 - Local school closures in Heinsberg.**
- 10/03 - Banned gatherings of more than 1000 people.
- **14/03 - School closures (effective date).**
- 16/03 - Shut borders with FR, CH, AT, DK and LU.
- 17/03 - State of emergency: closed sports and leisure facilities, closed non-essential shops, restrictions imposed on restaurants.
- 23/03 - Banned gatherings of more than 2 people. Closed all restaurants.

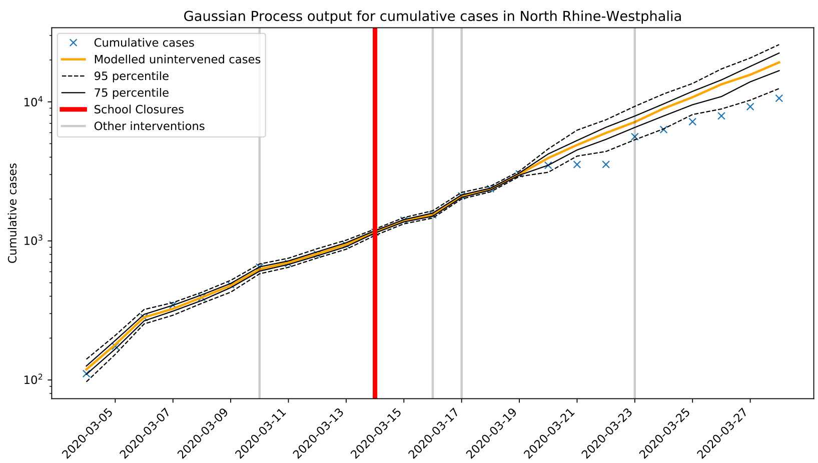

Figure S10. Modelled and observed cases in North Rhine-Westphalia.

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 clear decrease in growth rate in the period immediately following school closures, it is not possible to identify a single day on which this shift occurs. The GP model was unable to adequately predict the trajectory of cumulative hospitalisations in this period (S11). Furthermore, estimating growth rates over this period is problematic as the peak in incidence occurs soon after school closures (S12).

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

- 11/03 - Banned public gatherings of more than 100 people.
- 13/03 - Closed non-essential businesses.
- **13/03 - Closed educational establishments for students aged 16 or over.**
- **14/03 - Closed all schools (effective date).**
- 18/03 - Banned gatherings of more than 10 people. Closed shopping centres.

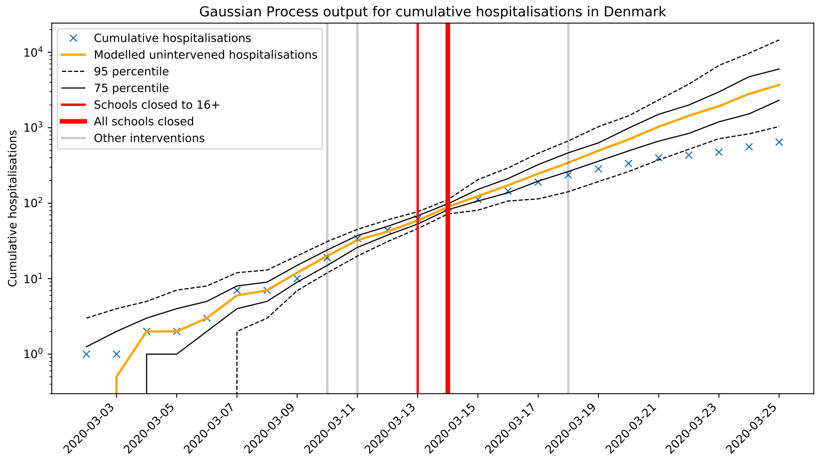

Figure S11. Modelled and observed cumulative hospitalisations in Denmark. The GP model has been fitted 5 days after school closures, but is unable to adequately predict the trajectory of cases. As such, it is not possible to estimate any lag period for the response.

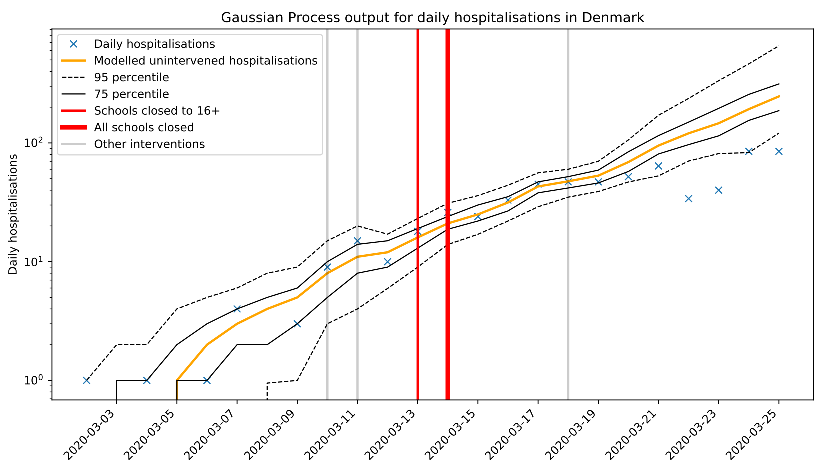

Figure S12. Modelled and observed incidence of hospitalisations in Denmark. The GP model incorrectly predicts the exponential growth seen in the period prior to closures to continue, making it difficult to identify a response to school closures.

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 (Figure S13).

For completeness, we fit the GP model to cumulative hospitalisations (Figure S13). However the short training period for the model resulted in a wide confidence interval for this prediction. 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:

- **12/03 - Closed all schools.**
- 12/03 - Closed non-essential businesses. Sports and cultural events cancelled. Restrictions on restaurants.
- 14/03 - Advised against foreign travel.
- 18/03 - Closed ports and airports to all repatriating citizens and imports.

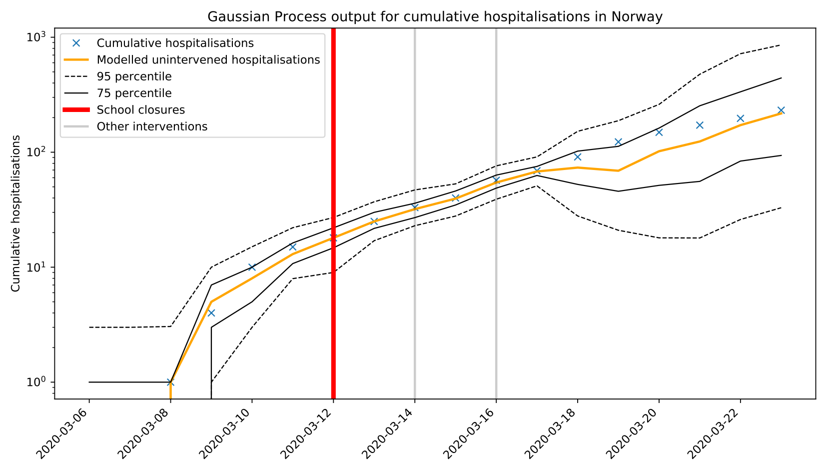

Figure S13. Modelled and observed cumulative hospitalisations in Norway. The model is able to reasonably predict the trend in cases for around 10 days after school closures, however the confidence in this prediction is very low.

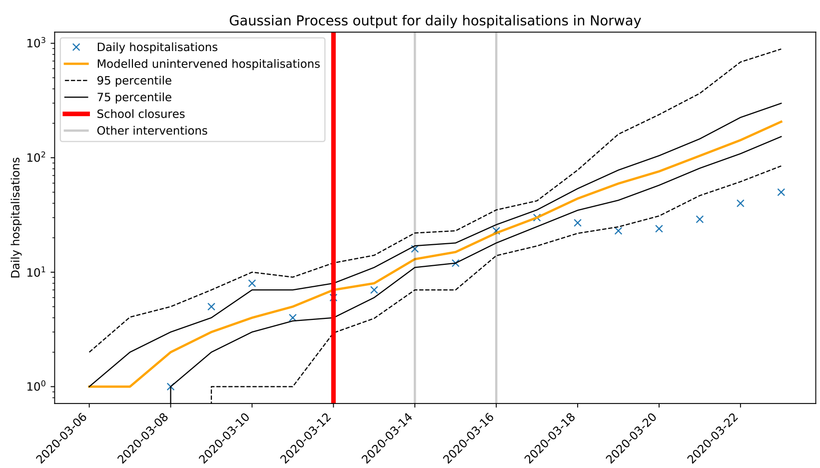

Figure S14. Modelled and observed hospitalisations in Norway. The GP model is less effective when dealing with the incidence data, failing to account for any points after the assumed lag period of 5 days. As such, it is impossible to estimate post response growth rates in Norway.

Supporting evidence for the impact of reopening schools

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

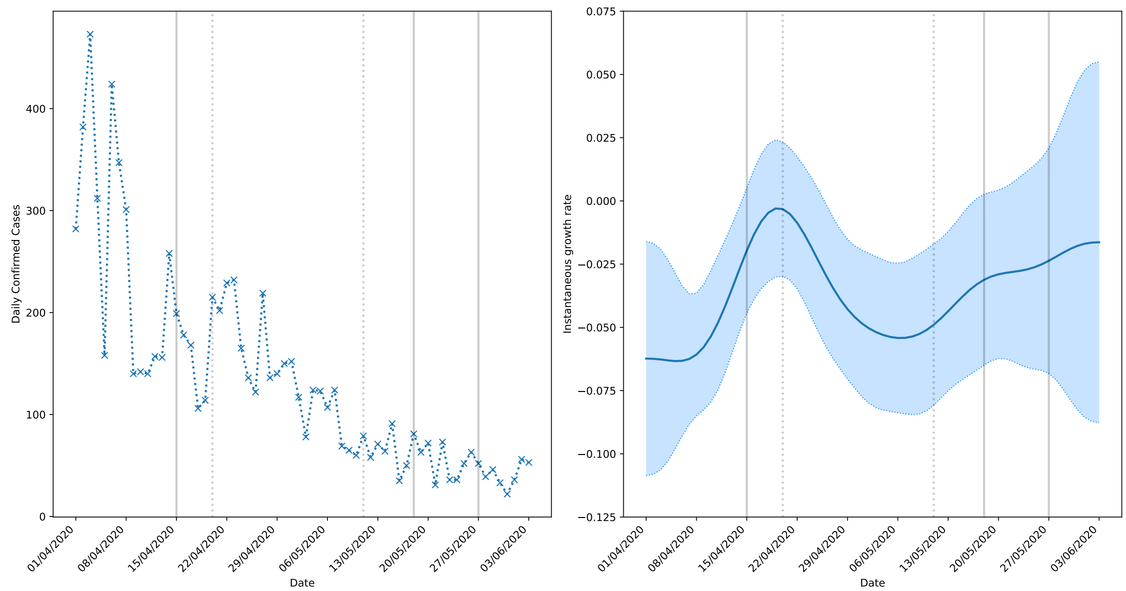

Figure S15. Reported daily confirmed cases in Denmark. New confirmed cases are reported left, and right shows the instantaneous growth rate (shaded regions are 95% confidence intervals). Solid vertical lines indicate when students returned to school, and dashed lines indicate other loosened measures. 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.

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 S2).

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 S2. A comparison of tests carried out among staff working in different stages of the Danish educational and childcare sector dated June 2^{nd}. 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.