ABSTRACT

Objectives COVID-19’s second wave started a debate on the potential role of schools as a primary factor in the contagion resurgence. Two opposite positions appeared: those convinced that schools played a major role in spreading SARS-CoV-2 infections and those who were not. We studied the growth rate of the total number of SARS-CoV-2 infections in all the Italian regions, before and after the school reopening (September–October 2020), investigating the hypothesis of an association between schools and the resurgence of the virus.

Methods Using a Bayesian piecewise linear regression to scrutinise the number of daily SARS-CoV-2 infections in each region, we looked for an estimate of a changepoint in the growth rate of those confirmed cases. We compared the changepoints with the school opening dates, for each Italian region. The regression allows to discuss the change in steepness of the infection curve, before and after the changepoint.

Results In 15 out of 21 Italian regions (71%), an estimated change in the rate of growth of the total number of daily SARS-CoV-2 infection cases occurred after an average of 16.66 days (95% CI 14.47 to 18.73) since the school reopening. The number of days required for the SARS-CoV-2 daily cases to double went from an average of 47.50 days (95% CI 37.18 to 57.61) before the changepoint to an average of 7.72 days (95% CI 7.00 to 8.48) after it.

Conclusion Studying the rate of growth of daily SARS-CoV-2 cases in all the regions provides some evidence in favour of a link between school reopening and the resurgence of the virus. The number of factors that could have played a role is too many to give a definitive answer. Still, the temporal correspondence warrants further systematic experiments to investigate on potential confounders that could clarify how much reopening schools mattered.

INTRODUCTION

We recognise that attending schools play a key role for younger generations in supporting their development toward shared societal values and in promoting their positive physical and mental well-being. Nonetheless, if the scientific community is to overcome the challenges posed by SARS-CoV-2, a serious assessment of the impact that educational settings may have had in the spread of the pandemic cannot take a back seat. Hence, the main motivation of our study.

As the summer of 2020 was coming to its end, with its relatively low number of COVID-19 cases in most western countries, many started arguing whether it was wise to normally restart school activities. During the first wave, in most nations schools were closed, as any other activity, and only some partially reopened them as the situation got under control and the lockdown was lifted. In this context, the large COVID-19 outbreak in a high school in Israel in May 2020 ignited the discussions about the role of schools in
the spread of the virus.\cite{1,2} Two opposing sides appeared quite clearly: on one side, those who considered schools to be a minor risk and the importance of school paramount; on the other, those who were concerned by the lack of clear data on the contagion dynamics in schools and were scared by the high number of asymptomatic cases in younger people. The same discussion emerged in all the other countries affected by the ongoing pandemic, with the two sides bringing to the table mostly the same arguments with the occasional country-specific remarks.

Most international literature that suggests the absence of considerable risk factors connected to schools focuses on the fact that children and adolescents seem to be the least affected by the virus, both in terms of the number of positive cases but also of symptoms and contagiousness. Ismail et al\cite{3}, studying UK schools, show how the incidence in students is not larger than the total incidence in the region and how the most cases inside schools are transmission between staff members. Similar studies targeting other countries draw similar conclusions, especially when they come to primary schools and kindergarten.\cite{4-9} The hypothesis that asymptomatic children could easily and unknowingly spread the virus in their families was rejected by Munro and Faust\cite{10} and by other similar analyses, based on argument that children have a lower susceptibility to the virus and thus play a lesser role in the transmission.\cite{11-13}

The researchers presenting the opposite hypothesis point to the weaknesses of many of the aforementioned studies, namely the very small samples and the fact that often the role of asymptomatic subjects is not tracked and considered properly. Many point out the correspondence between the insurgence of the second wave in many countries within 2–3 weeks from the school openings and point to the data that suggest a higher spread in the school-aged population in those months, especially in high school and university students.\cite{14,15} Flasche and Edmunds have responded to Ismail’s study saying that it was conducted with schools not fully populated underestimating the potential of children, especially in the age bracket 10–18 years.\cite{16} This group has seen a considerable increase in September, as did college students, and seemed to be a common source of SARS-CoV-2 infections in the households. Also Yamey and Walensky\cite{17} express their concerns for universities reopening, while Sebastiani and Palù\cite{18} studied the Italian situation and argued that the rise of new cases in September, with most SARS-CoV-2 infections happening in the household, was compatible with the hypothesis of school being a factor. While inside schools, measures were taken, they argued that outside contact was inevitable due to public transportation and social gatherings, so that young people spread the virus among themselves. Larosa et al\cite{19} conducted a study in the Reggio Emilia province (Emilia-Romagna region, Italy) showing that there were non-negligible clusters in the age bracket 10–18 years. They also suggested that more prompt isolation and testing could have hindered the spread, stressing how important timeliness is in this context. Despite their opposing positions, most researchers emphasise the need for the same measures: an active case finding approach with systematic and thorough testing of students and personnel.

Following this scientific debate, this work focuses on Italy, looking at the contagion curves and relating them to the dates schools opened in each of the 21 regions. Italy has faced a hard time recently, being hit by a second wave of COVID-19 cases that were bringing the healthcare system to its knees. This second wave started during the autumn, somewhere around the start of October 2020, and peaked in November when the government imposed a new form of lockdown with color-coded zones, based on risk in every region. September was a crucial period, as with the end of summer many activities were going back to normal, and the virus prevalence in the nation was quite low.\cite{20} In the first days of that month, a slight increase in the number of new cases was registered in most regions, probably due to the cross-regional movement for the summer holidays.\cite{21} This prompted many to warn of the arrival of the second wave, but the number of new cases stabilised in the coming days, and the growth was considered small in any case. School reopened between the third and the fourth week of September,\cite{22,23} while people had also already started going back to their offices and activities. Another thing to notice is the referendum of 20 September that in some regions corresponded with other elections for local government and senate representatives. While attendance was quite low, one could wonder the effects of such an event.

Because of the shocking scarcity of available and reliable data on SARS-CoV-2 infections inside Italian schools, we chose another perspective in order to investigate the hypothesis of an association between schools and the resurgence of the virus, by analysing the growth rate of the total number of SARS-CoV-2 infections in all the Italian regions, before and after the school reopening.

METHODS

Given the scarce availability of data collected in schools that could better describe their role in Italy (for the reasons stated in the Introduction section), we decided to work with the population-wide data at the regional level. We fitted a piecewise linear regression model where the dependent variable was only the number of new daily confirmed COVID-19 cases, and the independent variable was just the number of days since 1 September (until 31 October). The result is a model comprised of a changepoint and two segments, whose slopes represent the growth rate before and after the acceleration in the exponential growth causing the second wave; we passed from a stable situation with the exponent close to 0 which means little to no growth, to a situation with markedly positive exponent. To have a measurement of the uncertainty in our estimates, we decided to use a Bayesian framework for the regression as described in.\cite{24} Two transformations were applied to the initial data. First, we used
a 7-day rolling average as the raw data present a weekly periodicity due to the way COVID-19 tests are carried out and registered. Second, we applied a natural logarithm so that the exponential growth appears as an easily identifiable slope. Using the R package called mcp built on top of JAGS, we estimated a piecewise linear regression model for each region.

We modelled our dependent variable \( \ln(y) \) (the natural log of confirmed daily cases) as a normal distribution whose mean depends on the regression coefficients \( a_i \) and \( b_i \) (the intercept and angular coefficient) before a changepoint \( \tau \), and \( a_i \) and \( b_i \) after the changepoint, as represented in the formula below.

\[
\ln(y) \sim \mathcal{N}(a_i + xb_i, \sigma_i^2) \quad \forall \chi < \tau
\]

\[
\ln(y) \sim \mathcal{N}(a_i + xb_i, \sigma_i^2) \quad \forall \chi > \tau
\]

The model was fitted using a Markov Chain Monte Carlo method. Since the two lines are joined at the change-point of days \( DT_1 \) and \( DT_2 \), and \( b_i \) is not estimated as is bound to be \( b_i = \tau (b_i - b_i) + a_i \).

As described above, the intercept and slope priors used for starting the Bayesian estimation were chosen in Gaussian families, while for the changepoint a Uniform distribution was used, precisely as reported below:

\[
\tau \sim \mathcal{U}(\min(x), \max(x))
\]

\[
\sigma \sim \mathcal{N}(0,3sd(y))
\]

\[
b \sim t(0,3+sd(y),3)
\]

\[
a \sim t(0,3sd(y)/\max(x) - \min(x),3)
\]

The final estimates for all the parameters (the change-point, the two slopes, the initial intercept and the variance) are reported in Table 1 with the 95% CI. To give an idea of the increase in slope, we computed the number of days \( DT_1 \) and \( DT_2 \) necessary to observe a doubling in the number of new cases from the changepoint onwards, with both slopes (using the average angular coefficient), as shown in the following equations.

\[
DT_1 = \frac{\ln 2y_r - a_i}{b_i} - \tau ; \quad DT_2 = \frac{\ln 2y_r - \ln y_r}{b_i};
\]

where \( y_r = \exp(a_i + \tau b_i) \).

All the data used are available from the Italian government site: https://github.com/pcm-dpc/COVID-19. The results of all the computations are fully reproducible by using the method stated above. The authors are open to sharing the code with whoever reaches out by email.

**Patient and public involvement**

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**RESULTS**

Table 1 shows all the estimated quantities for each region. Similarly, figure 1 shows all the curves and regression lines. Out of the 21 Italian regions, 15 (71%) of them have a changepoint within 28 days from the date when the school opened. In particular, the average number of days between the opening and the changepoint is 16.66 days (95% CI 14.47 to 18.73). We believe this number is plausible in a scenario where one has to be exposed to the virus and then manifest symptoms in order to be tested, also considering that often children and adolescents are asymptomatic and that they only function as a vector to other people around them.

Looking at the estimated slopes of figure 1, we can have an idea of the strength of the increase. While one could comment on the difference in the angular coefficients, converting it into an angle, the number does not convey the idea very effectively. Translating it into the number of days required for a doubling in the growth rate is more interpretable. Of the 15 regions mentioned above, 4 had a slope that is null or slightly negative before the change, making this time infinite, but the fact that the trend was inverted to an average value of 6.23 days (95% CI 4.30 to 8.20) is significant enough by itself. The remaining 11 regions went from an average of 47.50 (95% CI 37.18 to 57.61) days at the rate before the changepoint to an average of 7.72 (95% CI 7.00 to 8.48) at the rate after the changepoint. Figure 1 and table 1 illustrate these results.

Of the six regions that break the pattern, two of them, P.A. Trento and P.A. Bolzano (often considered one region, called Trentino-Alto Adige) presented a changepoint more than 4 weeks after the school opening; they are indicated in blue in Table 1. On the other hand, the remaining four, in yellow in Table 1, begin their rapid increase before or in correspondence of the school opening. Both are displayed in figure 2.

All the models converge quite nicely, with narrow 95% CI, except for Basilicata, where the changepoint estimation is extremely wide, probably due to the high variability in the reported numbers.

**DISCUSSION**

The results highlight how the second wave in Italy started in the days between September and October, with a degree of variability. These changepoints are on average a couple of weeks after the school openings. Certainly, multiple confounding factors played a role in the acceleration of the SARS-CoV-2 infection, but our opinion is that schools are surely one of those, and the magnitude of their effects should be investigated more thoroughly.

In the short period that precedes the second wave, there were not many events that interested as many people as schools. The referendum and elections on 20 September almost coincide with school opening and did not have a large participation. Workers going back to their offices could also have played a major role, however by looking at the nationwide mobility report by Google news, we can see that in September the number of people moving to their workplace increased steadily from −30% to −20% with respect to the reference level before the pandemic. If we consider that in Italy, there are approximately 25 million workers, an increase of 10% would...
Table 1  Estimated parameters from the 21 Italian regions along with school opening date (Open), number of days between opening and changepoint dates (D) and the doubling time for the two slopes (DT).

<table>
<thead>
<tr>
<th>Region</th>
<th>Open</th>
<th>D</th>
<th>( \tau )</th>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( a_1 )</th>
<th>( \sigma )</th>
<th>( DT_1 )</th>
<th>( DT_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basilicata</td>
<td>24 Sep</td>
<td>−4</td>
<td>19.25 (0.01 to 51.02)</td>
<td>0.02 (−0.00 to 0.06)</td>
<td>0.06 (0.04 to 0.08)</td>
<td>1.15 (0.78 to 1.45)</td>
<td>0.35 (0.29 to 0.29)</td>
<td>30.10</td>
<td>11.98</td>
</tr>
<tr>
<td>Campania</td>
<td>24 Sep</td>
<td>−2</td>
<td>20.85 (18.81 to 22.86)</td>
<td>0.00 (−0.01 to 0.01)</td>
<td>0.07 (0.07 to 0.08)</td>
<td>5.07 (5.00 to 5.15)</td>
<td>0.08 (0.07 to 0.07)</td>
<td>∞</td>
<td>9.49</td>
</tr>
<tr>
<td>Abruzzo</td>
<td>24 Sep</td>
<td>0</td>
<td>22.54 (20.73 to 24.28)</td>
<td>−0.01 (−0.02 to 0.00)</td>
<td>0.10 (0.09 to 0.11)</td>
<td>2.73 (2.58 to 2.88)</td>
<td>0.18 (0.15 to 0.15)</td>
<td>∞</td>
<td>6.94</td>
</tr>
<tr>
<td>Sardegna</td>
<td>22 Sep</td>
<td>0</td>
<td>21.26 (17.65 to 26.73)</td>
<td>0.00 (−0.01 to 0.01)</td>
<td>0.05 (0.04 to 0.05)</td>
<td>3.94 (3.85 to 4.03)</td>
<td>0.08 (0.07 to 0.07)</td>
<td>∞</td>
<td>15.16</td>
</tr>
<tr>
<td>Puglia</td>
<td>24 Sep</td>
<td>8</td>
<td>30.92 (29.63 to 32.20)</td>
<td>0.01 (0.01 to 0.01)</td>
<td>0.07 (0.07 to 0.07)</td>
<td>4.13 (4.08 to 4.19)</td>
<td>0.08 (0.07 to 0.07)</td>
<td>69.76</td>
<td>9.76</td>
</tr>
<tr>
<td>Calabria</td>
<td>24 Sep</td>
<td>9</td>
<td>31.67 (29.68 to 33.65)</td>
<td>−0.01 (−0.01 to 0.00)</td>
<td>0.10 (0.09 to 0.11)</td>
<td>2.75 (2.62 to 2.88)</td>
<td>0.19 (0.16 to 0.16)</td>
<td>∞</td>
<td>6.75</td>
</tr>
<tr>
<td>Valle d’Aosta</td>
<td>14 Sep</td>
<td>13</td>
<td>26.22 (24.09 to 28.29)</td>
<td>0.00 (−0.02 to 0.01)</td>
<td>0.14 (0.13 to 0.15)</td>
<td>0.68 (0.49 to 0.87)</td>
<td>0.27 (0.22 to 0.22)</td>
<td>∞</td>
<td>4.95</td>
</tr>
<tr>
<td>Umbria</td>
<td>14 Sep</td>
<td>14</td>
<td>26.87 (25.72 to 28.09)</td>
<td>0.01 (0.00 to 0.01)</td>
<td>0.10 (0.10 to 0.10)</td>
<td>2.79 (2.72 to 2.87)</td>
<td>0.11 (0.09 to 0.09)</td>
<td>70.71</td>
<td>6.87</td>
</tr>
<tr>
<td>Friuli Venezia Giulia</td>
<td>16 Sep</td>
<td>16</td>
<td>30.91 (29.49 to 32.22)</td>
<td>0.01 (0.01 to 0.02)</td>
<td>0.09 (0.08 to 0.09)</td>
<td>2.96 (2.88 to 3.05)</td>
<td>0.13 (0.11 to 0.11)</td>
<td>33.96</td>
<td>6.45</td>
</tr>
<tr>
<td>Piemonte</td>
<td>14 Sep</td>
<td>16</td>
<td>28.58 (27.40 to 29.73)</td>
<td>0.02 (0.02 to 0.02)</td>
<td>0.11 (0.10 to 0.11)</td>
<td>3.99 (3.94 to 4.05)</td>
<td>0.08 (0.06 to 0.06)</td>
<td>61.36</td>
<td>8.94</td>
</tr>
<tr>
<td>Veneto</td>
<td>14 Sep</td>
<td>16</td>
<td>29.20 (27.09 to 31.62)</td>
<td>0.01 (0.01 to 0.02)</td>
<td>0.08 (0.07 to 0.08)</td>
<td>4.75 (4.68 to 4.82)</td>
<td>0.09 (0.07 to 0.07)</td>
<td>51.44</td>
<td>7.74</td>
</tr>
<tr>
<td>Lombardia</td>
<td>14 Sep</td>
<td>17</td>
<td>30.30 (29.56 to 31.02)</td>
<td>−0.01 (−0.01 to 0.00)</td>
<td>0.12 (0.12 to 0.13)</td>
<td>5.49 (5.44 to 5.55)</td>
<td>0.08 (0.07 to 0.07)</td>
<td>∞</td>
<td>5.56</td>
</tr>
<tr>
<td>Toscana</td>
<td>14 Sep</td>
<td>17</td>
<td>29.67 (28.81 to 30.51)</td>
<td>0.01 (0.01 to 0.01)</td>
<td>0.10 (0.10 to 0.10)</td>
<td>4.39 (4.33 to 4.45)</td>
<td>0.08 (0.07 to 0.07)</td>
<td>70.17</td>
<td>6.92</td>
</tr>
<tr>
<td>Emilia-Romagna</td>
<td>14 Sep</td>
<td>18</td>
<td>30.65 (30.11 to 31.23)</td>
<td>−0.01 (−0.01 to 0.01)</td>
<td>0.09 (0.09 to 0.09)</td>
<td>4.86 (4.83 to 4.90)</td>
<td>0.05 (0.04 to 0.04)</td>
<td>∞</td>
<td>7.67</td>
</tr>
<tr>
<td>Marche</td>
<td>14 Sep</td>
<td>19</td>
<td>32.35 (30.53 to 34.01)</td>
<td>−0.02 (0.01 to 0.02)</td>
<td>0.10 (0.09 to 0.11)</td>
<td>2.73 (2.63 to 2.83)</td>
<td>0.15 (0.12 to 0.12)</td>
<td>39.67</td>
<td>6.93</td>
</tr>
<tr>
<td>Liguria</td>
<td>14 Sep</td>
<td>20</td>
<td>32.74 (30.35 to 35.16)</td>
<td>0.02 (0.02 to 0.03)</td>
<td>0.08 (0.08 to 0.09)</td>
<td>3.85 (3.75 to 3.96)</td>
<td>0.15 (0.12 to 0.12)</td>
<td>27.91</td>
<td>8.29</td>
</tr>
<tr>
<td>Molise</td>
<td>14 Sep</td>
<td>22</td>
<td>35.16 (33.16 to 37.35)</td>
<td>0.02 (0.01 to 0.03)</td>
<td>0.12 (0.11 to 0.14)</td>
<td>0.72 (0.55 to 0.91)</td>
<td>0.29 (0.24 to 0.24)</td>
<td>19.63</td>
<td>9.67</td>
</tr>
<tr>
<td>Sicilia</td>
<td>14 Sep</td>
<td>22</td>
<td>34.52 (32.17 to 36.81)</td>
<td>0.04 (0.03 to 0.04)</td>
<td>0.07 (0.07 to 0.08)</td>
<td>3.81 (3.75 to 3.87)</td>
<td>0.09 (0.08 to 0.08)</td>
<td>36.39</td>
<td>5.71</td>
</tr>
<tr>
<td>Lazio</td>
<td>14 Sep</td>
<td>23</td>
<td>35.85 (34.69 to 37.01)</td>
<td>0.02 (0.01 to 0.02)</td>
<td>0.09 (0.09 to 0.10)</td>
<td>4.89 (4.85 to 4.94)</td>
<td>0.07 (0.05 to 0.05)</td>
<td>41.98</td>
<td>7.60</td>
</tr>
<tr>
<td>P.A. Bolzano</td>
<td>7 Sep</td>
<td>28</td>
<td>34.39 (28.46 to 41.99)</td>
<td>0.04 (0.03 to 0.05)</td>
<td>0.10 (0.07 to 0.12)</td>
<td>1.97 (1.83 to 2.12)</td>
<td>0.22 (0.17 to 0.17)</td>
<td>17.79</td>
<td>6.87</td>
</tr>
<tr>
<td>P.A. Trento</td>
<td>13 Sep</td>
<td>32</td>
<td>44.99 (42.97 to 47.17)</td>
<td>0.02 (0.02 to 0.03)</td>
<td>0.12 (0.09 to 0.14)</td>
<td>2.68 (2.55 to 2.81)</td>
<td>0.24 (0.20 to 0.20)</td>
<td>32.14</td>
<td>5.90</td>
</tr>
</tbody>
</table>

Between brackets are the 95% CI.
translate into 2.5 million more people circulating. If we compare this number with that of students and school personnel which is equal to approximately 11 million, we can conclude that perhaps, even when considering that both categories use public transportation heavily, schools could be more influential in spreading the virus.

The regions that do not follow the pattern may tell us something more. In a regionalised country like Italy, with a strong territorial differentiation, Trento and Bolzano could be outliers. They are two northern, neighbouring autonomous provinces, often considered a single region, which share characteristics of a higher care coverage of social and educational services that could set them apart from other regions. In the others, the change is before or coincidental with school opening, so any factor that could have ignited the second wave is to be researched outside of the school activity and all the other connected activities. The effect of schools, if any, would be absorbed in the inflation already in act. Being all four maritime regions, tourism could perhaps be one of the major

![Figure 1](https://example.com/figure1.png)

**Figure 1** Plots for the regions whose changepoint (blue with 95% CI area) is within 28 days since school openings (yellow). Black dots are the natural log of daily confirmed COVID-19 cases from 01 September 2020 to 31 October 2020. In green and red the regression lines using the mean parameters, before and after the changepoint. On the rightmost side of each plot, the y-axis is reported without the log transformation, to allow the reader to infer what the confirmed case rates were in each region.
Figure 2: The same plots of figure 1 but for the regions that do not exhibit the pattern. First four anticipate the start of the exponential growth before school opens, the last two take more than 28 days after that. On the rightmost side of each plot, the y-axis is reported without the log transformation, to allow the reader to infer what the confirmed case rates were in each region.
Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval Not applicable: neither humans nor animals nor personal data are being involved in this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available in a public, open access repository: https://github.com/pcm-dpc/COVID-19. All data and codes are available upon request to the corresponding author (email: luca.casini78@unibo.it).

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Marco Roccetti http://orcid.org/0000-0003-1264-8595
Luca Casini http://orcid.org/0000-0002-3468-6974

REFERENCES
21 Casini L, Roccetti M. A Cross-Regional analysis of the COVID-19 spread during the 2020 Italian vacation period: results from three computational models are compared. Sensors 2020;20:7319.