Age-dependency of the Propagation Rate of Coronavirus Disease 2019 Inside School Bubble Groups in Catalonia, Spain

Sergio Alonso, PhD,* Enric Alvarez-Lacalle, PhD,* Martí Català, MSc,*† Daniel López, PhD,*
Iolanda Jordan, MD, PhD,‡§¶ Juan José García-García, MD, PhD,§¶‖ Antoni Soriano-Arandes, MD, PhD,**
Uxue Lazcano, MSc,†† Pilar Sallés, MSc,‡‡ Marta Masats, MSc,‡‡ Julià Urrutia, MSc,‡‡ Anna Gatell, MD,§§
Ramon Capdevila, MD,¶¶ Pere Soler-Palacin, MD, PhD,** Quique Bassat, MD, PhD,¶‖‖***†††
and Clara Prats, PhD*†

Background: We analyzed contagions of coronavirus disease 2019 inside school bubble groups in Catalonia, Spain, in the presence of strong non-pharmaceutical interventions from September to December 2020. More than 1 million students were organized in bubble groups and monitored and analyzed by the Health and the Educational departments.

Methods: We had access to 2 data sources, and both were employed for the analysis, one is the Catalan school surveillance system and the other of the educational department. As soon as a positive index case is detected by the health system, isolation is required for all members of the bubble group, in addition to a mandatory proactive systematic screening of each individual. All infected cases are reported. It permits the calculation of the average reproductive number (R*), corresponding to the average number of infected individuals per index case.

Results: We found that propagation inside of the bubble group was small. Among 75% index cases, there was no transmission to other members in the classroom, with an average R* across all ages inside the bubble of R* = 0.4. We found a significant age trend in the secondary attack rates, with the R* going from 0.2 in preschool to 0.6 in high school youth.

Conclusions: The secondary attack rate depends on the school level and therefore on the age. Super-spreading events (outbreaks of 5 cases or more) in childhood were rare, only occurring in 2.5% of all infections triggered from a pediatric index case.

Key Words: coronavirus disease 2019, schools, transmission, reproductive number, children, adolescent

The need to implement nonpharmaceutical interventions (NPIs) to control the unrestrained transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), together with uncertainty regarding the potential of children to foster its transmission, led governments to hastily close schools globally during the first wave. Subsequently, there has been a discussion about the advisability of closing schools during the different stages in the evolution of the pandemic. Children, notably less affected by the health problems associated with this pandemic[1] and often infected asymptomatically,[2] appear to remain as susceptible or less to infection than adults[3] and thus could therefore, at least theoretically, still significantly contribute to overall community transmission. To what extent children can efficiently transmit the virus to their peers or other individuals, and whether children can act, as do adults, as superspreaders of the infection, remain to be clearly elucidated.[4]

School closure has significant direct and indirect consequences on the well-being of children and may enhance social inequities.[5,6] Recently, the European Centers for Disease prevention and Control recommended school closure as a last resource after the closure of all other activities except for essential services.[7] Furthermore, the Centers of Disease Control and Prevention, based in the United States, recommended that primary schools at least should be the last settings to close after all other NPIs have been applied.[8] Some countries, including Spain, have also considered schools as essential activities that should therefore remain open with strict NPIs.

In Spain, the use of masks in the school for students older than 6 is mandatory. Therefore, young children between 3 and 5 years are not required to use masks in schools. This measure has been combined with the creation of stable "bubble" groups, of a limited size (generally ~20–30 students and teachers), who remain closely together and have minimized interaction with other school members. The 2 measures are combined with standard hygienic measures and the promotion of indoor space ventilation, in addition to strict enforcement of nonschool attendance for any sick child. When one case was detected, because of a contact tracing

outside the school or the presence of symptoms, the whole bubble group was quarantined for 10 days, and all the individuals of the bubble group were tested to investigate potential intrabubble spread; see Figure 1 for an outline of the process of detection and measures taken. In Catalonia, the region in Spain with the closest school monitoring and surveillance, the results from the tests are used to determine the link among the cases inside a classroom and to differentiate index and secondary cases. The head of the school directly notifies all index cases and secondary cases to evaluate the size of the outbreak.

In this study, we analyzed these data from the Catalan educational system, which currently monitors 1.44 million students, teachers and other personnel working in 5104 schools, plus public health open data, with the objective of improving our understanding of secondary attack rates among children.

METHODS

Two data sources were used for the analysis. The first of these is the Catalan school surveillance system (Traçacovid) of the educational department[9] whose data are obtained through the report provided by each school director as soon as a positive index case is detected by the health system and communicated to the school. Once the screening of the remaining bubble members has been made, the school director needs to report the positive cases that have been detected through this surveillance system. These cases (both the index and the secondary ones) are linked to the corresponding bubble groups, allowing identification of the educational level at which each outbreak occurs. Second, we used surveillance data provided by the Catalan Agency for Health Quality and Evaluation (AQuAS), which are directly extracted from the health surveillance system. Data from AQuAS were grouped after the age stages that correspond with the educational cycles: preschool (0–2 and 3–5), primary (6–11), middle (12–15) and high school (16–17). These data were used to validate the incidence rates obtained from the schools' surveillance system (Traçacovid).[9] In this sense, it is worth mentioning that in Spain, only primary and middle school

are mandatory, while preschool and high school remain optional. Therefore, one would expect a high correspondence between the 2 sources at the primary and secondary age stages, while the correlation could be lower at the first and last stages.

The reference population for assessing the relative incidence in each educational level or cycle was obtained from open data of the educational department (www.educacio.gencat.cat) and is shown in Table 1. The information on reference populations for assessing the relative incidences among population age-stages was also provided by AQuAS, and the data are shown in Table 1.

This analysis focuses on infection data generated during the first trimester (September to December 2020) in preschools, primary, middle and high schools, which together account for around 1.09 million students. We did not consider data from professional cycles (+16 years old), because of the important proportion of older students which could distort the relation with the estimated age. We accessed the details on the school, grade and bubble group of each positive case notified by the school surveillance system (Table 1). We automatically extracted, for each bubble group, the list of positive results and sorted them by the test date. We checked each case sequentially and evaluated whether it corresponded to a new index case or if it was linked to a previous case. The condition for determining whether 2 cases were temporally linked was the difference between the 2 dates being less than 10 days. In this case, they were deemed to be associated with the same outbreak.

We aggregated data for each course level to obtain the temporal dynamics of the outbreak. For each course level, we calculated the number of infected individuals per outbreak and computed the reproductive number inside the bubble groups (R*) as the average number of secondary cases for each index case:

$$R^x = \frac{\sum S_i}{N}$$

where N is the number of index cases corresponding to a particular course level and S_i is the number of outbreaks with i secondary cases. The distribution of S_i is not homogeneous, and
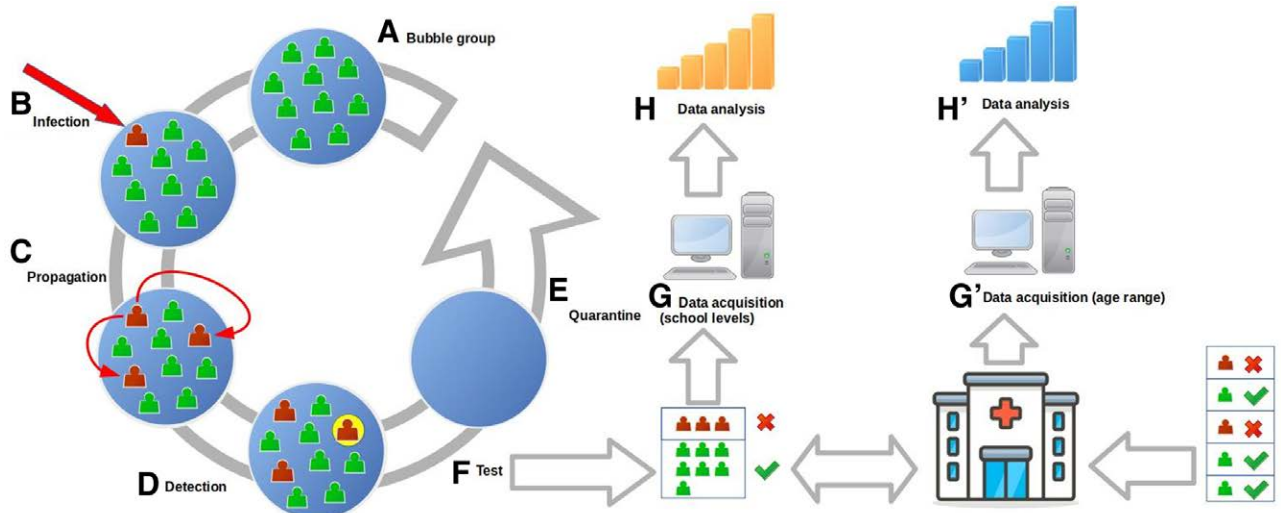
FIGURE 1. Outline of the protocols and data acquisition process. A: Bubble group made up of 10–30 students. B: Infection, outside, of an individual of the bubble group, (C) propagation of the disease inside the bubble group, (D) detection of one of the infected individuals, (E) quarantine of the whole bubble group for 10 days, (F) test of all the individuals of the group, (G) data acquisition of all the positives per group and school level, (H) data analysis of the propagation inside the bubble group per school level, (G) data acquisition for age ranges from the general data system, and (H) data analysis per age range.

TABLE 1. Organization of Levels and Distribution by Grade and Age of Cases and Reproductive Number

Level	Grade	Age†	Total Students	Index Cases	Total Cases	R*	Population in Each Age Range
Preschool	P3	3	63,857	658	823	0.19	
	P4	4	68,214	698	874	0.20	
	P5	5	70,878	857	1160	0.28	
		(3–5)	202.949				218.103
Primary school	1	6	72,503	856	1161	0.28	
	2	7	72,249	989	1326	0.28	
	3	8	76,530	1141	1621	0.35	
	4	9	78,577	1228	1756	0.34	
	5	10	80,751	1364	1994	0.37	
	6	11	81,714	1401	2167	0.44	
		(6–11)	462.324				495.828
Middle school	1	12	84,795	1421	2312	0.51	
	2	13	84,292	1400	2324	0.53	
	3	14	84,323	1425	2377	0.57	
	4	15	82,370	1542	2589	0.57	
		(12–15)	335.780				323.288
High school	1	16	31,588	914	1536	0.54	
	2	17	32,858	787	1334	0.58	
		(16–17)	98.263				148.709

†Corresponds to the age of a student on January 1, 2021 and not considering repeater students.

diverse statistical distributions can be employed to fit such distributions. Later, we employ the Poisson and the negative binomial distributions. First, the Poisson distribution describes random independent events and follows the next distribution:

$$P(S_i = X) = \frac{R^x e^{-R}}{X!}$$

where R is the mean value, which corresponds to R*, the average reproductive number. However, the previous distribution cannot model overdispersed data, to introduce such property we can employ the negative binomial distribution, limit of the Poisson distribution when the parameter R results from a distribution, which reads:

$$P(S_i = X) = \binom{X + r - 1}{r - 1}(k)^X(1 - k)^r$$

where the parameter k corresponds to the dispersion parameter and the parameter r reads:

$$r = \frac{R \cdot k}{1 - k}$$

where, as in the previous case, R is the mean value, which corresponds to R*, the average reproductive number.

Catalonia suffered the second wave of SARS-CoV-2 in autumn 2020. In late October, the cases per 100,000 inhabitants accumulated in the previous 14 days (A14) peaked at 847, the historical detected maximum. At that time, quarantines of positive cases and contacts in Catalonia lasted 10 days.

RESULTS

Figure 2 compares the daily new cases for the 2 data sources for the 4 different educational cycles. Despite the different data sources, a satisfactory overlay of both data is observed. As may be seen, correspondence of primary level data is extremely high. Preschool and high school surveillance curves are below those given by the health system, which agrees with the fact that these educational cycles are not compulsory. Therefore, it would be expected that the number of cases reported by the school surveillance system was lower than the number captured by the health system. Surprisingly, the number of cases in middle school reported by the educational surveillance system was higher than that reported by the health system. This could be explained by the existence of a certain percentage (estimated at 17%) of students ≥16 years old still enrolled in middle school, as they repeated one or more courses. Such students thus still appear in the middle school statistics in spite of being classified in the 16–17 age group by the health surveillance system.

The 2 data sources reported an important increase in incidence associated with age. Figure 3 shows the total number of students with a coronavirus disease 2019 diagnosis for each school level. The age-dependency was clear except for high school where the reduced number of students at such levels could be a differentiating factor.

We obtained the total number of index cases for each school level (Fig. 3A). There were no secondary cases in 75% ± 5% (78.5% primary, 71% secondary) of the index cases (Fig. 3A). Otherwise, a total of 20% of primary school and 25.5% of middle school index cases were linked to outbreaks with 2–4 secondary infected students. Finally, there were some outbreaks with a larger number of secondary cases, only occurring in 1.5% and 3.5% of the cases, respectively. In general, we noted that while 70% of the students were infected outside of the bubble group, around 30% of the students could have been infected inside (Fig. 3B). This percentage was associated with the age of the group, with 20% and 40% becoming infected within the bubble group for the youngest and the oldest grades, respectively. An equivalent description of such information is the comparison of the total number of cumulative cases and index cases during the period of study, shown in Figure 3C. An average of 1000–2000 students per course level was infected by SARS-CoV-2 during this period. The reduction in the number of infected kids for older ages was associated with a decrease in the number of students attending high school, which is not mandatory (Table 1). To remove the effect of the difference in population sizes with the course levels, we calculated the percentage of children as index or secondary case (Fig. 3D). The percentage of kids acting as index cases remained roughly constant at 1.7% for >9 years of age. In contrast, Figure 3D shows that the propagation inside the bubble was strongly associated with age because the percentage of secondary cases increased with age, either because children's
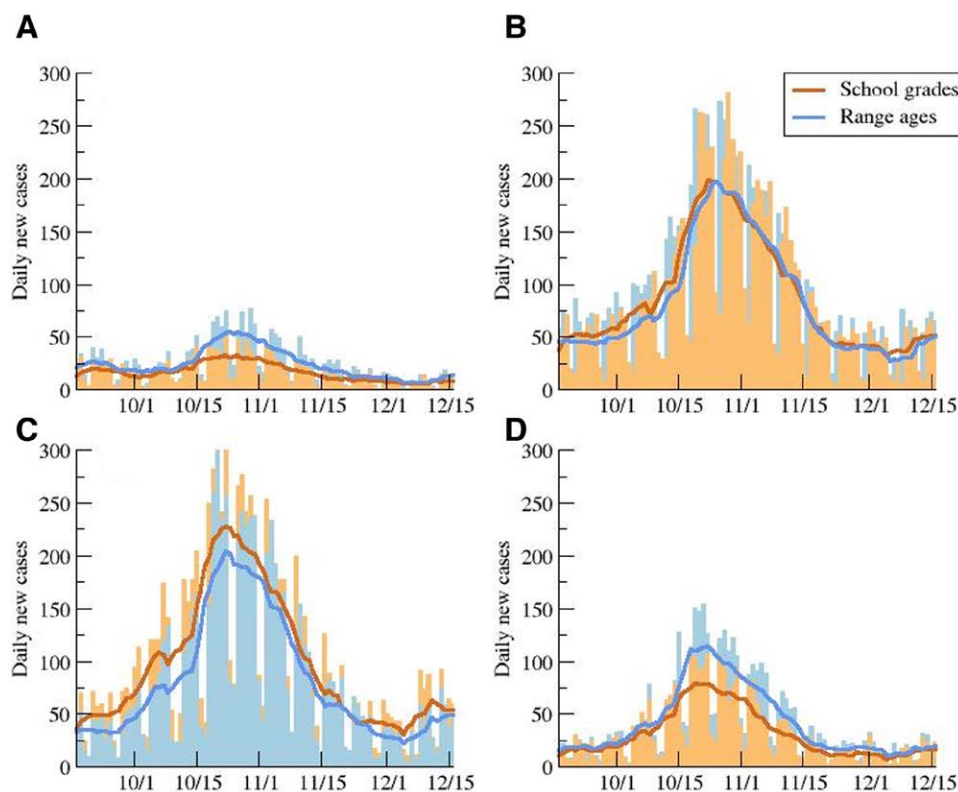
FIGURE 2. Evolution of daily new cases distributed by age. Daily new cases for preschool (A), primary (B), middle (C), and high (D) school, and 7-day averages obtained from the data in the schools by school years and from the official data by age ranges.

infectiousness genuinely increases with age or because they were in higher risk situations that could trigger transmission.

The reproductive number R represents the average number of individuals infected by an index case. We did not have access to data on all the infected individuals outside the bubble group, and therefore, the overall R cannot easily be calculated. However, we could measure the local reproductive number inside the bubble group R*; see Methods section. Figure 4A shows that the estimated value of R* for the different course levels was associated with age except for the number of kids repeating a course. The average value of R* was 0.40, and all course-related values of R* were within the range 0.2–0.6. The most remarkable result was the association of the linear growth of R* with the increasing age of the students. While the smallest value of 0.2 corresponded to preschool, the highest values, around 0.6, were related with high school and the last years of middle school.

The calculation of the reproductive number R* was obtained from the distribution of sizes for all the outbreaks observed for a certain course level; see Figure 4B–I for the cumulative data for preschool, primary, middle and high school. A sound approach for the distribution (Fig. 4B–E) is based on a completely random process of infection where, for a given probability of infection, the number of actual infected kids followed a Poisson distribution, which conveniently fits the data. The Poisson fits properly capture the large number of outbreaks with small numbers of secondary cases (1–4), which were actually the majority.

Figure 4B–I shows that the fit of a negative binomial distribution to the distribution of the outbreaks was adequate. The resulting values for R* did fit the data values properly, as shown by way

of comparison. Super-spreading events are thus key to reproducing the overall dynamics leading to a significant increase in R*. This can also be easily understood by the high clustering coefficient[11] which in our case corresponds to the 14% and 17% of kids in primary and middle school, respectively, who generated 80% of the transmissions inside the bubble group. Note that the great majority of index cases do not produce any transmission (see Fig. 3A). Therefore, super-spreading events among children remained rare.

DISCUSSION

The satisfactory crosscheck of the 2 data sources allows confirmation of the validity of the reported incident cases from the school system, permitting their use for a more thorough statistical analysis. The differences between the 2 data sources may be explained by the number of repeating students, which increases with the school level, and the nonobligatory nature of preschool and high school.

Some controversy still exists regarding the susceptibility of kids to becoming infected.[3,12] Low prevalence of SARS-CoV-2 antibodies in children and parents was found in southwest Germany, indicating that children did not drive the SARS-CoV-2 spread.[13] One possible explanation is related with an antibody-mediated immune response in children, similar to their confirmed infected parents and specific to SARS-CoV-2, without virological confirmation of infection.[14] The low values of R* we obtained for preschool period (3–5 years), with students not wearing masks inside the bubble group, are in concordance with previous studies.

The dependence of the susceptibility on age may be evaluated from the number of index cases in schools. Our analysis shows that
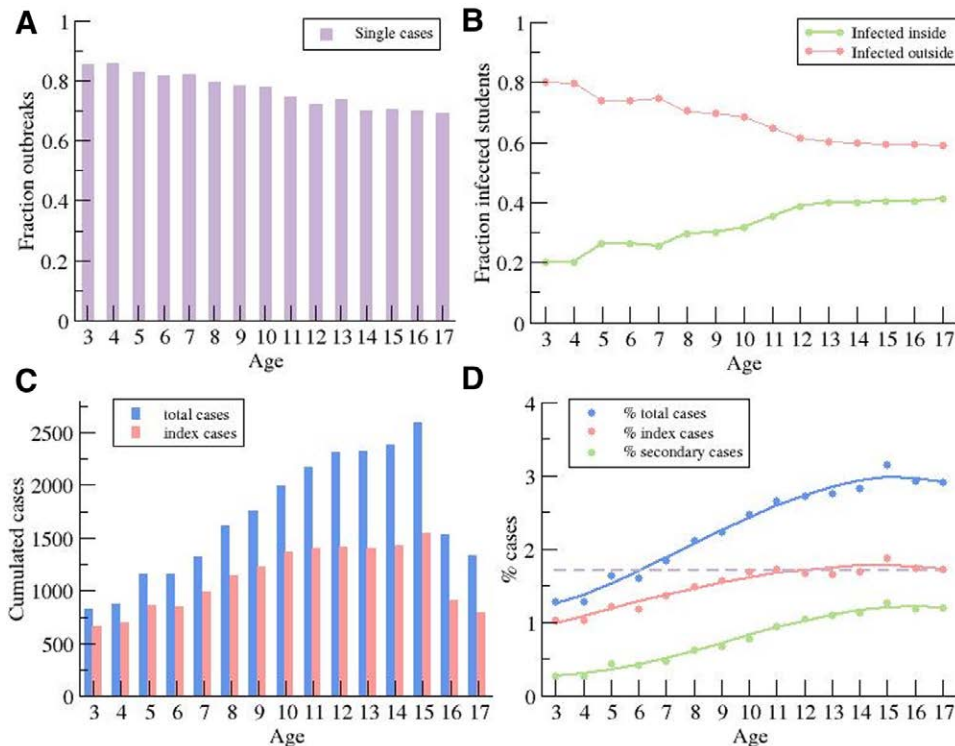
FIGURE 3. Distribution of daily new cases by age. A: Dependence of the fraction of index cases without secondary cases on the educational level. B: Fraction of students infected inside and outside of the bubble group, assuming no simultaneous infections by different index cases. C: Total number of cases and index cases for different educational level. D: Percentage of students infected during the 3 months (as reference for Catalonia 3% of the population were infected), percentage of case indexes, and percentage of secondary cases.

there is a constant percentage of index cases for ages older than 9, and this could indicate that the inclusion of index cases in the bubble group is independent of age (for age >9 years), and therefore, the probability of getting infected outside of school (typically from an adult at home)[15] is constant and independent for older children.[16] The total number of index cases for each level depends on the incidence outside of school; therefore, it indicates that for age >9 years, the probability of introducing the virus inside the bubble group depends only on the incidence outside.

The value of R* observed is in agreement with other values in the literature. For example, in another study, also in Catalonia, a value of R* = 0.3 was found in summer camps[17] when the weather was better and outdoor activities more common. Again in summer, but in England, larger values were obtained R* = 1[18] probably because of less common use of masks inside the groups. Low transmission and smaller values of R* were found in other studies.[19]

The estimated value of the reproductive number could be even lower for 2 reasons. First, because of the possible presence of infected but asymptomatic kids who do not transmit the virus inside the bubble group and who are not detected by an external contact tracing. In this case, the classmates would not be screened, and both the original case and the potential infected mates would remain undetected. Note that the probability of detection of an initial case in a bubble group, which triggers the group screening, increases with the number of infected individuals in the bubble. Therefore, it is possible that some outbreaks with small numbers of infected individuals or just a single index case may be more easily unnoticed. Second, some outbreaks may be caused by 2 individuals who are independently infected simultaneously, in particular in

cases of large community incidence, which our method is unable to discern.

The reproductive number inside the bubble group (R*) in Catalonian schools during the second wave of the epidemic has been rather small and increased importantly with age. This fact lends support to the idea that, when strict safety measures are implemented for students and staff, schools, especially primary schools, are not the drivers of the SARS-CoV-2 transmission. A similar phenomenon has been observed in schools in the United States[20] where masks were extensively employed. The number of outbreaks does depend on the underlying community incidence because, by definition, the index cases are always imported from outside the bubble group. However, the reproductive number does not substantially change during the period of analysis, and therefore, R* does not depend on the entry of new cases or local incidence. Finally, note that the data from September to December do not reflect the incremental proportion of the more infectious variants circulating,[21] as observed in 2021 in Spain.

Propagation is heavily affected by the presence of super-spreader events, which are responsible for a substantial increase in the value of R*. The use of the logarithm axis in Figure 4F–I permits us to visualize the distribution of outbreaks with large numbers of cases. This representation of the data shows the failure of the Poisson distribution in the description of the distribution of outbreaks with large numbers of secondary cases—greater than 4—and therefore explains the failure in the estimation of the value of R*. To recover the distribution for large number of secondary cases, we need a different distribution function—for example, the negative binomial distribution, which incorporates super-spreading
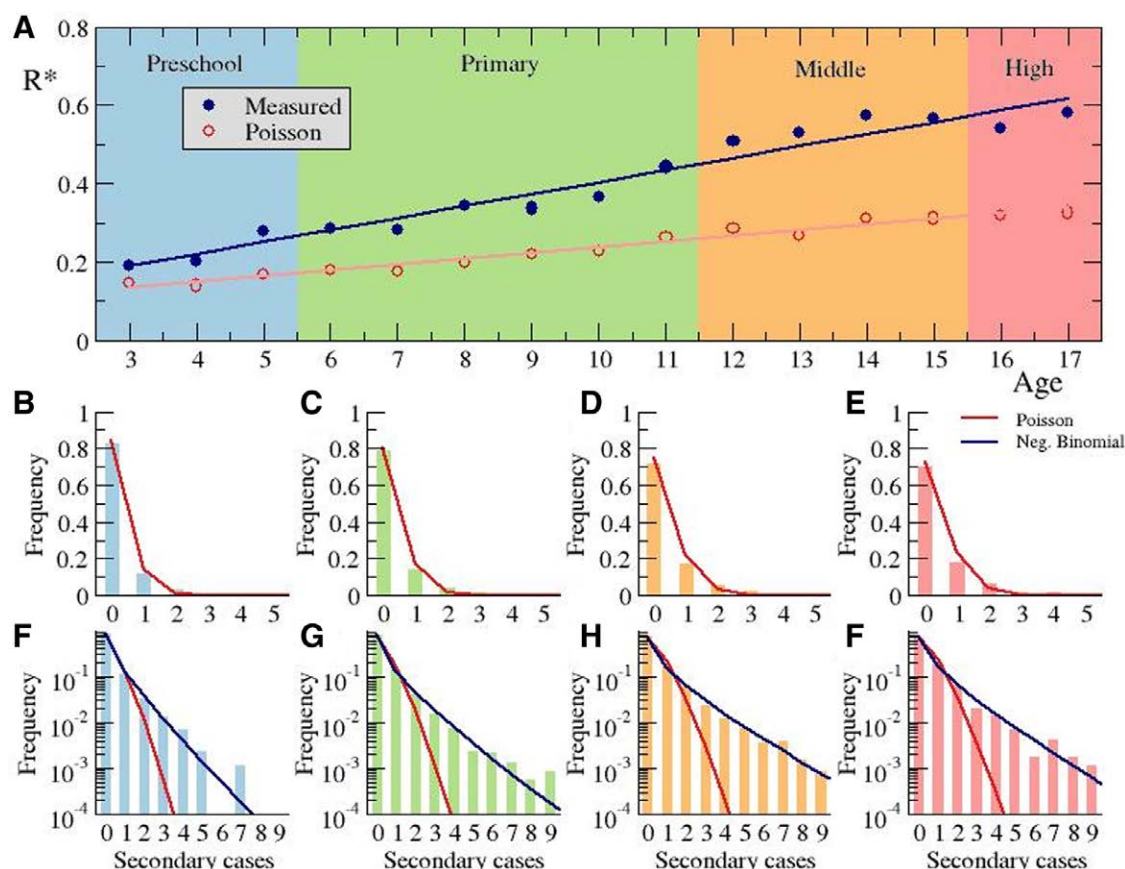
FIGURE 4. Calculation of the reproductive number. A: Dependence of the reproductive number on the educational level and the corresponding associated age. Frequency of the number of secondary cases per index case for preschool (B, F) primary (C, G), middle (D, H), and high (E, I) school, using linear (B–E) and logarithmic (F–I) scales. Poisson distribution is fitted employing a least square method for red lines in panels (B–I), and the negative binomial distribution is fitted employing maximum likelihood estimation method[10] for blue lines in panels (F–I) giving rise to values R = 0.27 and dispersion parameter k = 0.41 (F), R = 0.35 and dispersion parameter k = 0.32 (G), R = 0.55 and dispersion parameter k = 0.35 (H), R = 0.56 and dispersion parameter k = 0.41 (I).

situations. The presence of a small number of large clusters of infection is not a statistical fluctuation, and it may be related with unusual situations—for example, insufficiently ventilated classrooms or high-risk activities outside of school. Adherence to these measures seems, therefore, critical and may depend on the age.

In conclusion, after the validation of the data from the educational system, we obtained the propagation properties inside the bubble groups. We saw an increase in the propagation of the disease with the age of the bubble groups. Furthermore, the low reproductive number obtained indicates that schools are not the drivers of SARS-CoV-2 transmission, at least under the restriction measures employed during the time period which this study considered.

REFERENCES

1. Götzinger F, Santiago-García B, Noguera-Julián A, et al.; ptbnet COVID-19 Study Group. COVID-19 in children and adolescents in Europe: a multinational, multicentre cohort study. *Lancet Child Adolesc Health.* 2020;4:653–661.

2. Munro APS, Faust SN. COVID-19 in children: current evidence and key questions. *Curr Opin Infect Dis.* 2020;33:540–547.

3. Viner RM, Mytton OT, Bonell C, et al. Susceptibility to SARS-CoV-2 infection among children and adolescents compared with adults: a systematic review and meta-analysis. *JAMA Pediatr.* 2021;175:143–156.

4. Viner RM, Russell SJ, Croker H, et al. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *Lancet Child Adolesc Health.* 2020;4:397–404.

5. Viner RM, Bonell C, Drake L, et al. Reopening schools during the COVID-19 pandemic: governments must balance the uncertainty and risks of reopening schools against the clear harms associated with prolonged closure. *Arch Dis Child.* 2020;106:111–113.

6. Buonsenso D, Roland D, De Rose C, et al. Schools closures during the COVID-19 pandemic: a catastrophic global situation. *Pediatr Infect Dis J.* 2021;40:e146–e150.

7. European Centers for Disease Control and Prevention. COVID-19 in children and the role of school settings in COVID-19 transmission. 2020. Available at: https://www.ecdc.europa.eu/en/publications-data/children-and-school-settings-covid-19-transmission#no-link. Accessed July 15, 2021.

8. Transitioning from CDC's indicators for dynamic school decision-making to CDC's operational strategy for K-12 schools through phased mitigation to reduce COVID-19. Available at: https://www.cdc.gov/coronavirus/2019-ncov/community/schools-childcare. Accessed July 15, 2021.

9. Department of Education, Safe School Program. Traçacovid. Available at: http://educacio.gencat.cat/ca/actualitat/escolasegura. Accessed July 15, 2021.

10. Lloyd-Smith JO. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PLoS One.* 2007;2:e180.

11. Adam DC, Wu P, Wong JY, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat Med.* 2020;26:1714–1719.

12. Dattner I, Goldberg Y, Katriel G, et al. The role of children in the spread of COVID-19: using household data from Bnei Brak, Israel, to estimate the relative susceptibility and infectivity of children. *PLoS Comput Biol.* 2021;17:e1008559.

13. Toenshoff B, Müller B, Elling R, et al. Prevalence of SARS-CoV-2 infection in children and their parents in southwest Germany. *JAMA Pediatr.* 2021;175:586–593.

14. Tosif S, Neeland MR, Sutton P, et al. Immune responses to SARS-CoV-2 in three children of parents with symptomatic COVID-19. *Nat Commun.* 2020;11:5703.

15. Soriano-Arandes A, Gatell A, Serrano P, et al. Household SARS-CoV-2 transmission and children: a network prospective study [published online ahead of print March 12, 2021]. *Clin Infect Dis.* doi: 10.1093/cid/ciab228.

16. Brotons P, Launes C, Buetas E, et al. Kids corona study group, susceptibility to severe acute respiratory syndrome coronavirus 2 infection among children and adults: a seroprevalence study of family households in the Barcelona metropolitan region, Spain. *Clin Infect Dis.* 2020;72:e970–e977.

17. Jordan I, Fernandez de Sevilla M, Fumado V, et al. Transmission of SARS-CoV-2 infection among children in summer schools applying stringent control measures in Barcelona, Spain [published online ahead of print March 12, 2021]. *Clin Infect Dis.* doi: 10.1093/cid/ciab227.

18. Ismail SA, Saliba V, Lopez Bernal J, et al. SARS-CoV-2 infection and transmission in educational settings: a prospective, cross-sectional analysis of infection clusters and outbreaks in England. *Lancet Infect Dis.* 2021;21:344–353.

19. Goldstein E, Lipsitch M, Cevik M. On the effect of age on the transmission of SARS-CoV-2 in households, schools, and the community. *J Infect Dis.* 2021;223:362–369.

20. Hershow RB, Wu K, Lewis NM, et al. Low SARS-CoV-2 transmission in elementary schools — Salt Lake County, Utah, December 3, 2020–January 31, 2021. *MMWR Morb Mortal Weekly Rep.* 2021;70:442–448.

21. Davies NG, Nicholas G, Abbott S, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. *Science.* 2021;372:6538.