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

Haokun Yuan,[1*] Connor Reynolds,[1*] Sydney Ng,[2*] Wan Yang[1#]

[1]Department of Epidemiology, [2]Department of Biostatistics, Mailman School of Public Health, Columbia University

[*]Authors contributed equally

#Correspondence to:

Wan Yang

Department of Epidemiology

Mailman School of Public Health

Columbia University

722 W 168th Street, Room 514, New York, NY 10032

Phone: (212) 305-0421

Email: 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 key risk factors to better understand SARS-CoV-2 transmission in schools.

Methods

We collected records of 39 SARS-CoV-2 school outbreaks globally published through July 2021 and compiled information on hypothesized risk factors. We utilized the directed acyclic graph (DAG) to conceptualize risk mechanisms, used logistic regression to examine each risk-factor group, and further built multi-risk models.

Results

The best-fit model showed that the intensity of concurrent community transmission (adjusted odds ratio [aOR]: 1.2, 95% CI: 1.17 – 1.24, for each increase of 1 case per 10,000 persons per week), individualism (aOR: 1.72, 95% CI: 1.19 – 2.5, above vs. below the median) were associated higher risk, whereas preventive measures (aOR: 0.22, 95% CI: 0.17 – 0.29, distancing and masking vs. none) and higher population immunity (aOR: 0.28, 95% CI: 0.22 – 0.35) were associated with lower risk of SARS-CoV-2 transmission in schools. Compared to students in pre-schools, the aOR was 0.35 (95% CI: 0.23 – 0.54) for students in primary schools and 1.3 (95% CI: 0.9 – 1.88) for students in high schools.

Conclusions

Preventive measures in schools (e.g. social distancing and mask-wearing) and communal efforts to lower transmission and increase vaccination uptake (i.e. vaccine-induced population immunity) in the community 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, children, school, preventive measures, secondary attack rate

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 600 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 related factors (e.g. incidence in the community and population immunity cumulated over time) and further used regression models to examine key risk factors of having high SAR in schools. Consistent with previous work, we found the risk varied by school level, with lower risk among primary

school students than preschoolers and high schoolers. Accounting for school level, we found that implementation of preventive measures (distancing and mask wearing) in schools and higher population immunity were associated lower SAR in schools; in contrast, higher SARS-CoV-2 transmission in the community and higher level of individualism were associated with higher SAR in schools.

Methods

Data sources

Studies were searched for on the "Living Evidence for COVID-19" database [7], which retrieves articles from EMBASE via Ovid, PubMed, BioRxiv, and MedRxiv. Any article within this database was considered, from December 2019 up to July 28, 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 [8]. In total, 39 school outbreaks extracted from 26 articles were included in this analysis (see Fig 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 [9], 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 nine key components that may affect SARS-CoV-2 transmission in schools (Fig 2). Below we describe each of the nine components, rationale for inclusion, and related variables examined.

1) School types, based on studies indicating differential transmission risk among different age groups [10, 11]. 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 of the index cases 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) Preventive measures, which may reduce outbreak risk. Here we categorized this variable based on the implementation of mask-wearing and/or social distancing in schools, i.e., “No preventive measures” if neither measure was required, “Single preventive measure” if only one measure (i.e. distancing or masking) was required, and “Combined preventive measure” if both were required. Note that we were not able to test distancing and masking separately due to the small sample size of schools that required masking alone (n = 3).

134 4) Surveillance and/or testing policies implemented in schools. On the one hand, testing policies
135 could affect the reported values of SAR; for instance, testing of all contacts regardless of
136 symptoms may lead to identification of more infections including those asymptomatic and
137 increase the numerator of SAR. On the other, frequent testing of all if combined with school
138 closure may serve as a containment measure to reduce the risk of onward transmission and in
139 turn reduce SAR. Here we thus included the reported testing practices for contacts in the school
140 outbreak clusters as a categorical ordinal variable. Three types of testing were reported in the
141 school studies, including testing only the symptomatic, both symptomatic and some
142 asymptomatic, and all contacts. However, due to the small sample size in "only symptomatic"
143 (n = 3), we dichotomized surveillance to testing "only symptomatic or some asymptomatic" and
144 "all contacts" of an index case in each school cluster.
145
146 5) Seasonal changes such as humidity and temperature, which may affect the survival and
147 transmission of SARS-CoV-2. Here, we used specific humidity (a measure of absolute humidity)
148 to examine the potential impact from disease seasonality, as specific humidity and temperature
149 are highly correlated. Specifically, ground surface temperature and relative humidity for each
150 study location were extracted from the National Oceanic and Atmospheric Administration using
151 the "rnoaa" package [13]. Daily mean specific humidity in g H_20/kg air was then computed
152 based on the meteorological data using formula introduced by Bolton [14], and further
153 averaged over the corresponding study period.
154
155 6) Intensity of concurrent community transmission, which may increase the introduction of
156 infections into schools. To examine its impact, we included two measures, i.e., the weekly
157 COVID-19 case rate and weekly COVID-19-related death rate for the study area during the study
158 period; both measures were computed using data from the John Hopkins Coronavirus Resource
159 Center [15] and standardized by the corresponding population size (for non-US sites, country-
160 level data were used; and for US sites, county-level data were used). In addition, testing and
161 hence the case-ascertainment rate tended to increase over time, whereas infection-fatality risk
162 tended to decrease over time due to, e.g., more timely diagnosis and improved medical

163 treatments and management [16]. To account for such time-related impacts on the reported
164 case rates and death rates, we included an additional variable ("calendar period") whenever
165 either measure is included in the model. We dichotomized the "calendar period" variable,
166 based on whether a study was conducted before July 2020, i.e., roughly the end of the initial
167 pandemic wave when both testing capacities and medical treatments were improved.
168
169 7) Prior population immunity in the community. Population immunity gained from prior
170 infections or COVID-19 vaccination could lower population susceptibility and hence the risk of
171 SARS-CoV-2 in the community. As most school outbreaks included here occurred prior to the
172 rollout of mass-vaccination, population immunity at those times would mostly come from
173 natural infections (see Fig 1 for the timeline of each study, vs. earliest vaccination rollout for the
174 general population round spring 2021). Thus, here we used the cumulative COVID-19 case rate
175 (up to the mid-point of the corresponding study period) as a proxy to account for prior
176 population immunity.
177
178 8) Cultural climates, which "represent independent preferences for one state of affairs over
179 another that distinguish countries (rather than individuals) from each other" [17] and may
180 reflect the collective risk tendency of a population. The Hofstede's cultural dimensions theory
181 [17] included 6 related measures including individualism, masculinity, uncertainty avoidance,
182 long term orientation, and indulgence. In particular, individualism is defined as the degree of
183 interdependence of society maintains among its members. We reasoned that the individualism
184 measure would be most relevant to the level of compliancy to public health interventions and
185 in turn the risk of SARS-CoV-2 transmission. Thus, here we included individualism in our analysis
186 and dichotomized the reported values for each country [18]. Among all study sites included
187 here, the median of individualism scores was 76; thus, we coded those with a score >76 as
188 "Higher individualism" those with a score ≤76 as "Lower individualism".
189
190 9) Indicators of socioeconomic status such as national income that reflect a country's ability to
191 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 [19].

Statistical Analyses

Marginal analysis.

Due to the low number of observations (n = 39 school outbreaks), we conducted an initial analysis to test combinations of the DAG covariates described above. The goal was to examine the relationship between the 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 39 school outbreaks. *X* is one of the combinations of variables we examined as follows:

1) School type
2) Classroom size, adjusting for national income, seasonal changes, and cultural climate
3) Preventive measures, adjusting for seasonal changes and cultural climate
4) Seasonal changes
5) Community transmission (weekly death rate per 100,000 or weekly case rate per 10,000, i.e., only one measure is included, because these two measures are highly correlated), adjusting for cultural climate, population immunity and study period
6) Prior population immunity, adjusting for cultural climate
7) National Income

As noted above in the conceptual analysis, the type of surveillance policy implemented in schools could affect the reported SAR in both directions. Thus, we included surveillance type in all models. However, as a sensitivity analysis, we also tested each model without surveillance type included. Results for both versions are reported in Fig 3.

Multi-risk factor analysis

All seven variable groups described above were found to be associated with SAR in the marginal analysis (see Results). We thus tested models including different combinations of these variables to identify a multi-risk model that best explains the observed SAR. 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 tested all possible combinations of significant variables identified from the marginal analysis. 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:

$logit(SAR) \sim$ school type + preventative measures + surveillance + seasonal changes + weekly case rate + calendar period + population immunity + individualism

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 model results 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, for the best-performing multi-risk factor model, we additionally examined three other measures in representing the intensity of community transmission, in lieu of concurrent weekly case rate: 1) weekly death rate during the study period; 2) weekly death rate during the study period plus a 1-month extension; and similarly, 3) weekly case rate during the study period plus a 1-month

249 extension. The additional month added accounts for the time lag that occurs as death from
250 COVID-19 may take several weeks [20].
251
252 **Results**
253 *Summary statistics*
254 We identified 39 reported SARS-CoV-2 outbreaks in schools, totaling 1144 secondary cases in
255 children among 28826 contacts. These outbreaks occurred in 15 countries, spanning 4 WHO
256 regions including the Americas, Western Pacific, European, and Eastern Mediterranean Region.
257 Fig 1 shows the study site, school type, study period, and reported SARS-CoV-2 SAR for each
258 included outbreak. Table 1 shows the frequencies and summary statistics for SARS-CoV-2 SAR
259 and other variables included. While the reported SAR ranged from 0 to 100%, the majority of
260 schools reported very low SAR (median: 2%, interquartile range: 0 – 12%). Roughly even
261 proportion of different school types were included: 5 (13%) were pre-schools, 8 (21%) were
262 primary schools, 11 (28%) were high schools, and 15 (38%) were mixed schools. The majority of
263 schools tested all contacts of the index cases (25/39, or 64%); and the majority required at least
264 one preventive measure (28/39, or 72%).
265
266 *Marginal analysis*
267 The marginal analysis with or without adjusting for surveillance generated similar estimates
268 (Figure 3). Thus, below we present results adjusting for surveillance. This analysis identified
269 several associating factors that are directly related to schools, including school type, class size,
270 preventive measures, and seasonal changes. For school type, compared to pre-schools, being
271 in primary schools (aOR: 0.59, 95% CI: 0.43 – 0.82) or mixed schools (aOR: 0.51, 95% CI: 0.37 –
272 0.71) was associated with a lower risk of SARS-CoV-2 infection whereas being in high schools
273 (aOR: 1.9, 95% CI: 1.39 – 2.59) was associated with a higher risk of SARS-CoV-2 infection. For
274 the school physical setting measure, each 1-person increase in the national average class size
275 was associated with an increased risk of contracting SARS-CoV-2 in schools (aOR: 1.14, 95% CI:
276 1.11 – 1.17). Single (distancing or masking) and combined preventive measure (distancing and
277 masking) were both associated with a lower SAR in schools, with an aOR of 0.25 (95% CI: 0.19 –

278 0.34) and 0.22 (95% CI: 0.19 – 0.25), respectively. For disease seasonality, which could affect
279 the transmission in schools and the community in general, each 1 g/kg increase in specific
280 humidity was associated with a decreased risk of contracting SARS-CoV-2 in schools (aOR: 0.96,
281 95% CI: 0.93, 0.99).
282
283 In addition, the marginal analysis also identified several associating factors, indirectly related to
284 schools via the community/population. For the intensity of community transmission, both
285 higher COVID-19 case rate and death rate in the community were associated with an increased
286 risk of contracting SARS-CoV-2 in schools (aOR = 1.21, 95% CI: 1.18 – 1.24, for cases per 10,000
287 people per week; and aOR = 2.06, 95% CI: 1.92 – 2.21, for deaths per 100,000 people per week).
288 Individualism was included in four models based on the conceptual analysis (see Methods and
289 Fig 2); all four models showed that higher level of individualism was associated with an
290 increased risk (mean aOR ranged from 1.3 to 2.8 and all 95% CI had a lower bound >1).
291 Conversely, higher prior population immunity (using cumulative case rate per 100 people as a
292 proxy, aOR: 0.92, 95% CI: 0.87 – 0.97) and higher national income (aOR: 0.97, 95% CI: 0.96 –
293 0.98, per 1000 pound) were associated with a reduced risk.
294
295 *Multi-risk factor analysis*
296 Among all models tested (Table S1), the best-performing model with the lowest AIC included six
297 key groups of risk factors, namely, school type, preventive measures, seasonality, intensity of
298 community transmission, population immunity, and individualism, adjusting for surveillance.
299 Overall, these risk factors in combination were able to explain 60.6% of the variance in the
300 reported SARS-CoV-2 SAR (McFadden's pseudo R^2 = 0.606; Fig 4) [21, 22]. The estimated
301 adjusted ORs (aORs) for each risk factor are shown in Table 2 and Fig 5. The sensitivity analysis
302 shows consistent estimates across models using different measures of community transmission
303 (Table S2).
304
305 Consistent with the marginal analysis, the best-fit multi-risk factor model showed that higher
306 COVID-19 case rate in the community (aOR: 1.2, 95% CI: 1.17 – 1.24; for 1 additional case

307 reported among 10,000 people each week) and higher level of individualism (aOR: 1.72, 95% CI:
308 1.19 – 2.5; above vs. below the median) were associated with an increased risk of SARS-CoV-2
309 infection in schools. Conversely, both single (aOR: 0.22, 95% CI: 0.15 – 0.34) and combined
310 (aOR: 0.22, 95% CI: 0.17 – 0.29) preventive measure were associated a reduced risk of SARS-
311 CoV-2 infection in schools. In addition, both testing all contacts (aOR: 0.59, 95%: 0.50 – 0.70; vs.
312 only testing the symptomatic or some asymptomatic) and higher population immunity (aOR:
313 0.28, 95% CI: 0.22 – 0.35, using cumulative case rate per 100 people as a proxy) were also
314 associated a reduced risk of SARS-CoV-2 infection in schools. Compared to students in pre-
315 schools, the aOR was 0.24 (95% CI: 0.16 – 0.36) for students in primary schools, 1.3 (95% CI: 0.9
316 –1.88) for students in high schools, and 0.24 (95% CI: 0.16 – 0.36) for students in mixed schools.
317

318 **Discussion**

319 Leveraging available data on multiple reported SARS-CoV-2 outbreaks in schools and potential
320 risk factors, we have examined main factors associated with the risk of SARS-CoV-2
321 transmission in schools. Our analyses suggest that SARS-CoV-2 SAR in schools was associated
322 with both preventative measures in schools and population factors including the level of
323 community transmission, individualism, and population immunity, once adjusted for
324 surveillance and school type.
325

326 Foremost, we identified several population or community factors to be highly associated with
327 SARS-CoV-2 transmission in schools, all of which point to the importance of communal efforts
328 to collectively reduce the risk of transmission and protect children in schools. In particular, all
329 models (in both the marginal analysis and the multi-risk factor analysis) consistently showed
330 that higher level of individualism of the population was associated with higher SARs in schools.
331 This finding is consistent with a recent study linking collectivism (vs. individualism) to usage of
332 preventive measures like mask use during the COVID-19 pandemic [23]. Along similar lines, the
333 models associated higher transmission in the community with higher SARs in schools,
334 suggesting the potential community-to-school importation of cases and subsequent risk of
335 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, the models showed that higher population immunity, which could lower transmission overall, was associated with lower SARs in schools. 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 models estimated a substantial transmission reduction in schools when both distancing and mask-wearing were required (aOR: 0.22, 95% CI: 0.17 – 0.29) or when distancing alone was required [note the aOR for either measure alone was 0.22 (95% CI: 0.15 – 0.34), with the majority of schools in this category (8 of 11) requiring distancing alone]. Thus, both estimates indicate the importance of distancing. 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 (often of color) with underfunded and overcrowded schools [24, 25]. Furthermore, racially motivated structural factors prevent these disadvantaged communities from practicing social distancing policies outside of the school. For example, these communities tend to make up most of essential workers and thus have higher rates of transmission in their community [26], increasing potential introduction of infections into schools.

In comparison to pre-schools, students in both primary and mixed-level schools had a lower risk of SARS-CoV-2 infection whereas those in high schools had a higher risk. This finding is consistent with previous studies indicating the likely lower susceptibility to SARS-CoV-2

infection and transmissibility among young children [4, 5]. In addition, it is also likely in part due to the greater ability of older children to follow directions regarding preventive measures but with less compliance among high school students. Overall, these findings support flexible reopening policies for different levels of schools given the risk differences.

When schools test all contacts of the index case, this effectively functions as a control measure in that more cases will be detected and is more likely to result in a school closure. This may explain why the OR estimate for all contacts surveillance is protective, as the outbreak was identified and able to be contained before the asymptomatic individuals were able to infect others.

This study has several limitations. First, all school outbreaks included in this analysis occurred prior to the emergence and widespread circulation of the more transmissible SARS-CoV-2 variants of concern (e.g., the delta variant). We are thus unable to estimate variant-specific impacts. Nonetheless, even though the magnitude of impact may alter somewhat due to changes in circulating SARS-CoV-2 variants, the identified risk factors and their relative importance to school transmission likely would still hold given the robust risk mechanisms. Second, we were unable to estimate the impact of distancing and mask-wearing separately, due to the small sample size of schools that required masking alone (n = 3). Third, due to a lack of detailed information for each 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

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

Author contributions

WY conceived and supervised the study; HY, CR, and SN conducted the literature review and data extraction, and performed the statistical analysis; HY, CR, SN, and WY interpreted results and wrote the manuscript.

Conflict of interest

None.

References

1. Bayham, J. and E.P. Fenichel, *Impact of school closures for COVID-19 on the US health-care workforce and net mortality: a modelling study.* The Lancet Public Health, 2020. **5**(5): p. e271-e278.
2. Worby, C.J., et al., *On the relative role of different age groups in influenza epidemics.* Epidemics, 2015. **13**: p. 10-16.
3. UNESCO. *Education: From disruption to recovery*. Available from: https://en.unesco.org/covid19/educationresponse.
4. Viner, R.M., 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**(2): p. 143-156.
5. Otte Im Kampe, E., et al., *Surveillance of COVID-19 school outbreaks, Germany, March to August 2020.* Euro Surveill, 2020. **25**(38).
6. Last, J.M. and I.E. Association, *A Dictionary of Epidemiology*. 2001: Oxford University Press.

422 7. Covid-19 Open Access Project. *Living Evidence on COVID-19*. 2020; Available from:
423 https://ispmbern.github.io/covid-19/living-review/.
424 8. Xu, W., et al., *What is the evidence for transmission of COVID-19 by children in schools?*
425 *A living systematic review.* J Glob Health, 2020. **10**(2): p. 021104.
426 9. UNESCO Institute for Statistics, *International Standard Classification of Education ISCED*
427 *2011*. Montréal.
428 10. Davies, N.G., et al., *Age-dependent effects in the transmission and control of COVID-19*
429 *epidemics.* Nat Med, 2020. **26**(8): p. 1205-1211.
430 11. Galanti, M., et al., *Rates of asymptomatic respiratory virus infection across age groups.*
431 Epidemiol Infect, 2019. **147**: p. e176.
432 12. OECD, *Student-teacher ratio and average class size*. 2015.
433 13. Edmund;, H., S. Chamberlain;, and K. Ram;. *rnoaa: NOAA climate data from R.. R*
434 *package version 0.1.9.99.* 2014; Available from: https://github.com/ropensci/rnoaa.
435 14. Bolton, D., *The Computation of Equivalent Potential Temperature.* Monthly Weather
436 Review, 1980. **108**(7): p. 1046-1053.
437 15. Dong, E., H. Du, and L. Gardner, *An interactive web-based dashboard to track COVID-19*
438 *in real time.* The Lancet Infectious Diseases, 2020. **20**(5): p. 533-534.
439 16. James, N., M. Menzies, and P. Radchenko, *COVID-19 second wave mortality in Europe*
440 *and the United States.* Chaos: An Interdisciplinary Journal of Nonlinear Science, 2021.
441 **31**(3): p. 031105.
442 17. Hofstede, G., G.J. Hofstede, and M. Minkov, *Cultures and Organizations: Software of the*
443 *Mind, Third Edition*. 2010: McGraw-Hill Education.
444 18. Hofstede, G. *Hofstede insights: Country comparison* 2020; Available from:
445 https://www.hofstede-insights.com/country-comparison/.
446 19. World Inequality Lab, *World Inequality Database*. 2021.
447 20. Chen, Y., et al., *Impact of Fundamental Diseases on Patients With COVID-19.* Disaster
448 Med Public Health Prep, 2020. **14**(6): p. 776-781.
449 21. MITTLBÖCK, M. and M. SCHEMPER, *EXPLAINED VARIATION FOR LOGISTIC REGRESSION.*
450 Statistics in Medicine, 1996. **15**(19): p. 1987-1997.

451 22. McFadden, D., B.I.o.U. University of California, and R. Development, *Conditional Logit*
452 *Analysis of Qualitative Choice Behavior*. 1973: Institute of Urban and Regional
453 Development, University of California.
454 23. Lu, J.G., P. Jin, and A.S. English, *Collectivism predicts mask use during COVID-19.* Proc
455 Natl Acad Sci U S A, 2021. **118**(23).
456 24. Baker, B.D., et al., *School funding disparities and the plight of Latinx children.* education
457 policy analysis archives, 2020. **28**.
458 25. Anyane-Yeboa, A., T. Sato, and A. Sakuraba, *Racial disparities in COVID-19 deaths reveal*
459 *harsh truths about structural inequality in America.* J Intern Med, 2020. **288**(4): p. 479-
460 480.
461 26. Dorn, A.v., R.E. Cooney, and M.L. Sabin, *COVID-19 exacerbating inequalities in the US.*
462 The Lancet, 2020. **395**(10232): p. 1243-1244.
463 27. Danis, K., et al., *Cluster of Coronavirus Disease 2019 (COVID-19) in the French Alps,*
464 *February 2020.* Clinical Infectious Diseases, 2020. **71**(15): p. 825-832.
465 28. Heavey, L., et al., *No evidence of secondary transmission of COVID-19 from children*
466 *attending school in Ireland, 2020.* Eurosurveillance, 2020. **25**(21): p. 2-5.
467 29. Yung, C.F., et al., *Novel Coronavirus 2019 Transmission Risk in Educational Settings.*
468 Clinical Infectious Diseases, 2021. **72**(6): p. 1055-1058.
469 30. Macartney, K., et al., *Transmission of SARS-CoV-2 in Australian educational settings: a*
470 *prospective cohort study.* Lancet Child Adolesc Health, 2020. **4**(11): p. 807-816.
471 31. Torres, J.P., et al., *SARS-CoV-2 antibody prevalence in blood in a large school community*
472 *subject to a Covid-19 outbreak: a cross-sectional study.* Clinical Infectious Diseases,
473 2020.
474 32. Dawson, P., et al., *Pilot Investigation of SARS-CoV-2 Secondary Transmission in*
475 *Kindergarten Through Grade 12 Schools Implementing Mitigation Strategies - St. Louis*
476 *County and City of Springfield, Missouri, December 2020.* Mmwr-Morbidity and
477 Mortality Weekly Report, 2021. **70**(12): p. 449-455.
478 33. Dub, T., et al., *Transmission of SARS-CoV-2 following exposure in school settings:*
479 *experience from two Helsinki area exposure incidents.* medRxiv, 2020.

480 34. Fontanet, A., et al., *SARS-CoV-2 infection in primary schools in northern France: A*
481 *retrospective cohort study in an area of high transmission.* MedRxiv, 2020.
482 35. Gettings, J.R., et al., *SARS-CoV-2 transmission in a Georgia school district - United States,*
483 *December 2020-January 2021.* Clin Infect Dis, 2021.
484 36. Hershow, R.B., et al., *Low SARS-CoV-2 Transmission in Elementary Schools - Salt Lake*
485 *County, Utah, December 3, 2020-January 31, 2021.* Mmwr-Morbidity and Mortality
486 Weekly Report, 2021. **70**(12): p. 442-448.
487 37. Jordan, I., et al., *Transmission of SARS-CoV-2 infection among children in summer schools*
488 *applying stringent control measures in Barcelona, Spain.* Clinical Infectious Diseases,
489 2021.
490 38. Larosa, E., et al., *Secondary transmission of COVID-19 in preschool and school settings in*
491 *northern Italy after their reopening in September 2020: a population-based study.*
492 Eurosurveillance, 2020. **25**(49).
493 39. Mossong, J., et al., *SARS-CoV-2 transmission in educational settings during an early*
494 *summer epidemic wave in Luxembourg, 2020.* BMC Infect Dis, 2021. **21**(1): p. 417.
495 40. Ramirez, D.W.E., M.D. Klinkhammer, and L.C. Rowland, *COVID-19 Transmission during*
496 *Transportation of 1st to 12th Grade Students: Experience of an Independent School in*
497 *Virginia.* J Sch Health, 2021. **91**(9): p. 678-682.
498 41. Stein-Zamir, C., et al., *A large COVID-19 outbreak in a high school 10 days after schools'*
499 *reopening, Israel, May 2020.* Eurosurveillance, 2020. **25**(29): p. 2-6.
500 42. Volpp, K.G., et al., *Minimal SARS-CoV-2 Transmission After Implementation of a*
501 *Comprehensive Mitigation Strategy at a School - New Jersey, August 20-November 27,*
502 *2020.* Mmwr-Morbidity and Mortality Weekly Report, 2021. **70**(11): p. 377-381.
503 43. White, P., et al., *Children are safe in schools: a review of the Irish experience of*
504 *reopening schools during the COVID-19 pandemic.* Public Health, 2021. **195**: p. 158-160.
505 44. Cordery, R., et al., *Transmission of SARS-CoV-2 by children attending school. Interim*
506 *report on an observational, longitudinal sampling study of infected children, contacts,*
507 *and the environment.* medRxiv, 2021.

508 45. Buonsenso, D. and B. Graglia, *High rates of SARS-CoV-2 transmission in a high-school*
509 *class.* J Paediatr Child Health, 2021. **57**(2): p. 299-300.
510 46. Brandal, L.T., et al., *Minimal transmission of SARS-CoV-2 from paediatric COVID-19 cases*
511 *in primary schools, Norway, August to November 2020.* Euro Surveill, 2021. **26**(1).
512 47. Atherstone, C., et al., *SARS-CoV-2 Transmission Associated with High School Wrestling*
513 *Tournaments - Florida, December 2020-January 2021.* MMWR Morb Mortal Wkly Rep,
514 2021. **70**(4): p. 141-143.
515 48. Armann, J.P., et al., *Anti-SARS-CoV-2 IgG antibodies in adolescent students and their*
516 *teachers in Saxony, Germany (SchoolCoviDD19): very low seropraevalence and*
517 *transmission rates.* 2020.
518 49. Fontanet, A., et al., *Cluster of COVID-19 in northern France: a retrospective closed cohort*
519 *study.* 2020.
520 50. National Centre for Immunisation Research Surveillance, *COVID‐19 in schools and early*
521 *childhood education and care services–the Term 2 experience in NSW.* ncirs. org.
522 au/sites/default/files/2020-08/COVID-19% 20Transmission% 20in% 20educational%
523 20settings% 20in% 20NSW% 20Term% 202% 20report_0. pdf, 2020.
524 51. Desmet, S., et al., *No SARS‐CoV‐2 carriage observed in children attending daycare*
525 *centers during the intial weeks of the epidemic in Belgium.* Journal of Medical Virology,
526 2021. **93**(3): p. 1828-1831.
527

528 **Table and Figure Captions**

529 **Table 1.** Characteristics of school outbreaks and related risk factors.

530

531 **Table 2.** Results of the best-fit multi-risk factor model for the identification of factors associated
532 with SARS-CoV-2 SAR in schools. Adjusted odds ratio (aOR) estimates and 95% confidence
533 intervals are given from the logistic regression model including surveillance to control for
534 differences in testing school clusters, school type due to inconsistent reporting of age groups in
535 the literature, number of preventative measures implemented in schools, and characteristics of
536 the study sites (i.e., level of individualism, daily mean specific humidity, population immunity
537 and weekly case rates per 10,000).

538

539 **Figure 1.** SARS-CoV-2 school outbreak studies included in the analysis [27-51]. Each colored bar
540 represents an observed school outbreak; the school location is shown on the y-axis and study
541 period is shown by the position and length of the bar (see calendar time on the x-axis); school
542 type is shown in the panel title on the right; and reported secondary attack rate (SAR) is
543 indicated by the color of the bar (see the legend).

544

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

549

550 **Figure 3.** Odds ratio estimates from the marginal models. Left panel shows results from models
551 without adjusting for surveillance and right panel shows results from corresponding models
552 additionally adjusting for surveillance. Black points show the mean odds ratio estimates and
553 horizontal black bars show the 95% confidence intervals. The vertical black bar in each plot
554 indicates the null value of 1.0. Each set of models is delineated by alternating the shaded
555 regions.

556

Figure 4. 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 [22].

Figure 5. 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, 8 models for 7 groups of risk factors were tested (note weekly case rate and weekly death rate, both representing community transmission, were tested separately in two models). In the multi-risk factor analysis, 19 additional models with all other possible combinations of 6 major groups of risk factors tested significant (i.e., when a risk group was a subset of a larger one, only the larger risk group was tested). All models adjusted for surveillance. The best-performing model with the lowest AIC is bolded.

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

579 **Tables**

580 **Table 1.** Characteristics of school outbreaks and related risk factors.

Study Characteristics	**n (%), N = 39 in total**
Calendar Period	
Before July 2020	21 (54%)
After July 2020	18 (46%)
Surveillance	
Symptomatic or some asymptomatic	14 (36%)
All contacts	25 (64%)
School type	
Pre-school/ECEC	5 (13%)
Primary school	8 (21%)
High school	11 (28%)
Mixed school	15 (38%)
Preventive Measure	
No preventative measure	11 (28%)
Single preventative measure (Distancing or Mask-wearing)	11 (28%)
Combined preventative measure (Distancing and Mask-wearing)	17 (44%)
Higher individualism (>76)	18 (46%)
	Median (IQR)
Secondary attack rate	0.02 (0.00, 0.12)
Population immunity rate (per 100 people)	0.36 (0.03, 1.10)
Weekly case rate (per 10,000 people)	0.30 (0.03, 0.98)
Weekly death rate (per 100,000 people)	0.25 (0.03, 1.79)
Daily mean specific humidity (g/kg)	6.73 (4.14, 8.41)
National Income (thousand £)	41 (37, 53)
Average class size	20.1 (19.8, 24.0)

581

582

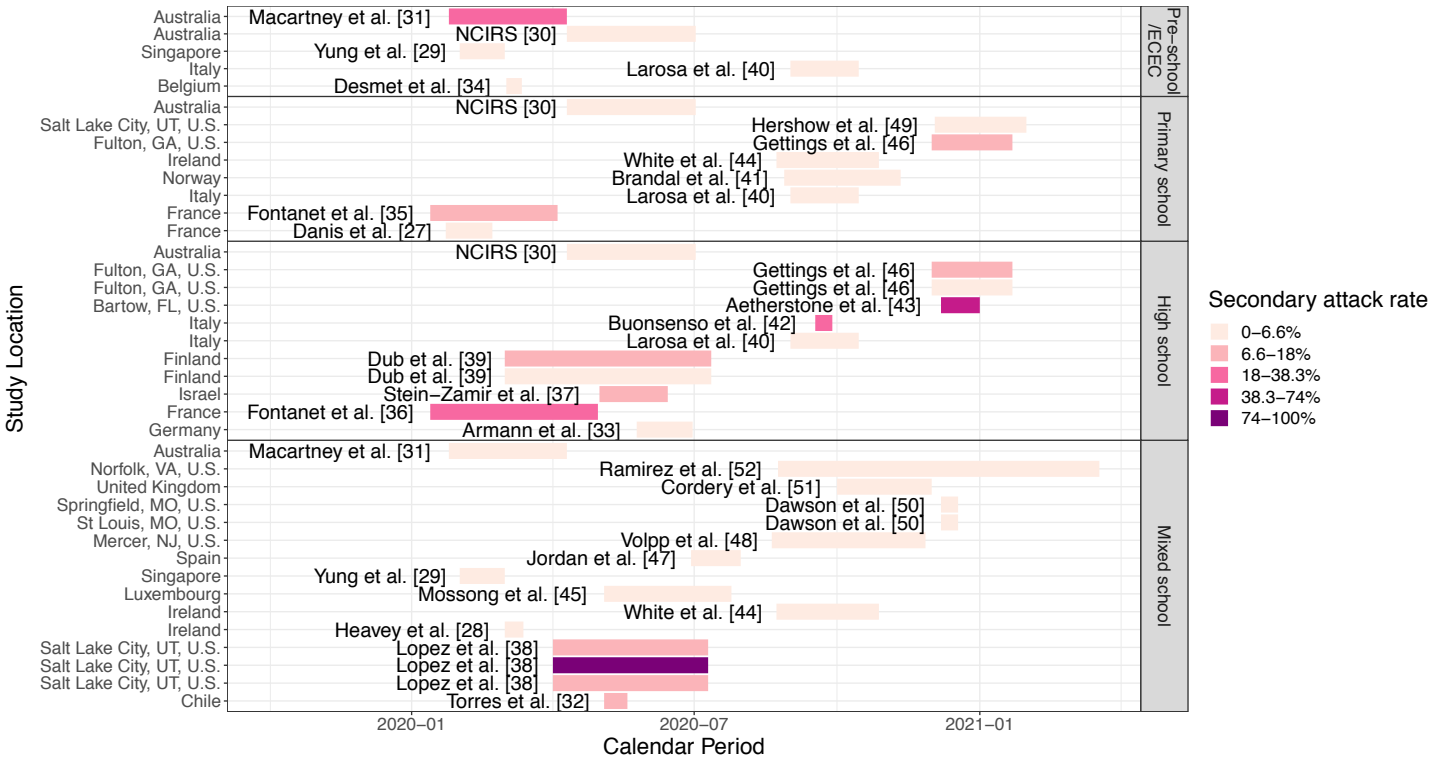
583 **Table 2.** Results of the best-fit multi-risk factor model for the identification of factors associated
584 with SARS-CoV-2 SAR in schools. Adjusted odds ratio (aOR) estimates and 95% confidence
585 intervals are given from the logistic regression model including surveillance to control for
586 differences in testing school clusters, school type due to inconsistent reporting of age groups in
587 the literature, number of preventative measures implemented in schools, and characteristics of
588 the study sites (i.e., level of individualism, daily mean specific humidity, population immunity
589 and weekly case rates per 10,000).

Variable	**aOR (95% CI)**
School type	
Mixed school	0.24 (0.16, 0.36)
High school	1.3 (0.9, 1.88)
Primary school	0.35 (0.23, 0.54)
Pre-school/ECEC	reference
Preventive measures	
Combined preventative measure (Distancing and Mask-wearing)	0.22 (0.17, 0.29)
Single preventative measure (Distancing or Mask-wearing)	0.22 (0.15, 0.34)
No preventative measure	reference
Seasonal changes	
Daily mean specific humidity	0.96 (0.93, 1)
Community transmission	
Weekly case rate	1.2 (1.17, 1.24)
Population immunity	
Population immunity	0.28 (0.22, 0.35)
Individualism	
Higher individualism	1.72 (1.19, 2.5)
Lower individualism	reference
Surveillance	
All contacts	0.59 (0.5, 0.7)
Symptomatic or some asymptomatic	reference
Calendar Period	
After July 2020	1.93 (1.45, 2.59)
Before July 2020	reference

590
591

592 **Figures**

593 **Figure 1.** SARS-CoV-2 school outbreak studies included in the analysis [27-51]. Each colored bar
594 represents an observed school outbreak; the school location is shown on the y-axis and study
595 period is shown by the position and length of the bar (see calendar time on the x-axis); school
596 type is shown in the panel title on the right; and reported secondary attack rate (SAR) is
597 indicated by the color of the bar (see the legend).



598

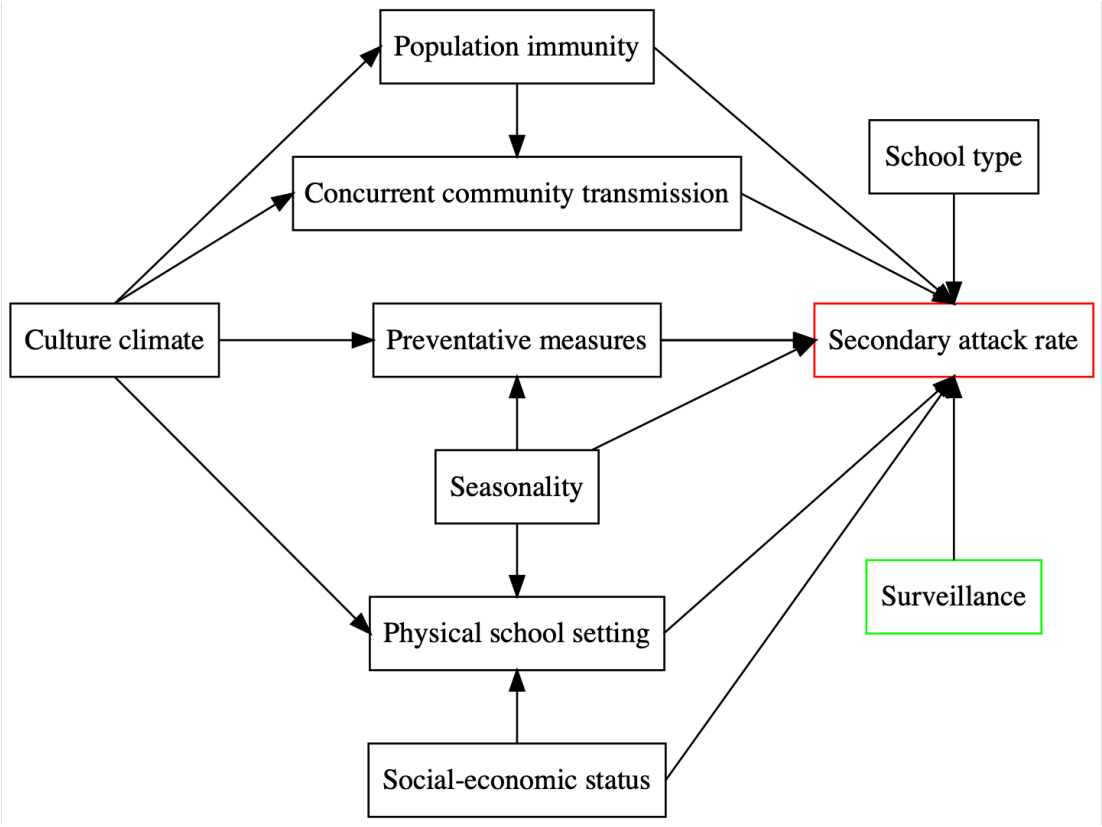
599

600

601

602

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



607
608

Figure 3. 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 indicates the null value of 1.0. Each set of models is delineated by alternating the shaded regions.

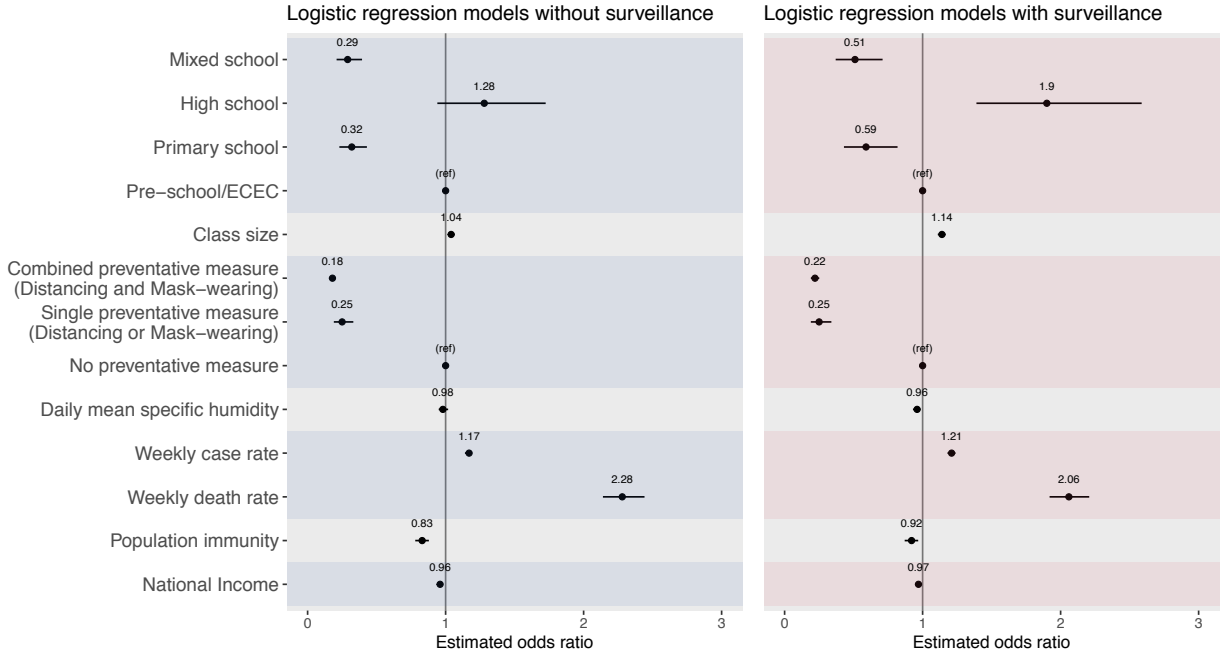


Figure 4. Model fit of the best-performing multi-risk model. Black dots show the fitted SAR for each study (y-axis), compared to the observed SAR (x-axis); the red bars around each point show the 95% confidence intervals of model-estimates. The McFadden's pseudo-R^2 is computed using Eqn. 30 in [22].

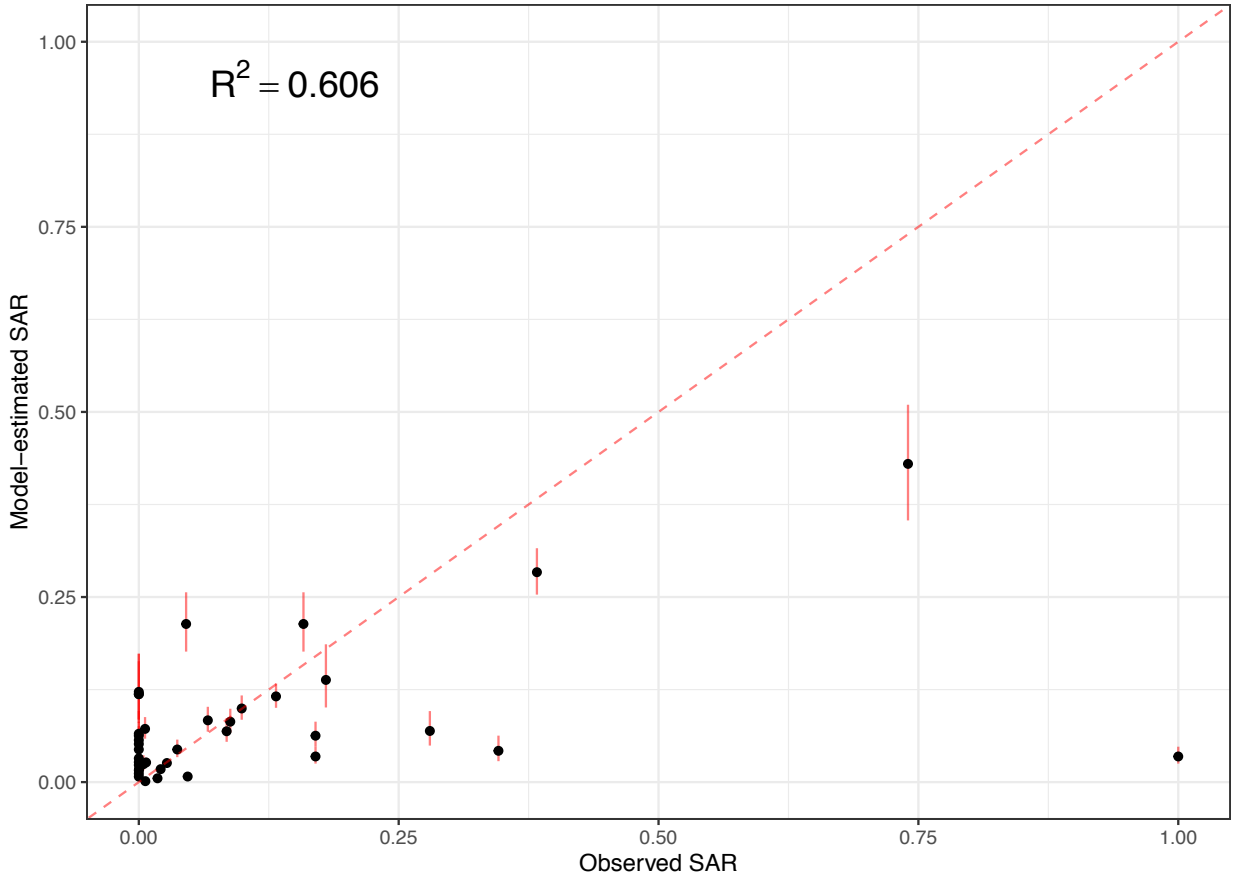
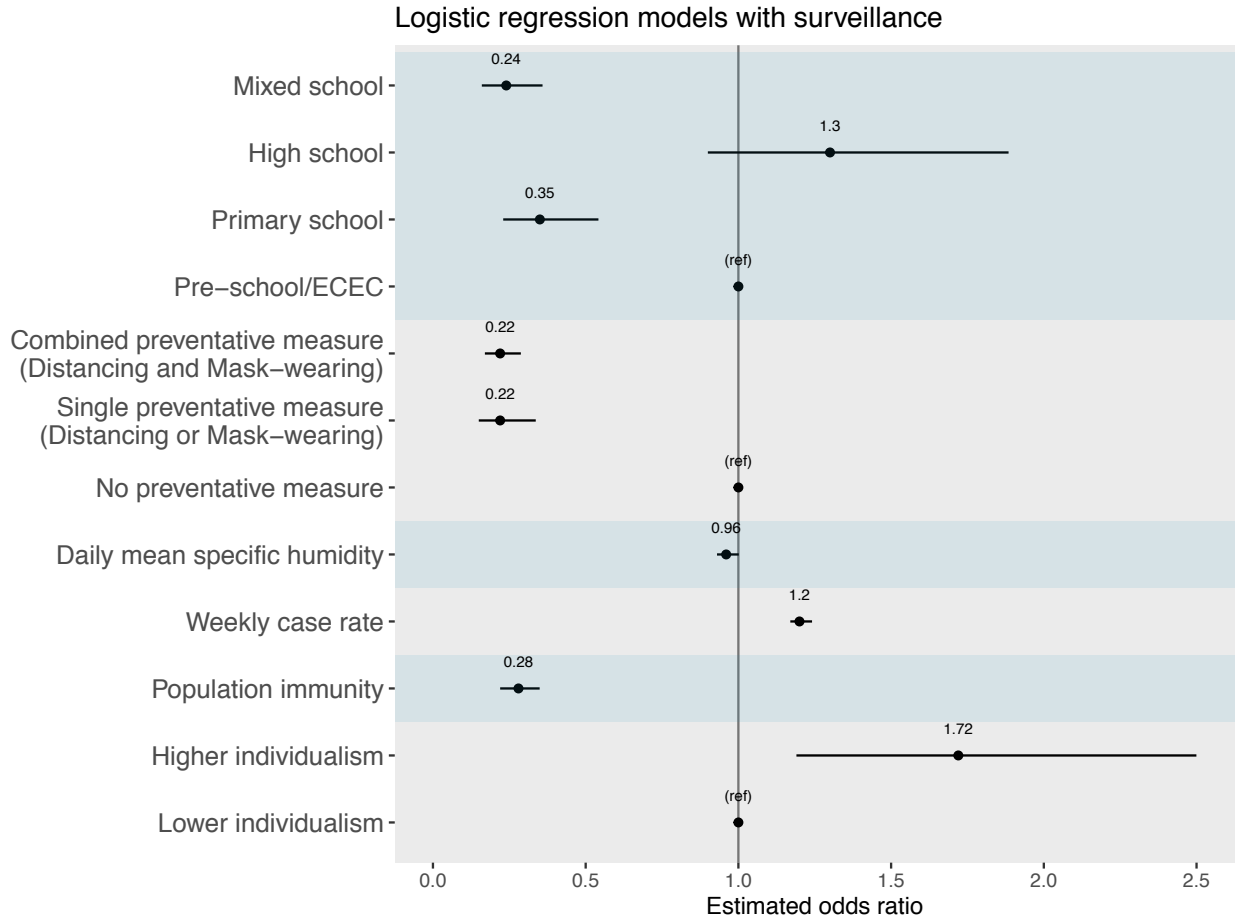


Figure 5. Odds ratio estimates from the best-performing multi-risk model. Black points show the mean odds ratio estimates and horizontal black bars show the 95% confidence intervals. The vertical black bar indicates the null value of 1.0. Each variable type is delineated by the shaded regions.



628 **Supplemental Tables**
629 **for**
630 **Factors affecting the transmission of SARS-CoV-2 in school settings**

631 **Table S1.** Performance of different models. In the marginal analysis, 8 models for 7 groups of
632 risk factors were tested (note weekly case rate and weekly death rate, both representing
633 community transmission, were tested separately in two models). In the multi-risk factor
634 analysis, 19 additional models with all other possible combinations of 6 major groups of risk
635 factors tested significant (i.e., when a risk group was a subset of a larger one, only the larger
636 risk group was tested). All models adjusted for surveillance. The best-performing model with
637 the lowest AIC is bolded.

Analysis	Model (all adjusted for surveillance)	AIC
Marginal analysis	Preventative measures, Daily mean specific humidity, Individualism	1,315
	Weekly case rate, Study period, Individualism, Population immunity	1,311
	Weekly death rate, Study period, Individualism, Population immunity	1,230
	Daily mean specific humidity	1,690
	Class size, National income, Daily mean specific humidity, Individualism	1,448
	School type	1,382
	National income	1,645
	Population immunity, Individualism	1,676
Multi-risk factor analysis	National income, Population immunity, Individualism	1,603
	School type, Population immunity, Individualism	1,378
	School type, National income	1,364
	School type, National income, Population immunity, Individualism	1,349
	Preventative measures, Daily mean specific humidity, Individualism, Population immunity	1,317
	Weekly case rate, Study period, Individualism, Population immunity	1,311
	Weekly case rate, Study period, Individualism, Population immunity, National income	1,298
	Preventative measures, Daily mean specific humidity, Individualism, National income, Population immunity	1,268
	Preventative measures, Daily mean specific humidity, Individualism, National income	1,266
	Preventative measures, Daily mean specific humidity, Individualism, Weekly case rate, Study period, Population immunity	1,175
	Preventative measures, Daily mean specific humidity, Individualism, Weekly case rate, Study period, Population immunity, National income	1,174

Analysis	Model (all adjusted for surveillance)	AIC
	Preventative measures, Daily mean specific humidity, Individualism, School type	1,065
	Preventative measures, Daily mean specific humidity, Individualism, School type, Population immunity	1,062
	Preventative measures, Daily mean specific humidity, Individualism, School type, National income	1,017
	Preventative measures, Daily mean specific humidity, Individualism, School type, National income, Population immunity	1,016
	Weekly case rate, Study period, Individualism, Population immunity, School type	961
	Weekly case rate, Study period, Individualism, Population immunity, School type, National income	961
	Preventative measures, Daily mean specific humidity, Individualism, Weekly case rate, Study period, Population immunity, School type, National income	808
	Preventative measures, Daily mean specific humidity, Individualism, Weekly case rate, Study period, Population immunity, School type	**807**

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

Variable		Without extension		With extension	
		aOR (95%CI)	AIC	aOR (95%CI)	AIC
case	Weekly case rate	1.2 (1.17, 1.24)	807.34	1.07 (1.05, 1.08)	924.49
	Single preventative measure (Distancing or Mask-wearing)	0.22 (0.15, 0.34)		0.16 (0.11, 0.22)	
	Combined preventative measure (Distancing and Mask-wearing)	0.22 (0.17, 0.29)		0.21 (0.16, 0.27)	
	Primary school	0.35 (0.23, 0.54)		0.32 (0.21, 0.47)	
	High school	1.3 (0.9, 1.88)		1.01 (0.72, 1.42)	
	Mixed school	0.24 (0.16, 0.36)		0.19 (0.13, 0.27)	
	All contacts	0.59 (0.5, 0.7)		0.78 (0.67, 0.91)	
	After July 2020	1.93 (1.45, 2.59)		2.13 (1.59, 2.84)	
	Population immunity	0.28 (0.22, 0.35)		0.61 (0.54, 0.69)	
	High individualism (>76)	1.72 (1.19, 2.5)		2.12 (1.54, 2.91)	
	Daily mean specific humidity	0.96 (0.93, 1)		0.87 (0.84, 0.91)	
death	Weekly death rate	1.73 (1.59, 1.89)	843.64	1.27 (1.18, 1.37)	940.31
	Single preventative measure (Distancing or Mask-wearing)	0.2 (0.14, 0.28)		0.19 (0.13, 0.27)	
	Combined preventative measure (Distancing and Mask-wearing)	0.18 (0.14, 0.23)		0.2 (0.15, 0.26)	
	Primary school	0.39 (0.25, 0.58)		0.54 (0.35, 0.82)	
	High school	0.89 (0.62, 1.27)		1.62 (1.13, 2.34)	
	Mixed school	0.28 (0.19, 0.41)		0.36 (0.24, 0.53)	
	All contacts	1.14 (0.95, 1.37)		0.86 (0.74, 1.02)	
	After July 2020	2.48 (1.9, 3.24)		2.55 (1.94, 3.35)	
	Population immunity	0.7 (0.64, 0.77)		0.78 (0.72, 0.85)	
	High individualism (>76)	1.95 (1.42, 2.67)		2.19 (1.6, 3)	
	Daily mean specific humidity	0.99 (0.95, 1.03)		0.95 (0.91, 0.99)	