Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections

W. J. EDMUNDS[1], C. J. O'CALLAGHAN AND D. J. NOKES

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK
(edmunds@oikos.warwick.ac.uk)

SUMMARY

Although mixing patterns are thought to be important determinants of the spread of airborne infectious diseases, to our knowledge, there have been no attempts to directly quantify them for humans. We report on a preliminary study to identify such mixing patterns. A sample of 92 adults were asked to detail the individuals with whom they had conversed over the period of one, randomly assigned, day. Sixty-five (71%) completed the questionnaire, providing their age, the age of their contacts and the social context in which the contacts took place. The data were analysed using multilevel modelling. The study identified, and allowed the quantification of, contact patterns within this sample that may be of epidemiological significance. For example, the degree of assortativeness of mixing with respect to age was dependent not only on the age of participants but the number of contacts made. Estimates of the relative magnitude of contact rates between different social settings were made, with implications for outbreak potential. Simple questionnaire modifications are suggested which would yield information on the structure and dynamics of social networks and the intensity of contacts. Surveys of this nature may enable the quantification of who acquires infection from whom and from where.

1. INTRODUCTION

Theory and observation have demonstrated the importance of mixing patterns on the transmission dynamics of respiratory infections such as measles and rubella (Fine & Clarkson 1982; Anderson & May 1985; Anderson & Grenfell 1986), influenza (Cliff & Haggett 1992) and of sexually transmitted diseases (STD) (Anderson *et al.* 1990; Gupta *et al.* 1989; Boily & Anderson 1991; Garnett & Anderson 1993, 1996). Arising from this is a requirement to quantify these mixing patterns if we are to fully understand the spread of these infectious agents. In the field of STD infection and control considerable effort has been put into directly measuring the rates of change of sexual partnerships and the structure and dynamics of sexual contact networks (Haraldsdottir *et al.* 1992; Woodhouse *et al.* 1994; Johnson *et al.* 1994; Garnett *et al.* 1996). Simultaneously, this has led to the development of suitable methodologies for data collection and interpretation, for example, those associated with contact tracing and questionnaire design. However, no similar effort has been channelled into directly quantifying the mixing behaviour responsible for infections transmitted by respiratory droplets or saliva, where a closeness of contact is required to effect transmission.

Instead, work has concentrated on indirectly assessing the patterns of contacts within and between age groups (in the form of a 'who acquires infection from whom' (WAIFW) matrix (Anderson & May 1985, 1991)). This technique involves defining an $n \times n$ matrix, the elements of which represent effective contacts between individuals in age group i and age group j (where effective contacts are those that are likely to result in the transfer of infection). Estimates of the effective contact rates are then derived from estimates of the instantaneous per-susceptible infection rate (force of infection) for each of the n age groups. There are severe limitations with this process. First, the method requires the investigator to make some pre-judgement on the mixing structure (see Anderson & May 1991). Second, the number of unique mixing rates is constrained by the number of age groups for which estimates of the force of infection have been made (that is, n rather than $n \times n$). This results in severe constraints on the matrix, such that the assumed structure is unlikely to accurately represent the true contact patterns in the community. Third, because the data used are typically the cumulative history of exposure (Grenfell & Anderson 1985), it is difficult to estimate effective contact rates when there are few susceptibles remaining. For example, in the case of predominantly childhood infections there is low confidence in values estimated for adults (at the extreme, if all individuals are infected by some age, then effective contact rates cannot be calculated for individuals in higher age classes). This situation becomes important if the age distribution of the susceptible population is altered, such as results from the introduction of mass infant vaccination (Nokes & Anderson 1992), whereupon it becomes very difficult to predict accurately the implications of such a change on infection in adulthood (Gay *et al.* 1995).

There is clearly a need to obtain direct estimates of the contact behaviour important to the transmission of aerosol- or saliva-borne close contact infections. Perhaps the major reasons for the absence of direct observational data are the difficulty in defining a suitable at-risk contact and the associated methodology for collecting such data. In spite of these difficulties we feel that exploratory attempts are justified in view of the inherent problems described above. In this study we take a simple definition for an at-risk contact; that of a two way conversation. By this we assert that if two individuals are close enough to have a conversation then they are probably close enough to transmit at least some of the common airborne infections. As will be discussed later, it is unlikely that this definition is suitable for all airborne infections of interest. However, it serves as a starting point, the advantage of which is that the concept of a conversation is well understood, easy to recall and record, and thus possible to collect from a study population.

The level of detail of interaction we attempt to collect in this study reflects some aspects of behaviour thought to be important in determining patterns of infection of typical childhood infections such as measles and rubella. We are therefore interested in variation in mixing within and between age groups (see Anderson & May 1985, 1991; Nokes *et al.* 1986, 1990; Babad *et al.* 1995; Gay *et al.* 1995), within and between households (Longini & Koopman 1982) and weekday to weekend variation (see Babad *et al.* 1995). Although also of importance (Fine & Clarkson 1982), we have not attempted to evaluate seasonal variation in contact rates. Such behaviours shown by school children are thought to dominate observed patterns of transmission for infections such as measles prior to mass vaccination. However, the introduction of mass vaccination against measles, mumps and rubella has resulted in marked changes in the age structure of susceptibility and infection and there is currently much interest in defining more accurately patterns of contact in the young adult age groups, e.g. university populations (Gay *et al.* 1995; Babad *et al.* 1995).

The aims of this study are twofold. First, to determine if the collection of direct observational data on contacts within an adult study population is practicable and second to assess if such information lends itself to meaningful analysis with results that illuminate the patterns of mixing which may influence airborne infection patterns. It is emphasized that this is a pilot study from which we hope to refine questionnaire design and methods of collection and analysis.

2. METHODS

(*a*) *Study participants*

A convenience sample of staff and students at two British universities and their family and friends was used. In total, 92 individuals were asked to complete the form and return them by post to the principal investigator. The survey was anonymous and no record of who returned the forms could be made.

(*b*) *Questionnaire*

The aims of the research were explained to the study participants. They were asked to complete a questionnaire which detailed their living arrangements (i.e. flat, bedsit, family home, etc.), how many people lived with them (defined as 'shared a kitchen with' them) and what their ages were. They were then asked to note down the age of each of the people they had a conversation with over the period of one, randomly assigned, day and in which social context that conversation took place (i.e. work/college, home, social, travel, etc.). If the study participant held a conversation with the same person in more than one of these contexts, then they were asked to record all of these contexts. If the study participants did not know the exact age of their contacts they were asked to provide an age range. The midpoint of this range was used for the purpose of analysis. The gender of the participants and their contacts was not recorded as serological profiles of airborne infections do not show a sex-related difference in infection rates. The participants were encouraged to fill out the questionnaire during the day of observation and to read through the form afterwards in an attempt to minimise recall bias. They were requested to provide us with any comments and suggestions.

A contact was defined as a two way conservation (at a distance which did not require raising the voice) in which at least two words were spoken by each party and in which there was no physical barrier between the two parties (such as security screens). Note that participants were not asked to record every conversation, but the different individuals they spoke with. There was therefore no record of the number of times a participant had a conversation with a single individual over the period of a day (other than the information provided by the number of different contexts that were recorded), nor did we record the length of any conversation. A day was defined as beginning when the participant got up and ending when they went to bed (rather than running from midnight to midnight).

(*c*) *Statistical analysis*

The data were entered into EPI INFO (Dean *et al.* 1994) and descriptive analysis was performed using this package and SPSS for Windows Release 6.0 (SPSS Inc., USA). As the data were organized in a hierarchical fashion, with contacts recorded within participants, who in turn were randomly allocated to days of observation, a more detailed analysis was undertaken in which generalized linear mixed models were prepared using MLn software (Multilevel Models Project, Institute of Education, University of London). Two models were developed investigating the association of covariates hypothesized to be of importance to patterns of mixing (and recorded at the various levels of observation) with (i) the number of contacts made, and (ii) the age of those contacts. Note, as a convenience sample was obtained, inference beyond the sampled population must be made with caution. Nevertheless, we believe that this method may provide a useful insight into the mixing patterns of secondary school children and university students, staff and their associates; as a number of recent outbreaks of measles and rubella have shown, such information is of growing importance in the vaccination era (e.g. Calvert *et al.* 1994).

3. RESULTS

(*a*) *Study participants and response to the questionnaire*

Sixty-five individuals (71%) completed the questionnaire. The age range of respondents was 22–66 years (mean age 31.9 years). These individuals shared a kitchen with a total of 176 others (a mean 'household' size of 2.7). Of the 65 respondents, 35 lived in a shared flat or house, 19 reported that they lived in a family home, seven lived in halls of residence, one in a bedsit and three reported other living arrangements.

None of the responders reported any difficulties filling out the form. The most common suggestion was that the exercise could be made simpler if we had provided them with an age range (0–5 years, 5–10 years, ...) for the age of their contacts. However, with respect to the instruction to record multiple contexts for repeated contact with the same individual, we felt that many participants did not observe this request, as 95% of contacts were reported in only one context (this may reflect the actual pattern of contact, though it seems unlikely). Therefore, for the purpose of modelling, those observations which recorded more than one context were transformed to a single/primary context by application of the following rules: (i) if the contact was recorded at home and elsewhere, then it was classified as a home contact; (ii) if the contact was recorded at work and elsewhere, but not home, then it was classified as a work contact. Owing to the relatively few multiple context contacts, these two simple rules were adequate.

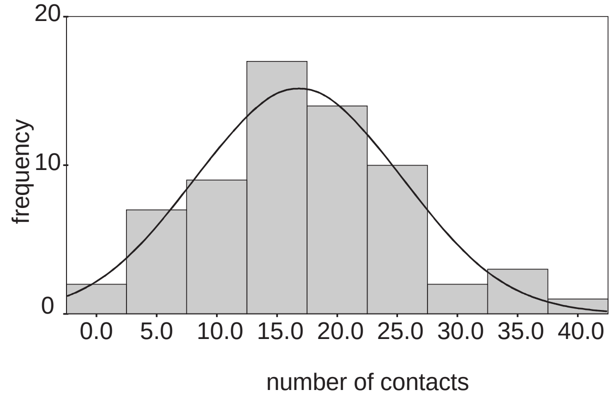

Figure 1. A frequency distribution of the number of contacts made by study participants over the period of one randomly assigned day. Sixty-five participants recorded a total number of 1093 contacts (mean 16.8 per day, range 0–40 per day, standard deviation 8.5). The solid line shows a normal distribution with the same mean and standard deviation. The midpoint of 5-year age classes is shown on the x-axis, such that the first class covers the range -2.5 to 2.5 years (effectively zero to 2.5 years).

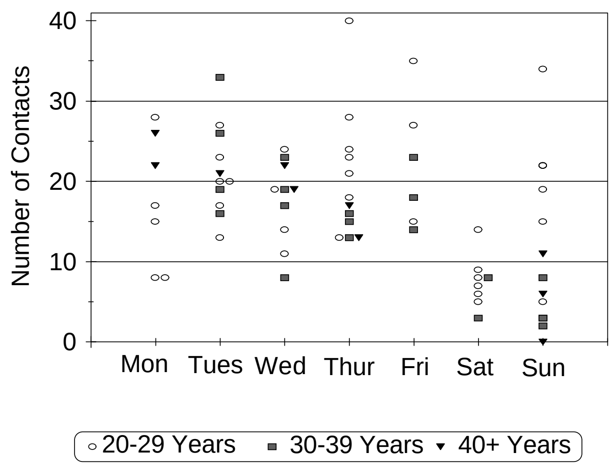

Figure 2. A scatter plot of the number of contacts by the day of observation and by the age group of the study participants. If two or more individuals had the same number of contacts on the same day, then they are shown alongside each other.

(*b*) *The daily number of contacts*

The 65 study participants had a total of 1093 contacts, yielding a mean of 16.8 contacts per day (range 0–40, standard deviation 8.5). Figure 1 shows a histogram of these contacts and compares this to a normal distribution with the same mean and variance. Despite the data being left censored, the number of contacts can still be approximated by a normal distribution (skewness = 0.29, kurtosis = -0.05, Shapiro–Wilk statistic; $W = 0.98$, $p = 0.55$) because of the high mean.

Simple variance component analysis using a two-level model (individual and day), in which the response variable was the number of contacts, showed that one quarter of the total observed variation in numbers of contacts could be attributed to variation between days on which the individuals filled out the survey (figure 2), while the remainder was accounted for by individual-level variation.

The results of the model of the number of contacts (table 1) indicate that there were no significant differences in the mean number of contacts between the different weekdays, irrespective of age of the participant (figure 2). Individuals assigned to Saturdays had a lower mean number of contacts than weekdays, again irrespective of age. With respect to individuals assigned to Sundays, older adults ($\geqslant 30$ years of age) had significantly fewer contacts than during the week, whereas the mean for younger adults (< 30 years) was not significantly different from the overall mean for weekdays (figure 2 and table 1). As anticipated, the effect of including these fixed covariates was to significantly reduce the variance observed at the level of the day of observation, such that this was actually estimated to be zero and subsequent mod-

Table 1. *Maximum likelihood parameter estimates for the most parsimonious model describing the number of contacts for study participants*

	parameter	estimate	standard error
fixed effects	intercept	19.46	0.85
	Saturday	−12.43	2.45
	Sunday* $\geqslant$ 30 years	−14.46	2.20
random parameters	$\sigma^2_{e\geqslant 30}$	24.61	6.46
	$\sigma^2_{e<30}$	55.55	13.09

els excluded level two (day) random effects. Variance estimated at the level of the study participants was also reduced by approximately 20%.

However, a plot of the standardized residuals versus age of the participant showed some evidence of heteroscedasticity, such that older participants appeared to have a smaller variance in the number of contacts than younger participants (data not shown). This was confirmed and accounted for by inclusion of separate random effect parameters for participants 30 years of age or older, and for participants less than this age, which resulted in significantly improved model fit ($p = 0.034$). That is, younger participants ($<$ 30 years) had a higher variance in the number of contacts than older participants. Further residual analysis suggested that the distributional assumptions were now met such that standardized residuals for both random parameters were approximately normally distributed (figure 3*a*, *b*).

In table 1, the model is defined as follows:

$$y_i = a + bx_i + e_{i<30} + e_{i\geqslant 30},$$

where y_i is the number of contacts for study participant i ($i = 1, \ldots, 65$), a is the constant intercept, equal to the average number of weekday contacts for all ages of participant, b is a vector of coefficients such that b_1 is equal to Saturday, equal to the average difference in number of contacts made on Saturday for all ages of participant, b_2 = Sunday* $\geqslant$ 30 years, equal to the average difference in number of contacts made on Sunday for participants 30 years of age or older, $e_{i<30} \sim N(0, \sigma^2_{e<30})$ is equal to residuals for participants less than 30 years of age

$$\mathrm{var}(e_{ij<30}) = \sigma^2_{e<30},$$

$e_{i\geqslant 30} \sim N(0, \sigma^2_{e\geqslant 30})$ is equal to residuals for participants 30 years of age or older

$$\mathrm{var}(e_{ij\geqslant 30}) = \sigma^2_{e\geqslant 30}.$$

In summary, younger participants ($<$ 30 years) had a higher variance in the number of contacts than older participants (table 1). Participants of all ages tended to have approximately 19 daily contacts during weekdays but only seven during Saturdays. Participants 30 years of age or greater had, on average, only five contacts on Sundays, whereas younger participants who were assigned Sundays had roughly the same number of contacts as during weekdays (table 1 and figure 2).

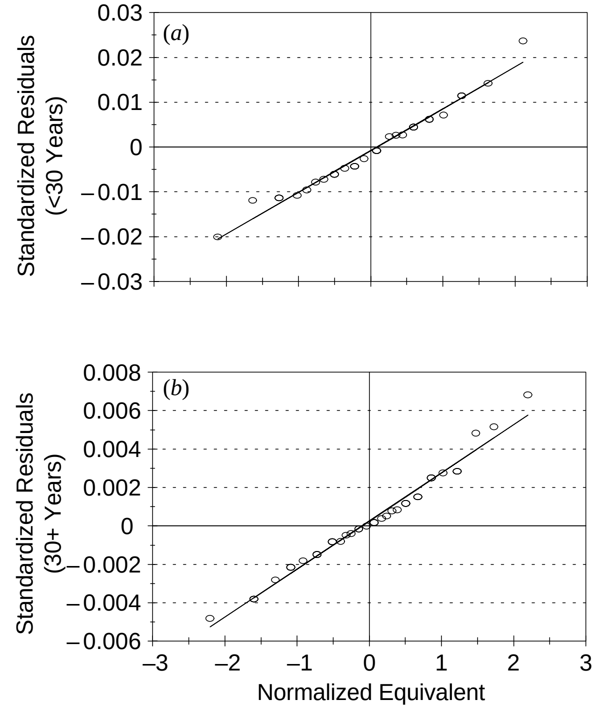

Figure 3. A plot of the cumulative distribution of the standardized residuals against their expected values from the standard normal distribution. The residuals are derived from the most parsimonious model of the number of contacts (table 1) in which a random effects term for participants less than 30 years of age is estimated (*a*); and a random effects term for participants equal or greater than 30 years of age is estimated (*b*). If the residuals are normally distributed then they should lie on a straight diagonal line (shown).

(*c*) *Social context of mixing patterns*

Figure 4*a*–*c* shows the number of contacts per person in the different social settings subdivided by the age of the study participants and the day that they were asked to complete the questionnaire (weekdays or weekends). With the exception of social contacts at the weekend it is clear from observation of figure 4 that the distribution of the context of contacts amongst the different age groups is similar. During weekdays, participants, irrespective of age, tended to have approximately 12–13 work related contacts, 2–4 social contacts, 1–2 home contacts and a few other contacts in shops or travelling per day. During the weekend, participants tended to have markedly fewer work related contacts (as expected) and roughly the same, or slightly fewer, contacts in the other contexts, with the exception of social contacts for younger participants ($<$ 30 years). These participants seemed to have approximately three more social contacts during the weekend than during the week.

(*d*) *The age of contacts*

Although the number and the contexts of the contacts seemed to be broadly consistent across the age groups (particularly for weekdays), the age distribution of these contacts appeared to vary (figure 5*a*–

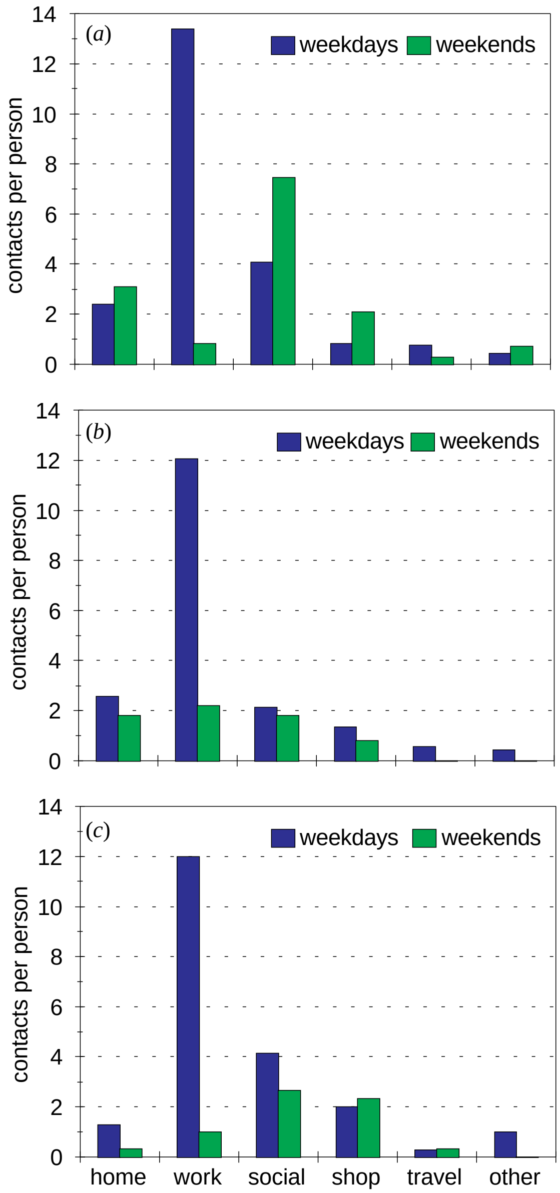

Figure 4. The number of contacts per person (per day) in the different social settings subdivided into weekdays and weekends and the age group of the study participants (20–29 year olds (*a*); 30–39 year olds (*b*); and 40+ year olds (*c*)). The number of contacts per person (per day) in each of the social classes for each of the age groups was calculated by dividing the total number of contacts occurring in the individual social settings at that particular time of the week (weekday or weekend) by the number of individuals in the age class assigned to that particular time of the week. The number of participants in the 20–29 year old age class was 36; 25 assigned weekdays, who had a total of 548 contacts and 11 weekends who had a total of 159 contacts. There were 19 participants in the 30–39 age class; 14 assigned weekdays, who had a total of 268 contacts and five assigned weekends who had 33 contacts. There were ten participants in the 40+ age group; seven assigned weekdays, these had 145 contacts, and three assigned weekends, these had a total of 20 contacts.

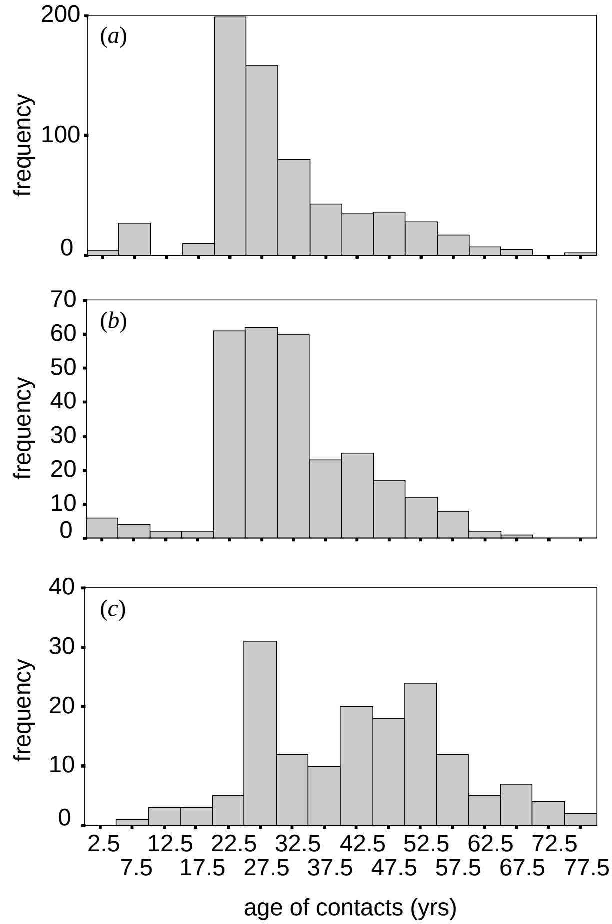

Figure 5. A frequency distribution of the age of the contacts of the participants in the survey, subdivided by the age group of the participants (20–29 year olds (*a*); 30–39 year olds (*b*); and 40+ year olds (*c*)). The midpoint of five year age classes is shown. Note the change in scale of the y axes which is mainly due to the fact that there were 36 participants aged 20–29, 19 aged 30–39, and ten aged 40+. The mean age of contacts of participants in the 20–29 year age group was 29.8 years (standard deviation 11.74, number of contacts 651). The mean age of contacts of participants in the 30–39 year age group was 31.2 years (standard deviation 10.98, number of contacts 285). The mean age of contacts of participants in the 40+ year age group was 41.3 years (standard deviation 14.55, number of contacts 157).

c). The mean age of contacts increased with the age of the participant, such that older participants ($\geqslant$ 40 years) had more contact with older adults than younger participants ($<$ 40 years). Older participants had a larger variability in the age of their contacts than younger participants. In addition to which, the shape of the distribution appeared to be different, with younger participants having a more skewed distribution in which the majority of contacts were nearer their own age, a few in older individuals and virtually none in children. On the other hand, older participants exhibited a more even distribution across the age range. The large number of contacts in the 25–30 year age group, which was consistent across the participant age range, most likely reflects the predominantly university setting of the study.

Modelling the age of contacts required the addition of another level, 'below' that of the participant; that is, the level of the contact (contacts being grouped within individuals which, in turn, are grouped within the day of observation). Simple variance component analysis using this three-level model (contact, participant and day), in which the response variable was the age of the contact, showed that the vast majority (83%) of the total observed variation could be attributed to within individual variability. A further 15% of the variation was accounted for by between individual variation while only 2% could be accounted for by variation between days of observation. That is, most of the variability in the age of contacts occurred within individuals; there were, however, some differences between individuals, but this does not seem to primarily depend on the day on which individuals were sampled.

The process of model evolution and residual analysis supported the previous observations concerning differences in variation and distribution, such that the most parsimonious model excluded a term for between day variation (as previously) and included a single random effect parameter at the level of the individual. However, the within-individual variation was best accounted for by means of a constant variance estimate for older participants ($\geqslant 40$ years) and a linearly increasing variance with respect to the total number of contacts for participants less than 40 years of age. That is, those participants less than 40 years of age with a small number of contacts, tended to have contacts close in age to their overall mean (which is similar to their own age). As the number of contacts increased in this participant group (< 40 years), the spread in age of the contacts tended to increase. On the other hand, older participants ($\geqslant 40$ years) tended to have a higher, and constant, variance in the age of their contacts, i.e. regardless of the number of contacts they had (table 2).

Table 2. *Maximum likelihood parameter estimates for the most parsimonious model describing the age of contacts of study participants*

	parameter	estimate	standard error
fixed effects	intercept	27.59	1.06
	work	4.34	1.05
	travel	10.53	2.81
	shop	3.78	1.60
	social	2.75	1.25
	other	10.72	2.40
	$\geqslant 40$ years	22.17	4.68
	$\geqslant 40$ years*work	−12.49	4.82
	$\geqslant 40$ years*travel	−0.54	11.16
	$\geqslant 40$ years*shop	−16.08	5.72
	$\geqslant 40$ years*social	−12.59	5.25
	$\geqslant 40$ years*other	−23.88	7.32
	Saturday	−3.57	1.76
	Sunday* < 40 years	−3.18	1.70
random parameters	σ_u^2	9.63	2.95
	$\sigma_{e\geqslant 40}^2$	190.50	21.96
	$\sigma_{e<40}^2$	48.36	11.99
	$\sigma_{ef<40}^2$	1.49	0.31

In table 2, the model is defined as follows:

$$y_{ij} = (a + bx_{ij}) + (u_j + e_{ij\geqslant 40} + e_{ij<40} + f_{ij<40}x_{ij<40}),$$

where y_{ij} is the age of contact i ($i = 1, \ldots, 40$) for study participant j ($j = 1, \ldots, 64$), a is the constant intercept, equal to the average age of weekday contacts made in the home for participants less than 40 years of age, b is a vector of coefficients such that b_1 is equal to work, equal to the average difference in age of contacts made at work, b_2 is equal to travel, equal to the average difference in age of contacts made while travelling, b_3 is equal to shop, equal to the average difference in age of contacts made while shopping, b_4 is equal to social, equal to the average difference in age of contacts made while socializing, b_5 is equal to other, equal to the average difference in age of contacts made in all other contexts, $b_6 \geqslant 40$ years, equal to the average difference in age of contacts for participants 40 years of age or older, $b_7 \geqslant 40$ years*work, equal to the average difference in age of contacts made at work for participants 40 years of age or older, $b_8 \geqslant 40$ years*travel, equal to the average difference in age of contacts made while travelling for participants 40 years of age or older, $b_9 \geqslant 40$ years*shop, equal to the average difference in age of contacts made while shopping for participants 40 years of age or older, $b_{10} \geqslant 40$ years*social, equal to the average difference in age of contacts made while socializing for participants 40 years of age or older, $b_{11} \geqslant 40$ years*other, equal to the average difference in age of contacts made in all other contexts for participants 40 years of age or older, b_{12} is equal to Saturday, equal to the average difference in age of contacts made on Saturdays, b_{13} = Sunday* < 40 years, equal to the average difference in age of contacts made on Sundays for participants less than 40 years of age, $u_j \sim N(0, \sigma_u^2)$ is equal to level 2 residuals for all participants

$$\text{var}(u_j) = \sigma_u^2,$$

$e_{ij\geqslant 40} \sim N(0, \sigma_{e\geqslant 40}^2)$ = level 1 residuals for participants 40 years of age or older

$$\text{var}(e_{ij\geqslant 40}) = \sigma_{e\geqslant 40}^2,$$

$e_{ij<40} \sim N(0, \sigma_{e<40}^2)$ is equal to level 1 residuals for participants less than 40 years of age, $f_{ij<40}$ is a random variable associated with $x_{ij<40}$ (number of contacts for participants less than 40 years of age)

$$\text{var}(e_{ij<40} + f_{ij<40}x_{ij<40}) = \sigma_{e<40}^2 + 2\sigma_{ef<40}x_{ij<40},$$

where $\sigma_{ef<40}$ is the covariance of $e_{ij<40}$ and $f_{ij<40}$.

With respect to fixed effect estimates, both parameters describing the context of the contact and interaction terms between age of the participant and context of each contact were found to be significant (table 2).

For those participants less than 40 years of age, the average age of contacts in the home during the week was 27.6 years. The mean age of their contacts in all other contexts was significantly higher during the week. For participants 40 years of age or greater, the mean age of their home weekday contacts was 49.7 years. However, the mean age of their contacts made during the week in all contexts other than travelling was significantly less than the mean age of their weekday home contacts (see table 2 for a more detailed breakdown of the findings).

With respect to weekend observations, the overall mean age of contacts for all participants, regardless of their age and context of the contact, was lower on Saturdays. For participants less than 40 years of age, this was also true for Sundays. However, the inclusion of these two parameters only marginally ($p = 0.039$) improved the model fit.

4. DISCUSSION

We have presented the results of a pilot study, the aim of which was to evaluate a method for providing useful information on the mixing patterns of adults which might lead to the spread of airborne infections. To enable quantification of these patterns it was necessary to define a contact event which not only had the potential to effect transmission, but would be understood and easy to record in a quantifiable way by study participants. We argued that the simplest event that conformed to these criteria was a two way conversation. More specifically, we defined that an at-risk contact had been made with another individual if at least one conversation was held with that person during the day. Clearly, to record each and every repeated contact with the same individual would be considerably more demanding on the recorder.

The very simplicity of this at-risk contact definition is the source of the merits and disadvantages of the method. The acceptance by 71% of the 92 individuals approached to complete the self questionnaire and the apparent lack of difficulty in its completion, suggest that such a measure of social contact is suitable for collection in large-scale surveys. With minor adjustments, the questionnaire would be suitable for adolescents as well as adults.

There are, however, numerous questions about the validity of such a contact definition as representing the true picture of contacts that might lead to transmission. For example, interpretation of the data as patterns of potentially effective contacts would require an implicit assumption of equal probability of transmitting infection irrespective of whether there was one or many conversations between an infectious individual and a susceptible individual during a day (this has parallels with many models of STD transmission, where a constant probability of transmission per partnership is assumed, irrespective of the duration of partnership or number of sex acts per partnership; see, for instance, Garnett & Anderson (1993)). The same condition applies to the duration of the conversation and the intensity of the contact (e.g. due to proximity of interaction); that is, we would need to assume there is a minimum contact necessary for transmission and that increasing the length or intensity of contact would have no bearing on the probability of transmission. Furthermore, the transmission of respiratory infections is not always going to be facilitated exclusively or predominantly by contaminated aerosols but perhaps by direct contact by contaminated hands, mouths or fomites. Further work is needed to investigate suitable at-risk contact definitions for a wide range of infections. In addition, other definitions would be required to elicit the patterns of mixing of younger children and infants. For instance, for infants we might modify our definition to include direct physical contact and require observers to make records of contact frequency.

Although the results derived from the convenience sample used in this study are not generalizable to the wider population, the study did identify some interesting patterns of contact which may have epidemiological implications if confirmed in a population-based survey. During the week the participants, regardless of age group, tended to have more than half of their contacts at work. Given a constant age- and setting-related prevalence of infection, our findings would suggest that during the week adults in this sample are approximately six times more likely to come into contact with an infectious individual at work than at home. This does not necessarily imply that they are six times more likely to be infected, however, as contacts at home might be more conducive to transmission than contacts elsewhere; that is, home contacts might be more prolonged and frequent than contacts elsewhere. One might imagine that the relative risk of developing an infection from a contact in the different social settings obeys the following pattern: home > work/social > background, where background contacts are those that occur in shops or while travelling, etc. Additionally, the turnover of these contacts is likely to vary, so for instance home and work contacts might be expected to be relatively stable through time, whereas background contacts turn over at a very high rate. These background contacts may act as links between social networks, alternatively individuals' networks may be so large that these background contacts are relatively unimportant (see Morris (1994), Morris *et al.* (1995) and Kretzchmar & Morris (1996) for a discussion of the importance of networks to the transmission of infectious diseases, though much of the discussion focuses on STDs). Further studies in which participants might be asked to detail every conversation over a (necessarily shorter) time period are needed to investigate the relative differences in frequency and duration of contacts in the different social settings. Additionally, a modification of the questionnaire to include retrospective questions, such as 'Did you speak to this person yesterday?' could provide useful information on the structure and dynamics of social networks.

Older participants ($\geqslant 40$ years of age) appeared to make contact with a larger age range than younger participants. One of the assumptions often used when modelling the age-specific rates of contact is that adults mix with themselves and all other age groups at the same rate (this assumption is often imposed rather than made by choice). Nevertheless, to a first approximation, older adults appear to act in roughly this way. However, younger adults do not. Mixing amongst adults less than 40 years of age appears to be largely assortative (within group), particularly if the individual's overall number of contacts is small. That is, if younger adults have a small number of contacts they tend to be within their own age group, as their number of contacts increases so does the age range of those contacts. If this pattern is confirmed then it could be that those individuals who tend to have a high number of contacts act as a bridge for infection between largely separate age groups. A similar conclusion could be drawn for mixing between the adult and child populations. There was little evidence of mixing with children except for those participants who either worked with children (one teacher filled out the survey) or who lived with them. Again, it could be that these individuals act as a gateway for infection between the adult and child populations.

Participants, of all age groups, who were assigned a weekday, had approximately 20 contacts per day (this is of clear importance in determining transmission rates). Participants appeared to have different patterns of mixing during the weekend, for instance on Saturdays participants typically had fewer contacts and the mean age of those contacts was lower (tables 1 and 2). This suggests that individuals contact different groups of people at the weekend compared to the week. Note that as the participants were randomly assigned a day of observation, we can tentatively conclude that the observed pattern of contacts broadly reflects the sampled population's individual contact patterns over the period of a week, though clearly longitudinal studies are needed to confirm this.

Finally, it is interesting to note that by defining and measuring at-risk contacts we might be able to identify and investigate the properties of social networks important for the transmission of infectious diseases. It is conceivable that an individual will have a single social network with different links between the nodes for different infections (see Wasserman & Faust (1994) for an introduction to social network analysis). For instance, an individual may have many links sufficient to spread influenza, but fewer of those links would be able to pass on an STD (see Woodhouse *et al.* (1994) for a study which attempted to measure different kinds of relationships (sexual, social, etc.) between individuals thought to be important for the spread of HIV).

The validation of this technique is a key area of interest for further investigation. Two approaches might be adopted. First, to investigate if there is a positive correlation between individuals' exposure to airborne pathogens, as shown by specific antibody conversion, and their contact patterns, as defined in this study. Further, indirect validation would be provided if models which utilise contact data generated by this technique were better able to capture observed patterns of infection than models that do not. In particular, we would wish to assess the capacity of these models to capture the temporal effects on age antibody prevalence and incidence data resulting from the introduction of a vaccination programme (an increasing amount of such data now exists for measles, mumps and rubella).

In summary, we believe that this simple questionnaire format has the potential to be used in large population surveys which may provide valuable information on the rates and patterns of mixing that may influence the spread of infectious diseases. Work is continuing in order to refine methods of data collection, extend the field of observation to other populations and validate the method. We believe that the use of data derived from similar surveys will help improve the current models of airborne infectious disease transmission and may provide an insight into different ways of modelling the transmission of such diseases.

W.J.E. is in receipt of a Wellcome Post-doctoral Health Services Research Training Fellowship (Grant Number 040952). C.J.O.C. is funded by ILRI/USAID. D.J.N. is a Royal Society University Research Fellow. The authors thank Graham Medley, Martin Cox and the other members of WUPERT for helpful discussions, Ilona Carneiro for organizing the sampling in Oxford and W.J.E.'s parents for trying out the original version of this questionnaire.

REFERENCES

Anderson, R. M. & Grenfell, B. T. 1986 Quantitative investigations of different vaccination policies for the control of congenital rubella syndrome (CRS) in the United Kingdom. *J. Hyg. Lond.* **96**, 305–333.

Anderson, R. M., Gupta, S. & Ng, W. 1990 The significance of sexual partner contact networks for the transmission dynamics of HIV. *J. AIDS* **3**, 417–429.

Anderson, R. M. & May, R. M. 1985 Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes. *J. Hyg. Camb.* **94**, 365–436.

Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans. Dynamics and control.* Oxford University Press.

Babad, H., Nokes, D. J., Gay, N., Miller, E., Morgan-Capner, P. & Anderson, R. M. 1995 Predicting the impact of measles vaccination in England and Wales: model validation and policy projection. *Epidemiol. Infect.* **114**, 319–44.

Boily, M. & Anderson, R. M. 1991 Sexual contact patterns between men and women and the spread of HIV-1 in urban centres in Africa. *IMA J. Math. Appl. Med. Biol.* **8**, 221–247.

Calvert, N., Cutts, F., Miller, E., Brown, D. & Munro J. 1994 Measles in secondary school children: implications for vaccination policy. *Commun. Dis. Rep.* R **4**, 70–73.

Cliff, A. D. & Haggett, P. 1992 *Atlas of disease distributions.* Oxford: Blackwell.

Dean, A. G., Dean, J. A., Coulombier, D. *et al.* 1994 Epi Info, version 6: a word processing, database, and statistics program for epidemiology on microcomputers. Atlanta, GA: Centers for Disease Control and Prevention.

Fine, P. E. M. & Clarkson, J. A. 1982 Measles in England and Wales. I. An analysis of factors underlying seasonal patterns. *Int. J. Epidemiol.* **11**, 5–14.

Garnett, G. P. & Anderson, R. M. 1993 Contact tracing and the estimation of sexual mixing patterns: the epidemiology of gonococcal infections. *Sex. Transm. Dis.* **20**, 181–191.

Garnett, G. P. & Anderson, R. M. 1996 Sexually transmitted diseases and sexual behaviour: insights from mathematical models. *J. Infect. Dis.* S **174**, 150–161.

Garnett, G. P., Hughes, J. P., Anderson, R. M. *et al.* 1996 Sexual mixing patterns of patients attending sexually transmitted diseases clinics. *Sex. Transm. Dis.* **23**, 248–257.

Gay, N. J., Hesketh, L. M., Morgan-Capner, P. & Miller, E. 1995 Interpretation of serological surveillance data for measles using mathematical models: implications for vaccine strategy. *Epidemiol. Infect.* **115**, 139–156

Grenfell, B. T. & Anderson, R. M. 1985 The estimation of age-related rates of infection from case notifications and serological data. *J. Hyg. Camb.* **95**, 419–436.

Gupta, S., Anderson, R. M. & May, R. M. 1989 Networks of sexual contacts: implications for the pattern of spread of HIV. *AIDS* **3**, 807–817.

Haraldsdottir, S., Gupta, S. & Anderson, R. M. 1992 Preliminary studies of sexual networks in a male homosexual community in Iceland. *J. AIDS* **5**, 374–381.

Johnson, A. M., Wadsworth, J., Wellings, K., Field, J. & Bradshaw, S. 1994 *Sexual attitudes and lifestyles.* Oxford: Blackwell.

Kretzschmar, M. & Morris, M. 1996 Measures of concurrency in networks and the spread of infectious disease. *Math. Biosci.* **133**, 165–195

Longini, I. M. & Koopman, J. S. 1982 Household and community transmission parameters from final distributions of infections in households. *Biometrics* **38**, 115–126

Morris, M. 1994 Epidemiology and social networks—modeling structured diffusion. In *Advances in social network analysis, research in the social and behavioral sciences* (ed. S. Wasserman & J. Galaskiewicz), pp. 26–52. London: Sage.

Morris, M., Zavisca, J. & Dean, L. 1995 Social and sexual networks—their role in the spread of HIV/AIDS among young gay men. *AIDS Education Prevention* **7**, ss24–ss35.

Nokes, D. J., Anderson, R. M. & Anderson, M. J. 1986 Rubella epidemiology in South East England. *J. Hyg. Camb.* **96**, 291–304.

Nokes, D. J., Morgan-Capner, P., Wright, J. & Anderson, R. M. 1990 Serological study of the epidemiology of mumps virus infection in north-west England. *Epidemiol. Infect.* **105**, 175–195.

Nokes, D. J. & Anderson, R. M. 1992 Mathematical models of infectious agent transmission and the impact of mass vaccination. *Rev. Med. Microbiol.* **3**, 187–195.

Wasserman, S. & Faust, K. 1994 *Social network analysis. Methods and applications.* Cambridge University Press.

Woodhouse, D. E. *et al.* 1994 Mapping social networks of heterosexuals at high risk for HIV infection. *AIDS* **8**, 1331–1336.

Received 27 February 1997; accepted 19 March 1997

As this paper exceeds the maximum length normally considered for publication in *Proceedings* B, the authors have agreed to make a contribution towards production costs.