Measuring social networks in British primary schools through scientific engagement

A. J. K. Conlan[1],*, K. T. D. Eames[2], J. A. Gage[3], J. C. von Kirchbach[4], J. V. Ross[5], R. A. Saenz[4] and J. R. Gog[4]

[1]*CIDC, Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK*
[2]*Infectious Diseases Epidemiology Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK*
[3]*Millennium Mathematics Project, and* [4]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK*
[5]*School of Mathematical Sciences, North Terrace Campus, The University of Adelaide, Adelaide, SA 5005, Australia*

Primary schools constitute a key risk group for the transmission of infectious diseases, concentrating great numbers of immunologically naive individuals at high densities. Despite this, very little is known about the social patterns of mixing within a school, which are likely to contribute to disease transmission. In this study, we present a novel approach where scientific engagement was used as a tool to access school populations and measure social networks between young (4–11 years) children. By embedding our research project within enrichment activities to older secondary school (13–15) children, we could exploit the existing links between schools to achieve a high response rate for our study population (around 90% in most schools). Social contacts of primary school children were measured through self-reporting based on a questionnaire design, and analysed using the techniques of social network analysis. We find evidence of marked social structure and gender assortativity within and between classrooms in the same school. These patterns have been previously reported in smaller studies, but to our knowledge no study has attempted to exhaustively sample entire school populations. Our innovative approach facilitates access to a vitally important (but difficult to sample) epidemiological sub-group. It provides a model whereby scientific communication can be used to enhance, rather than merely complement, the outcomes of research.

Keywords: schools; social networks; scientific outreach; epidemiology

1. INTRODUCTION

Schools are widely acknowledged to play an important role in the transmission of infectious diseases within human societies [1], evidenced by the association between the reproductive ratio of epidemic spread and the timing of school closures [2,3]. Despite this, surprisingly little quantitative data have been gathered on the social patterns of mixing within a school. In particular, no studies have attempted to gather network measures that are likely to be associated with the transmission of close contact infections such as measles, chicken pox and influenza from primary-school-aged children. Theoretical models for disease transmission based on divergent, even antithetical, assumptions concerning the heterogeneity of social contacts can nonetheless produce similar disease dynamics and equally well describe the available sources of epidemiological data [4]. However, the predictions of such models about the impact of

control measures such as vaccination can be as divergent as their underlying assumptions [3,5–7]. Through quantifying the patterns of social mixing within schools, we set out to provide an empirical base for the refinement of mechanistic transmission models, ultimately seeking to improve predictions for assessing potential control measures on epidemics of novel pathogens.

In recent years, questionnaire-based methods have achieved great success in quantifying the social patterns of mixing pertinent to the spread of infectious diseases [8,9]. The structure of children's social networks have also received a great deal of interest in the social sciences [10–14], particularly with respect to the influence of peer networks on educational achievement [15]. However, studies have had limited success in obtaining information about school-age children, having either small sample sizes or low response rates, or relying on parents to report their children's mixing patterns. A key challenge to collecting such data is obtaining access to the populations of interest and securing the active participation of the subjects, particularly young children. Care must also be taken in the design and interpretation of questionnaires for young children. Young children respond to questioning

in a markedly different way from adults [16], with the accuracy of their responses related in a complex way to age, cognitive development, context and the nature of the subject [17]. Several indirect methods have been developed to circumvent these issues, from using mobile phone data [18,19] to using data from websites that exploit a community ethos [20]. In this paper, we present results from a study taking a novel approach in which we used a scientific outreach programme to measure social networks in British primary schools (age 4–11).

We obtained access to our study population through the provision of enrichment material to older secondary school children—effectively using scientific engagement as a research tool. This methodology has allowed us to collect a dataset, which we believe constitutes the most detailed and rich picture of the social mixing patterns of school-age children collected to date. Although our primary motivation has been the measurement of social network measures associated with the risk of disease transmission, we hope that the methodology and dataset will be of significant interest beyond our own field to researchers within the social science and educational communities.

Epidemiologically, schools bring together large numbers of immunologically naive hosts at high density. Schools therefore constitute a core group with the potential to drive the transmission of infectious disease more widely within a community. The classic example of this is the dynamics of childhood diseases, most famously measles, where the timing of the opening and closing of school terms has a demonstrable impact on transmission rates within the community at large [2,21]. However, the importance of schools is not confined purely to diseases of childhood. Children are believed to play a key role in the transmission of acute respiratory illness [22], and in particular influenza [23,24]. School closures [25] and the targeted delivery of vaccine and anti-viral drugs to school children were much discussed during the recent H1N1 pandemic [26–28], with the former carried out in Mexico and Hong Kong [29,30].

Predictions for the efficacy of such interventions vary widely between different theoretical models [3,5–7], in part owing to our lack of understanding of social patterns of mixing that underlie the patterns of transmission we see in epidemiological data. A similar issue arises in the estimation of the basic reproductive ratio, R_0, from serological data. R_0, defined as the expected number of secondary cases in a fully susceptible population resulting from the introduction of a single infectious individual, can in principle be inferred from the age profile of susceptibility for endemic diseases [1]. However, in the absence of quantitative information on the structure of mixing between different age groups, the same serological data can support estimates of R_0 that differ by several orders of magnitude [4,31]. Thus, a key piece of missing information is a detailed quantitative description of how different age groups interact.

Recent large-scale questionnaire studies have provided the first quantitative data that begin to address these problems at a population level [9]. Here, we focus on the key epidemiological unit of primary schools. We apply techniques of social network analysis [10,11,32], with a particular emphasis on estimating how mixing differs within and between different classrooms in the schools.

Network methods have a long history of use in sociology and epidemiology [32,33]. Epidemiologically, networks have been particularly well used in understanding the spread of sexually transmitted infections [34,35], but they have also been applied to situations as diverse as tuberculosis [36], influenza [37], foot and mouth disease [38] and obesity [39]. Networks can be used to visualize the connections within a population; they help identify interactions and individuals key to the spread of infection; and they highlight behavioural heterogeneities and patterns of mixing between different population subgroups. High levels of social clustering within networks can slow the spread of infection, with implications for viral evolution [40]. Detailed network data are time-consuming to collect, but as studies continue to probe into the details of social interactions a greater amount of detail is emerging about the structure of human mixing patterns.

2. SCIENTIFIC OUTREACH AS A RESEARCH TOOL

Recruitment of schools took place in association with the established Motivate Project, part of the Millennium Mathematics Project at the University of Cambridge. Motivate runs a mathematical enrichment programme for schools, enhancing the mathematics curriculum through video conferences between researchers and schools. This approach allows researchers to interact with several schools simultaneously, and enables schools to share ideas quickly and easily. In the project described here, we worked with two groups of secondary schools over two academic years (2007–2008 and 2008–2009).

We developed a series of videoconferences, taking place over the course of an academic year, to share with schools the use of mathematics in understanding disease spread, and the methods and ethics of data collection and analysis. These video conferences allowed school pupils to explore simple epidemic models and to learn about some of the challenges involved in epidemic modelling. Central themes of these video conferences were the importance of primary school children in infection spread and the need for accurate social mixing data.

Over the course of the video conferences, the research team worked with the secondary school students to design a questionnaire that could be used to measure social mixing patterns in primary schools. The brief was to produce a questionnaire that could collect useful information and which was sufficiently straightforward that it could quickly be explained even to the youngest primary school children, the results of which would be discussed and analysed in further video conferences. Ideas were brought together over the course of several video conferences before a questionnaire design was finalized by the research team and consistently applied by all of the schools. The final questionnaire is included as electronic supplementary material.

(a) *Structure of the questionnaire*
The questionnaire asked primary school pupils a small number of questions about themselves (age, gender), their household (number and age of siblings) and their social contacts. Here, we present an analysis of the within-school mixing patterns collected through the survey, focusing in particular on changes with age and differences between boys and girls. This network analysis
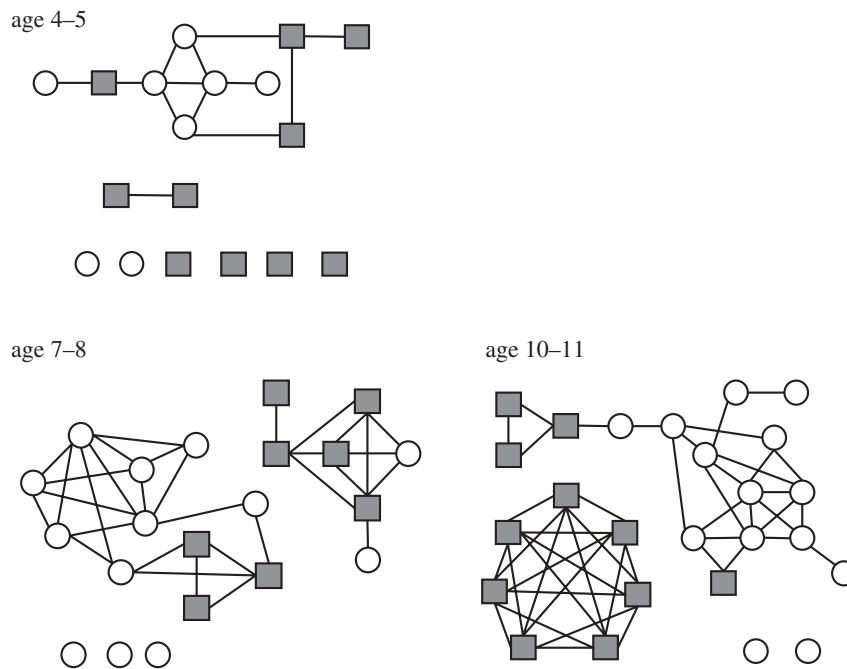
Figure 1. Sample networks of self-reported social contacts within a classroom. Networks of mutually reported social contacts for three representative classrooms taken from the same primary school demonstrating increasing complexity, cliquishness and gender segregation with age. Filled squares, girls; open circles, boys.

is based on the children's reported social contacts in answer to two key questions:

— 'Which pupils in your class do you spend the most time with?'
— 'Which pupils in other classes in the school do you spend the most time with?'

Students were asked to name up to six within-class contacts and up to four outside-class contacts.

(b) *Recruitment of the study population*

Six British secondary school groups (age 13–15) took part in the project over two academic years (2007–2008 and 2008–2009). The secondary school groups recruited local primary schools to be surveyed using schools with which they already had established links. Primary schools were therefore recruited on the basis of convenience, in terms of access and proximity to the participating secondary schools, rather than constituting a random selection. Likewise, recruitment of secondary schools was dependent on application by the school rather than selection by the research team. However, the use of video conferencing technology and a track record in recruiting schools to take part in mathematical enrichment events provided by the Motivate Project (one of the programmes that comprise the Millennium Mathematics Project, based on the Mathematics and Education Faculties of the University of Cambridge) enabled the recruitment of schools over a large geographical area and from a range of different communities. Participating schools ranged from small towns in rural Wales to inner-city schools in London and the north of England.

The secondary school groups taking part in this project obtained permission from head teachers to carry out the study in their local 'feeder' primary schools. Each secondary school group had to fit the project around the

constraints of their normal school curriculum, so the time available was limited. The secondary school students were encouraged to involve one or more local primary schools and to recruit, within those primary schools, a sample of classes that covered a range of year groups. Some larger primary schools contained several classes within each year group; in such cases, secondary schools were encouraged to recruit at least one class from each year group to take part. Informed consent was obtained from the parents of all primary and secondary pupils involved in the study, and from the pupils themselves. Video conferences were also held with many of the primary school children participating, both as science outreach and to explain the potential importance of the data they were being asked to provide. All data were anonymized prior to analysis.

The secondary school participants recruited a total of 11 primary schools, visited each school to explain the project to the primary school children and helped them to complete the questionnaires. A total of 1685 completed questionnaires were received from these primary schools, representing an 89.2 per cent recruitment rate, defined as the fraction of those primary school pupils whose classes were recruited that returned the questionnaires. Missing data were generally as a result of absence from school on the day the study took place.

3. RESULTS

Figure 1 shows sample networks collected from three classes within the same primary school, consisting of children in the age ranges of 4–5, 7–8 and 10–11. We plot a link between two individuals only if each names the other (so-called 'mutual links'). The data collected could be used to generate directed networks. However, these can be challenging to visualize clearly and are likely to be more susceptible to biases in the reporting of contacts
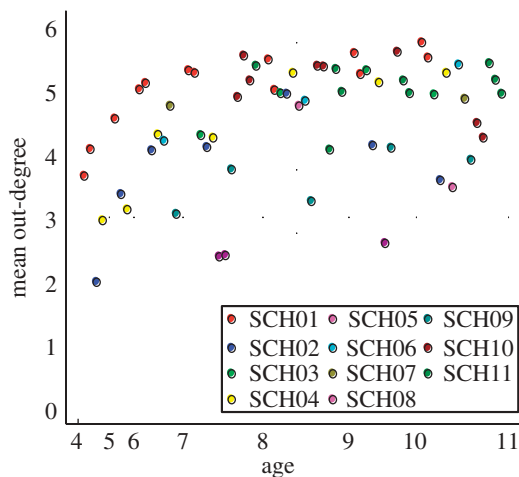
Figure 2. Mean out-degree (number of within-class contacts named per person) against age of school class. Data are plotted for each class in the survey, with the colour representing the school. The sample contains several classes within each age group, and the ordering of points within age groups is not significant: all classes plotted between ages 4 and 5, for example, have pupils aged between 4 and 5.

(an issue we shall discuss further below). Figure 1 displays several striking properties that we shall consider in turn: a large number of mutual links, segregation between boys and girls, and a high level of social clustering.

(a) *Mutual links*

A central issue with the interpretation of self-reported social contacts is quantifying the extent to which individuals can accurately identify their social contacts. The ability of young children to provide accurate information under questioning has received a great deal of attention, in particular with respect to assessing the validity of children's testimony in court cases [41]. On interview, young children often demonstrate a susceptibility to suggestion [41] and a tendency to manufacture responses to 'unanswerable' questions [42]. The forms of the questions posed are particularly important. Young children are less likely to provide fanciful answers to 'open' questions rather than those that are 'closed' categorically (yes/no) [42]. For simple objective questions, children are potentially capable of achieving similar levels of accuracy to adults [43], particularly with respect to objects or events that they have a particular interest in [41].

As previously described, rather than specifying a fixed number of choices or selecting contacts from a list, students were asked to name up to six within-class contacts and up to four outside-class contacts. If the true number of contacts exceeds these enforced limits, then the reported contacts will be censored. However, imposing no limit on the number of contacts introduces the risk that peer pressure may result in 'competitive naming' of contacts in order to appear more popular. There is also the possibility that 'closing' the question in this way may have encouraged respondents to provide additional spurious names beyond those they would subjectively rank as their *closest* contacts. The final design therefore sought to achieve a compromise between limiting the impact of these two potential sources of bias: censoring and over-reporting.

Plotting the out-degree by class suggests a trend towards older respondents reporting more contacts (figure 2). However, there is considerably more variation between classes than with respect to age. Because of the potential bias introduced by censoring, these mean out-degrees should be considered as a lower bound on the true values. Variation in the interpretation of the questionnaire questions by (primary school) respondents and the emphasis placed by different (secondary school) interviewers may have also led to biases in the criteria by which individuals prioritize their social contacts. For example, children who are perceived to be popular might be expected to attract links through aspiration of the responder rather than owing to an established relationship.

We attempt to control for, and quantify, these biases by considering *mutual* links, where pairs of individuals identify each other as social contacts [11]. It would be natural to expect that reciprocated links are more representative of *relationships* rather than *popularity*. Overall, 61 per cent of contacts in the data are mutual. The fraction of mutual links increases with the age of the respondent (figure 3). Even for the youngest age groups, the fraction of mutual links is significantly greater than we would expect to see purely by chance (figure 3a), suggesting that the reported contact networks are indeed evidence of established social structure.

(b) *Gender assortativity*

The majority of the within-class networks display a clear segregation of boys and girls according to the reported social contacts—indeed, in several classes there were no reported mutual links between girls and boys. The segregation becomes more striking for the older age groups (figure 4). This observed gender segregation contributes to the high number of mutual links, but even when taking the patterns of gender-based mixing into account there are significantly more mutual links observed than would be expected (figure 3b). This gender assortativity is therefore not sufficient to explain the excess number of mutual links, indicating once more that there is additional social structure in the reported contacts.

(c) *Social clustering*

Social clustering in within-class networks was measured by the clustering coefficient ϕ, defined as the number of triples divided by the number of triangles (or, more simply, 'the probability that two of my contacts know each other'). ϕ is observed to increase with age (figures 1 and 5). Given the high proportion of mutual links and gender assortativity, it is perhaps unsurprising that the within-class networks exhibit large numbers of 'cliques' (figure 1). To explore the extent to which clustering is driven by these known factors, we generated distributions of the clustering coefficient, ϕ, from random networks containing the same number of mutual links and the same patterns of gender-based mixing [33]. For the majority of our within-class networks, ϕ is significantly greater than is seen in these simulated networks (figure 5), suggesting that even these strong observed heterogeneities are not sufficient to explain the observed social structures.
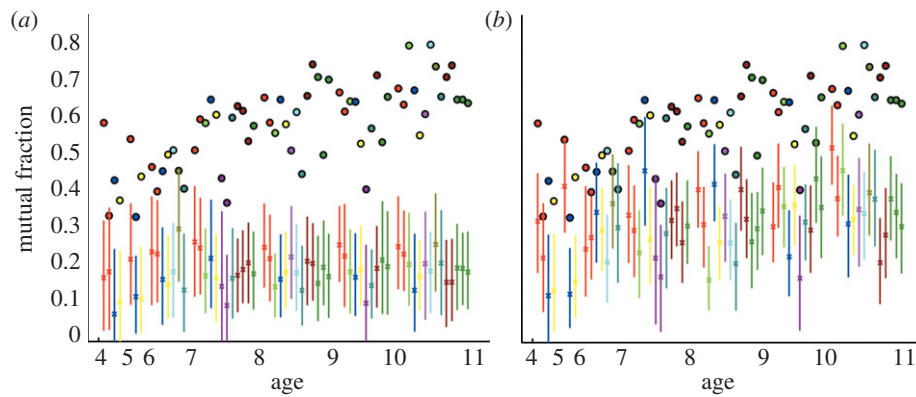
Figure 3. Fraction of mutual links against age of school class. A comparison of the data (circles) with the output of simulated networks of contacts within each class (mean (crosses) and central 99 percentiles (lines)). Colour scheme as in figure 2. (*a*) Simulated networks contain the observed number of individuals each with their observed out-degree. (*b*) Simulated networks include gender information; each contains the observed number of individuals with their observed number of contacts with boys and with girls (see Appendix for further details on generation of simulated networks).
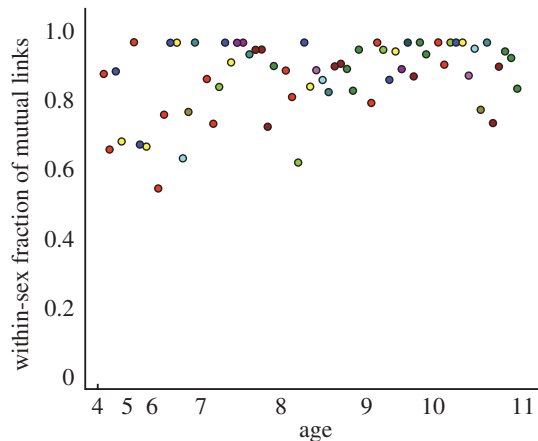

Figure 4. Gender assortativity against age of school class. The fraction of mutual links that connect individuals of the same gender. Colour scheme as in figure 2.

Figure 5. Clustering against age of school class. Clustering within the network of mutual links, measured by the clustering coefficient ϕ. A comparison of the data (circles) with the output of simulated networks of contacts within each class (mean (crosses) and central 99 percentiles (lines)). Colour scheme as in figure 2. Simulated networks contain the observed number of mutual links and the observed patterns of gender mixing (see Appendix for further details on generation of simulated networks).

(d) *Contacts between age groups*

Finally we consider the patterns of mixing between different classrooms. Because of the limits placed on the number of contacts listed, we cannot compare the relative proportions of within-class contacts to between-class contacts. However, in schools with more than one class in each year group, we can compare the fraction of reported between-class contacts that are within-year and between-year. Mixing between age groups is strongly assortative, with approximately 80 per cent of between-class contacts being reported within age group (figure 6).

4. DISCUSSION

We have successfully applied simple questionnaire-based methods to measure social networks within complete primary school classes, and in some cases complete schools, in Britain. The networks of reported social contacts demonstrate a great deal of complex structure, with important implications for the spread of infectious diseases and the interpretation of epidemiological data. There is a striking segregation between the genders, with very little reported social communication between boys and girls, and tight, internally connected cliques
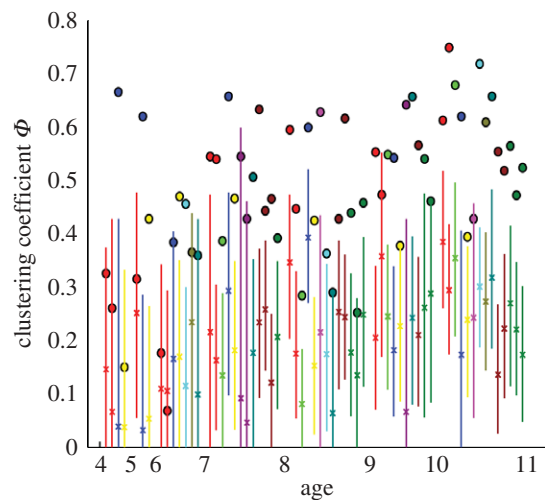
that become more evident in older groups. Sociological studies have consistently found similar patterns of sex bias and clustering in school peer networks [10–15]. However, no study to our knowledge has attempted to survey exhaustively entire school populations or considered the implications of detailed school mixing patterns for disease transmission.

Interpreting social contact networks in an epidemiological context can be delicate as there is no clear definition of what is likely to constitute an 'infectious contact'. The reported social contacts within our study are based on the qualitative judgement of the respondent as to 'which pupils . . . you spend the most time with'. We chose our question based on the assumption that the contacts children spend most time with are those for which there is the greatest potential for the transmission of disease. It would be natural to expect that a child's reported

	to						
from \	R	Y1	Y2	Y3	Y4	Y5	Y6
R	74.0	4.0	6.0	3.0	5.0	3.0	5.0
Y1	5.5	62.0	11.0	6.2	4.1	6.8	4.8
Y2	12.0	16.0	33.0	15.0	9.5	6.8	7.5
Y3	2.6	2.1	3.2	70.0	6.1	6.6	9.5
Y4	2.4	2.2	2.7	3.9	74.0	7.0	8.1
Y5	2.2	4.3	3.2	3.0	4.0	79.0	4.5
Y6	0.6	1.0	0.8	2.6	1.9	1.1	92.0

Figure 6. Reported contacts between cohorts. Reported contacts between different classes were classified according to the year group (R = reception, Y1 = year 1, etc.) of the student filling in the questionnaire ('from') and the year group of the named contact ('to'). These were normalized by the total number of reported contacts from each year group and are presented here as inset percentages; darker shading corresponds to a higher percentage. The data presented here are limited to the schools for which each year has multiple classes. For all years, the majority of reported contacts external to the respondents' own class are nonetheless within the same school year. Reported contacts are asymmetric, with a greater proportion of reported contacts directed to year groups senior to the respondent.

contact may vary depending on the semantic phrasing of this question. The fact that such a large number of the contacts were mutual links suggest that there was a high level of consistency in the interpretation of the question.

Not all primary school students could complete the questionnaire with equal ease; older primary school children were generally able to fill out the questionnaire with minimal assistance, but the youngest age groups required more individual attention from the secondary school students. By confining our analysis to mutually reported contacts, we hope to limit the potential for variations in the implementation of the questionnaire and subsequent interpretation by the respondents to bias our results. However, we must keep these subjective issues in mind when interpreting these data. For example, we cannot distinguish whether the increase in the fraction of mutual links with age is evidence of an increase in social structure or in the accuracy of respondents' ability to accurately report their social contacts. The reported number of contacts between age groups suggests that there are some age-dependent biases in reporting. Overall, children within the study were more likely to report contacts in older year groups rather than younger (figure 6). The question as to whether this was because of an aspirational selection of 'older' contacts (which was not reciprocated) or merely a function of the age-dependent accuracy of self-reported contacts would be an intriguing avenue for further study.

The phrasing of our question is subtly distinct from those previously used in social network studies in schools. From a sociological perspective, the duration of contacts is less interesting than understanding the nature of relationships and their formation [14]. A more typical question posed by previous studies would therefore be of the form of 'who would you *like* to spend the most time with' or 'name your *best* friend'. Although we have set out to measure a different aspect of children's social relationships, we acknowledge that a child's perception of whom they 'spend the most time with' may well be intermingled with who they 'would like to spend the most time with' [14].

These semantic issues could be addressed through gathering more detailed information than we have attempted here. For instance, respondents could be asked to classify the type of contacts in a 'contact diary'. Given our study population of mostly young children, adding extra levels of complexity to a survey would not necessarily meet with success. Ideally, it would be desirable to eliminate the subjective nature of these questions. Electronic methods of measuring social contact networks offer the potential to overcome this problem [44]. Furthermore, electronic methods could be used to quantify the structure of chance contacts on the playing fields and in the dining halls and corridors, which are likely to form an important role in disease transmission but are far less straightforward to capture through questionnaire methods. However, given the privacy issues associated with such intensive surveillance and the vulnerable nature of our study population, it is not certain that such methods would be acceptable to schools, parents or children, even though it may well form a fruitful area for further research.

The notion of a 'risky contact' varies greatly between different infectious agents. For short, acute respiratory infections, we might expect a classroom to constitute a single homogenously mixed unit even in the presence of the strong patterns of non-random mixing seen within these data. For such infections, therefore, it is the mixing between age groups that is more interesting and has not been studied to the same extent as within-class social networks. Our data verifies the long-held belief that mixing is strongly assortative between the age-groups in a school [1,45] on a finer scale than has been previously demonstrated [9].

These age-related mixing patterns are essential for the proper interpretation of epidemiological data for childhood infectious diseases such as measles, mumps and rubella, which are typified by a long-lasting immunity to reinfection after recovery. This is most easily characterized through the fundamental epidemiological concept of the basic reproductive ratio, R_0 [1]. Defined as the expected number of new infections on the introduction of a single infective agent into a fully susceptible population, R_0 sets the upper bound on the reproductive potential of a pathogen and, conversely, the effort required to bring it under control. The presence of immunity within a population serves to limit the rate of spread within a population, leading to a net, or effective, reproductive ratio (R) [1].

It is relatively straightforward to estimate R from serological data [4] or time series of case reports [2] set against the current background level of susceptibility within the community. However, if transmission rates are heterogeneous with respect to age, then the estimated value of R_0 depends critically on the architecture of the contacts between different age groups [4]. This has traditionally been accounted for by defining symmetrical

matrices that specify the form of transmission between age groups, styled as 'who aquires infection from whom' matrices [1]. In general, the more assortative that contacts are with respect to age, the larger the corresponding value for R_0 will be, based on the same value of R [31]. Our data provide a quantitative basis to support the selection of contact matrices and thus, hopefully, to improve predictions for R_0 for this important class of human diseases.

For infections transmitted through close social contacts, the clustering of individuals' contacts within the classroom is likely to become more important. Once again, the importance of these patterns lies in the interpretation of epidemiological data rather than in general terms of the rate of spread of an infection. An obvious example for further study would be the prevalence of head-lice within schools, which are notably found at higher rates in girls when compared with boys [46]. Given that intimacy is likely to aid the transmission of head-lice, the pronounced sex bias in reported contacts reported here might be relevant to understanding the transmission dynamics of ectoparasites in a school classroom. However, the signature for this would lie in the variability of the sex bias and the persistence of disease between different outbreaks rather than in a systematic bias.

The data obtained by this project required the investment of significant time and effort, on behalf of both the research team and the secondary school participants and their teachers. Indeed, this project was only possible owing to the involvement of large numbers of secondary school students in the motivation, design and implementation of the study. In itself, this may be considered to be a major limitation of the study. School children cannot be expected to rigorously apply a questionnaire or, in the case of the youngest primary school children, perform the same structured interview technique for every respondent. This introduces the possibility that variations between the different secondary school groups involved in the project may have led to unquantifiable biases in the collected data.

However, during the development of the project it became clear that performing the study in collaboration with secondary school students contributed several major advantages over more traditional 'researcher-driven' methodologies. The most important was simply the access to the study populations themselves. Surveying children is particularly difficult, as consent must be obtained from schools and families, and time taken to explain the methodology to teachers and parents. Recruiting our study populations indirectly through a public engagement project with secondary school children greatly facilitated this process. The existing links between the secondary schools and their feeder primary schools were the essential key to providing access. Both secondary and primary schools were keen to develop these links in an innovative way, which was further enhanced through the provision of on-site visits from the research team for some of the participating schools. The pre-existing video-conferencing infrastructure provided through Motivate allowed us to brief the teachers directly, who in turn provided the logistical support to deliver and process consent forms for the study. In return, the research team provided enrichment to the national maths and science curricula for the older children, and the experience of working on a genuine research project with a prestigious university.

The secondary students provided an important source of 'local knowledge' for the design and execution of the study. In many cases, the students revisited the primary schools that they themselves had attended, so they were acquainted with the school structure and culture before embarking on the data collection. Being closer in age and background to the respondents was also beneficial for establishing a rapport with the younger children who required more individual attention. This was a particular issue for some of the inner-city schools where children came from a variety of cultural and linguistic backgrounds.

The combined distributed effort of all of the secondary school participants has provided a dataset of social networks in schools that is unrivalled in its scope, size and detail. By combining our survey method with public engagement, we have exhaustively sampled 75 complete primary school classes from 11 primary schools in Britain with a high response rate (around 90% in most schools). The unique aspect to this approach is that part of the study population themselves—in the form of the secondary school participants—were involved in the planning, collation and analysis of the data. Epidemiology is a research area that directly concerns and affects the public. Mathematical models are becoming an ever more important part of the development of public health policies for the control of emerging infectious diseases. Explaining the mechanisms and assumptions inherent in these approaches to the public at large is vital if they are to make informed decisions about these important issues. In this project, we went further and used public engagement as a *methodology* to perform our research, in the process helping to inspire the next generation of epidemiologists and mathematical modellers. The work carried out by these engaged and motivated secondary school pupils was essential to the project, without which it could not have been a success.

APPENDIX

Materials developed during the project are available at http://motivate.maths.org/content/DiseaseDynamics.

(a) *Social network analysis*

The results of the questionnaire were described using methods of social network analysis. An adjacency matrix A was calculated for each participating class, such that $A(i,j) = 1$ if person i named person j as a contact and $A(i,j) = 0$ otherwise. Networks of contacts were plotted, allowing straightforward identification of heterogeneities within the network; out-degrees and in-degrees were calculated for each

individual. The fraction of mutual links was calculated as $\sum_{i,j} A(i,j)A(j,i) / \sum_{i,j: \text{person } j \text{ completed the survey}} A(i,j)$; the restriction in the sum in the denominator exists because a mutual link can only be observed between i and j if both individuals completed the survey. Within the networks, clustering was measured using the clustering coefficient, ϕ, defined as the number of triples divided by the number of triangles (or, more simply, 'the probability that two of my contacts know each other').

To explore the extent to which the measured networks differed from random graphs, simulated networks were generated. These simulated networks contained links placed at random, subject to certain constraints. When looking at the number of mutual links (figure 3a), the simulated network matched each node's observed out-degree, and, when considering gender mixing (figure 3b), each node's number of contacts with boys and with girls. When considering clustering (figure 5), the simulated networks contained the observed number of mutual links both within and between genders—in this instance, simulated networks contained only individuals who completed the survey (since non-responders could not, by definition, have any mutual links). Simulated networks were built link by link. When matching mutual links, pairs of individuals were picked at random and, when matching to out-degree, potential contacts were chosen at random for each individual one by one. In both cases, proposed links were only accepted provided that they did not violate the properties of the given data. Measures of network statistics—number of mutual links and clustering coefficient—were obtained from these simulated networks and compared with the values observed in the data. For each class and each comparison, 1000 simulated networks were generated.

(b) *Study population*

The complete dataset of reported contacts is presented as electronic supplementary material to this paper. Tables S1–S12 in the electronic supplementary material summarize this data in terms of the range, size and response rates for the schools and classes sampled over the 2 years of the project. A small number of mixed year group classes (one aged 5–7, two aged 7–9, three aged 9–11) were sampled, and these have been classified within the upper categories in the electronic supplementary material, tables S1 and S2. Some classes, particularly in schools 5, 7 and 8, were unusually small or poorly sampled. Six classes with fewer than 16 pupils and six classes with a response rate of lower than 75 per cent have been considered as outliers and omitted for the analyses presented in this paper; these 12 classes are spread across all age groups, with one aged 4–5, three aged 5–6, one aged 6–7, three aged 7–8, one aged 8–9 and three aged 9–10. These outliers are identified in tables S3–S12 of the electronic supplementary material. Versions of figures 2–5 including these outliers are also presented in the electronic supplementary material. Excluding the 12 outliers reduces the sample size to 1521 pupils, with a response rate of 92.4 per cent.

REFERENCES

1   Anderson, R. & May, R. 1991 *Infectious diseases of humans: dynamics and control.* New York, NY: Oxford University Press.

2   Fine, P. E. M. & Clarkson, J. A. 1982 Measles in England and Wales—I: analysis of factors underlying seasonal patterns. *Int. J. Epidemiol.* **11**, 5–14. (doi:10.1093/ije/11.1.5)

3   Cauchemez, S., Valleron, A.-J., Boëlle, P.-Y., Flahault, A. & Ferguson, N. M. 2008 Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature* **452**, 750–755. (doi:10.1038/nature06732)

4   Wallinga, J., Lévy-Bruhl, D., Gay, N. J. & Wachmann, C. H. 2001 Estimation of measles reproduction ratios and prospects for elimination of measles by vaccination in some Western European countries. *Epidemiol. Infect.* **127**, 281–295.

5   Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C. & Burke, D. S. 2006 Strategies for mitigating an influenza pandemic. *Nature* **442**, 448–452. (doi:10.1038/nature04795)

6   Milne, G. J., Kelso, J. K., Kelly, H. A., Huband, S. T. & McVernon, J. 2008 A small community model for the transmission of infectious diseases: comparison of school closure as an intervention in individual-based models of an influenza pandemic. *PLoS ONE* **3**, e4005. (doi:10.1371/journal.pone.0004005)

7   House, T. *et al.* 2009 Can reactive school closures help critical care provision during the current influenza pandemic? *PLoS Curr. Influenza* **23**, RRN1119.

8   Edmunds, W. J., O'Callaghan, C. J. & Nokes, D. J. 1997 Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc. R. Soc. Lond. B* **264**, 949–957. (doi:10.1098/rspb.1997.0131)

9   Mossong, J. *et al.* 2007 Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74. (doi:10.1371/journal.pmed.0050074)

10   Moreno, J. L., Jennings, H. H. & Stockton, R. 1943 Sociometry in the classroom. *Sociometry* **6**, 425–428. (doi:10.2307/2785221)

11   Ford, J. 1969 *Social class and the comprehensive school.* London, UK: Routledge & Kegan Paul Limited.

12   Nash, R. 1973 Clique formation among primary and secondary school children. *Br. J. Sociol.* **24**, 303–313. (doi:10.2307/588234)

13   Thorne, B. 1992 Girls and boys together… but mostly apart: gender arrangements in elementary schools. In *Education and gender equality* (ed. J. Wrigley), pp. 115–130. London, UK: The Falmer Press.

14   Gifford-Smith, M. E. & Brownell, C. A. 2003 Childhood peer relationships: social acceptance, friendships and peer networks. *J. School Psychol.* **41**, 235–284. (doi:10.1016/S0022-4405(03)00048-7)

15   Feiring, C. & Lewis, M. 1991 The development of social networks from early to middle childhood: gender differences and the relation to school competence. *Sex Roles* **25**, 237–253.

16   Aronsson, K. & Hundeide, K. 2002 Rational rationality and children's interview responses. *Hum. Dev.* **45**, 174–186. (doi:10.1159/000057073)

17   Goodman, G. S. & Schwartz-Henney, B. M. 1992 Why knowing a child's age is not enough: influences of cognitive, social and emotional factors on children's testimony. In *Children as witnesses* (eds H. Dent & R. Flin), pp. 15–32. Chichester, UK: John Wiley & Sons Ltd.

18   González, M. C., Hidalgo, C. A. & Barabási, A.-L. 2008 Understanding individual human mobility patterns. *Nature* **453**, 779–782. (doi:10.1038/nature06958)

19   Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. 2010 Limits of predictability in human mobility. *Science* **327**, 1018–1021. (doi:10.1126/science.1177170)

20   Brockmann, D., Hugnagel, L. & Geisel, T. 2006 The scaling laws of human travel. *Nature* **439**, 462–465. (doi:10.1038/nature04292)

21 Earn, D. J. D., Rohani, P., Bolker, B. M. & Grenfell, B. T. 2000 A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667–670. (doi:10.1126/science.287.5453.667)

22 Monto, A. S. 1994 Studies of the community and family: Acute respiratory illness and infection. *Epidemiol. Rev.* **16**, 351–373.

23 Viboud, C., Boëlle, P., Cauchemez, S., Lavenu, A., Valleron, A., Flahault, A. & Carrat, F. 2004 Risk factors of influenza transmission in households. *Int. Congr. Ser.* **1263**, 291–294. (doi:10.1016/j.ics.2004.01.013)

24 Reichert, T. A., Sugaya, N., Fedson, D. S., Glezen, W. P., Simonsen, L. & Tashiro, M. 2001 The Japanese experience with vaccinating schoolchildren against influenza. *N. Engl. J. Med.* **344**, 889–896. (doi:10.1056/NEJM200103223441204)

25 Hatchett, R. J., Mecher, C. E. & Lipsitch, M. 2007 Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc. Natl Acad. Sci. USA* **104**, 7582–7587. (doi:10.1073/pnas.0610941104)

26 Cauchemez, S., Ferguson, N. M., Wachtel, C., Tegnell, A., Saour, G., Duncan, B. & Nicoll, A. 2009 Closure of schools during an influenza pandemic. *Lancet Infect. Dis.* **9**, 473–481. (doi:10.1016/S1473-3099(09)70176-8)

27 Yang, Y., Sugimoto, J. D., Halloran, M. E., Basta, N. E., Chao, D. L., Matrajt, L., Potter, G., Kenah, E. & Longini Jr, I. M. 2009 The transmissibility and control of pandemic influenza A (H1N1) virus. *Science* **326**, 729–733. (doi:10.1126/science.1177373)

28 Wallinga, J., van Boven, M. & Lipsitch, M. 2009 Optimizing infectious disease interventions during an emerging epidemic. *Proc. Natl Acad. Sci. USA* **107**, 923–928. (doi:10.1073/pnas.0908491107)

29 Stern, A. M. & Markel, H. 2009 What Mexico taught the world about pandemic influenza preparedness and community mitigation strategies. *JAMA* **302**, 1221–1222. (doi:10.1001/jama.2009.1367)

30 Koonin, L. M. & Cetron, M. S. 2009 School closure to reduce influenza transmission. *Emerg. Infect. Dis.* **15**, 137–138. (doi:10.3201/eid1501.081289)

31 Diekmann, O., Heesterbeek, J. A. P. & Metz, J. A. J. 1990 On the definition and computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382.

32 Wasserman, S. & Faust, K. 1994 *Social network analysis: methods and applications*. Cambridge, UK: Cambridge University Press.

33 Keeling, M. J. & Eames, K. T. D. 2005 Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307. (doi:10.1098/rsif.2005.0051)

34 Klovdahl, A. S., Potterat, J. J., Woodhouse, D. E., Muth, J. B., Muth, S. Q. & Darrow, W. W. 2002 Social networks and infectious disease: the Colorado Springs study. *Soc. Sci. Med.* **38**, 79–88. (doi:10.1016/0277-9536(94)90302-6)

35 Bearman, P. S., Moody, J. & Stovel, K. 2004 Chains of affection: the structure of adolescent romantic and sexual networks. *AJS* **110**, 44–91.

36 Andre, M., Ijaz, K., Tillinghast, J. D., Krebs, V. E., Diem, L. A., Metchock, B., Crisp, T. & McElroy, P. D. 2007 Transmission network analysis to complement routine tuberculosis contact investigations. *AJPH* **97**, 470–477. (doi:10.2105/AJPH.2005.071936)

37 Eames, K. T. D., Webb, C., Thomas, K., Smith, J., Salmon, R. & Temple, J. M. F. 2010 Analysing the role of contact tracing in a suspected H7N2 influenza A outbreak in humans in Wales. *BMC Infect. Dis.* **10**, 141. (doi:10.1186/1471-2334-10-141)

38 Kao, R. R., Green, D. M., Johnson, J. & Kiss, I. Z. 2007 Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *J. R. Soc. Interface* **4**, 907–916. (doi:10.1098/rsif.2007.1129)

39 Christakis, N. A. & Fowler, J. H. 2007 The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370–379. (doi:10.1056/NEJMsa066082)

40 Read, J. M. & Keeling, M. J. 2003 Disease evolution on networks: the role of contact structure. *Proc. R. Soc. Lond. B* **270**, 699–708. (doi:10.1098/rspb.2002.2305)

41 Dent, H. 1992 The effects of age and intelligence on eye-witnessing ability. In *Children as witnesses* (eds H. Dent & R. Flin), pp. 1–13. Chichester, UK: John Wiley & Sons Ltd.

42 Waterman, A. H., Blades, M. & Spencer, C. 2001 Interviewing children and adults: the effect of question format on the tendency to speculate. *Appl. Congit. Psychol.* **15**, 521–531.

43 Goodman, G. S. & Reed, R. S. 1986 Age differences in eyewitness testimony. *Law Hum. Behav.* **10**, 317–332. (doi:10.1007/BF01047344)

44 Nicolai, T., Yonkei, E., Behrens, N. & Kenn, H. 2006 Exploring social context with the wireless rope. In *Lecture notes in computer science: on the move to meaningful internet systems*, pp. 874–883. Berlin, Germany: Springer.

45 Fine, P. E. M. & Clarkson, J. A. 1982 Measles in England and Wales—II: the impact of the measles vaccination programme on the distribution of immunity in the population. *Int. J. Epidemiol.* **11**, 15–25. (doi:10.1093/ije/11.1.15)

46 Counahan, M., Andrews, R., Bütner, P., Byrnes, G. & Speare, R. 2004 Head lice prevalence in primary schools in Victoria, Australia. *J. Paediatr. Child. Health* **40**, 616–619. (doi:10.1111/j.1440-1754.2004.00486.x)