

1
2
3
4
5
6
7
8
9
10
11
12

Quantifying threat from COVID-19 infection hazard in Primary Schools in England

Authors: Aspinall, W.P.^{1,2}, Sparks R.S.J.¹, Cooke, R.M.³, Scarrow, J.⁴

- 1. School of Earth Sciences, University of Bristol, Bristol BS8 1RJ, United Kingdom
- 2. Aspinall and Associates, Tisbury SP3 6HF, United Kingdom.
- 3. Institute of Applied Mathematics, Delft University of Technology, Building 28, Mourik Broekmanweg 6, 2628 XE Delft, Netherlands.
- 4. Departamento de Mineralogía y Petrología, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain.

Date of version (6th August 2020)

13 **Summary.** We have constructed a COVID-19 infection hazard model for the return of pupils to
14 the 16,769 state Primary Schools in England that takes into account uncertainties in model
15 input parameters. The basic probabilistic model estimates likely number of primary schools
16 with one or more infected persons under three different return-to-school circumstances. Inputs
17 to the infection hazard model are: the inventory of children, teachers and support staff; the
18 prevalence of COVID-19 in the general community including its spatial variation, and the ratio
19 of adult susceptibility to that of children. Three scenarios of inventory are: the counts on 1st
20 June when schools re-opened to Nursery, Reception, Year 1 and Year 6 children, when
21 approximately one-third of eligible children attended; a scenario assuming a full return of
22 eligible children in those cohorts; and a return of all primary age children, scheduled for
23 September. With a national average prevalence, we find that for the first scenario between 178
24 and 924 schools out of 16,769 in total (i.e. about 1% and 5.5% respectively) may have infected
25 individuals present, expressed as a 90% credible interval. For the second scenario, the range is
26 between 336 (2%) and 1873 (11%) schools with one (or more) infected persons, while for the
27 third scenario the range is 661 (4%) to 3310 (20%) schools, assuming that the prevalence is the
28 same as it was on 5th June. The range decreases to between 381 (2%) and 900 (5%) schools with
29 an infected person if prevalence is one-quarter that of 5th June, and increases to between 2131
30 (13%) and 9743 (58%) schools for the situation where prevalence increases to 4 times the 5th
31 June level. Net prevalence of COVID-19 in schools is reduced relative to the general community
32 because of the lower susceptibility of primary age children to infection. When regional
33 variations in prevalence and school size distribution are taken into account there is a slight
34 decrease in number of infected schools, but the uncertainty on these projected numbers
35 increases markedly. The probability of having an infected school in a community is proportional
36 to the local prevalence and school size. Analysis of a scenario equivalent to a full return to
37 school with an average national prevalence of 1 in 1700 and spatial prevalence variations,
38 estimated from data for late June, indicates 82% of infected schools would be located in areas
39 where prevalence exceeds the national average. The probability of having multiple infected
40 persons in a school increases markedly in high prevalence areas. Assuming national prevalence
41 characteristic of early June, individual, operational and societal risk will increase if schools re-
42 open fully in September due to both increases in numbers of children and the increased
43 challenges of sustaining mitigation measures. Comparison between incidents in primary schools
44 with positive tests in June and July and our estimates of number of infected schools indicates at
45 least an order of magnitude difference. The much lower number of incidents reflects several
46 factors, including effective reduction in transmission resulting from risk mitigation measure
47 instigated by schools.

48
49
50
51
52
53

54

55

56 **Introduction**

57 On 11 May 2020 Prime Minister Boris Johnson announced that selected primary age children
58 would return to school on 1st June in England. The returning cohort would include Reception,
59 Year 1 and Year 6, noting that there would also be Year 2 to 5 children of frontline workers and
60 those identified as vulnerable who were already being taught in many schools. Nursery age
61 children were also invited to return with some nurseries being part of Primary schools. The
62 devolved administrations decided at the time not to re-open schools. The Government
63 abandoned the proposal for a full return of Primary school children in England before the
64 summer holidays. There is now an expectation that schools will fully re-open in September.

65 The partial re-opening of Primary schools was widely debated with concerns from some
66 parents, teaching unions and teacher associations about the safety of the children and school
67 staff. There was also concern about the effect on infection rate in the wider community, for
68 example by triggering a second wave. While official guidelines for re-opening schools have been
69 issued by the Department of Education (DfE) and form the policy basis for risk mitigation in
70 schools, individual schools have adopted their own bespoke strategies (Sparks et al. 2020).
71 Some schools re-opened on 1st June, some delayed their re-start while, in other cases, schools
72 did not re-open under advice of local education authorities. Data from DfE indicates on 2nd July
73 that about 88% of state-funded Primary Schools had re-opened to some extent. The community
74 response has been variable and between 1st and 15th June 2020 approximately 35% of eligible
75 children returned. The numbers increased somewhat: 41% (year 1) and 49% (year 6) attended
76 on 2nd July. Between 18th May and 31st July 2020 there have been 247 COVID-19 related
77 incidents in schools of which 116 were tested to be positive test (PHE 2020).

78 This study is part of a quantitative hazard assessment for schools with respect to COVID-19. The
79 study seeks to enumerate potential infection levels in primary schools, with estimates of
80 attendant uncertainties. The study involves building a stochastic uncertainty model of hazard
81 which includes information on schools and epidemiological parameters that control occurrence
82 and transmission of COVID-19 within schools. Much of the information is derived from national
83 statistics on schools and information about the incidence and prevalence of COVID-19 in
84 England.

85 This study does not address the implications of opening schools for spread of the infection in
86 the wider community, which are being addressed by other modelling studies. Here we
87 distinguish between societal risk, operational risk and individual risk. The question of how
88 opening schools up affects the national and regional picture of infection is an example of
89 societal risk assessment; for example, by estimating the change in the reproduction number
90 (Panovska-Griffiths et al. 2020). Operational risk concerns a school identifying an infection or
91 the threat of an infection breakout. Mitigating steps might include asking confirmed or
92 suspected COVID-19 persons to self-isolate, classes to be sent home, or even schools to close
93 again, entirely. Individual risk concerns the probability that a child, teacher or other support

94 staff acquires infection as a consequence of the initially unknown presence of an infected
95 person in the school.

96 Here we have developed a model of infection hazard, which is defined as the probability of one
97 or more infectious person (child, teacher or support staff) being on the premises of a school on
98 any one day. The study uses a fast, stochastic uncertainty modelling tool (UNINET) to construct
99 an uncertainty model for infection hazard levels in primary schools in England. A model of
100 infection hazard is a necessary first step in developing a risk model which needs to include
101 infection transmission. The model is extended to take account of spatial variations in
102 prevalence to characterize infection hazard level in relation to geographical locations. Our
103 model is flexible and dynamic. Thus, in the future it will enable risk to be assessed for different
104 return-to-school scenarios, including secondary schools, for regional or local situations and
105 potentially for individual schools. The approach can also be modified for different kinds of
106 educational operation, such as Universities and Further Education colleges. The model can also
107 be used to estimate sampling power in schools in the context of an infection testing
108 programme.

109

110 **Hazard and risk terminology**

111 Different domains of science and society use terms like hazard and risk in different ways and in
112 different contexts. Risk can sometimes be discussed in the context of risk mitigation and safety,
113 whereas in, for example, the financial sector the risk is defined as expected monetary loss. In
114 the health and safety context risk is the chance of harm, whereas in some natural hazard
115 contexts threat is used to signal the potential of harm. Exposure is sometimes used to indicate
116 that an asset or person is in harm's way. Risk also pertains to different scales and kinds of loss
117 or harm. Risk management benefits from information on the hazard, the exposure, the
118 vulnerability to the hazard and its potential impact. Ultimately, quantitative risk assessment
119 involves calculating the probability of a loss or a harm. Here, we treat COVID-19 infection as a
120 hazard and the presence of infectious persons in schools as a threat that might lead to harm.

121

122 **Methodology**

123 Our computational platform is the UNINET LightTwist software ([https://lighttwist-
124 software.com/uninet/](https://lighttwist-software.com/uninet/)). UNINET is a standalone uncertainty analysis software with a focus on
125 dependence modelling for high dimensional statistical or probability distributions (Hanea et al.
126 2018). The program implements and computes continuous non-parametric Bayesian Belief
127 Network (BBN) models as the framework for probabilistic modelling. A BBN is a graphical
128 model, which represents joint distributions in an intuitive and efficient way. It encodes the
129 probability density (or mass) function of a set of variables by specifying a number of conditional
130 independence statements in the form of a directed acyclic graph. The structure of the model is
131 specified by the analyst to incorporate all essential input variables and conditionalizing factors
132 that are needed to quantify the problem and to enumerate output probability distributions.
133 The approach is quite general and allows traceable and defensible results to be generated very
134 efficiently as UNINET uses unique fast Bayesian updating algorithms. One of the other main

135 advantages of UNINET is that it can handle a large number of continuous variables and does not
136 require these to be represented as discretized distributions, as in many BBN programs.

137 An added benefit is that other factors, elements and complexities can be readily added to the
138 BBN model formulation, for example, by including details of school characteristics or age-
139 dependent transmission rates and, perhaps most pertinently, the BBN calculations can be
140 quickly updated if new or revised data are forthcoming. The model can be applied at different
141 granularities and at a regional or local level. As we have found with volcanic eruptions, this
142 latter capability is invaluable for real-time forecasting applications.

143 A full list of applications of UNINET to a wide diversity of problems in science, engineering and
144 public health can be found in the reference list on the LightTwist web site. We describe the
145 UNINET platform in Appendix I, explaining the essential calculations, the principles of
146 constructing a simple network as a calculator and provide one explicit example of calculations
147 with an analysis report.

148 149 **Data sources**

150 We extracted data on primary schools in England, including their size distributions, from DfE
151 links. Information figures for 2019 can be found at:

152 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/826252/Schools Pupils and their Characteristics 2019 Accompanying Tables.xlsx](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/826252/Schools_Pupils_and_their_Characteristics_2019_Accompanying_Tables.xlsx)

154 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/812539/Schools Pupils and their Characteristics 2019 Main Text.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/812539/Schools_Pupils_and_their_Characteristics_2019_Main_Text.pdf)

156 <https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2019>

158 <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics>

159 The basic facts about state-funded primary schools in England (2019) are: 16,769 schools;
160 4,727,090 pupils; and 216,500 teachers.

161 Data for other support staff to be found at:

162 <https://www.instituteforgovernment.org.uk/publication/performance-tracker-2019/schools>

163 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/811622/SWFC MainText.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/811622/SWFC_MainText.pdf)

165 <https://www.gov.uk/government/statistics/school-workforce-in-england-november-2018>

166 Table 1 shows the numbers of other staff as listed in these sources.

167

168 *Table 1. Data on other support staff in state-funded Nursery and Primary Schools in England (2018 data)*
169

Staff Categories	Number
Teaching assistants	176,679
Administrative staff	43,771
Technicians	1,573
Other support staff	15,240
Auxiliary staff	44,993
Headcount Third part staff	26,475
Total	308,731

170
171 The Government announcement for re-opening primary schools and nurseries on 1st June
172 stipulated that the returning children would be confined to Reception, Year 1 and Year 6. The
173 pupil database for January 2019 lists numbers of children by age, which does not exactly map
174 onto school years. We used these data to give numbers as follows: 618,700 (Reception),
175 636,664 (Year 1) and 620,053 (Year 6). A small proportion of Year 2 to 5 children of emergency
176 workers and children from vulnerable environments are present in school. Data for daily
177 national attendance of pupils, teachers, teaching assistants and ancillary staff in schools can be
178 found at:

179 [https://www.gov.uk/government/collections/attendance-in-education-and-early-years-](https://www.gov.uk/government/collections/attendance-in-education-and-early-years-settings-during-the-coronavirus-covid-19-outbreak)
180 [settings-during-the-coronavirus-covid-19-outbreak](https://www.gov.uk/government/collections/attendance-in-education-and-early-years-settings-during-the-coronavirus-covid-19-outbreak)

181 For epidemiological data we used a variety of sources. Prevalence data were obtained from
182 infection surveys published by the Office of National Statistics.

183 [https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsandd](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveyypilot/latest)
184 [iseases/bulletins/coronaviruscovid19infectionsurveyypilot/latest](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveyypilot/latest)

185 ONS state that the number of infections within the community population refers to private
186 residential households, but excludes those in hospitals, care homes or other institutional
187 settings. This measure of prevalence is suitable for our infection hazard model as those
188 presenting at school are not in the excluded group. However, some of these same people will
189 mix with family and friends in the excluded group, suggesting that the national prevalence
190 infection survey data are an underestimate.

191 Incidence data at UTLA level, to be used as a proxy for regional variations of prevalence, were
192 obtained from Public Health England.

193 <https://www.gov.uk/guidance/coronavirus-covid-19-information-for-the-public>

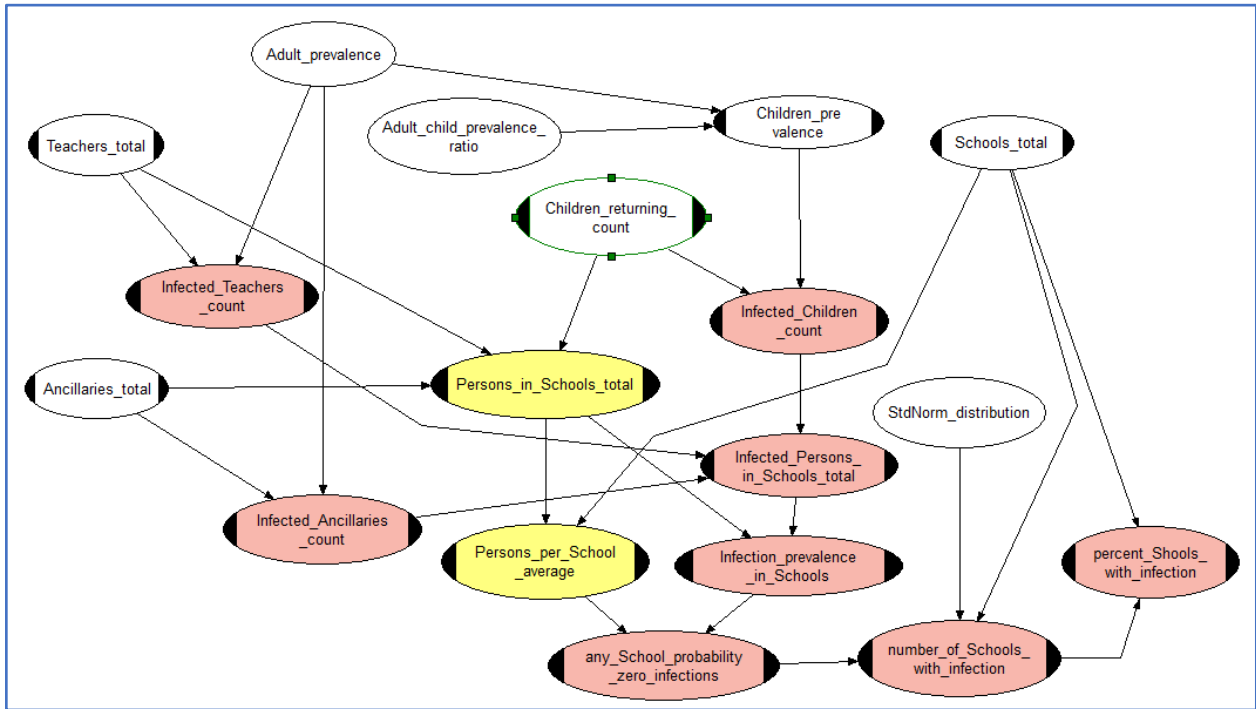
194 The incidence data were also converted into positive test numbers at UTLA level for the
195 purpose of estimating prevalence and attendant errors in larger scale regions. Population data
196 at UTLA level were obtained from ONS.

197
198

199 **Model Structure**

200 Figure 1 shows the basic Bayesian Belief Network established to estimate infection hazard in
201 schools. Input parameters include prevalence, the susceptibility ratio of children relative to
202 adults in the 20 to 60 year age range, and the inventory of children and staff in Primary Schools.
203 Outputs include the prevalence in schools, the number of schools with one or more infected
204 person on a given day and how the infection is spread between different cohorts of persons
205 (e.g. children and adults).

206



207 *Figure 1. Bayesian Belief Network to calculate schools COVID-10 infection hazard with prevalence*
208 *represented by a single random variable. In order to account for regional variations of prevalence later*
209 *BBN modelling introduced a range of random variable values for prevalence in different geographical*
210 *areas as depicted in Figure 3. Plain white nodes carry input variable uncertainty distributions, white*
211 *nodes with handles represent fixed (constant) input values. Yellow nodes are intermediate functional (i.e.*
212 *calculational) nodes. Pink nodes compute the required output variable probability distributions; some*
213 *these are also intermediate calculation steps, feeding uncertainty distributions into other output nodes.*

214

215 **Definitions**

216 Having an infectious person in a school poses a threat to individuals, the operation of the
217 school, which in extremis might have to close, and to the wider community. We term a school
218 of one or more infectious individual present as an “infected school.” The harm, potentially
219 caused by this threat, will depend on the context and a full risk assessment would involve
220 quantification of the probability of this harm occurring.

221

222 **Scenarios and model assumptions**

223 We adopt here a scenario approach in which some input parameters are either fixed or
224 represented by probability density functions reflecting informational uncertainty. The main
225 assumptions for a given scenario are the number of pupils, teachers, teaching assistants and
226 ancillary support staff present in schools. The scenario parameters are summarized in Table 2
227 and are now described.

228 *Scenario I* considers the situation on 1st June 2020, the first day of school return for selected
229 Year groups. Numbers of persons are obtained from the attendance data. The ONS published a
230 probability density function (*pdf*) for adult prevalence on 5th June 2020 (mean value expressed
231 as odds 1 in 1700) and this is used for *Scenario I*. *Scenario II* assumes attendance of all children
232 in the selected Year groups, adjusted for typical absenteeism (4%). Thus, *Scenario II* represents
233 the case had there been high compliance with the return to school on 1st June.

234 Looking forward to the full re-opening of schools in September, *Scenario IIIa* assumes the
235 return of all primary school children in all years (adjusted for typical absentee rates). Two
236 additional sub-scenarios are also considered for September: one in which the adult prevalence
237 is 4x greater than that given in the ONS *pdf* from the 5th June survey (*Scenario IIIb*) and another
238 for a drop to ¼ of that prevalence (*Scenario IIIc*). We note here that the prevalence in early
239 August is 1 in 1500, remaining close to the value of 5th June. These sub-Scenarios provide two
240 alternative but plausible ranges of adult prevalence reflecting the future level of, and
241 uncertainty in, prevalence in September. The numbers of persons in each *Scenario* are given in
242 Table 2.

243

244 Table 2. Inventory of persons present in Primary schools in the model cohorts for the three attendance
 245 level Scenarios. All numbers are for England and were extracted from the Government web site and
 246 relate to 2020. Scenario I: situation on 1st June 2020 Scenario II assumes attendance of all children in the
 247 selected Year groups, adjusted for typical absenteeism (4%). Scenarios IIIa assume a September return of
 248 all primary school children in all years, adjusted for typical absentee rates; two sub-scenarios were run
 249 for 4 times (Scenario IIIb) and ¼ (Scenario IIIc) the adult prevalence given in the ONS pdf from the 5th June
 250 survey.

	Scenario I	Scenario I	Scenario II	Scenario III (a, b, c)
	Attendance ¹	Numbers ^{2,3,4,5}	Numbers ^{3,5}	Numbers ³
Nursery	17.6%	55,720	297,930	297,930
Reception	29.2%	194,240	626,086	626,086
Year 1	24.5%	165,517	635,843	635,843
Year 2 to 5		174,938	174,938	2,327,686
Year 6	39.1%	264,871	650,322	650,322
Cohort 1	Rec + Yr1	415,471	1,559,859	1,559,859
Cohort 2	Yr6	439,809	825,260	2,978,008
Cohort 3	Teachers + TAs	235,000± 23,500	235,000± 23,500	377,450
Cohort 4	Support staff	70,000±7,000	70,000±7,000	126,720

251

252

Footnotes:

253

1. The Government Web Site reports 62% returns and describes a methodology to estimate the overall attendance. Note that some schools rationed attendance with groups of different children only attending on two days a week.
2. The numbers in Nursery have been increased by 2% to take account of children of critical workers and vulnerable children in each year groups. This estimate is given in the government web site.
3. These number for years 2 to 5 and for scenarios are reduced by 4% from the full inventory of pupils, teachers, teaching assistants and ancillary staff to take account of absentees.
4. The Government web site gives returning teachers, teaching assistants and ancillary staff for all schools but this includes secondary schools. Here we take the returning staff as 53% of the reported numbers.
5. In *Scenarios I and II* we were unable to obtain accurate national figures for Cohorts 3 and 4 because the June attendance information was not broken down between Primary and Secondary and seems to have included some private sector schools. Information from the 36 primary schools in the study gave an average of 65% return of teachers and 58% return of other staff, noting considerable variations between individual schools. Assuming the schools in the Royal Society network are a representative sample of Primary schools in England, we estimate 235,000 teachers and teaching assistants and 70,000 ancillary staff adjusting for the fact that 8% of schools had not re-opened. Teaching staff on rota during the week and part-time staff further increase the difficulty of determining precise numbers, so we have included +/-10% uncertainty on our estimates.

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

An initial and relatively simple baseline problem for modelling is to estimate the number of schools with at least one infectious pre-symptomatic or asymptomatic individual present. We assume here that once symptoms are recognized steps will be taken to remove the individual from the school and to identify possible contacts. A key issue here is that there is mounting evidence that children are less susceptible to COVID-19 infection than adults. The comprehensive modelling analysis of global data by Davies et al. (2020) led us to adopt a *pdf* for the ratio of adult to children's susceptibility with a median of 2.3 and 90% credible interval between 1.3 and 3.8 (see appendix 1). This range is within the results of other studies of

278 susceptibility ratio between adults and children (Jing et al 2020; Li et al. 2020; Mizumoto et al.
 279 2002; Zhang et al. 2020).

280 We divided the persons into four cohorts based on the children’s Year groups and staff roles.
 281 While a simplification, the cohorts are expected to have different contact characteristics, e.g.
 282 with respect to interactions between children, between children and adults and between
 283 adults. The cohorts are as follows: *Cohort 1* are Nursery, Reception and Year 1 children; *Cohort*
 284 *2* are Years 2-6 children, noting that for *Scenario 1* the Year 2-5 children are those of key
 285 workers and from vulnerable environments; *Cohort 3* are classroom teachers and teaching
 286 assistants; *Cohort 4* are non-teaching staff such as administrators, cooks, etc, who have more
 287 limited contact with children (see Table 1).

288 In order to develop the basic infection hazard model to take account of regional variations of
 289 prevalence, we used incidence data as a proxy for prevalence because there is no directly
 290 measured prevalence data at less than national scale. To first order, prevalence can be related
 291 to incidence since, at steady state, new infections are balanced by recovered individuals who
 292 are no longer infectious. Thus, prevalence can be scaled from incidence data. Table 3 displays
 293 the basis for our calibration of regional prevalence from incidence data at an Upper Tier Local
 294 Authority (UTLA) granularity.

295 *Table 3. Prevalence model derived from incidence data. Data are: weekly incidence (per 100,000) in*
 296 *prescribed bins from PHE data for week starting 25 June 2020 (Column 1); Population within UTLAs*
 297 *having incidence values in the bin range (column 2); population weighting for each bin (Column 3); total*
 298 *number of new cases (positive tests) for one week (Column 4); estimated number of infectious cases in*
 299 *each bin scaled to give bin prevalence of 1 in 1700 (Column 5); sample error based on positive test*
 300 *numbers (Column 6); estimated prevalence as odds (1 in X) in each bin (Column 7); estimated percentage*
 301 *of all infectious persons within each bin (Column 8). Final row are national averages of parameters.*
 302

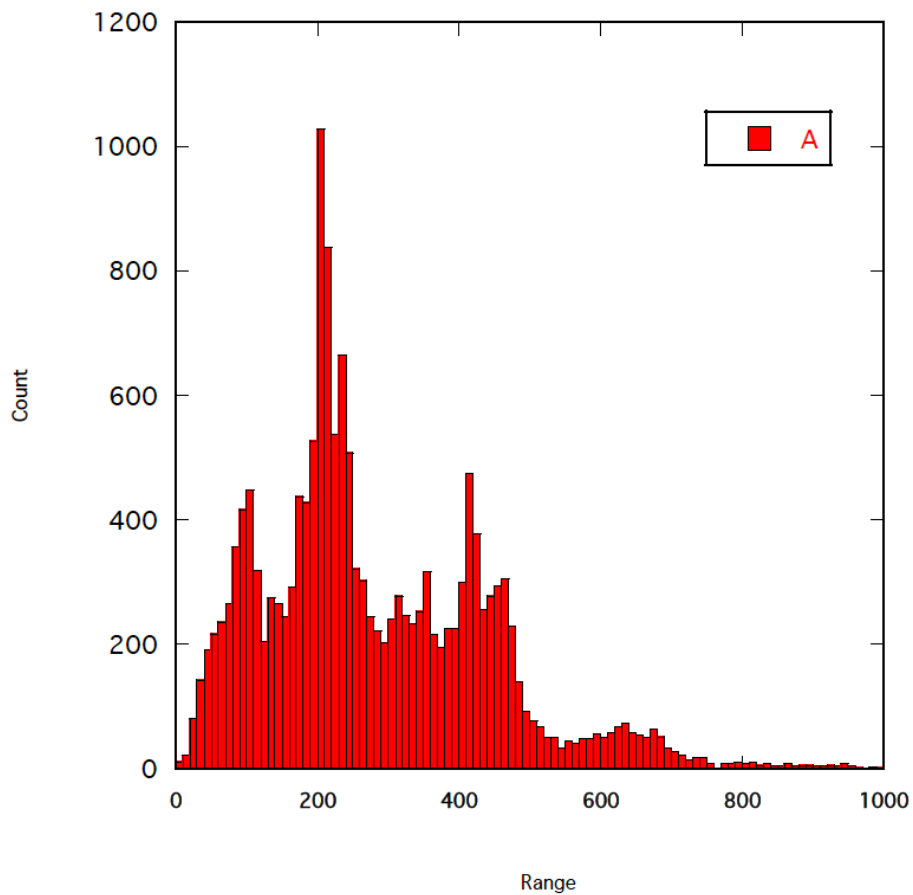
Weekly incidence	Population	Popn. Weight	New case numbers	Infectious persons	Sample error %	Equivalent Prevalence	% infected persons
0.3-1.0	3,965,491	0.07	29	180	18%	22,030	0.55%
1.0-2.9	10,525,879	0.19	193	1,195	7%	8,808	3.63%
3.0-9.9	23,641,804	0.42	1,372	8,492	2.5%	2,784	25.83%
10-29	15,128,342	0.27	2,245	13,896	1.8%	1089	42.26%
30-100	2,280,778	0.04	975	6,035	3.0%	378	18.35%
>100	355,218	0.01	498	3,083	4.3%	115	8.68%
	55,897,512	1	5312	32,881		1700	100.00%

303
 304 As with all proxies there are limitations and caveats. Incidence estimated from PCR testing will
 305 be underestimated significantly because not all those infected are tested. Further testing is
 306 focused more on symptomatic people so many asymptomatic persons may be missing from
 307 incidence data. To reconcile incidence and prevalence data from mid-June, for example, it is
 308 necessary to invoke a factor of about 6 in the under-reporting of incidence (Table 3). There will
 309 inevitably be significant uncertainties in this factor. However, creating an informative proxy
 310 model for regional prevalence variation does not require an accurate knowledge of how
 311 prevalence and incidence are related because incidence varies by about three orders of

312 magnitude at the UTLA scale. For example, in early July Leicester had a weekly incidence of 140
313 (per 100,000 people), whereas three other UTLAs reported zero incidence for the same week.

314 While subject to the above caveats, use of incidence as a proxy enables a simplified model of
315 spatial prevalence distribution to be constructed, scaled by spatial incidence distribution.
316 Incidence and population data from PHE and ONS were processed to create a coarsely binned
317 distribution of prevalence with population weights. Six bins are defined, divided by intervals of
318 weekly incidence per 100,000 persons: 0-0.29; 0.3-0.99; 1-2.9; 3-9.9; 10-29; 30-99; >100 (the
319 latter is only Leicester). Each bin represents an amalgamation of populations data and UTLA
320 incidence rates for the week starting 25th June 2020. Within each bin we estimated incidence
321 from the total number of positive tests and bin population. The error at the 95th confidence
322 level was calculated from the total number of positive tests sampled within each bin. Table 3
323 shows the data and the caption describes the basis for the model of regional prevalence.

324 A final embellishment added to the model explored the effects of school size distribution. Our
325 baseline scenarios assumed all primary schools had a mean pupil population of 282 and
326 averaged numbers of classroom and support staff based on the data in Table 1. The size
327 distribution of primary schools in England based on pupil numbers is wide and polymodal
328 (Figure 2).



329

330

Figure 2. Size distribution of Primary Schools in England based on pupil numbers.

331 Results

332

333 In the following findings, we report our results as values rounded to an appropriate number of
334 significant figures. Given the large uncertainties present in the results, exact values - as
335 provided automatically by the modelling program - can give a misleading impression of
336 precision and accuracy when output is reported without qualification. For brevity, we discuss
337 some of the results and compare *Scenarios* in terms of median and mean values. However, it is
338 essential that the large uncertainties in the model outputs are fully appreciated.

339

340 *Infection hazard levels*

341 We first present results for the distribution of infected schools for the three defined *Scenarios*.
342 Numbers of schools with one or more infected persons are listed in Table 4a and counts of
343 infected children, teachers and support staff, present in the national primary school system, are
344 given in Table 4b. Example model output distributions are shown in Figures 3a-c, illustrating
345 typical uncertainty spreads and some skewness in distribution shapes; long-tailed behavior, as
346 evinced here, could have implications for decision-making. Percentages of infected schools are
347 given. As expected, the relative number of schools with infected persons increases in
348 proportion to the increasing number of returning children and teachers.

349 The uncertainties are large and so only allow rather general inferences. Assuming a fixed
350 prevalence (5th June 2020 ONS data) under current conditions (*Scenario I*), a few hundred
351 schools are likely to be infected, while with full return (*Scenario IIIa*) the proportion increases
352 by a factor of slightly over 3.6x, in terms of mean numbers of schools involved. The increase in
353 number of persons between *Scenario I* and *Scenario III* (a factor of 5.6x) is greater than the
354 projected increase in the number of infected persons. This result is a consequence of the lower
355 susceptibility of children: the bulk of the increase in persons in school, between *Scenario I* and
356 *Scenario III*, is dominated by children, since about 65% of adult staff have already returned.

357 Prevalence in the community is a major factor in infection hazard. The large uncertainties in the
358 ONS national prevalence infection surveys make it difficult to infer trends. Prevalence has
359 fluctuated modestly over the last several weeks, declining through June and early July but
360 increasing during late July and early August at the time of writing. Thus, June 5th data appear
361 representative and are similar to early August. The effects of prevalence are clear from
362 comparing *Scenarios IIIa*, *IIIb* and *IIIc* (Table 4b). If continued public adherence to guidance is
363 maintained to September and prevalence drops to ¼ of current levels then, in terms of mean
364 values, approximately 2.5% of schools will have infected persons, compared to 10% at current
365 levels. However, a second wave, causing, say, a factor of 4 increase in prevalence, will increase
366 this estimate to about 30%. We stress that uncertainties are large so there is a 1 in 20 chance
367 that the percentages could be higher than the 95th percentile estimate reported in Table 4a (i.e.
368 58%). The prevalence in schools is lower than in the general community because children
369 dominate the school population. If the general community prevalence has a median value of 1
370 in 1700 then the median value of prevalence in schools is estimated as 1 in 2800 for *Scenario I*
371 and 1 in 3300 for *Scenario IIIa*. A major caveat, addressed below, is that prevalence may vary
372 considerably from region to region, or even from locality to locality, and there may be very

373 localised infection hot spots. This is one limitation to calculating – and relying on - averaged or
 374 overall national values.

375

376 *Table 4a Infection hazard results: number of state primary schools with one or more infected persons*
 377 *present (with percentage of all schools in parentheses).*

	Mean	Std dev	5% quantile	50% quantile	95% quantile
<i>Scenario I</i>	460 (2.7%)	244	178 (1.1%)	406 (2.4%)	924 (5.5%)
<i>Scenario II</i>	911 (5.4%)	503	336 (2.0%)	798 (4.8%)	1873 (11.1%)
<i>Scenario IIIa</i>	1635 (9.8%)	732	612 (3.6%)	1444 (8.6%)	3310 (19.8%)
<i>Scenario IIIb</i>	5255 (31%)	2350	2131(12.7%)	4863 (29.1%)	9743 (58.3%)
<i>Scenario IIIc</i>	437 (2.6%)	243	162 (0.9%)	381 (2.3%)	900 (5.3%)

378

379

380 *Table 4b. Estimated number of infected persons in primary schools in England, by return/attendance*
 381 *Scenarios*

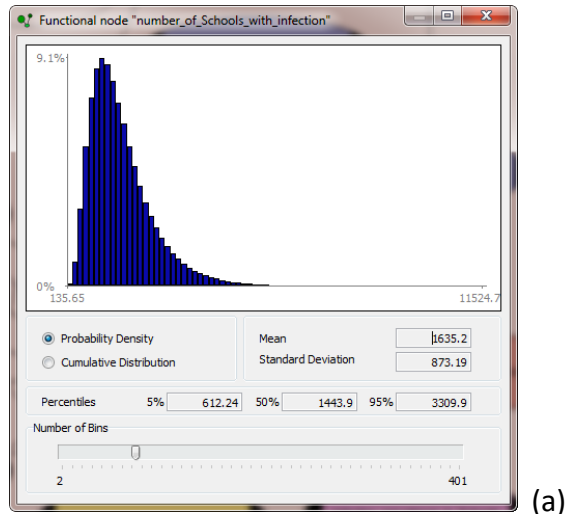
	Number	Std dev	5% quantile	50% quantile	95% quantile
<i>Scenario I</i>					
Children	267	166	88	227	582
Teachers & TAs	155	78	64	139	302
Ancillary staff	46	23	19	41	90
<i>Scenario II^a</i>					
Children	745	464	246	632	1624
<i>Scenario IIIa</i>					
Children	1417	883	468	1203	3090
Teachers & TAs	249	125	102	223	487
Ancillary staff	83	42	34	75	160
<i>Scenario IIIb</i>					
Children	5458	3558	1713	4572	12188
Teachers & TAs	960	511	372	847	193
Ancillary staff	322	171	125	284	648
<i>Scenario IIIc</i>					
Children	360	219	121	307	776
Teachers & TAs	63	31	27	57	121
Ancillary staff	21	10	9	19	41

382

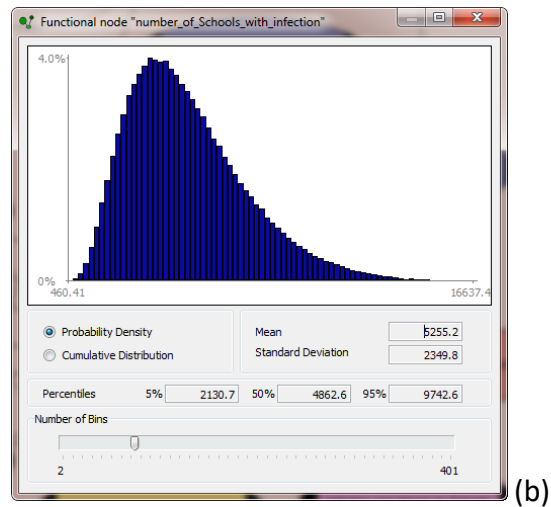
383

384

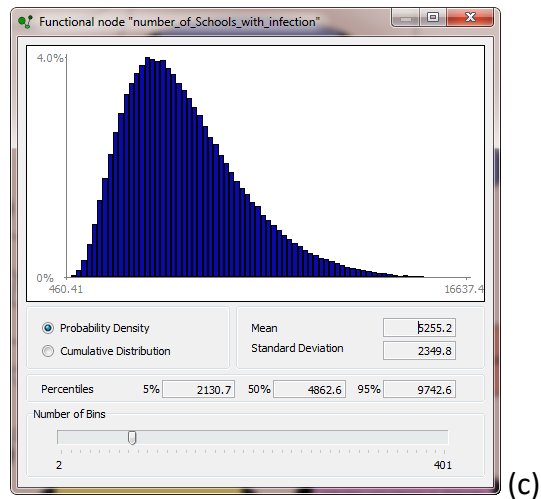
385



386



387



388 Figure 3. *Scenario IIIa*: probability density functions for number of Schools with infection, using the
389 model of Figure 1 (results should be read to nearest whole number – see Table 3a) (a) *Scenario IIIa*; (b)
390 *Scenario IIIb*; (c) *Scenario IIIc*.
391

392

393

394

395 Infection hazard for a particular school is a function, *inter alia*, of size. The average Primary
396 school in England hosts about 280 children. The percentages in Table 4a can be applied to a
397 school of this average size to give the probability that such a school will have an infected person
398 on any one day. The mean odds are 1 in 40 for *Scenario I* and 1 in 10 for *Scenario IIIa*. Using
399 children count as an approximate proxy for school size, the mean odds of there being an
400 infected person present in a small rural school with 60 children would be about 1 in 190 while,
401 for a large urban primary school of 700 children, the corresponding probability would be 1 in
402 16.

403 The number of infected persons in Table 4b add up to a slightly larger count (by a few) than the
404 number of schools in Table 4a. This difference is due to a small number of schools which will
405 have two or more infected persons present on the same day. Conditional on at least one
406 infected person being present, the probability of having two or more cases in one school of
407 average size is approximately 0.005 (i.e. 1-in-200), while for three or more it is about 0.00015
408 (1-in-6667). However, these probabilities only apply to infected persons drawn at random from
409 the wider community, assuming average national prevalence; the risk of having multiple
410 infected persons in a school is developed further, below, in relation to spatial prevalence
411 variations. These estimates do not take account of secondary infections occurring within the
412 school. Our analysis also assumes that all individuals (pupils, teachers and other staff) are
413 independent 'samples' in population statistics terms, and therefore their probabilities of
414 infection are uncorrelated. If there were two or more siblings in one school, say, then the joint
415 probability of them being infected together is increased relative to two unrelated or otherwise
416 unconnected members of a population (see Appendix 2).

417 Infection hazard does not equate directly to the extent that infection might spread, but we
418 expect that risk of an outbreak will be proportional to infection hazard. Between 18th May and
419 31st July 2020 there have been 247 COVID-19 related incidents in schools of which 116 were
420 tested to be positive test (PHE 2020). Details about the circumstances for those that tested
421 positive are not accessible, but the data provide a lower bound on the number of schools with
422 infectious persons present. There may be a significant number of asymptomatic persons
423 present and it also does not follow that having an infectious person in a school will lead to an
424 incident leading to a test, especially in the context of stringent risk mitigation measures being
425 applied in primary schools. The number of schools estimated to contain an infectious person
426 on a given day is thus much higher than the occurrences of reported cases with positive tests.

427 There are many reasons why infection may not be passed on, resulting in a much reduced
428 frequency of outbreaks and hence lower disease incidence. The infectiousness of the particular
429 individual may be low and they may start feeling ill outside school hours. Infectiousness is
430 thought to be at a maximum just before feeling ill (He et al. 2020; Ashcroft et al. 2020) and this
431 heightened state might be reached also outside school hours (e.g. weekend). In contrast,
432 asymptomatic carriers may not be recognized at all. The risk mitigation measures to reduce

433 contacts between people, ensure hygiene and isolate persons who display possible COVID-19
434 symptoms (Sparks et al. 2020) indicate they are effective in reducing opportunities for infection
435 spread.

436 It is tempting to scale up from the schools' actual breakout data during *Scenario I* to project
437 potential *Scenario III* incidents in a linear manner, using the relative exposed population ratio.
438 However, it seems more likely that the scaling will not be linear. In particular, risk mitigation
439 might be less strict and harder to maintain when many more children are back in school.

440
441 *Spatial variations in prevalence*

442 The results in Tables 4a and 4b, and discussed above, assume a constant prevalence across the
443 UK, but this is clearly not the case with localized hotspots, such as Leicester, Blackburn and
444 Bradford in July. Regional variations of prevalence in early July suggest a variation at the UTLA
445 (ONS upper tier local authority) scale of at least 350x. We have explored as a simple illustration
446 how this variation might affect the results. In the BBN, six half-log incidence bins with sampling
447 uncertainties are introduced, with separate contiguous weekly incidence rates, and these are
448 converted to equivalent prevalence distributions using a generic scaling factor (Table 3); the
449 prevalence values are the inverse of the odds of an infected person in a population. For each
450 bin we calculated from ONS data the total number of persons that tested positive in each bin to
451 estimate a sampling error.

452 We have modelled regional prevalence variation for a case equivalent to *Scenario IIIa* with a
453 uniform prevalence. Table 5a shows prevalence rates partitioned into six regions (Bins) with the
454 quantiles of the distributions ascribed to each Bin. Also shown are the relative population
455 weightings for each Bin. These BBN node histogram distributions are shown in Figure 4, with
456 means and standard deviations in the lower panels.

457 In terms of schools' infection hazard results, Table 5b shows the statistical characteristics of the
458 estimated number of state primary schools with one or more infected persons for the *Scenario*
459 *IIIa* base model versus the multi-prevalence model, outlined above (see also Figure 4); the
460 equivalent percentages of all schools are shown in parentheses. The prevalence model with
461 spatial variation projects slightly reduced number of infected schools compared with the
462 uniform prevalence model. Estimated number of infected persons in primary schools, by pupil,
463 teacher or ancillary staff, for *Scenario IIIa* base model versus multi-prevalence model are shown
464 in Table 5c. In terms of mean values, the numbers are almost identical. However, the variances
465 on the number of schools and infected numbers of persons in the spatial model are increased
466 relative to the base model with more marked skewness reflected in the lower median values
467 compared to the uniform prevalence model.

468
469
470
471

472 *Table 5a. Model spatial distribution of prevalence dividing England into six contiguous half-log*
 473 *prevalence regions, with weightings reflecting the relative population sizes of these regions. The net*
 474 *mean prevalence in the BBN, in odds terms, is slightly higher than the corresponding overall mean from*
 475 *the tabulated data (1-in-1631 -v- 1-in-1700). This is likely due to slight misfits when converting Bin raw*
 476 *data into separate statistical distributions.*
 477

Prevalence	Data mean	BBN mean	5th %ile	50th %ile	95th %ile	Population weighting
Bin 1	3.60E-05	3.48E-05	0	4.20E-05	6.12E-05	0.07
Bin 2	1.20E-04	1.17E-04	6.95E-05	1.12E-04	1.81E-04	0.19
Bin 3	3.42E-04	3.33E-04	1.94E-04	3.18E-04	5.24E-04	0.42
Bin 4	9.54E-04	9.48E-04	5.71E-04	9.10E-04	1.45E-03	0.27
Bin 5	2.59E-03	2.56E-03	1.59E-03	2.47E-03	3.85E-03	0.04
Bin 6	8.41E-03	8.41E-03	5.46E-03	8.41E-03	1.14E-02	0.01
Overall means	5.88E-04	6.08E-04				
Odds 1-in-	1700	1645				

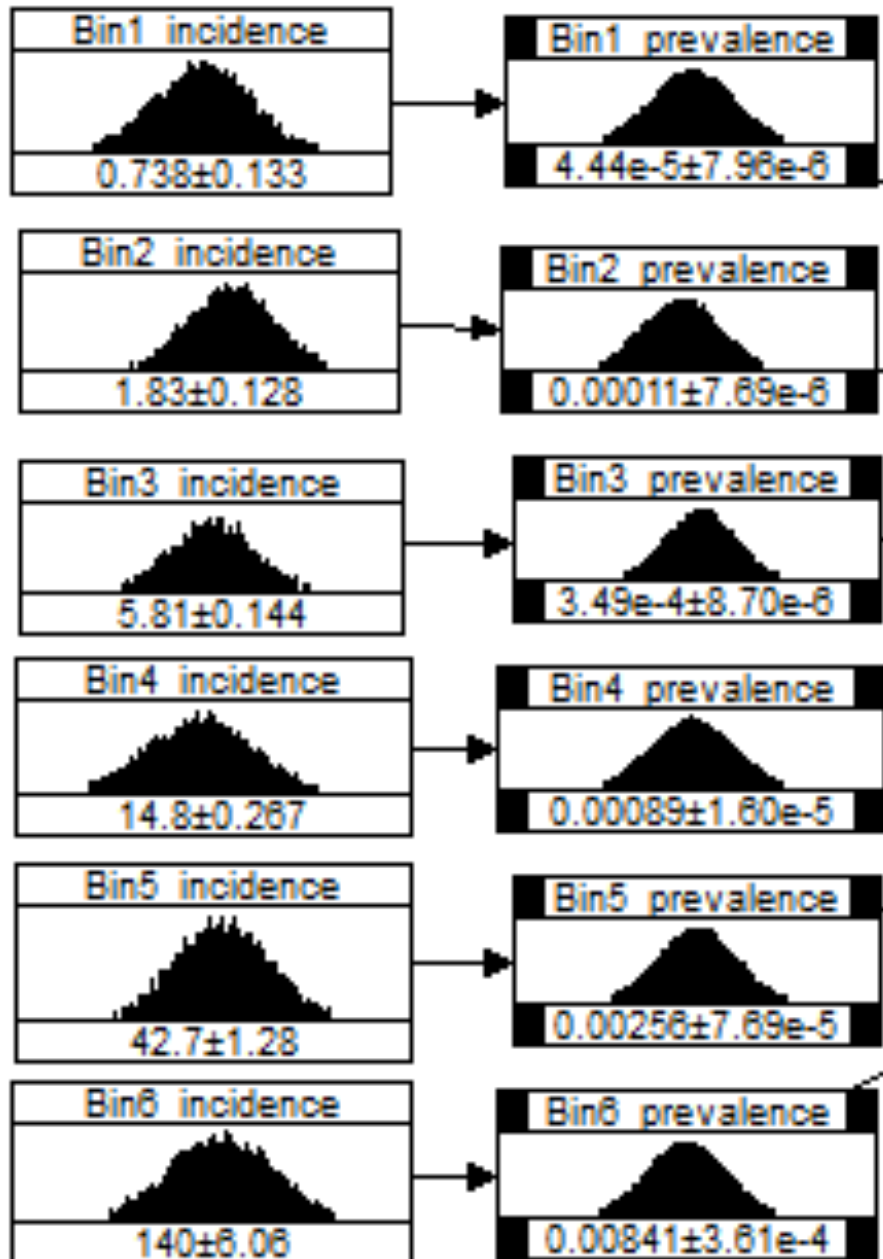
478 *Table 5b Infection hazard results: number of state primary schools with one or more infected persons*
 479 *present (with percentage of all schools in parentheses): Scenario IIIa base model versus multi-prevalence*
 480 *model versus multi-prevalence model including school size distribution (Figure 2).*
 481

	Mean	Std dev	5% quantile	50% quantile	95% quantile
Scenario IIIa	1635 (9.8%)	732	612 (3.6%)	1444 (8.6%)	3310 (19.8%)
Spatial prevalence	1458 (8.7%)	1873	122 (0.7%)	888 (5.3%)	4344 (25.9%)
School size distribution	1405 (8.4%)	1910	59 (0.35%)	754 (4.5%)	4910 (29.3%)

482
 483 *Table 5c. Estimated number of infected persons in primary schools: Scenario IIIa base model versus multi-*
 484 *prevalence model (Figure 3).*

	Number	Std dev	5% quantile	50% quantile	95% quantile
Scenario IIIa					
Children	1417	883	468	1203	3090
Teachers & TAs	249	125	102	223	487
Ancillary staff	83	42	34	75	160
Spatial prevalence					
Children	1415	2699	95	726	4167
Teachers & TAs	248	443	19	136	652
Ancillary staff	83	149	6	46	219

485
 486



488 Figure 4. Node distributions representing regional variations in incidence, converted
 489 prevalence and associated uncertainties used in the BBN infection hazard model (see Table 5a).
 490 These nodes replace the single distribution node “Adult-prevalence” in the basic BBN shown in
 491 Figure 1, above.

492 To model the distribution of infected schools in areas of different prevalence, we individually
 493 modelled each bin with school numbers proportioned to bin population. Table 6 shows the
 494 results comparing the bins by number of infected schools per 100,000 people. One major and
 495 unsurprising consequence of spatial variation of prevalence is that the number schools with

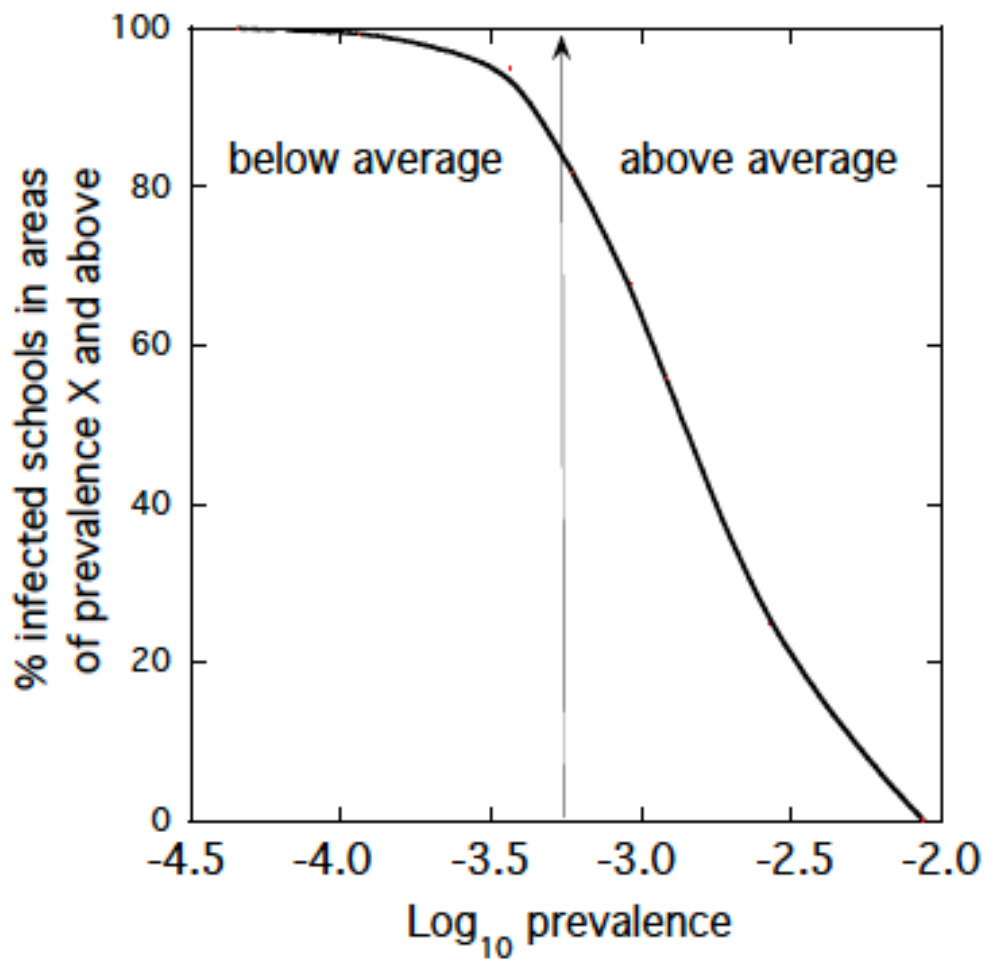
496 infectious individuals is strictly proportional to prevalence. The corollary of this result is a lower
 497 mean number of schools with infection (and lower numbers of infectious pupils, teachers and
 498 ancillaries) in areas of lower prevalence. Figure 5 displays the results as percent of infected
 499 schools that exceed a given prevalence value. For this illustrative scenario, 82% of infected
 500 schools are in areas of above average prevalence.

501
 502 *Table 6. Distributions for estimated average number of schools with one or more infected person as a*
 503 *function of spatial prevalence levels; calculations are based on the assumption that the total numbers of*
 504 *schools per prevalence bin are proportional to the corresponding population size. Different school size*
 505 *profiles within prevalence bins are not accounted for here, so the tabulated numbers of infected schools*
 506 *relate to an average sized school.*
 507

Prevalence	Mean no. infected schools	stddev	5th %ile	50th %ile	95th %ile	Infected schools per 100,000
Bin 1	8	+/- 3	5	8	13	0.2
Bin 2	55	+/- 15	35	53	83	0.5
Bin 3	377	+/- 98	244	363	559	1.6
Bin 4	592	+/- 147	391	572	862	3.9
Bin 5	221	+/- 47	153	216	306	9.7
Bin 6	121	+/- 15	96	121	145	34
Overall total	1374					

508
 509 One of the consequences of high prevalence in an area is to concentrate, in that area, the
 510 number of schools that have two more infectious persons on the same day. An approximate
 511 measure of the number of schools that have multiple infectious persons is the difference
 512 between the number of schools with one or more infected persons and the corresponding
 513 infected person numbers: the excess infectious cases indicate some schools in a high
 514 prevalence area can expect to have two or more cases at any one time. For the uniform model
 515 (*Scenario IIIa*), the difference between number of schools with an infected person and number
 516 of infected persons is 114, but for the spatial prevalence model the difference is 288. This result
 517 is significant in that there will be almost a factor of 3 increase in risk of an outbreak in those
 518 schools with multiple infected persons.

519
 520



522

523 Figure 5. Survivor function plot showing percentage of infected schools in areas of prevalence X

524 and above. The percentage of infected schools that occur at or above a fixed value of

525 prevalence can be read off the curve. The model is *Scenario IIIa* with regional prevalence

526 variation. The central vertical line is the national average prevalence assumed in the model.

527 82% of infected schools are located in areas with prevalence above the national average.

528

529

530 *Effects of school size and multiple infections*

531 Schools vary significantly in size (Figure 2). We expect that the probability of encountering one
532 or more infected persons in a school will depend on size. We therefore ran a model which
533 sampled from the national distribution of school sizes, rather than assuming a mean school size.
534 The model chosen was *Scenario IIIa* with spatial prevalence variation. The results are shown in
535 Table 5b. Including variation in school size reduces the number of infected schools from 1458
536 to 1405 while the uncertainty range increases slightly. The decrease is because of the higher
537 probability of an infectious person being encountered in a larger school. Since there is no
538 change in the overall number of infectious persons, the implication of a smaller mean number
539 of schools with infection is that there will more cases of multiple infections in some schools.

540 The discussion has so far focused on estimating the probabilities of encountering one or more
541 infected persons in school. Because there could be implications for transmissibility risk levels,
542 our analysis is extended to enumerate probabilities for 2, 3, 4 ... N infected persons per school.
543 For the model including the school size distribution and regional prevalence we find mean
544 proportions as follows for *Scenario IIIa*:

545	N = 0	Prob = 0.916
546	N = 1	Prob = 0.072
547	N = 2	Prob = 0.009
548	N = 3	Prob = 0.002
549	N = 4+	Prob = 0.001

550
551 By way of illustration, we contrast a small school (50 pupils; 10 teachers; 10 other staff), in an
552 area with the second lowest spatial mean prevalence, with a large school (900 pupils; 100
553 teachers; 100 other staff) in the second highest prevalence area (see Figure 3). The probability
554 of one infected person (pupil or adult) in the small school is 4.8×10^{-3} , for two or more the
555 probability is about 1.1×10^{-5} . In contrast, for the large school the probabilities of N infected
556 persons are:

557	N = 0	Prob = 0.26
558	N = 1	Prob = 0.35
559	N = 2	Prob = 0.24
560	N = 3	Prob = 0.11
561	N = 4	Prob = 0.04
562	N = 5+	Prob = 0.01

563 For a very big primary school in a high prevalence area, there is a 40% chance of having 2 or
564 more infected persons and about a 1% chance of encountering five or more infected persons in
565 the school at any one time.

566

567 *Comparison with observations*

568 Our model for *Scenario Ia* gives a mean of 460 (90% credible range 178 .. 924) infected schools
569 on any one day. Taking account of regional prevalence variations and school size distributions

570 this expected number would be somewhat reduced by about 15% to 390 (see Table 5b). With
571 time the numbers of schools in which an infected person has been present will slowly grow so
572 an implication is that the number of schools will be much higher after a few weeks. A quite
573 complicated calculation would be needed to infer what numbers of primary schools could have
574 been infected in June and July. It would need to consider both temporal and spatial variations
575 of prevalence and variations of how long an individual remains infectious. A very approximate
576 simplified calculation, however, gives a sense of the number. If infectiousness typically lasts a
577 week (He et al. 2020) then for fixed prevalence and a daily rate of new infected schools of 55
578 per day, the total number of schools that had been infected would be about 1925.

579 Between 18th May and 31st July 2020, 247 COVID-19 related incidents were reported in schools,
580 of which 116 were tested to be positive test (PHE 2020). Even allowing for the large
581 uncertainties, these numbers are below our estimate of the number of infected schools.
582 However, it seems likely that many infections were not detected because the infected persons
583 were asymptomatic or developed unreported symptoms outside school hours. The low number
584 could also reflect incomplete testing coverage. Further, the probability of transmission could
585 be significantly lower than 1. We have shown elsewhere (Sparks et al, 2020) that close contacts
586 between persons in schools were reduced in Primary schools by between 45% and 75%,
587 compared to normal Pre-COVID times, while other risk mitigation measure will have reduced
588 transmission significantly.

589

590 **Use of findings for informing policy decisions**

591 The main policy implications of our preliminary infection hazard results are that in England a
592 few hundred primary schools were expected to have infectious individuals within them in early
593 June, just after re-opening. With a partial return to school, social distancing measures are
594 relatively easy to comply with. On this basis, for a full return to school in September, similar
595 numbers of schools are expected to have infectious persons present on one day, but only
596 provided that prevalence falls to about $\frac{1}{4}$ of the 5th June level; on the other hand, numbers can
597 be expected to more than triple if prevalence stays at current levels.

598 If a national scale second wave comes with, for example, community prevalence increasing by a
599 factor of 4 compared to 5th June, several thousand schools may have infectious persons
600 present. This despite the fact the net prevalence in schools is likely to be much lower than in
601 the general community because of the high proportion of persons in schools being children
602 with lower infection susceptibilities. Infection hazard is proportional to school size but also
603 depends on the relative proportion of children and adults. Social distancing measures will be
604 much harder to implement with a full return to school.

605 A model which explores the effects of spatial prevalence variations finds that schools with
606 infected persons are preferentially concentrated in high prevalence communities. There is an
607 overall slight decrease in the number of infected schools nationally. However, in high
608 prevalence areas there is an approximately three-fold increase in likelihood of some schools
609 having more than one infectious person in attendance at any one time, even without
610 considering the effect of siblings and transmission leading to secondary infections. Infections
611 will tend to be concentrated in larger schools.

612 In the *Scenario IIIa* model with regional prevalence variations, the number of infected schools
613 per 100,000 people varies by a factor of over 350 between the lowest and highest prevalence
614 areas in England at the UTLA scale. This large variation suggests a strong policy focus on
615 mitigating risk in high prevalence areas. In our illustrative model, 82% of infected schools are
616 located in areas with prevalence above the national average. One option for government or for
617 a local authority is to consider stricter risk management regimes in schools where and when
618 local incidence rates exceed a threshold. To set such a threshold on a reasoned and defensible
619 basis, requires a formal quantitative risk assessment, with uncertainties properly incorporated.

620 Our results are also consistent with the effectiveness of risk mitigation measures taken in
621 schools to reduce transmission. The number of incidents in schools with positive tests are less
622 than the number of expected schools with infectious persons present by at least an order of
623 magnitude. Unfortunately, interpretation of the difference is confounded by other factors (e.g.
624 completeness of testing and role of asymptomatics). Nonetheless the reduction of contacts
625 within primary schools of between 45 and 75% (Sparks et al. (2020) and instigation of other
626 safety measure (e.g. handwashing and fastidious cleaning) supports risk mitigation as a
627 significant factor.

628

629 **Further work**

630 One factor not fully explored here is the size distribution of primary schools. We have assumed
631 that the national size distribution applies everywhere, but this size distribution is likely to vary
632 at the UTLA scale across the country, depending on the character and make-up of the local
633 community. Thus, our results can be anticipated to change somewhat when repeated at greater
634 granularity at Lower tier or school catchment scale, for example. A city like Leicester (a single
635 UTLA) could be modelled at much finer local scale to understand which individual schools
636 within that conurbation will have higher COVID-19 infection hazard. Our model captures the
637 first order picture but does not have the necessary spatial resolution to understand infection
638 hazard at local levels. Postcode-based incidence and prevalence data would allow modelling to
639 address risks at an individual school's catchment scale.

640 For incidence parameter uncertainty characterization, we adopted the assumption of lognormal
641 distributions for the different partitions of the UTLA geographic-referenced data. As always,
642 there are a number of reasons for settling on one preferred distribution type or another; in
643 many disciplines, when treating positive real variable data, a lognormal process is regarded as
644 the statistical realization of the multiplicative product of several factors, which is likely to be
645 true with incidence/prevalence data. The lognormal has the advantage of common statistical
646 usage, and understanding, hence we adopted here for our initial 'scoping' analyses. Other
647 distributions (e.g. Weibull; gamma) may represent more appropriate functional forms for these
648 particular data when aspects like goodness-of-fit are considered. We intend to explore
649 alternatives objectively in follow-on work, paying attention to upper tail fitting because, in a
650 safety-critical risk assessment, it is there that low probability, but potentially disastrous
651 outcomes may reside. The likelihood of occurrence of such threats warrants enumeration as
652 reliably as possible, in the face of many intrinsic uncertainties.

653 Our BBN model, as now set up, can also be configured straightforwardly to estimate infection
654 risk in schools of specific sizes (i.e. small, medium or large) and for individual schools, of given
655 size, in areas with known, specific prevalence rates. The model can be operationalized at
656 national and local levels to give near real-time infection hazard informed by weekly changes of
657 incidence and prevalence. The model is thus suitable for decision support with regard to
658 management strategies at school, city, local authority and national scales, and for informing
659 other research programmes, e.g. infection testing in schools.

660 This interim report only presents the results of an initial first stage analysis. The school infection
661 rate estimates presented above do not include the possibility of individual schools having
662 sibling pupils who may be simultaneously infectious; these would need to be modelled as non-
663 independent cases, i.e. correlated samples, within a school population. In similar vein, there
664 can be a number of schools with twins or triplets in the same class; this, too, would affect
665 potential infection and onward transmission rates. An initial scoping exercise, however,
666 suggests that co-sibling infection rates are low and, as the proportion of a school pupil
667 population that is represented by siblings is also small, the influence of children living at the
668 same home address is likely to be marginal for school population infection rates.

669 The current model is not a direct measure of outbreak risk in schools, although we surmise the
670 infection hazard is strongly correlated with outbreak risk. We have noted that the central
671 (mean) model estimates the daily presence of a few hundred infected primary schools in
672 England during June. Reported incidents with associated positive tests amounted to 116. While
673 hypothesizing that infections are likely to have been undetected or under-reported and while
674 transmission has been strongly suppressed by risk mitigation measures applied in the schools,
675 much more work needs to be done to understand the differences. An adjunct transmission
676 model will need to be added to our infection hazard to characterize the probability of detected
677 infections outbreak, given the presence of one or more infectious individuals in a school.

678 **Acknowledgements.** The study was part of the RAMP initiative of the Royal Society. We thank
679 Leon Danon and Jonty Rougier for commenting on an earlier version of this report.

680 **References**

681 Ashcroft P, Huisman JS, Lehtinen S, et al (2020) [COVID-19 infectivity profile Correction](#).
682 *arXiv:2007.06602v1 [q-bio.PEJ]*.

683 Hanea AM et al. (2010) Mining and visualising ordinal data with non-parametric continuous BBNs. *Comp*
684 *Stats Data Analysis* 54, 668-687.

685 He X, Lau EHY, Wu P, Deng, X., et al (2020) Temporal dynamics in viral shedding and transmissibility of
686 COVID-19. *Nature Medicine* 26, 672-675.

687 Jing, Q-L. et al. (2020) Household secondary attack rate of COVID-19 and associated determinants in
688 Guangzhou, China: a retrospective cohort study. *Lancet Infect. Dis.*, to appear. doi: 10.1016/S1473-
689 3099(20)30471-0.

690 Li, W. et al. (2020) Characteristics of Household Transmission of COVID-19. *Clin. Infect. Dis.* ciaa450, doi:
691 10.1093/cid/ciaa450

692 Mizumoto, K. et al (2020) Age specificity of cases and attack rate of novel coronavirus disease (COVID-
693 19). medRxiv preprint. doi: 10.1101/2020.03.09.20033142.

694 Panovska-Griffiths J, Kerr CC, Stuart RM, et al (2020) Determining the optimal strategy for reopening
695 schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19
696 epidemic wave in the UK:
697 a modelling study. Lancet Child Adolescent Health 2020 Published Online August 3, 2020
698 [https://doi.org/10.1016/S2352-4642\(20\)30250-9](https://doi.org/10.1016/S2352-4642(20)30250-9).

699 PHE (2020) Weekly Coronavirus Disease 2019 (COVID-19) Surveillance Report Summary of COVID-19
700 surveillance systems Weeks 28.

701 Sparks RSJ, Aspinall WP, Brooks-Pollock E, Cooke RM, Danon L, Barclay J, Scarrow JH and Cox JH (2020) A
702 novel approach
703 for evaluating contact patterns and risk mitigation strategies for COVID-19 in English Primary Schools
704 with application of Structured Expert Judgement. (in review)

705 Zhang, J. et al. (2020) Changes in contact patterns shape the dynamics of the COVID-19 outbreak in
706 China, Science 368 (2020), 1481-1486. doi: 10.1126/science.abb8001

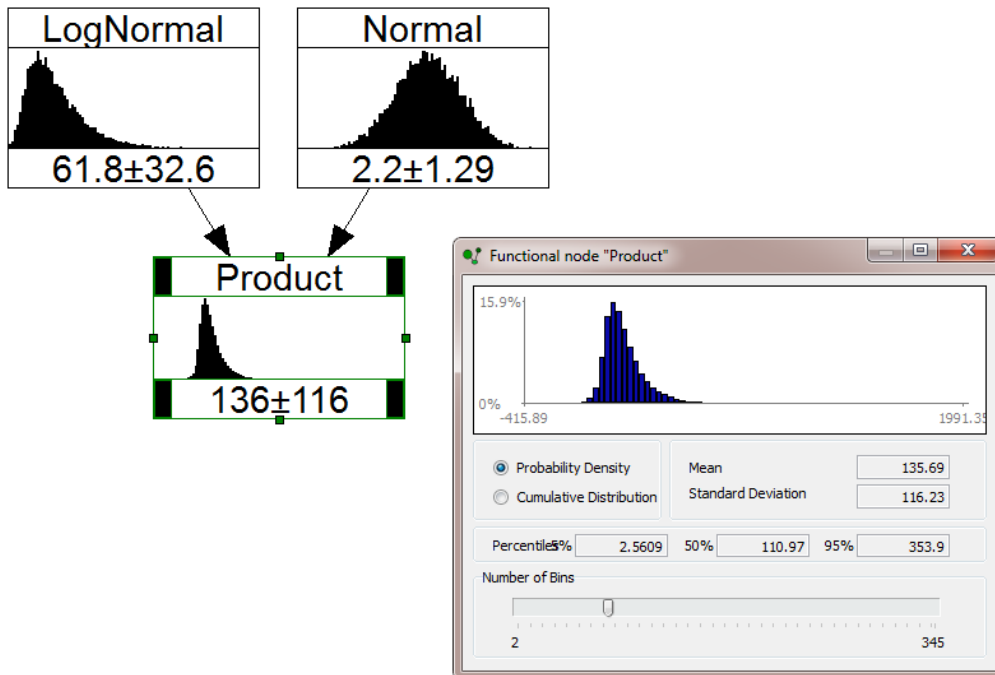
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721

722 **Appendix 1. Stochastic Modelling and Uncertainty Analysis with UNINET**

723 In the present study, we use UNINET in its basic stochastic uncertainty modelling mode, i.e. to
724 perform Monte Carlo re-sampling from random variables described by statistical distributions
725 or by uncertainties derived from expert judgement elicitation. (In this application, the program
726 is not used for Bayesian inference, although the model is represented by a directed acyclic
727 graph very similar to a Bayesian Belief Network, a BBN).

728 In the following simple illustrative model (Figure A1), there are two random variables, one a
729 Lognormal distribution, the other a Normal distribution, each with definable parameters. These
730 'parent' nodes are connected to a third 'child' node by influence arcs, denoting the child node
731 has dependencies on both parents. The child node is a functional or calculational node, that is,
732 it contains a formula for solving its relationship to the parents, in this case, the simple product
733 LogNormal * Normal.

734 The result of performing multiple re-sampling of the two parent node distributions and
735 calculating the product of these samples in the third node computes the full statistical
736 distribution of the product, given the uncertainties associated with the parent nodes (in this
737 model, 100,000 samples were used). UNINET can show graphically all node distribution shapes,
738 with their means and standard deviations displayed, as below, and can report more detailed
739 properties of the distributions, as in the lower right panel; the program can also export the
740 individual samples collected from the Monte Carlo re-sampling for post-processing in other
741 forms of analysis.



742
743
744 Figure A1. Simple UNINET model showing how variable uncertainty distributions can be
745 represented by probability nodes (e.g. LogNormal, Normal) and calculations can be made with
746 functional nodes containing equations or formulae (e.g. node Product). For a set number of

747 sampling iterations, the Product node computes the product of LogNormal * Normal and
748 creates a new distribution reporting its mean and quantiles.

749
750 The UNINET program includes implementations of several common forms of statistical
751 distribution, which can be uniquely parameterised, and has a full range of numerical operators
752 for performing stochastic calculations via functional node equations defined by the analyst. In
753 addition, the program has options for computing results when input variables are
754 conditionalized on specific point values or ranges of values. With these powerful capabilities,
755 UNINET provides an efficient, easily developed modelling tool for conducting sophisticated
756 exploratory uncertainty analysis in problems of high dimensionality.

757
758 The calculations being performed within the UNINET platform in this study would be
759 exceedingly simple if there were no uncertainty in the pertinent epidemiological parameters,
760 no spatial variations in prevalence and if all schools were exactly the same size with the same
761 proportions of children, teachers and other support staff. In addition, we need the total
762 number of schools (N_s), children (N_c) and adults (N_a) in the school system. Such a simple but
763 unrealistic case, however, does allow the essence of the calculations to be described. For
764 schools of identical characteristics the number of infectious persons present in schools would
765 be proportional to prevalence, p , and would depend on the susceptibility ratio of the adults
766 relative to children, S_r . For simplicity the susceptibility of an adult is taken as 1. The number of
767 infectious children, I_c , and infectious adults, I_a , can then be written as:

$$768 \quad I_a = pN_a \quad I_c = pN_c/S_r.$$

769
770 Even in this simplest of case the number of schools with an infectious person (I_s) would not
771 necessarily equal the sum of I_a and I_c because of the possibility that, by random chance, a few
772 schools might have 2 or more infectious people. Thus, we could expect $I_s \leq I_a + I_c$. For very small
773 $p \ll 1$, $I_s \sim I_a + I_c$. However, even here a computational model would need to be developed to
774 randomly assign infectious persons among the N_s schools and repeated calculations would
775 result in uncertainty for I_s .

776
777 However, the key epidemiological parameters have large uncertainties, while school sizes and
778 relative proportions of children and adults vary. Therefore, what is required is a high
779 dimensional computation enabled by the UNINET platform in which the uncertainty
780 distributions are sampled a very large numbers of times in order to make sure that tails of the
781 probability density functions (pdfs) are well sampled. Just as the inputs are uncertain so are the
782 outputs of I_s , I_a and I_c . Thus, stochastic uncertainty modelling of this kind is ubiquitous in
783 almost all risk-related areas of environmental sciences and in many engineering applications,
784 and this has been the case for several decades.

785
786 Complexity can be readily added to a BBN network. For example, school size distributions and
787 spatial variation in prevalence can be added to understand how infectious schools are
788 distributed geographically. A transmission model or a model of the presence of siblings in

789 schools can be added. Below we give a single example of calculations for one particular
790 scenario with constant inventory of children and adults in the schools.

791

792 Here we reproduce a comprehensive report, generated by UNINET, that records details of the
793 nodes that make up the BBN model underpinning this study, together with one set of mean
794 and quantile results, calculated by stochastic simulation within the program.

795

796 I. General information

797 1. Bbn file name: E:\Covid-19\Covid BBN\BBN schools Scenario 1 model 260620.uninet.

798

799 2. Number of samples per variable: 2000000.

800

II. Nodes information

1. [Children returning count](#)

- a) Description:
- b) Type: Functional: input data (see Table III).
- c) Dependence info: Child nodes
 - "Infected_Children_count", functional.
 - "Persons_in_Schools_total", functional.

2. [Infected Children count](#)

- a) Description:
- b) Type: Functional: $\text{Children_returning_count} * \text{Children_prevalence}$ (see Table III).
- c) Dependence info: Parent nodes
 - "Children_returning_count", functional.
 - "Children_prevalence", functional.
- d) Dependence info: Child nodes
 - "Infected_Persons_in_Schools_total", functional.

3. [Teachers total](#)

- a) Description:
- b) Type: Functional: input data (see Table III).
- c) Dependence info: Child nodes
 - "Infected_Teachers_count", functional.
 - "Persons_in_Schools_total", functional.

4. [Adult prevalence](#)

- a) Description:
- b) Type: Probabilistic: lognormal input data (see Table III).
- c) Dependence info: Child nodes
 - "Infected_Teachers_count", functional.

- "Infected_TA_count", functional.
 - "Infected_Ancillaries_count", functional.
 - "Children_prevalence", functional.
5. [Infected Teachers count](#)
- a) Description:
 - b) Type: Functional: $\text{Teachers_return} * \text{Adult_prevalence}$ (see Table III).
 - c) Dependence info: Parent nodes
 - "Adult_prevalence", probabilistic.
 - "Teachers_total", functional.
 - "Teachers_return", probabilistic.
 - d) Dependence info: Child nodes
 - "Infected_Persons_in_Schools_total", functional.
6. [Infected Persons in Schools total](#)
- a) Description:
 - b) Type: Functional: $\text{Infected_Teachers_count} + \text{Infected_Children_count} + \text{Infected_TA_count} + \text{Infected_Ancillaries_count}$ (see Table III).
 - c) Dependence info: Parent nodes
 - "Infected_Children_count", functional.
 - "Infected_Teachers_count", functional.
 - "Infected_TA_count", functional.
 - "Infected_Ancillaries_count", functional.
 - d) Dependence info: Child nodes
 - "Infection_prevalence_in_Schools", functional.
7. [Schools total](#)
- a) Description:
 - b) Type: Functional: input data (see Table III).
 - c) Dependence info: Child nodes
 - "Persons_per_School_average", functional.
 - "number_of_Schools_with_infection", functional.
 - "percent_Schools_with_infection", functional.
8. [any School probability zero infections](#)
- a) Description:
 - b) Type: Functional: $(1 - \text{Infection_prevalence_in_Schools})^{\text{Persons_per_School_average}}$ (see Table III).
 - c) Dependence info: Parent nodes
 - "Infection_prevalence_in_Schools", functional.
 - "Persons_per_School_average", functional.
 - d) Dependence info: Child nodes

- "number_of_Schools_with_infection", functional.

9. [Persons in Schools total](#)

- Description:
- Type: Functional: Teachers_return + Children_returning_count + TAs_total + Ancillaries_return (see Table III)
- Dependence info: Parent nodes
 - "Teachers_total", functional.
 - "Ancillaries_total", functional.
 - "Children_returning_count", functional.
 - "TAs_total", functional.
 - "Ancillaries_return", probabilistic.
 - "Teachers_return", probabilistic.
- Dependence info: Child nodes
 - "Infection_prevalence_in_Schools", functional.
 - "Persons_per_School_average", functional.

10. [Infection prevalence in Schools](#)

- Description:
- Type: Functional: Infected_Persons_in_Schools_total/Persons_in_Schools_total (see Table III).
- Dependence info: Parent nodes
 - "Persons_in_Schools_total", functional.
 - "Infected_Persons_in_Schools_total", functional.
- Dependence info: Child nodes
 - "any_School_probability_zero_infections", functional.

11. [Persons per School average](#)

- Description:
- Type: Functional: Persons_in_Schools_total / Schools_total (see Table III).
- Dependence info: Parent nodes
 - "Persons_in_Schools_total", functional.
 - "Schools_total", functional.
- Dependence info: Child nodes
 - "any_School_probability_zero_infections", functional.

12. [number of Schools with infection](#)

- Description: Binomial(#schools,(1 - any_School_probability_zero_infections')) is approx Normal with mean = Schools_total*(1- any_School_probability_zero_infections'), variance = any_School_probability_zero_infections*(1- any_School_probability_zero_infections)

b) Type: Functional: $\text{StdNorm_distribution} * ((1 - \text{any_School_probability_zero_infections}) * \text{any_School_probability_zero_infections})^{0.5} + \text{Schools_total} * (1 - \text{any_School_probability_zero_infections})$ (see Table III).

- c) Dependence info: Parent nodes
- "any_School_probability_zero_infections", functional.
 - "Schools_total", functional.
 - "StdNorm_distribution", probabilistic.
- d) Dependence info: Child nodes
- "percent_Schools_with_infection", functional.

13. [percent Schools with infection](#)

- a) Description:
- b) Type: Functional: $100 * \text{number_of_Schools_with_infection} / \text{Schools_total}$ (see Table III).
- c) Dependence info: Parent nodes
- "Schools_total", functional.
 - "number_of_Schools_with_infection", functional.

14. [Ancillaries total](#)

- a) Description:
- b) Type: Functional: input data (see Table III).
- c) Dependence info: Child nodes
- "Infected_Ancillaries_count", functional.
 - "Persons_in_Schools_total", functional.

15. [TAs total](#)

- a) Description: not used
- b) Type: Functional: not used
- c) Dependence info: Child nodes
- "Infected_TA_count", functional.
 - "Persons_in_Schools_total", functional.

16. [Infected TA count](#)

- a) Description: not used
- b) Type: Functional: not used
- c) Dependence info: Parent nodes
- "Adult_prevalence", probabilistic.
 - "TAs_total", functional.
- d) Dependence info: Child nodes
- "Infected_Persons_in_Schools_total", functional.

17. [Infected Ancillaries count](#)

- a) Description:
 - b) Type: Functional: $\text{Ancillaries_return} * \text{Adult_prevalence}$ (see Table III).
 - c) Dependence info: Parent nodes
 - "Ancillaries_total", functional.
 - "Adult_prevalence", probabilistic.
 - "Ancillaries_return", probabilistic.
 - d) Dependence info: Child nodes
 - "Infected_Persons_in_Schools_total", functional.
18. [Adult child prevalence ratio](#)
- a) Description:
 - b) Type: Probabilistic: lognormal (see Table III).
 - c) Dependence info: Child nodes
 - "Children_prevalence", functional.
19. [Children prevalence](#)
- a) Description:
 - b) Type: Functional: $\text{Adult_prevalence} / \text{Adult_child_prevalence_ratio}$ (see Table III).
 - c) Dependence info: Parent nodes
 - "Adult_child_prevalence_ratio", probabilistic.
 - "Adult_prevalence", probabilistic.
 - d) Dependence info: Child nodes
 - "Infected_Children_count", functional.
20. [Ancillaries return](#)
- a) Description:
 - b) Type: Probabilistic: Normal distribution (see Table III).
 - c) Dependence info: Child nodes
 - "Infected_Ancillaries_count", functional.
 - "Persons_in_Schools_total", functional.
21. [Teachers return](#)
- a) Description:
 - b) Type: Probabilistic: Normal distribution (see Table III).
 - c) Dependence info: Child nodes
 - "Persons_in_Schools_total", functional.
 - "Infected_Teachers_count", functional.

III. Simulation values

Node name	Mean	Std	5%-quantile	50%-quantile	95%-quantile
<u>Adult child prevalence ratio</u>	2.34e+0	7.77e-1	1.30e+0	2.22e+0	3.78e+0
<u>Adult prevalence</u>	6.60e-4	3.30e-4	2.71e-4	5.90e-4	1.28e-3
<u>Ancillaries return</u>	6.99e+4	7.00e+3	5.84e+4	6.99e+4	8.15e+4
<u>Teachers return</u>	2.35e+5	2.35e+4	1.96e+5	2.34e+5	2.73e+5
<u>Ancillaries total</u>	7.00e+4	0.00e+0	7.00e+4	7.00e+4	7.00e+4
<u>Children prevalence</u>	3.12e-4	1.94e-4	1.03e-4	2.65e-4	6.80e-4
<u>Children returning count</u>	8.55e+5	0.00e+0	8.55e+5	8.55e+5	8.55e+5
<u>Infected Ancillaries count</u>	4.62e+1	2.37e+1	1.85e+1	4.11e+1	9.10e+1
<u>Infected Children count</u>	2.67e+2	1.66e+2	8.84e+1	2.26e+2	5.81e+2
<u>Infected Persons in Schools total</u>	4.68e+2	2.55e+2	1.79e+2	4.11e+2	9.51e+2
<u>Infected Teachers count</u>	1.55e+2	7.95e+1	6.22e+1	1.38e+2	3.05e+2
<u>Infection prevalence in Schools</u>	4.03e-4	2.19e-4	1.54e-4	3.54e-4	8.19e-4
<u>Persons in Schools total</u>	1.16e+6	2.45e+4	1.11e+6	1.16e+6	1.20e+6
<u>Persons per School average</u>	6.91e+1	1.46e+0	6.67e+1	6.91e+1	7.16e+1
<u>Schools total</u>	1.67e+4	0.00e+0	1.67e+4	1.67e+4	1.67e+4
<u>Teachers total</u>	2.35e+5	0.00e+0	2.35e+5	2.35e+5	2.35e+5
<u>any School probability zero infections</u>	9.72e-1	1.46e-2	9.44e-1	9.75e-1	9.89e-1
<u>number of Schools with infection</u>	4.60e+2	2.44e+2	1.78e+2	4.06e+2	9.25e+2
<u>percent Schools with infection</u>	2.74e+0	1.46e+0	1.06e+0	2.42e+0	5.52e+0

Appendix 2. Effect of siblings on school infection hazard level estimates

One possible concern in relation to the occurrence of infections in schools is the role of siblings. If a child becomes infectious and has a sibling attending the same school, there is a significantly increased chance that the second sibling would also be infected. We perform a simple scoping calculation to understand the potential impact of co-infected siblings on school infection levels.

Numbers of siblings attending the same state primary school were determined by Nicole and Rabe (2014): they reported that 435,890 siblings in the age group 4 – 16 years lived at same address (including step- and half siblings). To put this in context, according to 2017 data in the ONS National Pupil Database, the total number of pupils aged 5 – 15 years was 7,100,054 pupils; the School Census reported 6,612,292 pupils for the same 2017 cohort. Thus, we adjust Nicole and Rabe's number by a factor of 10/12 to align it with the ONS age cohort. This leads to an estimated 363,242 sibling pupils in primary schools, or 5.5% of the total number of all pupils.

If we assume an average age difference of 3 years between siblings, then approximately half of all siblings will overlap with their other sibling in a primary school; we conclude that about 2.8% of primary school pupils are siblings attending the same school as their brother or sister. This means 1.4% of all pupils are one-half of a sibling pair. The question then arises, if the probability Sibling 1 is infected is identical with community prevalence, what is the probability Sibling 2 gets infected given their exposure is not an independent sample from a random community process. This probability is needed to compute the effect on infection hazard level due to having two (or more) overlapping, correlated siblings in the same school. Thus, we need to know the likelihood that a second sibling is infected GIVEN his/her sibling is infected.

If cross-sibling infection correlation were to approach 1, the effective prevalence rate for second siblings will be doubled, conditional on the first infection being a random community case. This said, if, as indicated above, only about 1.4% of all pupils are second siblings, with elevated effective prevalence rate, a small marginal increase in mean school infection hazard level is anticipated. Jing et al (2020) suggest sibling infection correlation rate is only about 0.15. Even if this value understates correlated sibling infections in British families, the "sibling effect" only impacts 95th percentile school infection rate estimates in our study.

Nicole P & Rabe (2014) Sibling spillover effects in school achievement. ISER Report No. 2014-40.

Jing et al (2020) Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. Lancet online.

Li W, Zhang B, Lu J et al The characteristics of household transmission of COVID-19. Clin Infect Dis. 2020. 2020