Effects of school-level and area-level socio-economic factors on elementary school student COVID-19 infections: a population-based observational study

Prachi Srivastava, Nathan T T Lau, Daniel Ansari, Nisha Thampi

ABSTRACT

Objectives To estimate the variability of the cumulative incidence of SARS-CoV-2 infections among elementary school students attributable to individual schools and/or their geographic areas, and to ascertain whether socio-economic characteristics of school populations and/or geographic areas may be predictive of this variability.


Setting 3994 publicly funded elementary schools in 491 forward sortation areas (designated geographic unit based on first three characters of Canadian postal code), Ontario, Canada, September 2020 to April 2021.

Participants All students attending publicly funded elementary schools with a positive molecular test for SARS-CoV-2 reported by the Ontario Ministry of Education.

Main outcome measures Cumulative incidence of laboratory-confirmed elementary school student SARS-CoV-2 infections in Ontario, 2020–21 school year.

Results A multilevel modelling approach was used to estimate the effects of socio-economic factors at the school and area levels on the cumulative incidence of elementary school student SARS-CoV-2 infections. At the school level (level 1), the proportion of the student body from low-income households was positively associated with cumulative incidence ($\beta=0.083$, $p<0.001$). At the area level (level 2), all dimensions of marginalisation were significantly related to cumulative incidence. Ethnic concentration ($\beta=0.454$, $p<0.001$), residential instability ($\beta=0.356$, $p<0.001$) and material deprivation ($\beta=0.212$, $p<0.001$) were positively related, while dependency ($\beta=-0.204$, $p<0.001$) was negatively related. Area-related marginalisation variables explained 57.6% of area variability in cumulative incidence. School-related variables explained 1.2% of school variability in cumulative incidence.

Conclusions The socio-economic characteristics of the geographic area of schools were more important in accounting for the cumulative incidence of SARS-CoV-2 elementary school student infections than individual school characteristics. Schools in marginalised areas should be prioritised for infection prevention measures and education continuity and recovery plans.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The study uses population data of all reported cases of COVID-19 school infections for all publicly funded elementary schools in Ontario, allowing for high confidence in the accuracy of the results.
- The multilevel modelling approach allows the study to statistically determine which school-level and area-level factors most likely predicted elementary student infections, by explicitly considering the nested nature of schools within geographic areas.
- As data on catchment area were not available, there remains a small degree of ambiguity in the results if student and school postal codes do not match.
- School-based race-based data were not publicly available.

INTRODUCTION

The SARS-CoV-2 pandemic has had an unprecedented effect on education, with negative outcomes disproportionately affecting marginalised groups.1–3 However, analyses of the potential differential effects of socio-economic factors on school infections are lacking. It is unclear whether school-related infections are more likely to occur in schools and/or areas with aggravated indicators of disadvantage. To extend equity-focused analyses, we investigate which school populations were more vulnerable to SARS-CoV-2 infections, situating this study in Ontario, Canada. To the best of our knowledge, this is the first study to examine school-level and area-level socio-economic factors on school SARS-CoV-2 infections.

Ontario has the largest cohort of elementary and secondary students in Canada. Furthermore, the province had the longest period of school closures in the country, at an average of 26 weeks between March 2020 and June 2021,4 and among the longest compared with regional averages for North America and Europe.4 School closures elsewhere have...
been disproportionately experienced in marginalised areas.\(^6\) Area-level school closure data are unavailable for Ontario. However, the overall prevalence of SARS-CoV-2 infections has been disproportionately concentrated in neighbourhoods with lower-income and racialised groups in Canada,\(^9\) and in its largest city, Toronto, Ontario.\(^7\) As schools are nested within specific geographic areas, this raises critical health and social equity concerns affecting children and youth.

Given the well-established literature on the nested nature of schools in specified geographic areas and resulting equity effects,\(^8\)–\(^10\) we sought to understand the relationship between school-level and area-level socio-economic factors on the cumulative incidence of laboratory-confirmed SARS-CoV-2 student infections within and across publicly funded elementary schools in Ontario for the 2020–21 school year. Publicly funded schools claim 94% of all enrolment in the province.\(^11\)\(^,\)\(^12\)

The COVID-19 School Dashboard was developed as a data visualisation platform to enable analysis between systems-level education data and school-related infections.\(^13\) The underlying data integrate all officially reported cases in publicly funded schools in Ontario, with school demographic and geolocation data. Knowing the geographic distribution of all school cases, we investigated the effects of school-level and area-level socio-economic factors on the cumulative incidence of laboratory-confirmed SARS-CoV-2 elementary school student infections.

We aimed to ascertain whether the proportion of students from marginalised socio-economic backgrounds in an individual school had an effect on the cumulative incidence of elementary school student infections in that school. We further sought to determine whether the proportion of households from marginalised backgrounds in a defined area had an effect on the cumulative incidence of elementary school student infections across schools in that area.

**METHODS**

We performed multilevel modelling analyses to determine whether, and to what extent, socio-economic factors within individual schools and/or factors common to schools within defined geographic areas predicted cumulative incidence. The estimated multilevel models had schools (school level, level 1) nested in forward sortation areas (FSAs) (area level, level 2), that is, the designated geographic unit based on the first three characters of the Canadian postal code. In principle, catchment areas structure how students are sorted into public schools in Ontario. Ideally, we would have had access to school catchment boundaries, however, they were not publicly available for all schools and school boards. The FSA was the most comparable unit available.

**Setting**

We included all 491 of the 513 FSAs in Ontario containing publicly funded elementary schools (all schools with the grade range from junior kindergarten up to grade 8 inclusive, excluding secondary schools). The analysis included 3994 publicly funded elementary schools. We could not include private schools because neither data on private school-related infections nor on private school demographics were publicly released.

We focused on elementary schools to minimise contextual variation in pedagogical strategies and associated face-to-face interactions. During periods that Ontario schools were open for in-person instruction in 2020–21, elementary schools typically instituted full-day face-to-face classes. This was in contrast to an adapted instructional model in secondary schools that employed a combination of reduced and alternative days of in-person and virtual instruction. Children were ineligible for the paediatric vaccine during this time.

Available datasets did not specify student age. Generally, students in Ontario enter junior kindergarten during the year of their fourth birthday. However, junior kindergarten and kindergarten are not compulsory. Students must enrol in grade 1 in the year of their sixth birthday. Given school enrolment cut-off dates, this may mean the age range spans from the youngest of 3.67 years in junior kindergarten (those born in December) to approximately 14 years for the oldest in grade 8. The exact distribution of ages is unknown.

**Patient and public involvement**

Patients were not involved in the design or data collection for this study. We used existing publicly available public health and relevant demographic and educational data-sets for analysis as described below.

**Data sources**

We used the latest relevant publicly available data at the time of analysis. We obtained data on all laboratory-confirmed cases of SARS-CoV-2 school-related infections in Ontario and school administrative and student demographic data via the Government of Ontario Open Data Directive.\(^14\) We used the *Schools with Recent COVID-19 Cases (27 April 2021 update)* dataset to extract laboratory-confirmed student cases of SARS-CoV-2 in schools.\(^15\) The latest school-level demographic data were available for the 2019–20 school year, which we extracted from the *School Information and Student Demographics (29 June 2021 update)* dataset.\(^16\) We used the Ontario Marginalisation Index (ON-Marg) data (based on the 2016 census) for indicators of socio-economic marginalisation at the area level.\(^17\) Finally, we obtained data on FSA population size from the 2016 census conducted by Statistics Canada.\(^18\)

**Variables**

**Dependent variable**

*Cumulative incidence of elementary school student SARS-CoV-2 infections*

The dependent variable is the cumulative incidence of laboratory-confirmed elementary school student SARS-CoV-2 infections in the 2020–21 school year. The number...
of confirmed elementary and secondary school infections was reported every weekday by the Ontario Ministry of Education. Case counts were disaggregated in the original dataset as student, staff and unidentified. This analysis includes elementary student cases only.

We inferred and computed the cumulative number of student cases. The active cases dataset included the total number of active cases per day but did not specify new cases. We computed new cases as additional cases of student infections on a given day as compared with the previous day’s total. We computed the cumulative number of new cases over the complete span of recorded time in the dataset (11 September 2020 to 15 April 2021). Schools in Ontario closed for spring break from 12 to 16 April 2021 and did not reopen for in-person instruction for the reminder of the school year ending in June 2021. The computed cumulative case count was divided by the number of enrolled students per school and multiplied by 100 to arrive at the cumulative incidence.\(^3\) Results reflect the number of cases per 1000 students in the school.

**School-level (level 1) predictors**

The choice of school-level predictors was guided by the literature on pandemic school closures and education inequities,\(^1\) \(^3\) and limited to what was available in the school demographic dataset. School-level predictors were the proportion of low-income households, language mismatch (mismatch of student first language with school medium of instruction) and low parental education. The percentage of the student body that was recent immigrant from non-English or non-French countries was available but highly correlated with language mismatch, thus excluded. The school demographic dataset did not have race-based data.

**Low-income households**

We extracted the proportion of students in a school belonging to low-income households from the school demographic dataset, that is, the percentage of students in the school from households with income below the after-tax low-income measure threshold as defined by Statistics Canada. Each school estimated this proportion by using postal code data of the student body and cross-referencing it with income data from the 2016 census.\(^23\) \(^24\)

**Language mismatch**

We calculated language mismatch as the percentage of the student body whose first language did not match the school medium of instruction (ie, percentage of students whose first language was not English in English-medium schools; percentage of students whose first language was not French in French-medium schools). Data on the first language of the student body and the medium of instruction (English or French) were extracted from the school demographic dataset.

**Low parental education**

The proportion of students in a school from households with low parental education was extracted from the school demographic dataset, defined therein as the percentage of students with parents who did not have a degree, diploma or certificate.

**Area-level (level 2) predictors**

Area-level predictors were the population of the FSA, the FSA average of the three school-level predictors above and the four marginalisation indicators from ON-Marg. Area means of the three school-level predictors were also considered as potential predictors. However, they were highly correlated with ON-Marg variables raising multicollinearity concerns, and thus, excluded from the final model (see online supplemental material S1 and table 1 for calculations).

**FSA population size**

We extracted the population size for the relevant FSAs from the 2016 census to increase compatibility with available ON-Marg data, which were derived from the 2016 census.

**ON-Marg Index indicators**

The ON-Marg is an area-based index that combines a number of demographic indicators into four dimensions of marginalisation: residential instability, material deprivation, dependency and ethnic concentration. We used all four indicators (see online supplemental material S2 and table 2 for description of constituent items).

**Analysis**

We estimated two multilevel models as part of the main analysis (see online supplemental material S3 for model specifications and online supplemental material S4 and table 3 for additional analyses). All predictors were standardised (M=0, SD=1) prior to analysis to remove non-essential multicollinearity.\(^23\) All analyses were conducted using Mplus V.8.3, with the maximum likelihood estimator with robust SEs. Missing data were addressed using full information maximum likelihood.\(^24\) All data and code for the reported analysis are available via open access.\(^23\) \(^25\)

Model 1 was an intercept-only model estimating the variability of the cumulative incidence of SARS-CoV-2 elementary school student infections at the school level (level 1) and the area level (level 2). This enables us to calculate the intraclass correlation (ICC), that is, the percentage of total variability in the cumulative incidence of student infections that can be accounted for by the geographic area (ie, FSA) of schools. Consequently, the higher the ICC, the more likely that two schools within the same geographic area will have similar cumulative incidences of student infections. This information is pertinent as it informs policymakers the degree to which elementary school student infections may be linked with the geographic location of schools.

Model 1 explores the degree to which variability of the cumulative incidence of SARS-CoV-2 elementary school
Table 1  Ontario public elementary school student population and area descriptive statistics for the 2020–2021 school year

<table>
<thead>
<tr>
<th>Student characteristics</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>% Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative incidence SARS-CoV-2 elementary school student infections (per 1000 students)</td>
<td>5.36</td>
<td>2.70</td>
<td>8.10</td>
<td>3.52</td>
<td>24.45</td>
<td>0.00</td>
</tr>
<tr>
<td>Low-income households (%)</td>
<td>18.89</td>
<td>15.00</td>
<td>11.16</td>
<td>1.15</td>
<td>1.39</td>
<td>3.18</td>
</tr>
<tr>
<td>Low parental education (%)</td>
<td>6.03</td>
<td>5.00</td>
<td>6.72</td>
<td>2.11</td>
<td>8.18</td>
<td>3.18</td>
</tr>
<tr>
<td>Language mismatch (%)</td>
<td>23.10</td>
<td>15.00</td>
<td>23.11</td>
<td>1.21</td>
<td>−0.16</td>
<td>3.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Area characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size of FSA</td>
<td>38446.5</td>
<td>33027.00</td>
<td>23188.14</td>
<td>1.12</td>
<td>1.16</td>
<td>0.07</td>
</tr>
<tr>
<td>Residential instability*</td>
<td>−0.22</td>
<td>−0.32</td>
<td>0.57</td>
<td>0.85</td>
<td>1.28</td>
<td>1.75</td>
</tr>
<tr>
<td>Material deprivation*</td>
<td>−0.18</td>
<td>−0.24</td>
<td>0.5</td>
<td>0.67</td>
<td>0.71</td>
<td>1.75</td>
</tr>
<tr>
<td>Dependency*</td>
<td>−0.06</td>
<td>−0.07</td>
<td>0.42</td>
<td>0.24</td>
<td>0.82</td>
<td>1.75</td>
</tr>
<tr>
<td>Ethnic concentration*</td>
<td>0.3</td>
<td>−0.04</td>
<td>1.01</td>
<td>1.14</td>
<td>0.25</td>
<td>1.75</td>
</tr>
</tbody>
</table>

*The four ON-Marg variables reflect factor scores constructed from principal component analysis. Factor scores are standardised (mean=0, SD=1) when the full Canada index is used. Observed deviations are due to aggregating ON-Marg score values to the FSA unit and using only factor scores from Ontario.

FSA, forward sortation area; ON-Marg, Ontario Marginalisation Index.

student infections may be attributable to individual schools or to geographic areas. At the school level, low-income households, language mismatch and low parental education were used as predictors. Model 2 explores the degree to which school and area characteristics may be predictive of this variability. To this end, model 2 includes predictors at the school and area levels. FSA population, residential instability, material deprivation, dependency and ethnic concentration were used as area-level predictors. School-level predictors were group-mean centred to remove between-area variations that would have otherwise been attributed to the school-level.

**RESULTS**

**Elementary school student population and area characteristics**

Table 1 presents the characteristics of the Ontario public elementary school student population and the FSAs included in analysis. All independent variables deviate moderately from normality (skewness <2, kurtosis <7). However, we observe high kurtosis for the cumulative incidence of student infections. This indicates that the cumulative incidence of student infections in most schools centered around the mean, with large fluctuations at extreme values.

**Predictive effects of socio-economic characteristics of school student body and geographic area on cumulative incidence**

Table 2 presents estimated models 1 and 2. Model 1 tests whether there may be meaningful differences in the cumulative incidence of SARS-CoV-2 elementary school student infections across FSAs. Results indicate that there were indeed meaningful between-FSA differences. Variability at the area level (ICC) accounted for 15.5% of total variability. That is, 15.5% of the school variability of student infections is accounted for by the geographic area of the school. This suggests modest between-area variability of school student infections.

Model 2 examines whether socio-economic factors and dimensions of marginalisation may predict the cumulative incidence of SARS-CoV-2 elementary school student infections. Results show that the proportion of the student body from low-income households was the only significant factor at the school level (level 1). It was positively related to cumulative incidence (β=0.083, p<0.001). This suggests that when between-FSA variability is controlled for, schools with a higher proportion of student bodies from low-income households were associated with higher relative cumulative incidences of student infections within the school. However, school-level variables accounted for only 1.2% of school-level variability.

The association with area marginalisation variables was strong, accounting for 57.6% of variability at the area level (level 2). All four indicators of marginalisation were statistically significant at this level. While ethnic concentration (β=0.454, p<0.001), residential instability (β=0.356, p<0.001) and material deprivation (β=0.212, p<0.001) were positively related to the cumulative incidence of student infections, dependency (β=0.204, p<0.001) was negatively related. This suggests that, on average, schools in FSAs with relatively higher levels of ethnic concentration, residential instability and material deprivation would have had higher cumulative incidences of student infections, while schools in FSAs with higher dependency would have had lower cumulative incidences (see online supplemental material S5 and figures 1–4 for visualisations of the relationship between each area marginalisation indicator and the cumulative incidence).

We performed sensitivity analysis to ascertain whether the observed relationships in model 2 may be due to how the
cumulative incidence variable was operationalised. Alternative models 1 and 2 were estimated with the cumulative number of cases per school as the outcome variable for comparison. Results from the alternative analysis were similar to the results presented. This suggests that the current results are robust to differences in operationalisation (see online supplemental material S6 for sensitivity analysis).

**DISCUSSION**

Our analyses uncovered several important findings. First, a substantial proportion of variation in the cumulative incidence of SARS-CoV-2 elementary school student infections can be explained by the geographic area of schools. Second, the proportion of the student body from low-income households in individual schools was associated with the cumulative incidence of elementary school student infections, although weakly. Finally, area indicators of marginalisation were strongly associated with geographic variability in the cumulative incidence of elementary school student infections.

From a policy perspective, understanding the most predictive level of analysis can allow for greater precision in implementing school-based infection prevention measures and education continuity and recovery strategies. A multilevel modelling approach afforded us the opportunity to statistically determine which of the multiple school-level and area-level factors most likely predicted infections, by explicitly considering the nested nature of schools within specific geographic areas. The study strongly suggests that area characteristics, rather than factors unique to individual schools may have driven the cumulative incidence of SARS-CoV-2 elementary school student infections. Our results show that most of the variance in infections could be explained by area-level predictors.

Our findings are consistent with public health reporting in Ontario. Among school-aged SARS-CoV-2 cases, the most diverse neighbourhoods had rates that were approximately 3.5 times higher than those in the least diverse neighbourhoods, and rates were 1.6 times higher in the most deprived, compared with the least deprived, neighbourhoods. Results of our analysis show the effects of the embeddedness of schools within communities. Thus, student susceptibility to SARS-CoV-2 must be viewed through a social determinant framework for public health interventions to prevent onward transmission in schools and their communities.

Finally, increasing inequities in prevalence have also been noted over time, and associated with households with lower incomes, crowding and essential workers. Household size has been shown to be associated with increased odds of a positive test result for SARS-CoV-2, likely related to more intense exposure compared with other settings. Moreover, households had increased risk of infection associated with increasing number of children in the household. Interestingly, in our analysis, schools in areas with higher dependency (defined as

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Standardised coefficients, SEs and CIs for the multilevel models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
</tr>
<tr>
<td></td>
<td>Beta (SE) (95% CI)</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.688 (0.111)*** (1.468 to 1.908)</td>
</tr>
<tr>
<td>School-level (level 1) predictors</td>
<td></td>
</tr>
<tr>
<td>Low-income households</td>
<td>–</td>
</tr>
<tr>
<td>Language mismatch</td>
<td>–</td>
</tr>
<tr>
<td>Low parental education</td>
<td>–</td>
</tr>
<tr>
<td>Area-level (level 2) predictors</td>
<td></td>
</tr>
<tr>
<td>Population of FSA</td>
<td>–</td>
</tr>
<tr>
<td>Residential instability</td>
<td>–</td>
</tr>
<tr>
<td>Material deprivation</td>
<td>–</td>
</tr>
<tr>
<td>Dependency</td>
<td>–</td>
</tr>
<tr>
<td>Ethnic concentration</td>
<td>–</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
</tr>
<tr>
<td>Variance within schools</td>
<td>55.690 (5.557)*** (44.687 to 66.693)</td>
</tr>
<tr>
<td>Variance between FSAs</td>
<td>10.242 (1.403)*** (7.464 to 13.020)</td>
</tr>
<tr>
<td>ICCschool</td>
<td>0.845</td>
</tr>
<tr>
<td>ICCFSA</td>
<td>0.155</td>
</tr>
<tr>
<td>R²school</td>
<td>–</td>
</tr>
<tr>
<td>R²FSA</td>
<td>–</td>
</tr>
</tbody>
</table>

**P<0.05; **p<0.01; ***p<0.001. Standardised coefficients are presented.
FSA, forward sortation area; ICC, intraclass correlation.
households with individuals aged 65 years and older, 14 years and younger and/or adults not participating in the labour force) had relatively lower cumulative incidences of student infections. This warrants further study to determine potential factors associated with dependency, such as reduced mobility and contacts outside the home-related and age-related prioritised vaccine eligibility.

**Equity-related implications**

There are equity-related implications of the analysis. While school closure data were not available, the implications of higher cumulative incidence of student infections in schools in marginalised areas is that those school populations would have likely sustained relatively more disruption at multiple levels due to necessary isolation measures, compared with schools in other areas: individual (students excluded), classroom (cohort dismissals) and school (closures). This raises concerns. Systematic reviews of studies on COVID-19-related school closures indicate significant learning loss even with the provision of emergency virtual instruction, the bulk of studies showing greater losses among students from lower-income and other marginalised groups.\(^3\)\(^4\) Additionally, an earlier systematic review of extended education disruption during COVID-19 outbreaks showed substantial well-documented harms, such as aggravated mental health effects and protection and social welfare concerns, disproportionately borne by socially disadvantaged groups.\(^3\)\(^4\) Thus, from a policy perspective, our findings suggest that schools in more marginalised areas should be prioritised for resources to reduce the risk of infection and associated education disruption.

**Limitations of study**

Lack of publicly available datasets on the prevalence and duration of localised school closures prevented a more fine-grained analysis of the cumulative incidence variable. The lack of publicly available race-based school-level data underscores the need for these data to be included in the school demographic dataset for more direct precision rather than inferred variables on ethnic composition. Nonetheless, the datasets we used were the most complete we could publicly obtain and are consistent with data used by decision-makers. Moreover, our findings may underestimate the extent of variability predicted by area-level socio-economic factors due to the exclusion of private schools in official data, and the use of FSAs rather than school catchment boundaries.

**CONCLUSIONS AND POLICY IMPLICATIONS**

The socio-economic characteristics of the geographic location of schools were much more strongly associated with the cumulative incidence of elementary school student infections than individual school characteristics. Elementary schools in marginalised areas in Ontario were more negatively affected. Furthermore, schools with a higher proportion of students from lower-income households had a higher cumulative incidence of SARS-CoV-2 student infections, although this relationship was relatively weak. Given the inequitable effects of protracted school closures, schools in marginalised areas should be prioritised for infection prevention and mitigation measures and in education continuity and recovery plans.  

**REFERENCES**
