Factors affecting the transmission of SARS-CoV-2 in school settings

Connor Reynolds, MPH[1*]; Sydney Ng, MPH[2*]; Wan Yang, PhD[1#]
[1]Department of Epidemiology, [2]Department of Biostatistics, Mailman School of Public Health, Columbia University;
[*]Authors contributed equally;
[#]Correspondence to: wy2202@columbia.edu

Abstract

Background

Several studies have reported SARS-CoV-2 outbreaks in schools, with a wide range of secondary attack rate (SAR; range: 0-100%). We aimed to examine the key risk factors to better understand transmission in school settings.

Methods

We collected records of SARS-CoV-2 school outbreaks globally published from January 2020 to January 2021 and compiled information on hypothesized risk factors. We utilized the directed acyclic graph (DAG) to conceptualize the risk mechanisms, used logistic regression to examine each risk-factor group, and further built multiple variable models based on the marginal analysis. Adjusted odds ratios (aOR) and 95% confidence intervals (CI) were calculated.

Results

From 17 relevant articles, 26 school clusters were included for analysis. The best-fit model showed that the intensity of community transmission (aOR: 1.26; 95% CI: 1.22 - 1.30, for each increase of 10 cases per 100,000 persons per week), social distancing (aOR: 0.26; 95% CI: 0.18 - 0.37), mask-wearing (aOR: 0.52; 95% CI: 0.35 - 0.78) were associated the risk of SARS-CoV-2 infection in schools. Compared to students in pre-schools, the aOR was 0.12 (95% CI: 0.07 - 0.19) for students in primary schools and 1.31 (95% CI: 0.93 - 1.87) for students in high schools.

Conclusions

Preventive measures in both schools (e.g. social distancing and mask-wearing) and communities (additionally, vaccination) should be taken to collectively reduce transmission and protect children in schools. Flexible reopening policies may be considered for different levels of schools given their risk differences.

Keywords

COVID-19, transmission, children, school, secondary attack rate

Key messages

- We collected published records of COVID-19 school outbreaks globally to investigate the considerable heterogeneity in secondary attack rates (SAR) reported from school outbreaks and compiled information regarding potential risk factors.
- Higher community death and case rates are associated with higher SARs in children in school settings.
- Mask-wearing and social distancing are associated with lower SARs.
- When compared to pre-schools and early childhood education centers, primary schools have lower rates of transmission of SARS-CoV-2 however high schools have higher rates.

Introduction

Since the early stages of the COVID-19 pandemic, concerns have been raised about the impact of schools on community transmission and the well-being of students and staff, as well as the impact on the schedules of healthcare workers concerning childcare [1]. Out of an abundance of caution and fear that the SARS-CoV-2 virus would spread rapidly in schools much like influenza pandemics [2], countries globally decided to suspend in-person classes and begin online instruction. By April 2020, over 1.5 million students worldwide were affected by school closures in response to the COVID-19 pandemic [3]. In contrast to influenza pandemics where children are the key drivers of transmission, studies have indicated that children are likely less susceptible to SARS-CoV-2 infection, tend to experience less severe disease when infected, and likely have lower transmissibility [4, 5]. Given this new evidence, schools in many places have gradually reopened since the summer of 2020, while implementing varying level of preventive measures (e.g., mask wearing, distancing, limiting the number of students, rotating schedules, and viral testing) to reduce risk of transmission. Given these circumstances, the risk of SARS-CoV-2 outbreaks in school settings may differ substantially across space and time. Indeed, several studies have examined school outbreaks of COVID-19 and reported secondary attack rates (SAR) – i.e., the proportion of infected contacts of an index case out of all contacts of that index case [6] – ranging from 0 (i.e., no secondary infections) to 100% (i.e., infections among all contacts). However, this discrepancy is still not fully understood and a better understanding can inform better preventive measures for future outbreaks not limited to COVID-19 or school settings.

To identify the main factors that determine the transmission of SARS-CoV-2 in schools and inform strategies to prevent future school outbreaks, here we examined the associations between SARS-CoV-2 SAR in children and various potential risk factors. We compiled data from relevant studies in the literature reporting SARS-CoV-2 SAR in schools and for posited risk factors (e.g. incidence in the community) and further used regression models to examine key

risk factors of having high SAR in schools. We found that appropriate distancing and mask wearing, school levels, and the magnitude of SARS-CoV-2 transmission in the community are key determinants of SAR in school settings.

Methods

Data sources

Studies were searched for on the "Living Evidence for COVID-19" database [24], which retrieves articles from EMBASE via Ovid, PubMed, BioRxiv, and MedRxiv. Any article within this database was considered, from December 2019 up to February 17, 2021. The search terms used include "transmission AND (school OR schools)" or "transmission AND children." When titles and abstracts were identified as being potentially relevant, the articles were read to determine if the outbreaks took place in a school setting and the number of infections and contacts were reported. In addition, we extracted 11 observations included in a systematic review of evidence regarding the ability of children to transmit SARS-CoV-2 in schools [7]. In total, 17 studies were included in this analysis (see Appendix 1).

Relevant data, as deemed by an initial conceptual analysis using the directed acyclic graph (DAG; see details below), were taken from the articles identified above. These included the time period of the study, study design, location, season, age of children, whether the study happened during a lockdown, type of school according to the International Standard Classification of Education [8], reported SARs, number of contacts of the index case, testing method (PCR vs. serology), level of surveillance (all contacts, some contacts, only symptomatic), and whether masks and social distancing were required. In addition, we compiled additional data for potential risk or confounding factors of SARS- CoV-2 transmission in schools for each identified study as detailed in the next section.

Conceptual analysis and variable coding

The unit of analysis was individual study and in cases where several school types were covered in one study the data were stratified by those school types. We first conducted a conceptual analysis using the DAG and identified eight key components that may affect SARS-CoV-2 transmission in schools (Fig 1). Below we describe each of the eight components, rationale for inclusion, and related variables examined.

1) School types, based on studies indicating differential transmission risk among different age groups [20, 22]. Here, we examined this factor as a categorical variable including 4 levels, i.e. pre-school or early childhood education center (ECEC), primary school, high school, and mixed-level school. The first three levels were per reports in the included school studies. For studies that examined several types of school but did not report school type specific SARs, we assigned

them to a “mixed-level school” category. For example, if a study gave the overall SAR combining a pre-school and a primary school it was given the value “mixed-level school.”
SARS-CoV-2 SAR among children in school settings is the number of infected contacts divided by the total number of contacts at each school.

2) Physical school settings such as student density in the classroom and ventilation systems that may affect the intensity of school contact and clearance of air. As it is difficult to obtain information related to ventilation settings, here we included class size in our analysis based on the average number of students per classroom in each country, as reported by the Organisation for Economic Co-operation and Development (OECD) [12].

3) Cultural climates, which “represent independent preferences for one state of affairs over another that distinguish countries (rather than individuals) from each other” [9] and may reflect the collective risk tendency of a population. The Hofstede’s cultural dimensions theory [9] included 6 related measures including individualism, masculinity, uncertainty avoidance, long term orientation, and indulgence. In particular, individualism is defined as the degree of interdependence of society maintains among its members. We reasoned that the individualism measure would be most relevant to the level of compliancy to public health interventions and in turn the risk of SARS-CoV-2 transmission. Thus, here we included individualism in our analysis and dichotomized the reported values for each country into high (≥ 50) or low.

4) Preventive measures, which may reduce outbreak risk. Here we included both mask-wearing and social distancing. For these variables, we dichotomized based on as the use of face masks (or social distancing) by either students, staff, or both. For instance, for mask-wearing, if either group (students or staff) were required to wear face masks onsite, the school would receive a value of 1 and 0 otherwise.

5) Surveillance and/or testing policies implemented in schools. On the one hand, testing policies could affect the reported values of SAR; for instance, testing of all contacts regardless of symptoms may lead to identification of more infections including those asymptomatic and increase the numerator of SAR. On the other, frequent testing of all if combined with school closure may serve as a containment measure to reduce the risk of onward transmission and in turn reduce SAR. Here we thus included the reported testing practices for contacts in the school outbreak clusters as a categorical ordinal variable. Three types of testing were reported in the school studies, including testing only the symptomatic, both symptomatic and some asymptomatic, and all contacts. However, due to the small sample size in “only symptomatic” (n = 3), we dichotomized surveillance to testing “only symptomatic or some asymptomatic” and “all contacts” of an index case in each school cluster.

6) Community transmission intensity, which may increase the introduction of infections into school settings. To examine its impact, we included two measures, i.e., the weekly case rate per 100,000 for the study area for the study period and weekly death rate per 100,000 from COVID-19; both measures were computed using data from the John Hopkins Coronavirus Resource Center [10] and standardized by overall country population. See "Sensitivity analysis" below for justification of these choices.

7) Seasonal changes such as humidity and temperature, which may affect the survival and transmission of SARS-CoV-2. Here, we used specific humidity (a measure of absolute humidity) to examine the potential impact from disease seasonality, as specific humidity and temperature are highly correlated. Specifically, ground surface temperature and relative humidity for each study location were extracted from the National Oceanic and Atmospheric Administration using the "stationaRy" package [11]. Daily mean specific humidity in g H_20/kg air was then computed based on the temperature and relative humidity data and was averaged over the study period.

8) Indicators of socioeconomic status such as national income that reflect a country's ability to mobilize resources to fight COVID-19. As such, we included measured national income for each study in our analysis; specifically, national income is measured as the gross domestic product (GDP) subtracting capital depreciation and adding net foreign income, using data from the World Inequality Database [13].

Statistical Analyses

Marginal analysis.

Due to the low number of observations (n = 26 schools), we conducted an initial analysis to test combinations of the DAG covariates described above. The goal was to examine the relationship between SAR and only one group of variables at a time and then include the most relevant predictors into the final model based on this analysis. For each test, we used a logistic regression model of the following form:

$$logit(SAR) \sim X$$

where logit is the log-odds (i.e., $log\left(\frac{p}{1-p}\right)$, with p as the probability of event) and SAR represents the SAR as reported from each of the 26 schools. X is one of the combinations of variables we examined as follows:

1) Preventive measures (mask-wearing and social distancing) adjusting for seasonal changes
2) Community transmission (weekly death rate per 100,000 and weekly case rate per 100,000)
3) Seasonal changes
4) Classroom size, adjusting for national income
5) School type

Multi-risk factor analysis

All five variable groups were found to be associated with SAR in the marginal analysis (see Results). We thus tested models including different combinations of these variables. For all models, we included surveillance type to account for potential biases in reporting including missing asymptomatic infections, which would underestimate SAR. We also assessed for confounding between our variables of interest and SARS-CoV-2 SAR (see adjustments specified above). This procedure began with including all of the significant variables from the marginal analysis and removing one at a time to see the effect of its removal on model fit. We then evaluated and selected the most parsimonious model with the best fit based on the Akaike information criterion (AIC; Table S1). The best performing model took the following form:

$logit(SAR) \sim$ mask-wearing + social distancing + weekly case rate + school type + surveillance

All statistical analyses were performed in RStudio, a user interface for R (R Foundation for Statistical Computing, Vienna, Austria). All models were fitted using the "glm" function from the built-in "stats" library in R.

Sensitivity analysis

We tested different measures of community transmission, to examine the robustness of our models to potential biases due to variations in case-ascertainment, mortality risk, and delay in event occurrence (e.g. from infection to death) and reporting. Specifically, we additionally examined two mortality-related measures in representing the intensity of community transmission: 1) weekly death rate for the study period and 2) weekly death rate for the study period plus one month. The additional month added accounts for the time lag that occurs as death from COVID-19 may take several weeks [14]. Risk factors identified using these two measures were similar to those in our main analysis (Table S2). As the risk of death due to COVID-19 reduced substantially over time [15], likely due to improved medical treatment and management, such a change may bias the death rates recorded over time for the included studies spanning the initial phase in March 2020 to a later phase in Dec 2020. For this reason, we used weekly case rate per 100,000 to represent the intensity of community transmission in our main analysis.

Results

Summary statistics

We identified 26 reported SARS-CoV-2 outbreaks in schools, totaling 630 secondary cases in children among 8322 contacts. These outbreaks occurred in 12 countries, spanning 4 WHO regions including the Americas, Western Pacific, European, and Eastern Mediterranean Region.

228 Table 1 shows the frequencies and summary statistics for SARS-CoV-2 SAR and other variables
229 included. SARS-CoV-2 SAR and weekly cases per 100,000 population were both right skewed.
230 While the reported SAR ranged from 0 to 100%, the majority of schools reported very low SAR
231 (median: 0.7%, interquartile range: 0 – 17%). The majority of school settings included in these
232 studies required mask-wearing (19/26, or 73%) and distancing (14/26, or 54%). Around half of
233 the studies tested all contacts of the index case (12/26, or 46%). The studies covered roughly
234 even proportion for different school types: 5 (19%) were pre-schools, 9 (35%) were primary
235 schools, 7 (27%) were high schools, and 5 (19%) were mixed schools.
236
237 *Marginal analysis*
238 The marginal analysis (Figure 2), adjusting for surveillance, indicated that, for the intensity of
239 community transmission, each 10-unit increase in weekly case rate was associated with an
240 increased risk of contracting SARS-CoV-2 in schools [odds ratio (OR): 1.11; 95% confidence
241 interval (CI): 1.08 - 1.14]; and each 10-unit increase in weekly death rate was associated with an
242 increased risk of contracting SARS-CoV-2 in schools (OR: 1.47; 95% CI: 1.40 - 1.54). For
243 preventive measures, social distancing (OR: 0.21, 95% CI: 0.16 - 0.29) was associated with a
244 lower SAR in schools. For school type, compared to pre-schools, being in primary schools (OR:
245 0.69, 95% CI: 0.49 - 0.98) or mixed schools (OR: 0.69; 95% CI: 0.49 - 0.98) was associated with a
246 lower risk of SARS-CoV-2 infection whereas being in high schools (OR: 1.96; 95% CI: 1.44 - 2.69)
247 was associated with a higher risk of SARS-CoV-2 infection. For the school physical setting
248 measure, each 1-person increase in the national average class size was associated with an
249 increased risk of contracting SARS-CoV-2 in schools (OR: 1.26; 95% CI: 1.21 - 1.30). For disease
250 seasonality, each 1 g H_20/kg increase in specific humidity was associated with a decreased risk
251 of contracting SARS-CoV-2 in schools (OR: 0.88, 95% CI: 0.85, 0.92). Estimates were similar
252 when the model did not adjust for surveillance (Figure 2).
253
254 *Multi-risk factor analysis*
255 Among all models tested (Table S1), the best-performing model with the lowest AIC included
256 three key groups of risk factors, namely, intensity of community transmission, preventive
257 measures, and school type, adjusting for surveillance. Overall, these risk factors in combination
258 were able to explain 57.7% of the variance in the reported SARS-CoV-2 SAR (McFadden's
259 pseudo R^2 = 0.577; Fig 3) [21, 23]. Fig 4 and Table 2 show the estimated adjusted ORs (aORs)
260 for each risk factor.

261
262 The multi-risk factor model showed that every 10-unit increase in COVID-19 cases rate in the
263 community was associated with an increased risk of SARS-CoV-2 infection in schools (aOR: 1.26;
264 95% CI: 1.22 - 1.30). Conversely, both social distancing (aOR: 0.26; 95% CI: 0.18 - 0.37) and

mask-wearing (aOR: 0.52; 95% CI: 0.35 - 0.78) were associated a decreased risk of SARS-CoV-2 infection in schools. Compared to students in pre-schools, the aOR was 0.12 (95% CI: 0.07, 0.19) for students in primary schools, 1.31 (95% CI: 0.93, 1.87) for students in high schools, and 0.26 (95% CI: 0.18, 0.39) for students in mixed schools.

Discussion

Leveraging available data on multiple reported SARS-CoV-2 outbreaks in schools and potential risk factors, we have examined main factors associated with the risk of SARS-CoV-2 transmission in schools. Our analyses suggest that the most significant predictors of SARS-CoV-2 SAR in schools are the level of community transmission (measured as weekly COVID-19 case rate) and preventative measures, once adjusted for surveillance and school type.

The model identified community transmission as the most significant variable associated with the risk of SARS-CoV-2 transmission in schools. In particular, the multi-risk factor model estimated that, in a community of 100,000 persons, every 10 additional cases per week in the community increases the odds of infection in schools by 26%. This finding highlights the potential community-to-school importation of cases and subsequent risk of outbreak in schools. As such, care must be applied when reopening or operating schools in areas with high levels of community transmission. In addition, reversing some original fears about school-to-community SARS-CoV-2 transmission, it is likely that the community transmission drives outbreaks in school, not the reverse. Further, with the availability of COVID-19 vaccines, predominantly to adults at present, it is paramount that all eligible adults get vaccinated promptly to lower the risk of transmission in the community and in turn to provide indirect protection to children via the increased population immunity.

Our model suggests that distancing policies have the greatest impact on reducing transmission in schools (aOR: 0.26, 95% CI: 0.18 - 0.37). This finding is likely a combined outcome of reduced number of contacts and reduced short-range transmission when social distancing policies were followed. Maintaining distance has necessitated fewer people in a room at the same time, leading to fewer contacts. In addition, the increased personal space in classroom enables students to avoid the likely higher viral concentration within short-range of the emitter (either via aerosols, droplets, or in combination) when far apart.

Nevertheless, it is important to note that social distancing measures may be more difficult to achieve fully in disadvantaged communities with underfunded and overcrowded schools, further exacerbating the COVID-19 crises in these communities, usually of color [16, 17]. Furthermore, racially motivated structural factors prevent these disadvantaged communities from practicing social distancing policies outside of the school. For example, these communities

303 tend to make up most of essential workers and thus have higher rates of transmission in their
304 community [19], increasing potential introduction of infections into schools.
305
306 Another preventive measure, mask-wearing, was also found to be significantly protective
307 against SARS-CoV-2 infection in schools. While the association was not as substantial as for
308 social distancing, a reduction by half (aOR: 0.52; 95% CI: 0.35 - 0.78) is still considerable. In
309 addition to mitigating the risk of infection, this protective association may also come from
310 masks reducing the emission of viral particles of SARS-CoV-2 infected wearers [18]. This latter
311 means of protection is particularly important in settings like schools, as infections in children
312 are more likely to be asymptomatic or mild and thus more difficult to identify [23].
313 Nevertheless, concerns have been raised regarding the potential developmental and health
314 harms of masks on very young children. Given this trade-off, more effective use of other
315 preventive measures such as vaccination of all those eligible especially care takers, improved
316 ventilation, and physical distancing may serve to reduce the need for long-term use of masks
317 among very young children for their protection.
318
319 In comparison to pre-schools, students in both primary and mixed-level schools had a lower risk
320 of SARS-CoV-2 infection whereas those in high schools had a higher risk. This finding is
321 consistent with previous studies indicating the likely lower susceptibility to SARS-CoV-2
322 infection and transmissibility among young children [23]. In addition, it is also likely in part due
323 to the greater ability of older children to follow directions regarding preventive measures but
324 with less compliance among high school students. The similarity of the estimates between
325 primary school and mixed-level schools is likely due to the number of primary schools that were
326 folded into the mixed-level school category – three of the five schools that made up this group
327 included primary schools. Overall, these findings support flexible reopening policies for
328 different levels of schools given the risk differences.
329
330 When schools test all contacts of the index case, this effectively functions as a control measure
331 in that more cases will be detected and is more likely to result in a school closure. This may
332 explain why the OR estimate for all contacts surveillance is protective, as the outbreak was
333 identified and able to be contained before the asymptomatic individuals were able to infect
334 others.
335
336 This study has several limitations. First, the analyses included a limited number of studies (n =
337 26), due to the inclusion criteria and the dearth of published reports on this topic to date.
338 Future research with additional relevant studies may enhance the statistical power of inference
339 for the risk factors examined here. Second, due to a lack of detailed information for each
340 specific school setting, we used proxy measures in the analyses (e.g., class size at the national

level was used rather than for each reporting school), which may have limited the ability of the models to identify the association of these factors with SARS-CoV-2 transmission risk. Similarly, due to the lack of data, we were not able to examine other key factors such as ventilation in classrooms, social economic status of individual students and their households, and potential differences in susceptibility and transmissibility by age group. Future work with comprehensive study designs and data collection is warranted to provide further insights into how infections, not limited to SARS-CoV-2, spread in schools and the broad, bi-directional impact of school and community transmission. This would be invariable to inform better strategies to combat future infectious disease outbreaks.

Funding

WY was supported by the National Institute of Allergy and Infectious Diseases (AI145883).

Author contributions

CR and SN contributed to study design, conducted the literature review and data extraction, and performed the statistical analysis; WY conceived and supervised the study; CR, SN, and WY wrote the manuscript.

Conflict of interest

None.

References

1. Bayham J, Fenichel EP. The impact of school closure for COVID-19 on the US healthcare workforce and the net mortality effects [Internet]. medRxiv; 2020. p. 2020.03.09.20033415. Available from: https://doi.org/10.1101/2020.03.09.20033415
2. Worby CJ, Chaves SS, Wallinga J, Lipsitch M, Finelli L, Goldstein E, et al. On the relative role of different age groups in influenza epidemics. Epidemics [Internet]. 2015 Dec;13:10–6. Available from: /pmc/articles/PMC4469206/ /pmc/articles/PMC4469206/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4469206/
3. UNESCO I of S. Education: From disruption to recovery [Internet]. 2021. Available from: https://en.unesco.org/covid19/educationresponse
4. Viner RM, Mytton OT, Bonell C, Melendez-Torres GJ, Ward J, Hudson L, et al. Susceptibility to SARS-CoV-2 Infection among Children and Adolescents Compared with Adults: A Systematic Review and Meta-Analysis [Internet]. Vol. 175. American Medical Association; 2021. p. 143–56. Available from: https://jamanetwork.com/

5. Otte im Kampe E, Lehfeld A-S, Buda S, Buchholz U, Haas W. Surveillance of COVID-19 school outbreaks, Germany, March to August 2020. Eurosurveillance [Internet]. 2020 Sep;25(38):2001645. Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.38.2001645
6. CDC D of SE&PD. Principles of epidemiology in public health practice, lesson 3: Measures of risk [Internet]. 2012. Available from: https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section2.html
7. Xu W, Li X, Dozier M, He Y, Kirolos A, Lang Z, et al. What is the evidence for transmission of COVID-19 by children in schools? A living systematic review [Internet]. Cold Spring Harbor Laboratory Press; 2020 Oct p. 2020.10.11.20210658. Available from: https://www.medrxiv.org/content/10.1101/2020.10.11.20210658v1 https://www.medrxiv.org/content/10.1101/2020.10.11.20210658v1.abstract
8. UNESCO I of S. International standard classification of education (ISCED) [Internet]. 2011. Available from: http://uis.unesco.org/en/topic/international-standard-classification-education-isced
9. Hofstede I. Hofstede insights: Country comparison [Internet]. 2020. Available from: https://www.hofstede-insights.com/country-comparison/the-usa/
10. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infectious Dis. 2020 Jan;
11. Iannone R. stationaRy: Detailed meteorological data from stations all over the world [Internet]. 2020. Available from: https://CRAN.R-project.org/package=stationaRy
12. Organisation for Economic Co-operation and Development. Education indicators in focus [Internet]. 2012. Available from: http://www.oecd.org/education/skills-beyond-school/EDIF%202012--N9%20FINAL.pdf
13. Alvaredo F, Chancel L, Piketty T, Saez E, Zucman G. World inequality database [Internet]. 2018. Available from: wid.world
14. Chen Y, Li T, Ye Y, Chen Y, Pan J. Impact of fundamental diseases on patients with COVID-19. Disaster Medicine and Public Health Preparedness. 2020;
15. James N, Menzies M, Radchenko P. COVID-19 second wave mortality in europe and the united states. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2021;
16. Baker BD, Srikanth A, Cotto Jr R, Green III PC. School funding disparities and the plight of latinx children. education policy analysis archives. 2020;28(135):n135.
17. Anyane-Yeboa A, Sato T, Sakuraba A. Racial disparities in COVID-19 deaths reveal harsh truths about structural inequality in america. Journal of internal medicine. 2020;288(4):479–80.
18. Asadi S, Cappa CD, Barreda S, Wexler AS, Bouvier NM, Ristenpart WD. Efficacy of masks and face coverings in controlling outward aerosol particle emission from expiratory activities. Scientific reports. 2020;10(1):1–3.

415 19. Dorn A van, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. The
416 Lancet. 2020 Apr;395(10232):1243–4.
417 20. Davies NG, Klepac P, Liu Y, Prem K, Jit M, Eggo RM. Age-dependent effects in the
418 transmission and control of COVID-19 epidemics. Nature medicine. 2020
419 Aug;26(8):1205-11.
420 21. Mittlböck M, Schemper M. Explained variation for logistic regression. Statistics in
421 medicine. 1996 Oct 15;15(19):1987-97.
422 22. Galanti M, Birger R, Ud-Dean M, Filip I, Morita H, Comito D, Anthony S, Freyer GA,
423 Ibrahim S, Lane B, Matienzo N, et al.. Rates of asymptomatic respiratory virus infection
424 across age groups. Epidemiology & Infection. 2019;147.
425 23. McFadden, D., *Conditional logit analysis of qualitative choice behavior*, in *Frontiers in*
426 *econometrics*, P. Zarembka, Editor. 1973, Academic Press: New York. p. 105 - 142.
427 24. COVID-19 Open Access Project. (2021). *Living Evidence on COVID-19*. Retrieved
428 from https://ispmbern.github.io/covid-19/living-review/
429

Table and Figure Captions

Table 1. Characteristics of collected data from literature review.

Table 2. Results of the best-fit multi-risk factor model for the identification of factors associated with SARS-CoV-2 SAR in schools. Adjusted odds ratio (aOR) estimates and 95% confidence intervals are given from the logistic regression model including surveillance to control for differences in testing school clusters, school type due to inconsistent reporting of age groups in the literature, mask-wearing and social distancing for preventative measures, and weekly case rates per 100,000.

Figure 1. Directed acyclic graph (DAG) describing the relationship among variables. This DAG represents the meaningful relationships between the variables relevant to SARS-CoV-2 SAR among children in school settings and informs all further analyses. The outcome measurement, SAR, is presented in red while risk factors are in black and surveillance in green.

Figure 2. Odds ratio estimates from the marginal models. Left panel shows results from models without adjusting for surveillance and right panel shows results from corresponding models additionally adjusting for surveillance. Black points show the mean odds ratio estimates and horizontal black bars show the 95% confidence intervals. The vertical black bar in each plot indicate the null value of 1.0. Each set of models is delineated by alternating the shaded regions.

Figure 3. Model fit of the best-performing multi-risk model. Black dots show the fitted SAR for each study (y-axis), compared to the observed SAR (x-axis); the red bars around each point show the 95% confidence intervals of model-estimates. The McFadden's pseudo-R^2 is computed using Eqn 30 in [23].

Figure 4. Odds ratio estimates from the best-performing multi-risk model. Black points show the mean odds ratio estimates and horizontal black bars show the 95% confidence intervals. The vertical black bar indicates the null value of 1.0. Each variable type is delineated by the shaded regions.

Supplemental Tables

Table S1. Performance of different models. In the marginal analysis, 5 models for 5 groups of risk factors are tested. In the multi-risk factor analysis, 11 models with all possible combinations of 4 groups of risk factors are tested (i.e., excluding models with only one group tested in the marginal analysis). All models adjusted for surveillance. Note that although class size was identified as a significant risk factor in the marginal analysis, further including this variable in

468 the multi-risk models gave impossible values for parameter estimates; we therefore excluded
469 class size in the multi-risk model analysis. The best-performing model with the lowest AIC is
470 bolded.
471
472 **Table S2.** Sensitivity analysis for both cases and deaths, with and without a 1-month extension
473 to the study time period.
474

475 **Tables**

476 **Table 1.** Characteristics of collected data from literature review

Variable	Summary
Surveillance	n (%)
All contacts	12 (46.2%)
Some asymptomatics	11 (42.3%)
Only symptomatics	3 (11.5%)
School type	n (%)
Mixed	5 (19.2%)
Secondary	7 (26.9%)
Primary	9 (34.6%)
Pre-school/ECEC	5 (19.2%)
Mask-wearing	n (%)
Yes	7 (26.9%)
No	19 (73.1%)
Social distancing	n (%)
Yes	12 (46.1%)
No	14 (53.9%)
High individualism (≥ 50)[a]	n (%)
	22 (88%)
Secondary attack rate	median (IQR)[b]
	0.007 (0 - 0.17)
Weekly case rate (per 100,000)	median (IQR)
	7.42 (0.29, 26.4)
Daily mean specific humidity (g H2O/kg air)	median (IQR)
	8.00 (4.95, 9.83)
Average class size	median (IQR)
	21.55 (20.00, 23.80)
National income (in thousands)	median (IQR)
	38,942 (37,248, 50,752)

[a] Individualism was accounted for in 25 studies, had no significant effect and was not included for final study.

[b] IQR, interquartile range

477

478

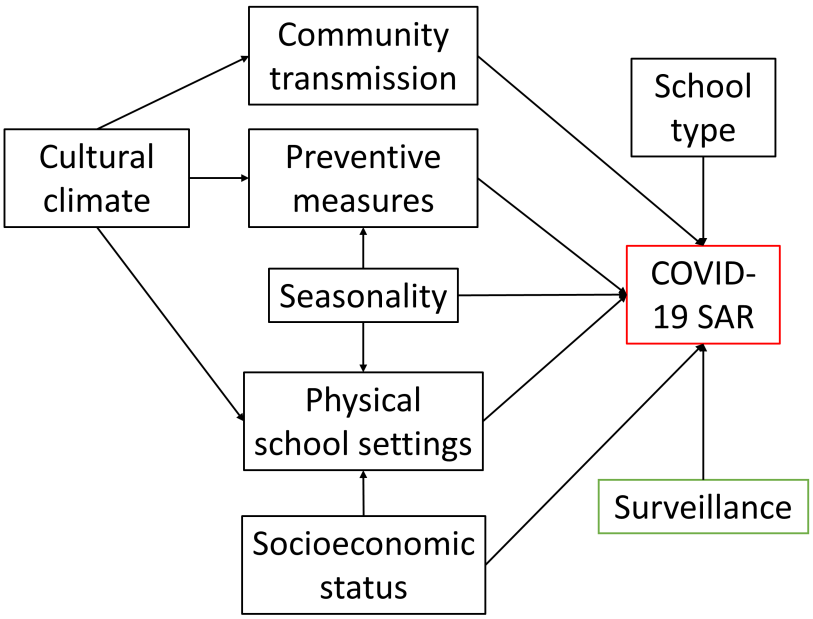
479 **Table 2.** Results of the best-fit multi-risk factor model for the identification of factors associated
480 with SARS-CoV-2 SAR in schools. Adjusted odds ratio (aOR) estimates and 95% confidence
481 intervals are given from the logistic regression model including surveillance to control for
482 differences in testing school clusters, school type due to inconsistent reporting of age groups in
483 the literature, mask-wearing and social distancing for preventative measures, and weekly case
484 rates per 100,000.

Variable	**aOR (95% CI)**
Surveillance	
All contacts	0.542 (0.454, 0.646)
Only symptomatics or some asymptomatics	(reference)
School Type	
Mixed	0.261 (0.176, 0.388)
Secondary	1.310 (0.925, 1.860)
Primary	0.115 (0.068, 0.196)
Pre-school/ECEC	(reference)
Mask wearing	
Yes	0.523 (0.351, 0.779)
No	(reference)
Social distancing	
Yes	0.260 (0.184, 0.368)
No	(reference)
Weekly case rate (per 100,000)	
	1.260 (1.220, 1.300)

485
486

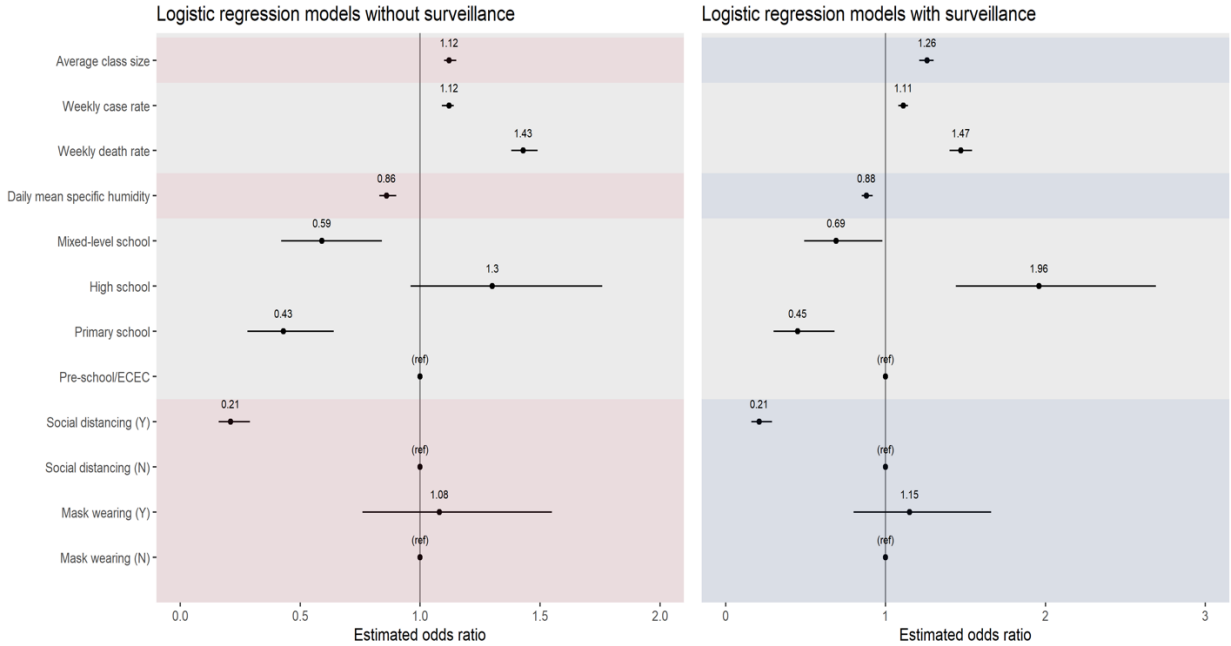
487 Figures

488 **Figure 1.** Directed acyclic graph (DAG) describing the relationship among variables. This DAG
489 represents the meaningful relationships between the variables relevant to SARS-CoV-2 SAR
490 among children in school settings and informs all further analyses. The outcome measurement,
491 SAR, is presented in red while risk factors are in black and surveillance in green.



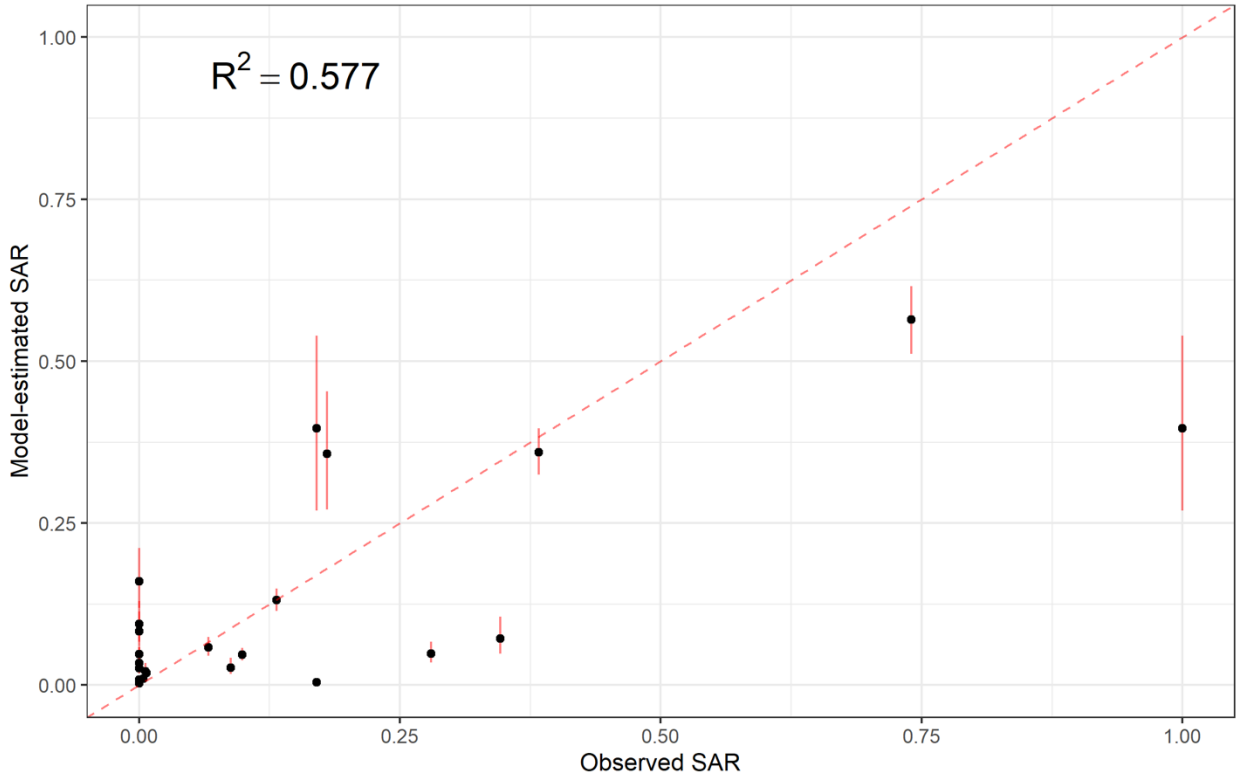
492
493

494 **Figure 2.** Odds ratio estimates from the marginal models. Left panel shows results from models
495 without adjusting for surveillance and right panel shows results from corresponding models
496 additionally adjusting for surveillance. Black points show the mean odds ratio estimates and
497 horizontal black bars show the 95% confidence intervals. The vertical black bar in each plot
498 indicate the null value of 1.0. Each set of models is delineated by alternating the shaded
499 regions.

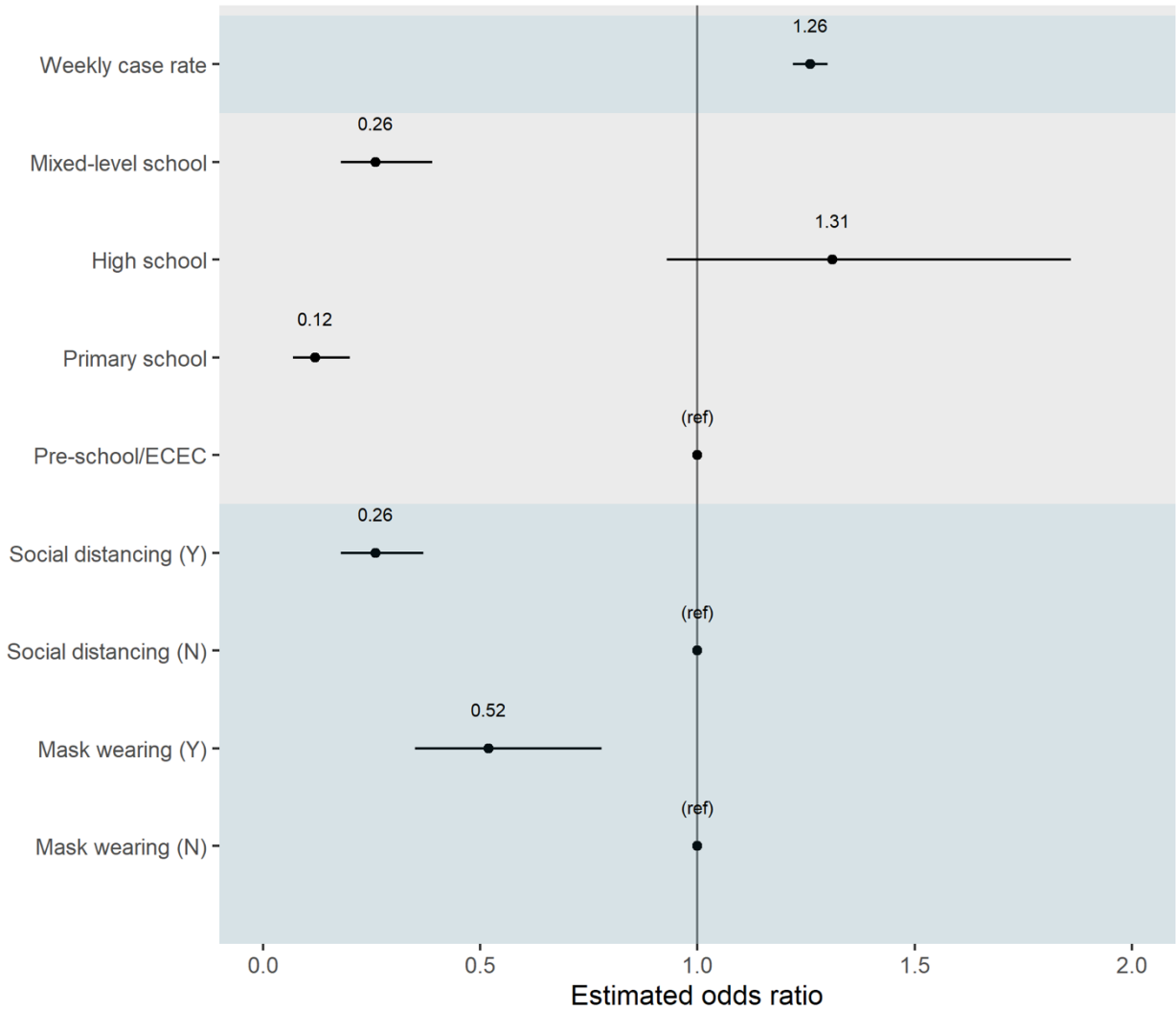


500
501

Figure 3. Model fit of the best-performing multi-risk model. Black dots show the fitted SAR for each study (y-axis), compared to the observed SAR (x-axis); the red bars around each point show the 95% confidence intervals of model-estimates. The McFadden's pseudo-R^2 is computed using Eqn 30 in [23].



507 **Figure 4.** Odds ratio estimates from the best-performing multi-risk model. Black points show
508 the mean odds ratio estimates and horizontal black bars show the 95% confidence intervals.
509 The vertical black bar indicates the null value of 1.0. Each variable type is delineated by the
510 shaded regions.



511

512

Supplemental Materials
for
Factors affecting the transmission of SARS-CoV-2 in school settings

Connor Reynolds, MPH; Sydney Ng, MPH; Wan Yang, PhD

This document includes: 1) Appendix 1 listing all SARS-CoV-2 school outbreak studies included in our analysis; and 2) all supplemental tables.

Appendix 1: Studies Included

1. J. P. Armann, M. Unrath, C. Kirsten, et al. "Anti-SARS-CoV-2 IgG antibodies in adolescent students and their teachers in Saxony, Germany (SchoolCoviDD19): very low seropraevalence and transmission rates." 2020.
2. A. Fontanet, R. Grant, L. Tondeur, et al. "SARS-CoV-2 infection in primary schools in northern France: A retrospective cohort study in an area of high transmission." In: *medRxiv* (Jun. 2020), p. 2020.06.25.20140178. DOI: 10.1101/2020.06.25.20140178. <URL: https://doi.org/10.1101/2020.06.25.20140178>.
3. C. Atherstone, M. Siegel, E. Schmitt-Matzen, et al. "SARS-CoV-2 Transmission Associated with High School Wrestling Tournaments Florida, December 2020-January 2021." In: *Morbidity and Mortality Weekly Report* 70.4 (2021), p. 141.
4. L. T. Brandal, T. S. Ofitserova, H. Meijerink, et al. "Minimal transmission of SARS-CoV-2 from paediatric COVID-19 cases in primary schools, Norway, August to November 2020." In: *Eurosurveillance* 26.1 (2021), p. 2002011.
5. D. Buonsenso and B. Graglia. "High rates of SARS-CoV-2 transmission in a high-school class." In: *Journal of Paediatrics and Child Health* (2021).
6. K. Danis, O. Epaulard, T. Benet, et al. "Cluster of coronavirus disease 2019 (COVID-19) in the French Alps, February 2020." In: *Clinical Infectious Diseases* 71.15 (2020), pp. 825-832.
7. S. Desmet, E. Ekinci, I. Wouters, et al. "No SARS-CoV-2 carriage observed in children attending daycare centers during the first weeks of the epidemic in Belgium." In: *medRxiv* (May. 2020), p. 2020.05.13.20095190. DOI: 10.1101/2020.05.13.20095190. <URL: https://doi.org/10.1101/2020.05.13.20095190>.
8. T. Dub, E. Erra, L. Hagberg, et al. "Transmission of SARS-CoV-2 following exposure in school settings: experience from two Helsinki area exposure incidents." In: *medRxiv* (Jul. 2020), p. 2020.07.20.20156018. DOI: 10.1101/2020.07.20.20156018. <URL: https://doi.org/10.1101/2020.07.20.20156018>.
9. A. Fontanet, L. Tondeur, Y. Madec, et al. "Cluster of COVID-19 in Northern France: A Retrospective Closed Cohort Study." In: *SSRN Electronic Journal* (Apr. 2020), p.

2020.04.18.20071134. ISSN: 1556-5068. DOI: 10.1101/2020.04.18.20071134. <URL: https://doi.org/10.1101/2020.04.18.20071134>.

10. L. Heavey, G. Casey, C. Kelly, et al. “No evidence of secondary transmission of COVID-19 from children attending school in Ireland, 2020.” In: *Eurosurveillance* 25.21 (May. 2020), p. 1. ISSN: 15607917. DOI: 2807/1560-7917.ES.2020.25.21.2000903. <URL: /pmc/articles/PMC7268273/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7268273/>.
11. National Centre for Immunisation Research and Surveillance. *COVID-19 in schools the experience in NSW*. 2020. <URL:https://www.ncirs.org.au/sites/default/files/2020-04/NCIRS%20NSW%20Schools%20COVID_Summary_FINAL%20public_26%20April%202020.pdf}.>
12. E. Larosa, O. Djuric, M. Cassinadri, et al. “Secondary transmission of COVID-19 in preschool and school settings after their reopening in northern Italy: a population-based study.” In: *medRxiv* (2020).
13. A. S. Lopez, M. Hill, J. Antezano, et al. “Transmission Dynamics of COVID-19 Outbreaks Associated with Child Care Facilities” Salt Lake City, Utah, April“July 2020.” In: *MMWR. Morbidity and Mortality Weekly Report* 69.37 (Sep. 2020), pp. 1319-1323. ISSN: 0149-2195. DOI: 10.15585/mmwr.mm6937e3. <URL: http://www.cdc.gov/mmwr/volumes/69/wr/mm6937e3.htm?s{_}cid=mm6937e3{_}w>.
14. K. Macartney, H. E. Quinn, A. J. Pillsbury, et al. “Transmission of SARS-CoV-2 in Australian educational settings: a prospective cohort study.” In: *The Lancet Child and Adolescent Health* 4.11 (Nov. 2020), pp. 807-816. ISSN: 23524642. DOI: 10.1016/S2352-4642(20)30251-0.
15. C. Stein-Zamir, N. Abramson, H. Shoob, et al. “A large COVID-19 outbreak in a high school 10 days after schools’ reopening, Israel, May 2020.” In: *Eurosurveillance* 25.29 (Jul. 2020), p. 2001352. ISSN: 1560-7917. DOI: 10.2807/1560-7917.ES.2020.25.29.2001352. <URL: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.29.2001352>.
16. J. P. Torres, C. Piñera, V. De La Maza, et al. “SARS-CoV-2 antibody prevalence in blood in a large school community subject to a Covid-19 outbreak: a cross-sectional study.” In: *Clinical Infectious Diseases* (Jul. 2020). ISSN: 1058-4838. DOI: 10.1093/cid/ciaa955. <URL:https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciaa955/5869860>.
17. C. F. Yung, K. Kam, K. D. Nadua, et al. “Novel coronavirus 2019 transmission risk in educational settings.” In: *Clinical Infectious Diseases* 72.6 (2020), p. 1055-1058. <URL: https://academic.oup.com/cid/article/72/6/1055/5862649>

Supplemental Tables

Table S1. Performance of different models. In the marginal analysis, 5 models for 5 groups of risk factors are tested. In the multi-risk factor analysis, 11 models with all possible combinations of 4 groups of risk factors are tested (i.e., excluding models with only one group tested in the marginal analysis). All models adjusted for surveillance. Note that although class size was identified as a significant risk factor in the marginal analysis, further including this variable in the multi-risk models gave impossible values for parameter estimates; we therefore excluded class size in the multi-risk model analysis. The best-performing model with the lowest AIC is bolded.

Analysis	Model	AIC
Marginal analysis	Mask wearing, distancing, seasonal changes, and surveillance	1106.60
	Weekly cases, weekly deaths, surveillance	890.34
	Specific humidity and surveillance	1278.58
	Class size, national income, and surveillance	1028.87
	School type and surveillance	1177.23
Multi-risk factor analysis (all adjusted for surveillance)	Mask wearing, distancing, and specific humidity	1106.59
	School type, and specific humidity	1035.03
	Mask wearing, distancing, and weekly cases	950.51
	Mask wearing, distancing, and school type	795.92
	Specific humidity and weekly cases	1106.99
	School type and weekly cases	893.05
	Mask wearing, distancing, specific humidity, and school type	794.60
	Mask wearing, distancing, specific humidity, and weekly cases	951.39
	Specific humidity, school type, and weekly cases	794.85
	Mask wearing, distancing, weekly cases, school type, and specific humidity	598.01
	Mask wearing, distancing, weekly cases, and school type	**596.33**

598 **Table S2.** Sensitivity analysis for both cases and deaths, with and without a 1-month extension
599 to the study time period.

	With extension		Without extension	
Cases	*Variable*	*aOR (95% CI)*	*Variable*	*aOR (95% CI)*
	mask wearing	0.559 (0.365-0.858)	mask wearing	0.523 (0.351-0.779)
	distancing	0.126 (0.0869-0.183)	distancing	0.260 (0.184-0.368)
	weekly cases	2.12 (1.88-2.40)	weekly cases	9.83 (7.21-13.4)
	surveillance	0.499 (0.417-0.597)	surveillance	0.542 (0.454-0.646)
	primary school	0.264 (0.171-0.406)	primary school	0.115 (0.0678-0.196)
	high school	1.46 (1.03-2.09)	high school	1.31 (0.925-1.86)
	mixed school	0.193 (0.130-0.288)	mixed school	0.261 (0.176-0.388)
	AIC	642.04	AIC	596.33
Deaths	*Variable*	*aOR (95% CI)*	*Variable*	*aOR (95% CI)*
	mask wearing	0.74 (0.50-1.11)	mask wearing	0.96 (0.64-1.44)
	distancing	0.28 (0.2-0.404)	distancing	0.23 (0.16-0.319)
	weekly deaths	16.7 (10.2-27.3)	weekly deaths	21.5 (11.1-41.7)
	surveillance	1.02 (0.82-1.28)	surveillance	1.01 (0.79-1.30)
	primary school	0.27 (0.18-0.42)	primary school	0.29 (0.19-0.439)
	high school	1.15 (0.82-1.62)	high school	0.73 (0.5-1.06)
	mixed school	0.28 (0.19-0.409)	mixed school	0.26 (0.18-0.384)
	AIC	663.71	AIC	714.65

600
601
602