

RESEARCH

Open Access

# Quantifying heterogeneous contact patterns in Japan: a social contact survey



Lankeshwara Munasinghe, Yusuke Asai and Hiroshi Nishiura\* 

## Abstract

**Background:** Social contact surveys can greatly help in quantifying the heterogeneous patterns of infectious disease transmission. The present study aimed to conduct a contact survey in Japan, offering estimates of contact by age and location and validating a social contact matrix using a seroepidemiological dataset of influenza.

**Methods:** An internet-based questionnaire survey was conducted, covering all 47 prefectures in Japan and including a total of 1476 households. The social contact matrix was quantified assuming reciprocity and using the maximum likelihood method. By imposing several parametric assumptions for the next-generation matrix, the empirical seroepidemiological data of influenza A (H1N1) 2009 was analysed and we estimated the basic reproduction number,  $R_0$ .

**Results:** In total, the reported number of contacts on weekdays was 10,682 whereas that on weekend days was 8867. Strong age-dependent assortativity was identified. Forty percent of weekday contacts took place at schools or workplaces, but that declined to 14% on weekends. Accounting for the age-dependent heterogeneity with the known social contact matrix, the minimum value of the Akaike information criterion was obtained and  $R_0$  was estimated at 1.45 (95% confidence interval: 1.42, 1.49).

**Conclusions:** Survey datasets will be useful for parameterizing the heterogeneous transmission model of various directly transmitted infectious diseases in Japan. Age-dependent assortativity, especially among children, along with numerous contacts in school settings on weekdays implies the potential effectiveness of school closure.

**Keywords:** Epidemic, Epidemiological model, Mathematical model, Cumulative incidence, Influenza

## Background

The epidemiological dynamics of directly transmitted infectious diseases frequently exhibit highly heterogeneous patterns [1]. In particular, the transmission of acute infectious diseases tends to depend on the age of both primary cases and their contacts [2], indicating a critical need to account for age-dependent heterogeneity [3]. The basic reproduction number in such a heterogeneously mixing population is derived as the largest eigenvalue of the age-dependent next-generation matrix [1, 4], and its computation with  $n$  heterogeneously mixing age groups requires quantification of  $n \times n$  elements of the matrix. However, it is frequently the case that the degrees of freedom during the statistical estimation are limited (e.g., limited by  $n$  different values of input data) [5].

Because use of the so-called WAIFW (who acquires infection from whom) matrix serves as an approximation of the actual heterogeneous contact pattern [2], this has recently attracted the attention of researchers, and conducting a social contact survey appears to greatly assist in reducing uncertainty with respect to heterogeneous contact patterns and quantifying a WAIFW matrix [6, 7]. In many instances, participants in social contact surveys, who are mostly recruited via convenience sampling, are asked to report the number of contacts they experience on a given day (e.g., a weekday or weekend day) with someone in the same or a different age group, the social setting of that contact, the type of contact (e.g., exchanging a few sentences or physical bodily contact), the duration of contact, and so on. A landmark study reporting the social contact patterns among eight different European countries was conducted as part of the so-called POLYMOD, a European

\* Correspondence: [nishiurah@med.hokudai.ac.jp](mailto:nishiurah@med.hokudai.ac.jp)  
Graduate School of Medicine, Hokkaido University, Kita 15 Jo Nishi 7 Chome, Sapporo, Japan



Commission project [6]. Starting with the first study published in 2008, similar surveys have been increasingly conducted across the world, mostly using a written diary [2, 6, 8, 9] or remote sensing device [10–12]. Additional survey methods have also been reported, including the analysis of time-use data [13] and mobile phone network data [14]. Such empirical data appear to be an essential back-up for infectious disease modelling in the event of emerging infectious disease epidemics, such as the Ebola virus disease epidemic in West Africa in 2014–2016 [15, 16].

POLYMOD surveys have taken place in East and Southeast Asian countries and regions, including China, the Hong Kong Special Administrative Region, Japan, Taiwan, Thailand, and Vietnam [8, 9, 17–23]. Whereas the findings have mostly echoed those reported in Western countries (e.g., strong age assortativity, decreased contacts with age, and increased contacts with household size) [8], these studies can be expected to provide key information to parameterize infectious disease transmission in these countries. An existing study in Japan by Ibuka et al. [17] also successfully identified characteristics similar to those stated above; however, that study indicated a weak predictability of influenza using a contact matrix alone. Also, people aged 20–29 years were considered to have been under-represented in the published survey. The need remains for a similar data collection method but one that addresses the issue of validation using infectious disease data, possibly with better representation of the Japanese population.

The present study aimed to conduct a contact survey in Japan, offering estimates of contacts according to 5-year age groups, and validating the matrix using a seroepidemiological dataset of influenza.

## Methods

### Data collection

We conducted an internet-based questionnaire survey covering all 47 prefectures in Japan. Area sampling was conducted using the population size in Japan, determining the sample size by age and prefecture to be proportional to the actual age- and prefecture-specific distributions. The survey was conducted via a private company that has a collection of monitors to act as potential respondents across Japan, and published studies that rested on the internet survey of the same company are found elsewhere [24–26]. The present survey started with an advertisement that was notified among all registered monitors, and of these, a total of 1476 households have voluntarily decided to participate in the survey. As a consequence, a total of 2271 respondents were conveniently sampled to achieve a proportional sample of 0.015% in each 5-year age stratum and prefecture. The survey was conducted from 28 October to 1 December

2014, with manual validation of unclear responses continuing until 9 March 2015. Demographic variables other than age and prefecture of residence included sex, occupation and household size. Each respondent recorded all contacts during one weekday (Wednesday) and one weekend day (Sunday). A contact was defined as an exchange of three Japanese sentences or a physical touch on the skin. Each survey respondent was requested to keep a diary, recording each contactor's age, sex, type of contact (i.e., conversation or physical), and the duration and location of contact. The questionnaire format was the same as that used elsewhere [6].

### Contact matrix

According to the ages of each respondent–contactor pair, reported contacts were grouped into 15 discrete age groups (0–4, 5–9, ..., 65–69, and 70 years or older). These constituted the so-called social contact matrix  $\{m_{ij}\}$  representing the rate of contact between an individual in age group  $j$  with individuals in age group  $i$  on a given day. To quantify  $\{m_{ij}\}$ , we adopted an assumption of reciprocity, i.e., the mean number of contacts that an individual in age group  $i$  experiences with individuals in age group  $j$  is equal to the number that an individual in age group  $j$  experiences with individuals in age group  $i$ . Thus, we adjusted the asymmetry of the contact rate using the age-dependent population size. The mean contact rate  $\{m_{ij}\}$  was estimated using the maximum likelihood method, as previously reported [7].

### Validation using influenza data

A key assumption when using the social contact matrix to model age-dependent heterogeneity of infectious disease transmission is that the age-dependent transmission matrix containing the contact frequency is informed by the social contact matrix. To capture age-dependent heterogeneity, we used the so-called next-generation matrix,  $K = \{k_{ij}\}$ , comprising the average number of secondary infections in age group  $i$  that are caused by a single infectious individual in age group  $j$  in a fully susceptible population. The matrix describes the per-generation heterogeneity for the risk of infection by age; in the simplest terms, it may be parameterized using the social contact matrix as  $k_{ij} = qm_{ij}$ , where  $q$  is a constant that can be interpreted as the disease-specific infectivity [27]. Alternatively, one can parameterize  $K$  as  $k_{ij} = s_i m_{ij}$  if the susceptibility varies with age group  $i$ , and similarly,  $k_{ij} = u_j m_{ij}$  if the infectiousness of age group  $j$  matters biologically.

Once the parametric assumption of the next-generation matrix was determined, we quantified the next-generation matrix by analysing the empirical seroepidemiological dataset of influenza A (H1N1) 2009 collected by the government of Japan [28] and estimating the basic

reproduction number,  $R_0$ , which is interpreted as the average number of secondary cases generated by a typical single primary case in a fully susceptible population and is calculated as the largest eigenvalue of the next-generation matrix [1, 3, 4]. Depending on the parametric assumption, we estimated the constant  $q$  or age-dependent susceptibility  $s_i$  to scale the next-generation matrix. To do so, we analysed the age-specific final size distribution (or serologically determined cumulative incidence) of influenza A (H1N1) 2009, using the difference in the age-specific seroprevalence between the 2009 and 2010 surveys. In 2009 and 2010, a total of 6626 and 6539 samples were collected, respectively, from all age groups. In each year, the survey took place from June to September; thus, the surveys in 2009 and 2010 contained the first wave of the pandemic within the inter-survey period. To determine seropositivity, we imposed a cut-off level of 1:20 as the default [29]. Additional file 1: Table S1 shows the dataset used to validate the importance of using the social contact matrix. Subtracting the positive fraction in 2009 from that in 2010, we obtained the empirical value of cumulative incidence  $z_i$ , which satisfies the so-called age-dependent final size equation:

$$z_i = 1 - \exp\left(-\sum_j z_j k_{ij}\right)$$

where  $i$  and  $j$  denote age groups and  $\{k_{ij}\}$  denotes the next-generation matrix. Iteratively solving this equation, we optimized the likelihood function

$$L(\theta; \mathbf{n}, \mathbf{m}) = \prod_i \binom{n_{10,i}}{m_{10,i}} (z_{9,i} + z_i)^{m_{10,i}} (1 - z_{9,i} - z_i)^{n_{10,i} - m_{10,i}}$$

where  $n_{10,i}$  and  $m_{10,i}$  represent the observed total and positive serological samples in age group  $i$  in 2010, respectively,  $z_{9,i}$  is the positive fraction of age group  $i$  in 2009, and  $\theta$  is the population parameter. We parameterized the next-generation matrix in various ways and tested the goodness of fit referring to the Akaike information criterion (AIC). As possible candidate matrices that do not rest on the social contact matrix, we tested the fully parameterized next-generation matrix including the homogeneous mixing assumption ( $k_{ij} = R$ ), where  $R$  is a constant and the separable mixing assumption ( $k_{ij} = a_i a_j$ ), comparing their fit against those using the social contact matrix. The 95% confidence intervals (CI) of parameters were computed using the profile likelihood.

#### Ethical considerations

The purpose of the study was explained to participants, and they were ensured that the extent of the use of survey information would be limited to the present study. Informed consent was obtained via a website, and

participants had the right to withdraw at any time during the study period. The Medical Ethics Committees of the Graduate School of Medicine, The University of Tokyo approved this study (approval ID: 10478). As for the seroepidemiological data, the present study used publicly available data [28]. The datasets had already been fully anonymized and did not include any identifiable information. Thus, ethical approval was not required for the analysis of seroepidemiological data.

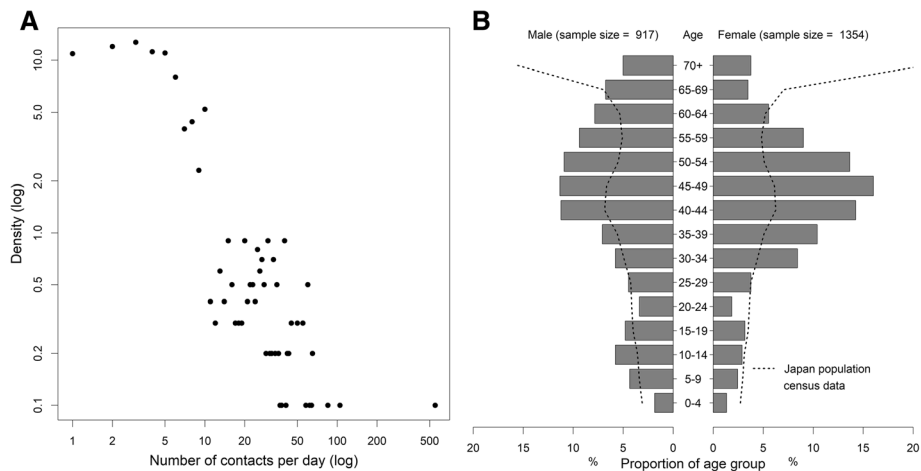
#### Data sharing policy

Seroepidemiological surveillance data can be accessed online via a linked URL [28]. A summary of the contact datasets presented in this study can be obtained from the corresponding author upon request.

#### Results

In total, the reported number of contacts on weekdays was 10,682 and that on weekend days was 8867. Figure 1a shows the distribution of the daily rate of contacts per person, revealing a right-skewed distribution that could potentially be approximated to follow a power law, which can be confirmed by approximately a linear distribution by log-log plot. Due to area sampling with a proportional age sampling distribution, the samples of our survey qualitatively captured the essential part of the ages that were observed, whereas the ages among elderly respondents were relatively under-sampled (Fig. 1b). Due to the internet survey that starts with recruitment of registered monitors that are dominated by housewives, female slightly dominated respondents (59.6%). Common occupations included office workers (25.4%), housewives (23.9%), part-time worker (12.4%), retired persons (4.9%), primary school children (4.5%), and self-employed persons (4.2%). The household size ranged from 1 to 10 with the mean and standard deviation of 3.2 and 0.2 persons, respectively, which are consistent with the household size of the entire Japan.

Of contacts on weekdays and weekend days, 27.8 and 34.9% were classified as physical contacts, respectively. On weekdays, the highest average number of physical contacts per person (5.36) was seen in the age group 0–4 years, and the highest number of non-physical contacts (5.67) was seen in respondents aged 15–19 years. The lowest average number of physical contacts (0.69) was found among those aged 65–69 years and that of non-physical contacts (0.12) was among small children aged 0–4 years (Additional file 2: Table S2). These patterns were maintained on weekend days (Additional file 3: Table S3). The difference between weekdays and weekend days was highlighted by the location of contact, i.e., 40% of weekday contacts took place at school or the workplace, but this declined to 14% on the weekend (Additional file 4: Figure S1). On the contrary,



**Fig. 1** Distributions of contact frequency and age of sample population in Japan. **a** Log-log plot of contact frequency distribution. Logarithm of the proportion of the sample population was taken against the number of contacts (contact frequency) per day. **b** Age distribution of the study sample (bars) by age group and sex. Dashed lines represent the age distributions of the entire population of Japan as of 1 November 2016, overlaid with the sample population

<b>A</b> AgeGroup	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70 +
0-4	1.91	0.27	0.07	0.02	0.01	0.22	0.31	0.22	0.22	0.09	0.10	0.04	0.10	0.05	0.03
5-9	0.29	4.54	0.57	0.04	0.04	0.11	0.35	0.45	0.36	0.10	0.07	0.04	0.05	0.04	0.02
10-14	0.08	0.59	4.48	0.44	0.01	0.05	0.07	0.29	0.44	0.30	0.12	0.06	0.04	0.03	0.03
15-19	0.02	0.05	0.48	4.34	0.36	0.09	0.05	0.08	0.26	0.49	0.32	0.08	0.02	0.05	0.03
20-24	0.01	0.04	0.02	0.36	2.20	0.45	0.16	0.11	0.12	0.22	0.48	0.24	0.07	0.02	0.06
25-29	0.28	0.13	0.05	0.09	0.46	0.91	0.54	0.34	0.24	0.17	0.32	0.53	0.23	0.10	0.04
30-34	0.45	0.48	0.09	0.07	0.19	0.61	0.81	0.53	0.36	0.28	0.30	0.41	0.43	0.19	0.06
35-39	0.36	0.68	0.43	0.11	0.15	0.43	0.59	0.85	0.59	0.33	0.27	0.28	0.26	0.22	0.12
40-44	0.42	0.66	0.78	0.41	0.19	0.37	0.48	0.71	0.91	0.62	0.42	0.36	0.23	0.21	0.11
45-49	0.17	0.18	0.52	0.77	0.34	0.26	0.36	0.39	0.61	0.80	0.61	0.36	0.23	0.10	0.18
50-54	0.16	0.11	0.17	0.43	0.62	0.41	0.33	0.27	0.35	0.52	0.78	0.50	0.33	0.18	0.16
55-59	0.06	0.06	0.10	0.10	0.30	0.63	0.43	0.26	0.28	0.29	0.47	0.67	0.48	0.14	0.14
60-64	0.17	0.08	0.05	0.03	0.09	0.29	0.47	0.26	0.20	0.20	0.33	0.51	0.75	0.40	0.16
65-69	0.17	0.08	0.05	0.04	0.03	0.16	0.26	0.28	0.22	0.11	0.23	0.19	0.51	0.79	0.34
70 +	0.12	0.08	0.14	0.14	0.22	0.15	0.20	0.38	0.29	0.46	0.49	0.47	0.49	0.82	1.26

<b>B</b> AgeGroup	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70 +
0-4	0.42	0.36	0.08	0.02	0.10	0.17	0.30	0.27	0.24	0.05	0.09	0.05	0.16	0.06	0.07
5-9	0.39	1.31	0.54	0.05	0.02	0.11	0.30	0.44	0.39	0.12	0.05	0.07	0.08	0.12	0.07
10-14	0.09	0.56	1.85	0.30	0.02	0.02	0.07	0.34	0.51	0.31	0.09	0.02	0.08	0.05	0.07
15-19	0.02	0.06	0.33	2.23	0.35	0.07	0.07	0.07	0.23	0.52	0.35	0.09	0.04	0.04	0.06
20-24	0.11	0.02	0.02	0.35	1.69	0.26	0.19	0.11	0.08	0.23	0.46	0.21	0.15	0.11	0.08
25-29	0.22	0.13	0.03	0.07	0.27	0.76	0.38	0.18	0.11	0.13	0.24	0.46	0.21	0.07	0.09
30-34	0.44	0.41	0.09	0.06	0.22	0.43	0.68	0.38	0.18	0.15	0.21	0.27	0.36	0.20	0.19
35-39	0.44	0.67	0.50	0.10	0.14	0.23	0.42	0.80	0.40	0.19	0.18	0.16	0.24	0.22	0.10
40-44	0.46	0.70	0.89	0.37	0.13	0.16	0.24	0.48	0.80	0.45	0.28	0.19	0.20	0.19	0.14
45-49	0.09	0.21	0.52	0.82	0.35	0.19	0.20	0.22	0.45	0.59	0.45	0.23	0.10	0.13	0.18
50-54	0.14	0.07	0.13	0.46	0.60	0.30	0.23	0.18	0.24	0.38	0.78	0.34	0.31	0.14	0.17
55-59	0.08	0.11	0.03	0.12	0.26	0.55	0.29	0.16	0.15	0.18	0.32	0.60	0.46	0.13	0.15
60-64	0.25	0.12	0.12	0.05	0.19	0.26	0.40	0.24	0.16	0.08	0.31	0.48	0.64	0.35	0.17
65-69	0.12	0.23	0.10	0.07	0.14	0.12	0.28	0.28	0.20	0.14	0.18	0.18	0.44	0.82	0.36
70 +	0.36	0.32	0.33	0.23	0.31	0.34	0.65	0.31	0.36	0.47	0.53	0.49	0.53	0.87	0.94

**Fig. 2** Contact matrix on weekdays and weekend days. **a** Average weekday and **b** weekend contact rate with discrete gradations. Age-dependent contact heterogeneity is approximately captured by these matrices. The number in each cell represents the contact rate per person

household contacts accounted for 35 and 50% of contacts on weekday and weekend days, respectively.

An age-dependent contact matrix by 5-year age groups is shown in Fig. 2. The assortativity coefficient, measured using Newman's assortativity index or equivalently, Pearson's moment correlation, was 0.27 and 0.18 for weekday and weekend contacts, respectively, clearly indicating that the mixing pattern was highly age assortative. From this difference and the data in Additional files 2 and 3: Tables S2 and S3, it can be seen that the assortative contacts stemmed from contacts taking place at schools or workplaces. In addition to the contacts within similar age groups, we identified the relatively high contact between children and those aged from 25 to 40 years, representing household contacts. The density of household contacts was elevated on weekend days.

Figure 3 shows contacts classified into those in households and the community. The non-household contacts indicate that assortative or within-age group contacts dominated the overall contacts (Fig. 3a). However, in households, contacts between parents and children dominated the overall contacts (Fig. 3b). Contacts in the community are likely to be non-physical, whereas those in households are frequently physical contacts.

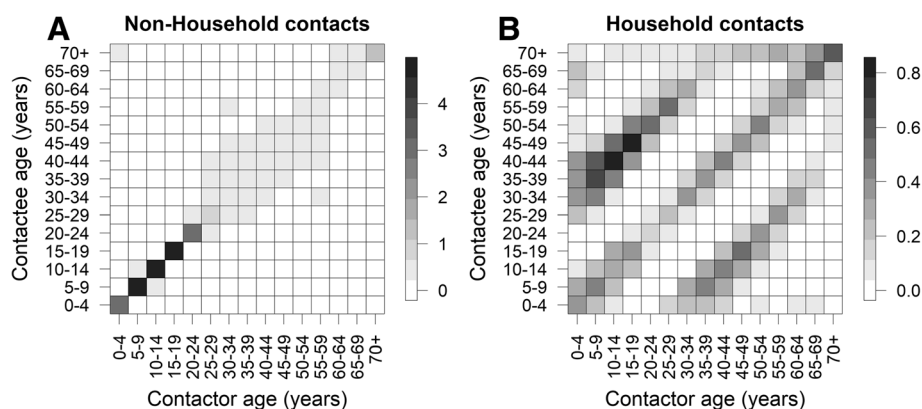
Analysing the difference in the seropositive proportions between 2009 and 2010 (Additional files 1: Table S1), we estimated the next-generation matrix using the above-mentioned weekday contact matrix. Figure 4 shows a comparison of the observed fraction infected against the predicted cumulative incidence, using a variety of contact heterogeneity methods. The highest proportion infected was seen in adolescents aged 10–14 years (64.0%) and the lowest proportion was among elderly people aged 60–64 years (13.5%). The overall cumulative incidence, weighted by age-specific population, was 36.7%. Table 1 shows results

from the model comparison. Without accounting for age-dependent heterogeneity,  $R_0$  was estimated at 1.25 (95% CI: 1.24, 1.26), but the AIC was as large as 819. When we accounted for age-dependent heterogeneity without using the contact matrix, the AIC was reduced, but it remained unclear whether the estimated heterogeneity sufficiently captured the age-dependent heterogeneous transmission patterns. Using the social contact matrix, but without adjusting age-dependent susceptibility, the AIC was elevated to 458. Using the matrix and adjusting for age-dependent susceptibility, a minimum AIC of 209 was obtained, and  $R_0$  was estimated to be 1.45 (95% CI: 1.42, 1.49). Owing to maximum use of the degrees of freedom, AIC for separable mixing and that using the contact matrix with adjustment of age-dependent susceptibility were the same to the second decimal place, but the latter yielded the minimum value with greater precision. On the other hand, when infectiousness was varied with age, the resulting improvement of AIC was not as large as that of age-dependent susceptibility.

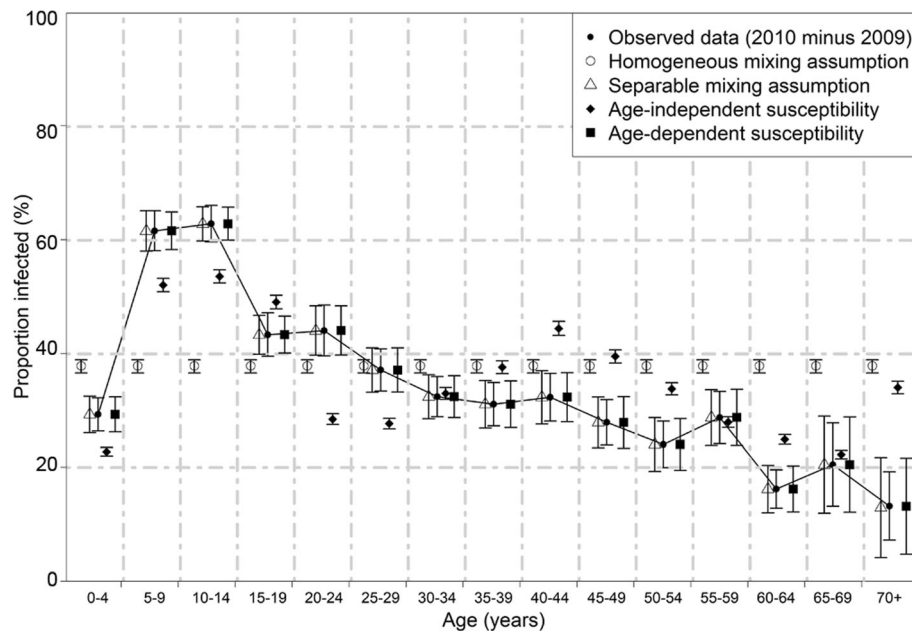
Figure 5 shows the estimated age-dependent relative susceptibility, indicating a critical need to not rely on the reported contact; age-dependent susceptibility must be considered, to capture the observed age-dependent heterogeneous patterns of transmission. Larger estimates of relative susceptibility were obtained among those aged from 20 to 29 years compared with children, perhaps reflecting limited representation of the social contact among adults to capture the actual transmission of influenza. There were large variations in relative susceptibility estimates among elderly, but it is plausible that those variations were induced by sampling error.

## Discussion

In the present study, we performed a social contact survey in Japan, quantifying an age-dependent contact



**Fig. 3** Contact matrix within the household and in the community (weekday contact). Colour bars indicate the mean number of contacts. **a** Non-household contact matrix represents the estimated mean number of contacts per day between respondents (i.e., survey participant) and persons other than their household members. **b** Household contact matrix represents the estimated mean number of contacts per day between respondents (i.e., survey participants) and their household members. Household is defined as the same unit of living space, and household members are the individuals who share that living space, regardless of blood relationship



**Fig. 4** Comparison between observed and estimated age-specific proportions of infected individuals during 2009 influenza A (H1N1) pandemic. Age-specific proportions of infection, or the so-called population attack rate or final size, during the 2009 influenza A (H1N1) pandemic, illustrated by age. Estimates were obtained by imposing various assumptions of age-dependent contact patterns, including homogeneous (or random) mixing, separable mixing (i.e., contributions of contactor and contactee are separable), age-independent susceptibility (i.e., the contact matrix was used, but the entire next-generation matrix was assumed proportional to that matrix), and age-dependent susceptibility (i.e., contact matrix plus age-dependent susceptibility per contact were estimated)

matrix by age group and also examining the characteristics according to location of contact. Weekday contacts were frequently seen in schools or workplaces whereas weekend contacts mostly took place in households. We identified strong age-dependent assortativity, especially among children, implying the potential effectiveness of school closure. A high rate of contact was also identified between school-age children and the age groups of their parents or guardians. The contact matrix was used to quantify heterogeneous transmission pattern by age, and we validated its usefulness by analysing seroepidemiological datasets of influenza A (H1N1) 2009. Using the social contact data, we obtained the minimum AIC, yielding an estimate of  $R_0$  for Japan at 1.45 (95% CI: 1.42, 1.49).

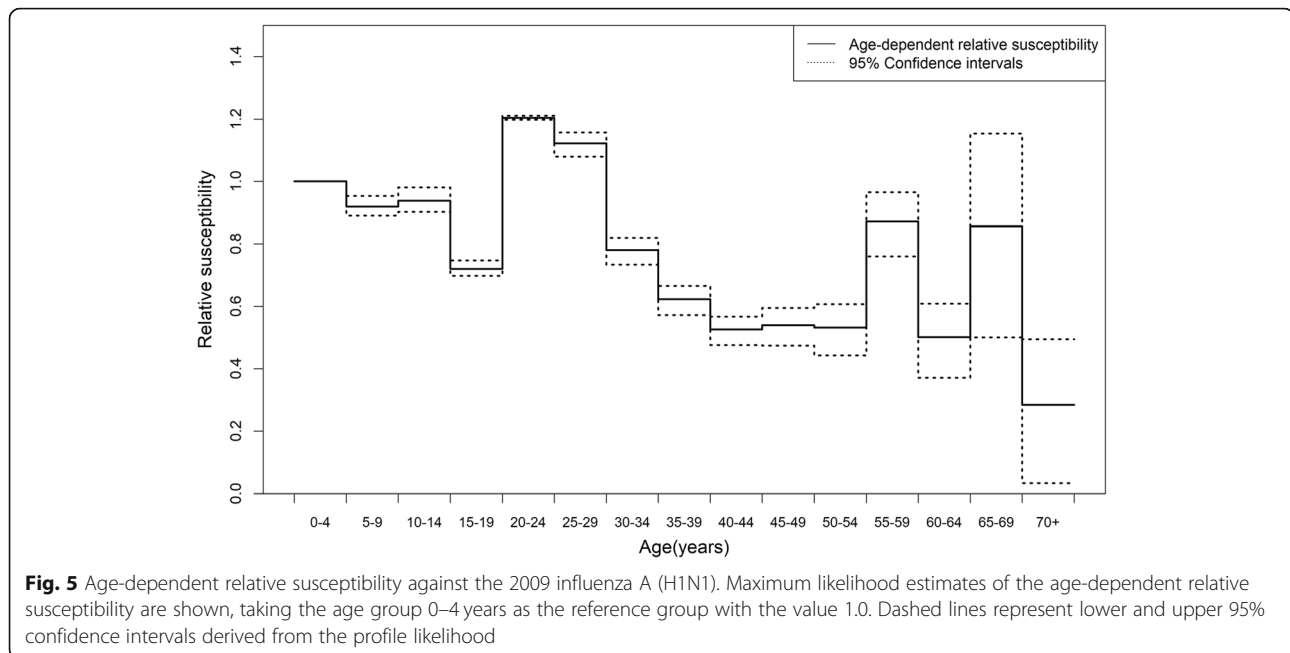
There have been numerous studies conducted on social contact [2, 6–13, 17–23, 30–60], including an earlier study in Japan [17] and a systematic review [30]. In the present study, we aimed to validate social contacts in Japan using infectious disease data. We specifically investigated age-dependent cumulative incidence using seroepidemiological data and estimating a next-generation matrix, which would be key to real-time interpretation of epidemiological dynamics [61–64]. We thereby demonstrated the importance of accounting for age-dependent susceptibility. It is important to consider variations by survey setting and method. In fact, depending on geographic location, a statistical study using a Bayesian hierarchical model demonstrated that the assortative contact pattern can vary by location of contact [31] (e.g., age-dependent assortativity in workplaces was

**Table 1** Comparison of model fit and estimated parameters

Contact pattern		$R_0^\ddagger$ (95% confidence interval)	Number of parameters	AIC <sup>†</sup>
Use social contact matrix	Plus age-independent susceptibility	1.34 (1.33, 1.35)	1	457.5
	Plus age-dependent susceptibility	1.45 (1.42, 1.49)	15	208.6
Separable mixing		1.40 (1.38, 1.44)	15	208.6
Homogeneous mixing		1.25 (1.24, 1.26)	1	818.9

<sup>†</sup> Akaike information criterion, calculated as the 2NLL + 2param where NLL and param represent negative log-likelihood and the number of parameters, respectively

<sup>‡</sup> Basic reproduction number,  $R_0$ , interpreted as the average number of secondary cases generated by a single primary case in a fully susceptible population. When the age-dependent heterogeneity is taken into account,  $R_0$  is calculated as the dominant eigenvalue of the next-generation matrix.



the least assortative, and inter-generational contact was more frequent in Asia). Keeping in mind the possibility of such variations, the present study focused on the use of social contact data that were validated using epidemiological data of influenza, which has been missing in the literature.

The present study findings support known evidence from the POLYMOD surveys. First, in our study and also in other published studies, age-dependent contact was highly assortative, and quantifying this pattern was key to appropriately estimating age-dependent interventions, including age-dependent supplementary immunization. Second, physical and non-physical contacts were surveyed, and we observed that physical contacts were common in households whereas non-physical contacts were frequent in schools and workplaces. Physical contact may be the key to understanding transmission via contact (e.g., household transmission of Ebola virus disease), whereas non-physical contact may mirror the pattern of respiratory diseases. Different patterns between weekends and weekdays highlight the possible effectiveness of school closure on influenza transmission [32]. These are useful notions for parameterizing models to predict future incidence of infectious diseases [5].

Compared with Ibuka et al. [17], there are three major differences in our study. First, the contact matrix was validated using infectious disease data. Second, we performed age-dependent area sampling. Sampling appropriate households according to age has been identified as key for successfully capturing overall patterns of contact [37]. Third, with respect to the overall frequency of contact, the absolute number of contacts in the present

study was smaller than that in Ibuka et al. [17]; the earlier study implemented a diary-based survey whereas ours relied on daily input of contacts via a website. As indicated elsewhere [37], a paper-based diary is more efficient for recording contacts than an electronic device. Improvement of the sampling scheme and data collection method remain open research subjects.

Whereas in the present study, we successfully quantified the social contact matrix, three technical issues of the social contact matrix must be noted, from the viewpoint of contact networks. First, whereas we devised the approximate matrix without individual identity, it would be useful to study how individual-based contacts affect infectious disease transmission by investigating detailed topological features as well as dynamic behaviour of the network. In addition, contact duration was not incorporated into the estimated contact matrix, which is our future aim. Second, conventional survey-based contact data collection can mostly offer only static network data and may be subject to human error, including reporting and recall biases. Moreover, such surveys may be time consuming and costly. Establishing a low-cost alternative method using novel or existing technologies is warranted [14, 16].

Four study limitations must be noted. First, the sample size of the present study was limited, with fewer than 3000 participants, smaller than that of Ibuka et al. [17]. Second, whereas we explicitly chose to use area sampling, sampling was conducted by prefecture; owing to the proportional nature, we were unable to compare differences between urban and rural locations. Third, the actual definition of contact can be very broad; however,

it must be remembered that social contact surveys involve arbitrariness in defining the contact in each case. Fourth, the survey included reporting bias, e.g., respondents were mostly sampled from among homemakers registered with an internet survey company, and the extent of responses was dependent on the respondent-driven sampling of household members and respondents' computer literacy.

We must emphasize that the present study successfully adds validated evidence of the social contact matrix to the literature, identifying common features in the published POLYMOD surveys. These datasets will be useful for parameterizing the heterogeneous transmission model of infectious diseases in Japan in the future.

## Conclusions

In the present study, we conducted a social contact survey in Japan, validating an age-dependent contact matrix according to age group and investigating the characteristics according to location of contact. Strong age-dependent assortativity was identified, especially among children, and a high rate of contact was also identified between school-age children and the age groups of their parents/guardians. Survey datasets will be useful for parameterizing the heterogeneous transmission model of various directly-transmitted infectious diseases in Japan.

## Additional files

**Additional file 1: Table S1.** Calculation of the cumulative incidence of 2009 pandemic influenza A (H1N1). (DOCX 14 kb)

**Additional file 2: Table S2.** Summary statistics of weekday contacts in Japan. (DOCX 16 kb)

**Additional file 3: Table S3.** Summary statistics of weekend contacts in Japan. (DOCX 16 kb)

**Additional file 4: Figure S1.** Proportion of physical and non-physical contacts by contact duration and location of contact (weekdays). A) Proportion of physical and non-physical contacts that took place in different locations within a day. B) Proportion of physical and non-physical contacts per day against contact duration. (TIFF 219 kb)

## Abbreviations

AIC: Akaike Information Criterion; CI: Confidence interval

## Acknowledgements

LM acknowledges Dr. Shinya Tsuzuki for his support in planning the analysis. We thank Analisa Avila, ELS, of Edanz Group ([www.edanzediting.com/ac](http://www.edanzediting.com/ac)) for editing a draft of this manuscript.

## Funding

LM's study was supported by the Japan Society for the Promotion of Science (JSPS) Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers. HN received funding support from the Japan Agency for Medical Research and Development (JP18fk0108050); Japan Society for the Promotion of Science KAKENHI (Grant Numbers 16KT0130, 17H04701, 17H05808 and 18H04895); Health and Labour Sciences Research Grant (H28-AIDS-General-001); the Inamori Foundation, the Telecommunication Advancement Foundation; and the Japan Science and Technology Agency (JST) CREST program (JPMJCR1413). The funders had no

role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Availability of data and materials

Seroepidemiological surveillance data can be accessed online using a linked URL [28]. A summary of the contact datasets that were reported in the present study can be obtained from the corresponding author upon request.

## Authors' contributions

HN conceived the study and collected the data. HN and LM conceptualized the study design, formulated the mathematical model, and performed statistical analyses. YA assisted with the statistical analyses using R software. HN and LM drafted an early version of the manuscript, and LM drafted the figures. All authors provided comments on the revised manuscript and approved the final version of the manuscript.

## Authors' information

The authors are experts with an interest in infectious disease epidemiology and also theoretical epidemiology, and the corresponding author is the chairperson and team leader of the Department of Hygiene, Hokkaido University Graduate School of Medicine.

## Ethics approval and consent to participate

The purpose of the study was explained to participants and they were ensured that the extent of use of the survey information was limited to the present study. Informed consent was obtained via the internet webpage, and participants had the right to withdraw at any time during the study period. The Medical Ethics Committees at the Graduate School of Medicine, The University of Tokyo approved this study (approval ID: 10478). As for the seroepidemiological data, we used publicly available data in the present study [28]. The datasets had already been fully anonymized and did not include any identifiable information. Thus, ethical approval was not required for the analysis of seroepidemiological data.

## Consent for publication

Not applicable – all details relating to participants were de-identified prior to inclusion in this study.

## Competing interests

The authors declare that co-author H. Nishiura is the Editor-in-Chief of *Theoretical Biology and Medical Modelling*. This does not alter the authors' adherence to all policies of *Theoretical Biology and Medical Modelling* on sharing data and materials.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 10 December 2018 Accepted: 5 March 2019

Published online: 20 March 2019

## References

1. Diekmann O, Heesterbeek JA, Metz JA. On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations. *J Math Biol.* 1990;28:365–82.
2. Edmunds WJ, O'Callaghan CJ, Nokes DJ. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc Biol Sci.* 1997;264:949–57.
3. Farrington CP, Kanaan MN, Gay NJ. Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *J R Stat Soc Ser C.* 2001;50:251–92.
4. Diekmann O, Heesterbeek JA, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. *J R Soc Interface.* 2010;7:873–85.
5. Grenfell BT, Anderson RM. The estimation of age-related rates of infection from case notifications and serological data. *J Hyg.* 1985;95:419–36.
6. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Hejblum J, Sadkowska-Todys M, Rosinska M, Edmunds WJ. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* 2008;5:e74.



7. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am J Epidemiol*. 2006;164:936–44.
8. Leung K, Jit M, Lau EHY, Wu JT. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Sci Rep*. 2017;7:7974–86.
9. Read JM, Lessler J, Riley S, Wang S, Tan LJ, Kwok KO, Guan Y, Jiang CQ, Cummings DA. Social mixing patterns in rural and urban areas of southern China. *Proc R Soc Lond Ser B*. 2014;281:20140268.
10. Barrat A, Cattuto C, Tozzi AE, Vanhems P, Voirin N. Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clin Microbiol Infect*. 2014;20:10–6.
11. Mastrandrea R, Fournet J, Barrat A. Contact patterns in a high school: a comparison between data collected using wearable sensors, Contact Diaries and Friendship Surveys. *PLoS One*. 2015;10:e0136497.
12. Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH. A high-resolution human contact network for infectious disease transmission. *Proc Natl Acad Sci U S A*. 2010;107:22020–5.
13. Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong J, Edmunds WJ. Using time-use data to parameterize models for the spread of close-contact infectious diseases. *Am J Epidemiol*. 2008;168:1082–90.
14. Wesolowski A, Buckee CO, Engo-Monsen K, Metcalf CJE. Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data. *J Infect Dis*. 2016;214(suppl\_4):S414–20.
15. Krauer F, Gsteiger S, Low N, Hansen CH, Althaus CL. Heterogeneity in district-level transmission of Ebola virus disease during the 2013–2015 epidemic in West Africa. *PLoS Negl Trop Dis*. 2016;10:e0004867.
16. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big Data for Infectious Disease Surveillance and Modeling. *J Infect Dis*. 2016;214(suppl\_4):S375–9.
17. Ibuka Y, Ohkusa Y, Sugawara T, Chapman GB, Yamin D, Atkins KE, Taniguchi K, Okabe N, Galvani AP. Social contacts, vaccination decisions and influenza in Japan. *J Epidemiol Community Health*. 2016;70:162–7.
18. Horby P, Pham QT, Hens N, Nguyen TT, Le QM, Dang DT, Nguyen ML, Nguyen TH, Alexander N, Edmunds WJ, Tran ND, Fox A, Nguyen TH. Social contact patterns in Vietnam and implications for the control of infectious diseases. *PLoS One*. 2011;6:e16965.
19. Fu YC, Wang DW, Chuang JH. Representative contact diaries for modeling the spread of infectious diseases in Taiwan. *PLoS One*. 2012;7:e45113.
20. Kwok KO, Cowling B, Wei V, Riley S, Read JM. Temporal variation of human encounters and the number of locations in which they occur: a longitudinal study of Hong Kong residents. *J R Soc Interface*. 2018;15.
21. Chen SC, You ZS. Social contact patterns of school-age children in Taiwan: comparison of the term time and holiday periods. *Epidemiol Infect*. 2015;143:1139–47.
22. Stein ML, van Steenbergen JE, Chanyasanha C, Tipayamongkhogul M, Buskens V, van der Heijden PG, Sabaiwan W, Bengtsson L, Lu X, Thorson AE, Kretzschmar ME. Online respondent-driven sampling for studying contact patterns relevant for the spread of close-contact pathogens: a pilot study in Thailand. *PLoS One*. 2014;9:e85256.
23. Stein ML, van der Heijden PGM, Buskens V, van Steenbergen JE, Bengtsson L, Koppeschaar CE, Thorson A, Kretzschmar MME. Tracking social contact networks with online respondent-driven detection: who recruits whom? *BMC Infect Dis*. 2015;15:522–34.
24. Mizumoto K, Yamamoto T, Nishiura H. Age-dependent estimates of the epidemiological impact of pandemic influenza (H1N1–2009) in Japan. *Comput Math Methods Med*. 2013;2013:637064.
25. Nishiura H, Oshitani H. Household transmission of influenza (H1N1–2009) in Japan: age-specificity and reduction of household transmission risk by zanamivir treatment. *J Int Med Res*. 2011;39:619–28.
26. Nishiura H, Oshitani H. Effects of vaccination against pandemic (H1N1) 2009 among Japanese children. *Emerg Infect Dis*. 2011;17:746–7.
27. Nishiura H, Chowell G, Safan M, Castillo-Chavez C. Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza a (H1N1) 2009. *Theor Biol Med Model*. 2010;7:1–9.
28. National Institute of Infectious Diseases, Japan. National Epidemiological Surveillance of Vaccine-Preventable Diseases (NESVPD). In: Tokyo: National Institute of Infectious Diseases; 2018. Available from: <https://www.niid.go.jp/niid/ja/yosoku-index.html>.
29. Endo A, Ejima K, Nishiura H. Capturing the transmission dynamics of the 2009 Japanese pandemic influenza H1N1 in the presence of heterogeneous immunity. *Ann Epidemiol*. 2018;28:293–300.
30. Hoang TV, Coletti P, Melegaro A, Wallinga J, Grijalva C, Edmunds J, Beutels P, Hens N. A systematic review of social contact surveys to inform transmission models of close contact infections. *bioRxiv*. 2018:292235. <https://doi.org/10.1101/292235>.
31. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput Biol*. 2017;13:e1005697.
32. Hens N, Ayele GM, Goeyvaerts N, Aerts M, Mossong J, Edmunds JW, Beutels P. Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. *BMC Infect Dis*. 2009;9:187–99.
33. Kretzschmar M, Mikolajczyk RT. Contact profiles in eight European countries and implications for modelling the spread of airborne infectious diseases. *PLoS One*. 2009;4:e5931.
34. Mikolajczyk RT, Kretzschmar M. Collecting social contact data in the context of disease transmission: prospective and retrospective study designs. *Soc Netw*. 2008;30:127–35.
35. Rolls DA, Geard NL, Warr DJ, Nathan PM, Robins GL, Pattison PE, McCaw JM, McVernon J. Social encounter profiles of greater Melbourne residents, by location—a telephone survey. *BMC Infect Dis*. 2015;15:494–505.
36. Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. *PLoS One*. 2014;9:e104786.
37. McCaw JM, Forbes K, Nathan PM, Pattison PE, Robins GL, Nolan TM, McVernon J. Comparison of three methods for ascertainment of contact information relevant to respiratory pathogen transmission in encounter networks. *BMC Infect Dis*. 2010;10:166–77.
38. Danon L, House TA, Read JM, Keeling MJ. Social encounter networks: collective properties and disease transmission. *J R Soc Interface*. 2012;9:2826–33.
39. Eames KTD, Tilston NL, Edmunds WJ. The impact of school holidays on the social mixing patterns of school children. *Epidemics*. 2011;3:103–8.
40. Read JM, Eames KT, Edmunds WJ. Dynamic social networks and the implications for the spread of infectious disease. *J R Soc Interface*. 2008;5:1001–7.
41. van de Kasstele J, van Eijkeren J, Wallinga J. Efficient estimation of age-specific social contact rates between men and women. *Ann Appl Stat*. 2017;11:320–39.
42. Bernard H, Fischer R, Mikolajczyk RT, Kretzschmar M, Wildner M. Nurses' contacts and potential for infectious disease transmission. *Emerg Infect Dis*. 2009;15:1438–44.
43. Glass LM, Glass RJ. Social contact networks for the spread of pandemic influenza in children and teenagers. *BMC Public Health*. 2008;8:61–76.
44. DeStefano F, Haber M, Currivan D, Farris T, Burrus B, Stone-Wiggins B, McCalla A, Guled H, Shih H, Edelson P, Wetterhall S. Factors associated with social contacts in four communities during the 2007–2008 influenza season. *Epidemiol Infect*. 2011;139:1181–90.
45. Potter GE, Handcock MS, Longini IM, Halloran ME. Estimating within-school contact networks to understand influenza transmission. *Ann Appl Stat*. 2012;6:1–26.
46. Smieszek T. A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread. *Theor Biol Med Model*. 2009;6:25–35.
47. Eames KTD, Tilston NL, Brooks-Pollock E, Edmunds WJ. Measured dynamic social contact patterns explain the spread of H1N1v influenza. *PLoS Comput Biol*. 2012;8:e1002425.
48. Jackson C, Mangtani P, Vynnycky E, Fielding K, Kitching A, Mohamed H, Roche A, Maguire H. School closures and student contact patterns. *Emerg Infect Dis*. 2011;17:245–7.
49. Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LD, Bekker LG, Wood R. Social mixing patterns within a south African township community: implications for respiratory disease transmission and control. *Am J Epidemiol*. 2011;174:1246–55.
50. Béraud G, Kazmierczak S, Beutels P, Levy-Bruhl D, Lenne X, Mielcarek N, Yazdanpanah Y, Boëlle PY, Hens N, Dervaux B. The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. *PLoS One*. 2015;10:e0133203.
51. Potter GE, Smieszek T, Sailer K. Modeling workplace contact networks: the effects of organizational structure, architecture, and reporting errors on epidemic predictions. *Netw Sci*. 2015;3:298–325.
52. Vino T, Singh GR, Davison B, Campbell PT, Lydeamore MJ, Robinson A, McVernon J, Tong SYC, Geard N. Indigenous Australian household structure:

- a simple data collection tool and implications for close contact transmission of communicable diseases. *PeerJ*. 2017;5:e3958.
53. Dodd PJ, Looker C, Plumb ID, Bond V, Schaap A, Shanaube K, Muyoyeta M, Vynnycky E, Godfrey-Faussett P, Corbett EL, Beyers N, Ayles H, White RG. Age- and sex-specific social contact patterns and incidence of mycobacterium tuberculosis infection. *Am J Epidemiol*. 2015;183:156–66.
  54. Grijalva CG, Goeyvaerts N, Verastegui H, Edwards KM, Gil AI, Lanata CF, Hens N. A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru. *PLoS One*. 2015;10:e0118457.
  55. Kiti MC, Tizzoni M, Kinyanjui TM, Koech DC, Munywoki PK, Meriac M, Cappa L, Panisson A, Barrat A, Cattuto C, Nokes DJ. Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. *EPJ Data Sci*. 2016;5:21–42.
  56. Strömgren M, Holm E, Dahlström Ö, Ekberg J, Eriksson H, Spreco A, Timpka T. Place-based social contact and mixing: a typology of generic meeting places of relevance for infectious disease transmission. *Epidemiol Infect*. 2017;145:2582–93.
  57. van Hoek AJ, Andrews N, Campbell H, Amirthalingam G, Edmunds WJ, Miller E. The social life of infants in the context of infectious disease transmission; social contacts and mixing patterns of the very young. *PLoS One*. 2013;8:e76180.
  58. Aiello AE, Simanek AM, Eisenberg MC, Walsh AR, Davis B, Volz E, Cheng C, Rainey JJ, Uzicanin A, Gao H, Osgood N, Knowles D, Stanley K, Tarter K, Monto AS. Design and methods of a social network isolation study for reducing respiratory infection transmission: the eX-FLU cluster randomized trial. *Epidemics*. 2016;15:38–55.
  59. Melegaro A, Del Fava E, Poletti P, Merler S, Nyamukapa C, Williams J, Gregson S, Manfredi P. Social contact structures and time use patterns in the Manicaland province of Zimbabwe. *PLoS One*. 2017;12:e0170459.
  60. Ajelli M, Litvinova M. Estimating contact patterns relevant to the spread of infectious diseases in Russia. *J Theor Biol*. 2017;419:1–7.
  61. Yamamoto N, Ejima K, Nishiura H. Modelling the impact of correlations between condom use and sexual contact pattern on the dynamics of sexually transmitted infections. *Theor Biol Med Model*. 2018;15:6.
  62. Sakamoto Y, Yamaguchi T, Yamamoto N, Nishiura H. Modeling the elevated risk of yellow fever among travelers visiting Brazil, 2018. *Theor Biol Med Model*. 2018;15:9.
  63. Nishiura H, Tsuzuki S, Yuan B, Yamaguchi T, Asai Y. Transmission dynamics of cholera in Yemen, 2017: a real time forecasting. *Theor Biol Med Model*. 2017;14:14.
  64. Nishiura H, Endo A, Saitoh M, Kinoshita R, Ueno R, Nakaoka S, Miyamatsu Y, Dong Y, Chowell G, Mizumoto K. Identifying determinants of heterogeneous transmission dynamics of the Middle East respiratory syndrome (MERS) outbreak in the Republic of Korea, 2015: a retrospective epidemiological analysis. *BMJ Open*. 2016;6:e009936.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

