

OPEN

# Patterns of human social contact and contact with animals in Shanghai, China

Juanjuan Zhang<sup>1</sup>, Petra Klepac<sup>2</sup>, Jonathan M. Read<sup>3</sup>, Alicia Rosello<sup>2</sup>, Xiling Wang<sup>1</sup>, Shengjie Lai<sup>1,4,5</sup>, Meng Li<sup>1</sup>, Yujian Song<sup>1</sup>, Qingzhen Wei<sup>1</sup>, Hao Jiang<sup>1</sup>, Juan Yang<sup>1</sup>, Henry Lynn<sup>1</sup>, Stefan Flasche<sup>2</sup>, Mark Jit<sup>2,6,7</sup> & Hongjie Yu<sup>1\*</sup>

East Asia is as a principal hotspot for emerging zoonotic infections. Understanding the likely pathways for their emergence and spread requires knowledge on human-human and human-animal contacts, but such studies are rare. We used self-completed and interviewer-completed contact diaries to quantify patterns of these contacts for 965 individuals in 2017/2018 in a high-income densely-populated area of China, Shanghai City. Interviewer-completed diaries recorded more social contacts (19.3 vs. 18.0) and longer social contact duration (35.0 vs. 29.1 hours) than self-reporting. Strong age-assortativity was observed in all age groups especially among young participants (aged 7–20) and middle aged participants (25–55 years). 17.7% of participants reported touching animals (15.3% (pets), 0.0% (poultry) and 0.1% (livestock)). Human-human contact was very frequent but contact with animals (especially poultry) was rare although associated with frequent human-human contact. Hence, this densely populated area is more likely to act as an accelerator for human-human spread but less likely to be at the source of a zoonosis outbreak. We also propose that telephone interview at the end of reporting day is a potential improvement of the design of future contact surveys.

Most of the major global infectious disease threats of the last decade (including epidemics of pandemic influenza A/H1N1, Middle East respiratory syndrome coronavirus and Ebola) have emerged from pathogens crossing the species barrier from animals to humans. A principal hotspot for such disease emergence is the megacities of East Asia<sup>1</sup>. For these diseases, successful establishment in human populations is dictated by both animal-to-human (most notably with domesticated animals raised in agricultural settings such as poultry, swine and cattle) and human-to-human contact<sup>2</sup>. However, studies measuring and characterizing contacts on human-animal interface and subsequent transmission among humans are scarce. For emerging pathogens with established human-human transmission, the spread or even outbreak in humans will be driven by patterns of human encounters, which can also determine the effectiveness of interventions against them (e.g. vaccination, contact tracing and social distancing)<sup>3</sup>. Patterns of human encounters are location specific, therefore improving local intervention strategies relies on understanding local patterns of social contact<sup>4–8</sup>. There are still, however, relatively few empirical studies of age-specific social contacts which integrate location, animal contact data, and the interaction between human-human and human-animal contact. Furthermore, an evaluation of public health strategies that is based on social contact information derived from a different country that might not be representative of local mixing and may result in erroneous conclusions<sup>9</sup>.

China is a large, diverse country with rapid economic development, high urbanization, and frequent interaction between humans and animals. The richest and most populous city is Shanghai with a population of 24 million. We used two types of questionnaire delivery (self-completed and telephone interview) to quantify social contact patterns in Shanghai City, which is a hub for spreading infectious diseases due to its high population density and connectivity of the air transportation network.

<sup>1</sup>School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai, China. <sup>2</sup>Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Public Health, London School of Hygiene and Tropical Medicine, London, UK. <sup>3</sup>Centre for Health Informatics, Computation and Statistics, Lancaster Medical School, Lancaster University, Lancashire, UK. <sup>4</sup>WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, UK. <sup>5</sup>Flowminder Foundation, Stockholm, Sweden. <sup>6</sup>Modelling and Economics Unit, Public Health England, London, UK. <sup>7</sup>School of Public Health, University of Hong Kong, Hong Kong, China. \*email: [yhj@fudan.edu.cn](mailto:yhj@fudan.edu.cn)

Large population-based surveys of age-specific social contacts exist for several European countries<sup>10–14</sup> and are increasingly conducted in low- and middle-income Asian (Thailand, Vietnam and Indonesia)<sup>15,16</sup> and African countries (Zimbabwe and South Africa<sup>17–19</sup>, Kenya<sup>20</sup>, Uganda<sup>21</sup>). There have also been similar surveys in Russia<sup>22</sup>, Australia<sup>23,24</sup>, Peru<sup>25</sup>, Japan<sup>26</sup>, Taiwan<sup>27</sup>, southern China (Guangzhou<sup>28</sup>, Hong Kong<sup>29–31</sup>). Data from such contact studies are increasingly made freely available for download<sup>32</sup>. These surveys have their limitations, so recently there have been proposals for improved designs that would measure social contact patterns with less recall bias. Recent technological developments allow the measurement of human proximity and social interactions using wearable sensors, but such surveys are limited to specific settings, such as households<sup>33</sup> or schools<sup>34</sup> and require high levels of participation within such settings to succeed. Some studies encouraged participants to record each contact as it occurred<sup>11,17,20,29</sup>, but such prospectively recording has been found difficult to do (e.g. less than 5% participants completed the questionnaire prospectively in a study by Leung *et al.*<sup>29</sup>). Two most common modes of data collection are self-reporting (in form of diary, for example) and interviewer-led (the interviewer may record information in a diary for the subject), typically collecting information retrospectively.

A possible way to minimize recall bias is to perform interviews on the day of reporting. Comparing results from such an interview to purely self-reported data allows assessment of the bias in studies that rely entirely on self-reporting without any memory aids. Such studies are rare and to our knowledge none assess consistency between self-reporting studies and telephone interviews. Another gap we identified involves characterizing and measuring human-animal contact. We identified a single study that considers human-human and human-animal contact patterns in Belgium<sup>35</sup>. However, detailed contact characteristics such as human-human contact duration, number of contact settings, human-to-animal touching duration, and contact matrix were not reported.

Our current study has three aims: (i) quantify local human-human (H-H) and human-animal (H-A) contacts in an urban setting in China, (ii) explore the interaction between H-H and H-A contact, and (iii) assess consistency between two different modes of data-collection, self-reporting and telephone interview-led studies.

## Methods

**Study site.** The survey was carried out between December 2017 and May 2018 in Shanghai, a city in the southeast of China (Supplementary Fig. S1). The central urban districts of Shanghai were selected as our study sites, representing typical urban areas with high income and high population density (Supplementary Fig. S2). The 7 central urban districts are defined as areas within the outer ring road of Shanghai<sup>36</sup>, comprising a total population size of 7.3 million and an area of 289 square kilometers.

**Sampling.** We sampled forty neighborhoods using multi-stage stratified sampling proportional to population size. A convenience sample of 25 households per neighborhood was selected so as to be broadly representative of the whole population of the neighborhood in terms of geographical distribution (Supplementary Fig. S3). One person per household was invited to participate in our study until we met our predefined target sizes equally balanced by age group and gender (Supplementary Text S1 for sampling details).

Our target sample size was 1000. Individuals of all ages who had been living in the selected neighborhoods for longer than two weeks and did not intend to move from Shanghai in the following two weeks were considered eligible for inclusion in the study (Supplementary Text S1 for population details).

**Data collection.** Participants were asked to complete a questionnaire on their contact behavior for a randomly assigned day of the week. The questionnaire (see Supplementary Text S13) consisted of three sections, including general information, human-human and human-animal contact on the assigned day. General information comprised respondent demographics, household composition, daily travel, and daily contact. In the H-H contact section, in line with POLYMOD<sup>10</sup>, H-H contact was defined as either, (1) a two-way conversation with three or more words in the physical presence of another person (conversational contact), or (2) physical skin-to-skin contact (e.g. a handshake, hug, kiss or contact sports). Participants were requested to record each contact they made on the assigned day (during the 24 hours before going to bed at the participants' normal bedtime), including the age and sex of the contact, whether contact was conversational or physical, the contact duration, the contact setting (e.g. home, work, etc.) and the frequency with which the respondent usually contacts the individual (e.g. daily, weekly, etc.).

In the H-A contact section, participants were first requested to record the species and the number of animals they owned. Animal ownership was defined as having an animal in the household in which they spend the majority of their time (i.e. living together in the house) as an alternative to touching, given that both are proxies for events that could lead to animal-human transmission. We only considered physical contact with animals, where contact was defined as touching at least one living animal. Participants were asked to record every type of animal they contacted on the assigned day, including the species, the number of animals, the contact duration, the contact setting, and the frequency with which they usually contacted the animal. At the end of the questionnaire, participants were also asked to report whether or not their assigned day was a typical one, as the factor may have affected their social contact patterns.

Large numbers of contacts are difficult to record individually, which can lead to under-reporting. We define “group contact” as contact with a group of at least 20 individuals (e.g. as may be experienced by students, doctors, or those participating in a social event where one meets a lot of people). Contacts reported as individuals (i.e., not group contact) were termed “individual contact”. The questionnaire allowed participants to report a maximum number of 40 individual contacts.

Regarding the mode of data collection, there were two options for participants to choose freely: self-reporting and telephone interview. Participants were encouraged to select the telephone interview by answering a phone call from our trained investigator before their going to bed and recounting their contacts on the assigned day (whereby the investigator would complete the diary on their behalf). Alternatively, they could personally fill in a

paper-based questionnaire (with attached instruction sheet). Participants who chose to self-report were encouraged to record their contacts prospectively (i.e. as they happened), and to complete the rest of the questionnaire before going to bed on the assigned day (Supplementary Text S2 for data collection details). Here we assumed that going to bed at night is sufficiently close to the end of the day and the end of the day's social activities. Accordingly, the time period recorded was the 24 hours before going to bed.

**Statistical analysis.** *Demographic, travel and temporal factors.* Four explanatory factors were considered when describing human social contact, animal ownership and contact. (1) Demographic and socioeconomic characteristics of the participant, specifically gender, age, work type, education level, individual income, household size, and number of years of living in Shanghai. (2) Whether participants travelled daily out of the subdistrict, or only occasionally. (3) Whether the reported contact day was a weekday or at a weekend, and whether the day was considered a typical day by the participant (had a similar life pattern to that of at least four days per week). Finally, (4) other survey-related factors included the mode of data collection (self-reporting or telephone interview), and population density level of the subdistrict (low: under 20,000 people per km<sup>2</sup>, moderate: 20,001–50,000 people per km<sup>2</sup>, high: over 50,001 people per km<sup>2</sup>).

To test for a difference in contact patterns under two different modes of data collection, we first applied propensity score matching to reduce confounding. The propensity scores were estimated for each participant using logistic regression, with the mode of data collection as response variable and other factors mentioned above as independent variables (“MatchIt” package in R 3.5.0).

*Characteristics of human-human and human-animal contact patterns.* Given the matched sample, we used weighted generalized additive mixed models (GAMMs) to assess the effect of the mode of data collection on the number of contacts, the number of individual contacts, the probability of reporting group contact, the contact duration in hours and the number of contact settings, whilst controlling for the animal ownership and other covariates mentioned above. Of all the models, we fitted penalized splines to explore potential nonlinear relationships of continuous participant age to the response variables, as it could be significant as found in a study by Kwok *et al.*<sup>31</sup>. The sampling structure of our study allowed us to include the random intercept effect in our models to consider the proportion of variance in response variables attributable to intra- and inter-district variation. Estimating of the contact duration and the number of settings is provided in Supplementary Text S3. Similarly, we used the generalized additive mixed models to model the probability of reporting animal ownership and animal contact. Penalized splines were used to explore potential nonlinear relationships of continuous participant age and specially the number of H-H contacts to the response variables. Model details and variables selection are provided in Supplementary Text S4.

*Human-human and human-animal contact matrix.* We established 17 age classes (0–35 m, 3–6 y, 7–9 y, 10–14 y, 15–19 y, 20–24 y, 25–29 y, 30–34 y, 35–39 y, 40–44 y, 45–49 y, 50–54 y, 55–59 y, 60–64 y, 65–69 y, 70–74 y, 75 y and over) to build our age-specific H-H contact matrix, aiming to estimate the age-specific contact rate per person per day (Supplementary Text S3 for the estimating of the age of contacts). An original, symmetric and population age-weighted contact matrix was built using the “socialmixr” package in R 3.5.0. We used the  $q$  indices, representing departures from proportionate mixing and ranging from zero (proportionate) to one (fully assortative), and the bootstrapped 95% confidence intervals to assess the degree of age assortativity<sup>37</sup> (Supplementary Text S5). H-A contact matrices were built in a similar manner. The bivariate smoothing approach was used to further estimate the social contact matrix<sup>10,38,39</sup> (Supplementary Text S6).

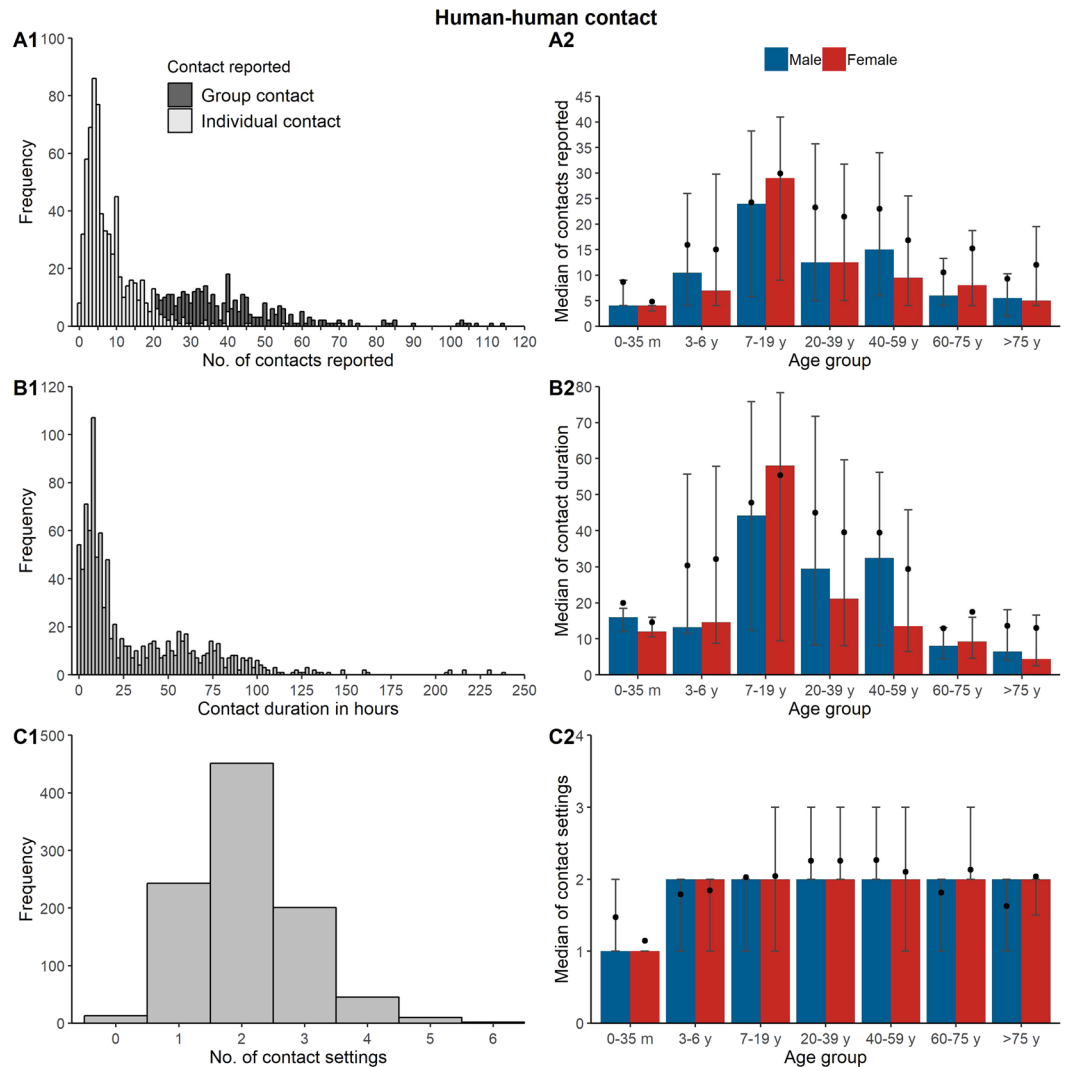
*Hypothesis testing for mode of data collection.* We used bootstrap method to test the null hypothesis that telephone interviewed respondents would report more contacts by computing the difference between human contact matrices from self-reporting and telephone interviewed respondents (Supplementary Text S7).

*Ethical approval.* The study was approved by the ethical review committee of the School of Public Health at Fudan University, Shanghai, China (Ref: 2018-01-0659S). All methods were performed in accordance with relevant guidelines and regulations. Informed consent was obtained from all subjects (from a parent/guardian if participant was below 18 years of age).

## Results

**Demographic characteristics of participants.** Overall, 1000 individuals from 1000 households were recruited, of whom 965 individuals were included in the final analysis (Supplementary Fig. S4). There was no statistical difference in gender between our sample and the central urban areas of Shanghai population (Shanghai census in 2017), but a statistical difference was found in terms of age (Supplementary Table S1). We found no difference in the age distribution and sex of participants between the participants recruited and the effective participants (Supplementary Table S2).

**Characteristics of human-human and human-animal contact patterns.** We recorded 18,116 contacts in total from 474 male respondents and 491 female respondents, including 6,953 (38.4%) individual contacts, 10,240 (56.5%) contacts reported as group contacts, and 923 (5.1%) contacts that people estimate they had left out (Supplementary Text S13; questions 17–18.). The distribution of the number of contacts was highly right-skewed with a median of 10 contacts (Fig. 1 and Supplementary Table S3). Participants aged 7–19 years reported the greatest number of contacts. Similar distribution was found for the contact duration with a median of 15.7 hours (Supplementary Fig. S5 for the estimating of the duration of group contact). The number of contact settings was also right-skewed with a median of two settings, and participants aged younger than 3 years had the



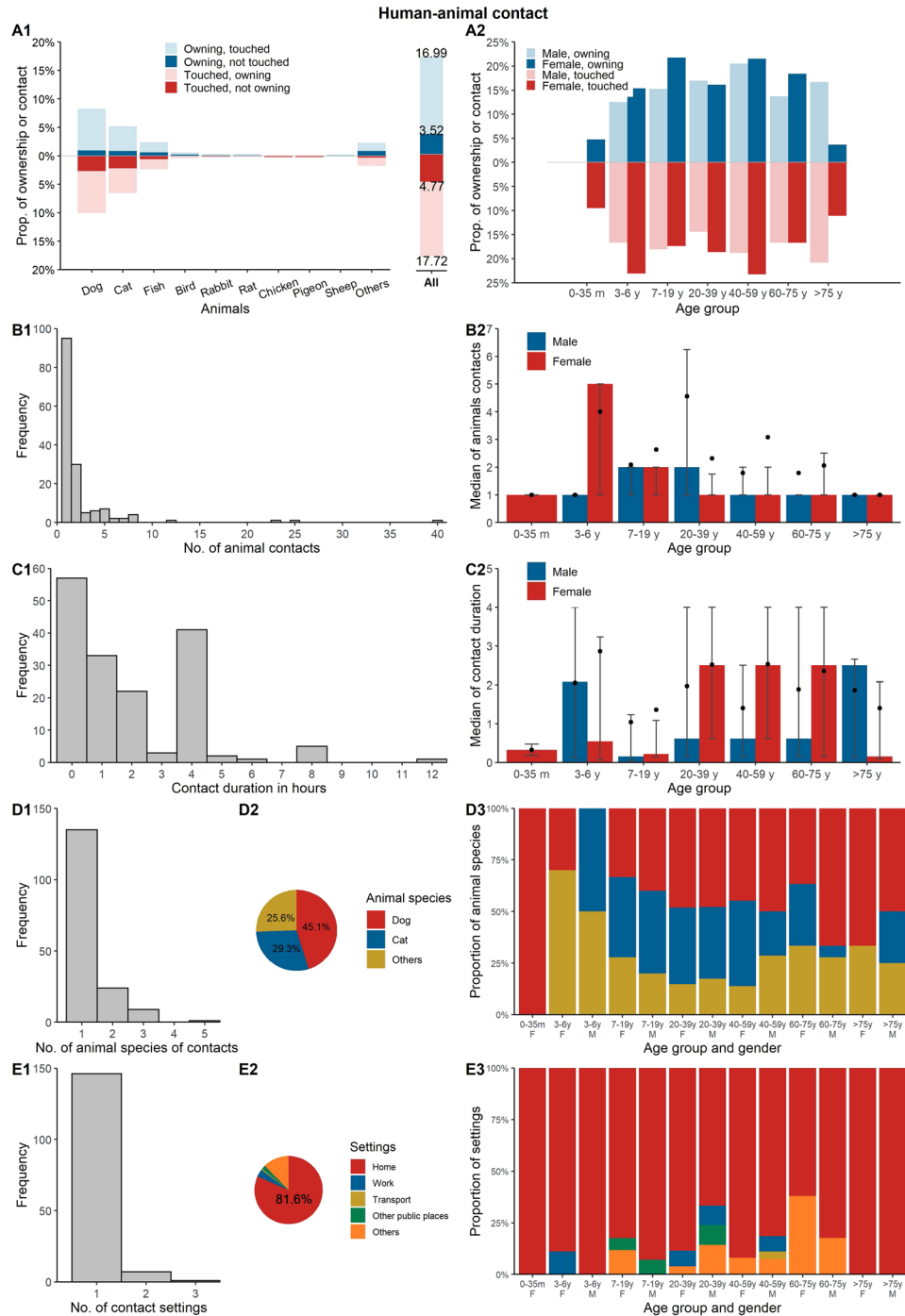
**Figure 1.** Distribution of human-human contact patterns (including the number of contacts, contact duration, and number of settings). The error bars in the right panels correspond to 25% and 75% quantiles, and the solid points correspond to the means.

smallest number of settings. There was little difference in the number of contacts, duration, and the number of settings reported in terms of participant gender. More results about individual and group contacts are provided in Supplementary Text S8, Supplementary Fig. S6, and Supplementary Table S3. Physical contacts and other contact characteristics are described in Supplementary Figs S7 and S8.

Of the 965 respondents, 164 (17.0%) and 171 (17.7%), respectively, owned and had contact with at least one animal. Most of the reported animals were companion animals: 148 (15.3%) and 159 (16.5%) participants owned and had contact with pets (not including poultry or livestock), respectively, and most of which were dogs and cats. None of the participants reported owning poultry and only 0.2% of them had a contact with poultry. Only 0.1% of participants reported owning livestock, and none of them had contact with livestock (Supplementary Table S4). Only 3.52% of respondents owned animals but had no contact with animals on the assigned day, while 4.77% respondents had contact with animals that were not their own (Fig. 2A1). Respondents aged 0–35 months and those aged over 75 were least likely to own or have contact with animals, especially if they were male (Fig. 2A2). More results about H-A contact patterns are described in Supplementary Text S9, Supplementary Tables S4 and S5 and Supplementary Figs S9 and S10.

Of the 965 respondents, 194 (20.1%) reported owning or having contact with animal(s). The characteristics of such participants interacting with animals and their H-H contact patterns are shown in Supplementary Table S6 and Supplementary Fig. S11. Such participants had more social contacts (median: 16.5) and longer contact duration (median: 17.4), especially for children aged 3–6 and males aged 20–59, which motivated us to explore the interaction of H-H and H-A contact patterns.

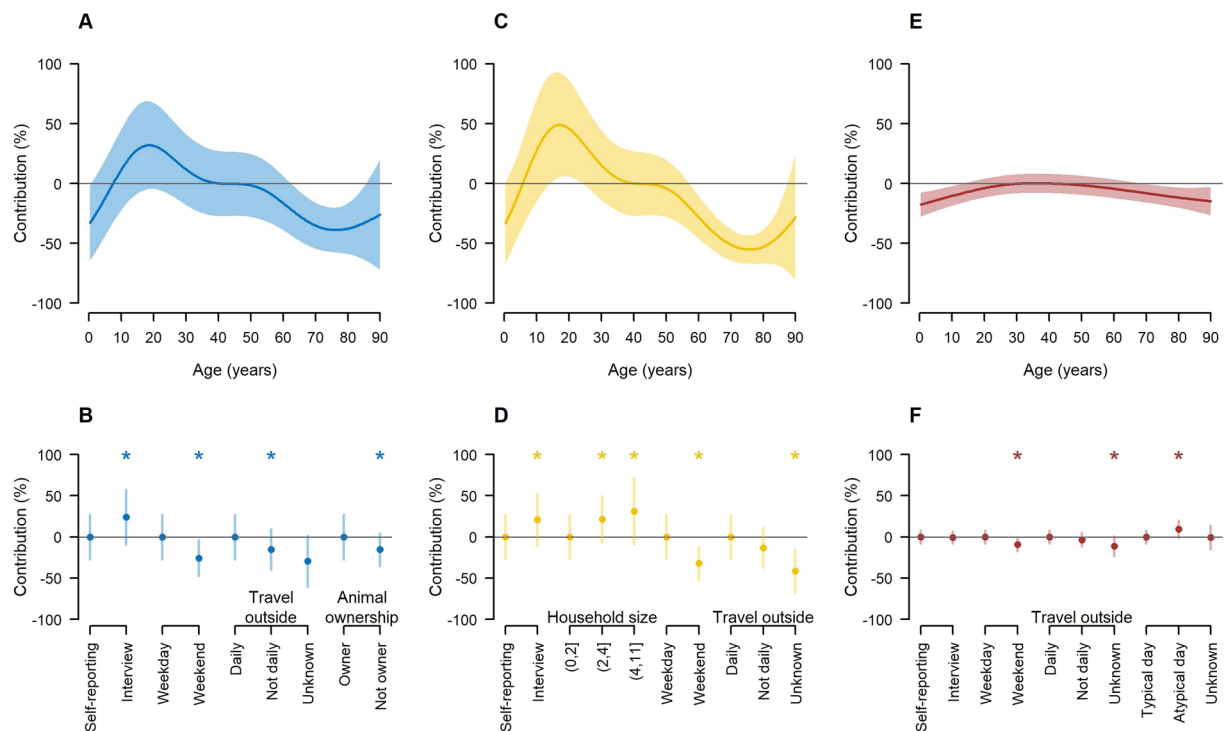
**Factors associated with human-human and human-animal contact patterns.** Participants were matched for potential explanatory factors associated with number of contacts using propensity score matching (Supplementary Table S7, Supplementary Fig. S12), resulting in 772 matched respondents (386 for each mode of



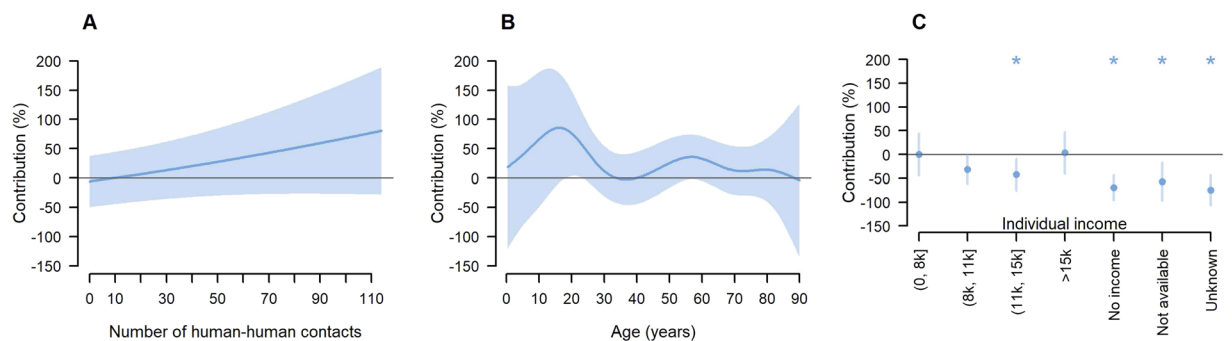
**Figure 2.** Distribution of human-animal contact patterns (including the probability of reporting animal ownership or animal contact, number of animal contacts, contact duration, number of species, and number of settings). (A1) Show the proportion of owners touching their own animals, the proportion of owned animals not touched, the proportion of touched animals that are their own and the proportion of touched animals not owned, respectively. All of the age groups and gender in the right panels refer to the humans. The error bars correspond to 25% and 75% quantiles, and the solid points correspond to the means.

data collection). Using matched data, the median of total contacts, contact duration and settings in H-H contact were 10, 15.2, and 2, respectively (Supplementary Table S8). The distribution of contact characteristics showed similar patterns as those before matching (Supplementary Fig. S13).

For the H-H contacts, we used matched data to fit GAMMs (Fig. 3, Supplementary Table S9). In the final selected models, a significantly greater number of contacts was associated with telephone interview. Workdays, daily out-of-subdistrict travel, and animal owners were associated with a significantly greater number of contacts than weekends, not daily out-of-subdistrict travel, and not animal owners. We found a significant nonlinear



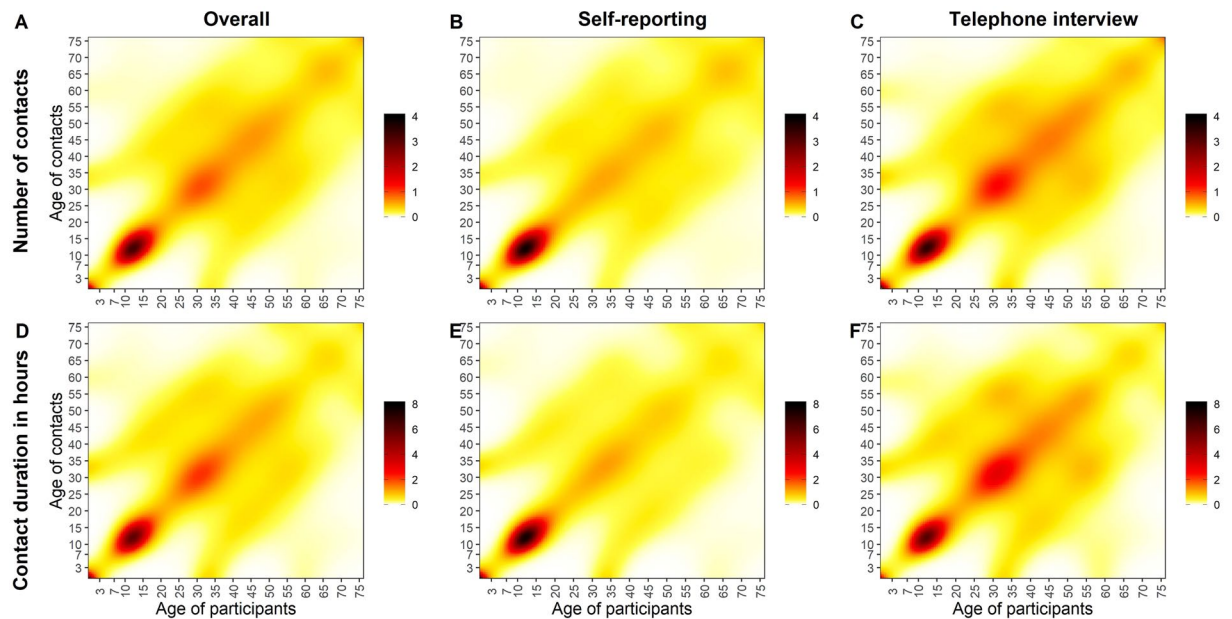
**Figure 3.** Estimates of percentage contribution of variables in human-human regression models. (A,B) Correspond to the predicted number of contacts; (C,D) correspond to the contact duration; (E,F) correspond to the number of contact settings. The predicted values are relative to a 40 years old person with other characteristics referring to the first level of each factor. 95% Confidence intervals are denoted by a shaded region as (A,C,E), or error bars as (B,D,F). Significant covariates of the regression models (at the 5% level) are denoted by stars.



**Figure 4.** Estimates of percentage contribution of variables in the human-animal contact regression model. The predicted values are relative to a 40 years old person reporting the annual income of (0, 8000], and making contacts with 10 people. 95% confidence intervals are denoted by a shaded region as (A,B), or error bars as (C). Significant covariates of the regression models (at the 5% level) are denoted by stars.

association between the number of contacts and the participant age, with the largest number of contacts in those aged 18. Telephone interview was associated with a significantly longer contact duration. Living in a household with more than 2 people and workdays were associated with a significantly longer contact duration. A similar nonlinear relationship was found between contact duration and participant age. We found no significant association between the number of settings and modes of data collection. Workdays and atypical days were associated with a significantly greater number of contact settings. The number of contact settings was also significantly associated nonlinearly with participant age, with a greatest number associated with ages 30–50 and a decline below or above that. More results about individual and group contact are given in Supplementary Fig. S14, Supplementary Table S9, and Supplementary Text S8.

For the H-A contacts, the number of social contacts was positively associated with the probability of having contact with animal(s), when controlling for the age and annual income of participants, with the largest probability associated with around 15 years and 60 years (Fig. 4, Supplementary Table S9). Participants reporting having



**Figure 5.** Smoothed human-human contact matrix. (A–C) Show the predicted average number of total contacts per day per participant and that stratified by different modes of data collection; (D–F) show the predicted average contact duration per day per participant and that stratified by different modes of data collection.

no income was significantly associated with the lowest probability of animal contact, while participants reporting an annual income of over \$15k had the highest probability of animal contact. Similar results were found for animal ownership (Supplementary Fig. S15).

**Human-human and human-animal contact matrix.** For the H-H contact (Supplementary Fig. S16), the degree of assortative mixing for the contact matrix of the number of contacts was calculated as  $q = 0.716$  (bootstrapped 95%CI: (0.613, 0.779), self-reporting: 0.749 (0.611, 0.832), telephone interview: 0.568 (0.393, 0.744)), showing that the matrix was strongly assortative. With the smoothed matrices (Fig. 5), the dominant feature was the strong diagonal element, indicating that participants in all age groups were inclined to mix assortatively by age, which was most pronounced in those aged 7–20, and least pronounced in those aged 25–55. Two parallel secondary diagonals starting around 30–40 for both contacts and participants were offset from the central diagonal likely capturing contacts that parents make with their children and vice versa. Compared to self-reporting, telephone interviewing was associated with a greater number of contacts for participants aged 20–60, but smaller for participants aged 3–20. The contact matrix of contact duration had a similar pattern to that of number of contacts. The estimating of the age of group contact is provided in Supplementary Fig. S5. Contact matrices stratified by different characteristics are shown in Supplementary Figs S17 and S18.

For the H-A contact (Supplementary Fig. S16), the number of animals owned or contacted with did not vary substantially across age groups, but participants aged 3–7, 15–35, and above 50 years had longer animal contact duration compared to other age groups. H-H and H-A contact matrices of reported contacts are shown in Supplementary Tables S10 and S11.

For the interaction between H-H and H-A contact, Supplementary Figs 19 and 20 showed the H-H contact matrices stratified by whether had animal ownership or animal contact. For the number of contacts, the contact matrix for participants not interacting with animals was more strongly assortative ( $q = 0.73$ ) than that interacting with animals ( $q = 0.65$ ), especially for those aged 7–20. But animal ownership or contact was associated with a greater number of contacts in those aged 25–55. The contact matrix of contact duration had a similar pattern to that of number of contacts. H-H contact matrices for such participants interacting with animals are shown in Supplementary Table S12.

**Hypothesis testing for mode of data collection.** We compared the age-specific contact matrices of observed contacts via self-reporting and telephone interview. No significant difference was found for most of the elements (Supplementary Figs S21 and S22). For participants aged 20–60, telephone interviewing was associated with greater number of contacts, which was consistent with what we found in the smoothed contact matrices (Fig. 5).

In addition to comparing the similarity of contact matrices for the two modes of data collection, we evaluated differences in respondent self-ratings about the quality of data collection. We found that the telephone interview participants rated the quality of their responses across a number of dimensions higher than self-reporting participants (Supplementary Text S10, Supplementary Table S13).

## Discussion

We have conducted the contact study of H-H and H-A contacts in Shanghai City, a high-income densely populated megacity in Eastern China. We recorded a mean of 18.7 contacts, 32.1 contact hours, and 2.1 contact settings per participant, where the average number of contacts by telephone interview (19.3) was significantly higher than that recorded by self-reporting (18.0). The H-A contact (especially poultry) was rare, although the probability of owning or having contact with animal(s) was positively associated with the number of social contacts.

Our study is the first contact study in any part of mainland China apart from Guangdong. Nevertheless, it is in broad agreement with other studies in China as well as in the wider region. The average number of contacts per person in our study was in broad agreement with studies from Guangdong<sup>28</sup> (18.56) and Hong Kong<sup>40</sup> (18.0–18.6), while it was larger than those from Taiwan<sup>27</sup> (12.5) and Japan<sup>26</sup> (15.3) (Supplementary Table S14). We identified three studies in Hong Kong presented different results (18.0–18.6 in a study by Kucharski *et al.*<sup>40</sup>, 12.5 in a study by Kwok *et al.*<sup>31</sup>, and 8.1 in a study by Leung *et al.*<sup>29</sup>). Kucharski *et al.* included group contacts to capture large numbers of contacts. Kwok *et al.* also permitted the reporting of group contacts, but it was a retrospective design, where the interview day was assigned within 4 days after the reporting day. Leung *et al.* did not define group contacts, possibly contributing to underreporting. The number of contact settings in our study was slightly lower than the 2.9 contact locations in the Kwok *et al.* study, probably because the Kwok *et al.* study measured the number of geographical locations but we measured the number of contact settings that could not accurately determine the total number of different locations the participant encountered people at. We observe similar age-assortativity to many studies in other countries or regions, but with higher number of contacts made among 20–60 years (Supplementary Fig. S23). Similar to a large number of other studies<sup>41</sup>, we also identified age, day type (e.g., weekday/weekend), and household size as determinants of H-H contact patterns. In addition, we identified other factors having a significant influence on H-H contact pattern, such as whether the day was considered a typical day by the participant, animal ownership, and whether participants travelled daily out of the subdistrict, or only occasionally. The link between travel and contact patterns would provide additional insights on whether or not infection could be imported.

Our study collected details about important H-A contact patterns, including animal species, number of animals, contact duration, contact setting and contact frequency of contacts with different types of animals, explored the association between H-H and H-A contact patterns, and quantified the H-H contact patterns for such participants owning or touching animals. This is particularly important because only one other study<sup>35</sup> has examined both H-H and H-A contacts. This study sampled respondents in Belgium, and the pattern of animal contacts differed substantially from ours. In particular, both animal ownership and animal contact were much higher in Belgium (ownership: 51% pets, 15% poultry and 5% livestock; touch: 46% pets, 2% poultry and 2% livestock) compared to Shanghai (15.3% pets, 0% poultry and 0.1% livestock) and 17.7% (16.5% pets, 0.2% poultry and 0% livestock). Although both studies were based in high-income urban settings, the lower animal contact in Shanghai (especially poultry), may stem from the closure of live poultry markets due to the spread of influenza A(H7N9) in Shanghai since 2013, and also from the differences in both cultural norms and population density in Europe compared to Asia. For instance, population density is 25,260 people per km<sup>2</sup> in central urban areas of Shanghai compared to 6,790 people per km<sup>2</sup> in Brussels. Both studies found that animal owners were likely to have more social contacts, and we still identified linear relationship between the number of social contacts with the probability of owning or touching animals. 18.4% of the animal contact in our study occurred out of the home setting, which may explain why more animal contacts were associated with more social contacts. From a socioeconomic perspective, participants reporting highest annual incomes (i.e. over \$15k) had the highest probability of owning animals or touching with animals, after adjusting for participant age. We also quantified the H-H contact patterns, including the H-H contact matrices, for such participants interacting with animals, which provided important parameters to evaluate the potential for crossing the animal human barrier and the potential of outbreak in human populations.

Such information is important since Eastern China has been identified as hotspot for emergence of zoonotic infectious diseases<sup>1</sup> and Shanghai has been one of the potential epidemic centers in China as the first case of influenza A(H7N9) was found in Shanghai<sup>42</sup>. Our study results suggest that data on animal contacts from other settings cannot be extrapolated to Asia. In particular, we find that there is low contact between humans and poultry in Shanghai. However, we also found that people with more animal contacts are likely to have more social contacts. The potential for crossing the animal human barrier is limited but once that is done, the potential for transmission is enormous. In addition, our work helps understand the transmission dynamics of diseases transmitted by dogs and cats, such as rabies. In China, 92% human deaths of rabies were caused by animal bites, where wounds from domesticated animals (dogs and cats) were responsible for more than half of the rabies cases<sup>43</sup>, and the east and south regions of China (including Shanghai) were more seriously affected compared with other regions<sup>44</sup>. Hence the human-animal contact patterns observed in this study would be very important evidence for further research on epidemic dynamics at the human-animal interface, animal control and vaccination programs, which could greatly impact public health.

Another important feature of our study is the use of multiple modes of data collection and questionnaire types. A few other studies have also used multiple modes and these have found important differences in responses when using different modes<sup>10,14,45</sup>. Several different modes of data collection used in previous studies included self-completed paper contact diaries<sup>10,14,29,45</sup>, electronic contact diaries<sup>24</sup>, online contact diaries<sup>11,14,29</sup>, and interview-led contact diaries<sup>28,31</sup> (Supplementary Table S14). Perfect recall of all encounters is impossible, especially for short-duration or large-size encounters. Some studies have found electronic self-completed contact results to perform similarly to paper diaries<sup>14</sup>, while others have found more contacts are recorded using paper diaries<sup>24,29</sup>. Web-based questionnaires are very helpful in large-scale surveys, but are probably associated with lower response rates compared to paper-based questionnaires<sup>41</sup>. A recent systematic review<sup>41</sup> has shown that interviews are typically better at capturing all contacts compared to self-reporting. In our study, we found that



more contacts are recorded via telephone interview at the end of the reporting day compared to self-reporting, which is likely to result from less recall bias. However, there has not been any evidence to answer whether interviews conducted face-to-face or via telephone yield different results. Based on our experience in this survey, we believe face-to-face interview will provide more accurate measurements, though at the expense of greater human and material resources than telephone interview. We considered this kind of telephone interview-led method as a potential improvement of the design of future contact surveys, despite of the following limitations. First, participants got called at a pre-agreed time period after dinner and before going to bed (i.e. close to the end of the survey day rather than the real end of the day); different calling times may possibly interfere with the recording of social contacts. While we balanced the two modes of data collection via propensity score matching, there may still be other factors influencing a participants' choice of interview mode that we didn't measure. For example, individuals who are busy and have many social activities may be more likely to want an interview by telephone rather than self-reporting, in the belief that there would be less time burden for participation. Finally, similarly to other forms of interviews, telephone interviews could introduce interviewer bias (due to different interviewers), despite standardized training of interviewers. We identified heterogeneity in reported rates between different interviewers, though sensitivity analysis suggests individual interviewers have little influence on our primary contact rate estimates (Supplementary Text S12).

Besides collection of detailed data about animal contacts, our study also contains a number of other important strengths. Firstly, we used a spatial sampling design, sampling from the Chinese administration units (district-subdistrict-neighborhood) so as to represent the population of the central urban of Shanghai. This study also proposed a novel data collection method and evaluated the ability to record contact behavior via self-reporting compared to interview-led methods (and thus, was able to compare these methods). A telephone interview at the end of the reporting day by a trained interviewer elicited a greater number of contacts than self-reporting; therefore, we believe this methodology can substantially help to reduce recall bias.

There are some limitations to this study. We asked respondents about the contact setting rather than the actual geographical locations where the contact occurred. Similar to many other contact studies<sup>10,12,13,18,21,28,31,35,40</sup>, we allowed respondents to group large numbers of contacts found in the same location in a single "group contact", but we only asked about the size of the group and the age range of the group contacts without other contact details, such as gender, age, physicality, setting, and duration of each contact. Thus, we used statistical models to estimate the age and duration of each group contact. The number of group contacts that could be reported was limited to one; similarly, the questionnaire limited the reporting of individual contacts to 40. This may have led to some censoring and under-reporting of contacts, though the majority of participants reported fewer than 40 contacts (Supplementary Fig. S6).

The human-human contact in this megacity is very frequent but that contact with animals, and poultry in particular, is surprisingly low. This would imply that the most likely source for a zoonosis lies outside in the more rural areas. We propose that a separate study with a similar design should look at this.

## Data availability

The data collected as part of this survey will be made available to the scientific community via the zenodo data repository as part of a social contact data collection initiative [www.socialcontactdata.org](http://www.socialcontactdata.org).

Received: 22 May 2019; Accepted: 29 September 2019;

Published online: 22 October 2019

## References

- Allen, T. *et al.* Global hotspots and correlates of emerging zoonotic diseases. *Nature communications* **8**, 1124, <https://doi.org/10.1038/s41467-017-00923-8> (2017).
- Lloyd-Smith, J. O. *et al.* Epidemic Dynamics at the Human-Animal. *Interface. Science* **326**, 1362, <https://doi.org/10.1126/science.1177345> (2009).
- Pourbohloul, B. *Modeling infectious diseases in humans and animals.* (Princeton University Press, 2008).
- Basta, N. E., Chao, D. L., Halloran, M. E., Matrajt, L. & Jr, I. M. L. Strategies for Pandemic and Seasonal Influenza Vaccination of Schoolchildren in the United States. *American Journal of Epidemiology* **170**, 679–686, <https://doi.org/10.1093/aje/kwp237> (2009).
- Medlock, J. & Galvani, A. P. Optimizing influenza vaccine distribution. *Science* **325**, 1705, <https://doi.org/10.1126/science.1175570> (2009).
- Patel, R., Jr, L. I., Halloran & Finding, M. E. optimal vaccination strategies for pandemic influenza using genetic algorithms. *Journal of Theoretical Biology* **234**, 201–212, <https://doi.org/10.1016/j.jtbi.2004.11.032> (2005).
- Wallinga, J., Teunis, P. & Kretzschmar, M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology* **164**, 936–944, <https://doi.org/10.1093/aje/kwj317> (2006).
- Arthur, R. F., Gurley, E. S., Salje, H., Bloomfield, L. S. & Jones, J. H. Contact structure, mobility, environmental impact and behaviour: the importance of social forces to infectious disease dynamics and disease ecology. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **372**, <https://doi.org/10.1098/rstb.2016.0454> (2017).
- Prem, K., Cook, A. R. & Jit, M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput Biol*, <https://doi.org/10.1371/journal.pcbi.1005697> (2017).
- Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *Plos Medicine* **5**, e74, <https://doi.org/10.1371/journal.pmed.0050074> (2008).
- Klepac, P., Kissler, S. & Gog, J. Contagion! The BBC Four Pandemic – The model behind the documentary. *Epidemics*, <https://doi.org/10.1016/j.epidem.2018.03.003> (2018).
- Beraud, G. *et al.* The French Connection: The First Large Population-Based Contact Survey in France Relevant for the Spread of Infectious Diseases. *PLoS One* **10**, e0133203, <https://doi.org/10.1371/journal.pone.0133203> (2015).
- Danon, L., House, T. A., Read, J. M. & Keeling, M. J. Social encounter networks: collective properties and disease transmission. *Journal of the Royal Society Interface* **9**, 2826, <https://doi.org/10.1098/rsif.2012.0357> (2012).
- Beutels, P., Shkedy, Z., Aerts, M. & Van Damme, P. Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. *Epidemiol Infect* **134**, 1158–1166, <https://doi.org/10.1017/S0950268806006418> (2006).

15. Organization, W. H. Report on the WHO Quantitative Immunization and Vaccines Related Research (QUIVER). Advisory meeting. Geneva, 5–7 October 2010 (2011).
16. Horby, P. *et al.* Social Contact Patterns in Vietnam and Implications for the Control of Infectious Diseases. *Plos One* **6**, e16965, <https://doi.org/10.1371/journal.pone.0016965> (2011).
17. Melegaro, A. *et al.* Social Contact Structures and Time Use Patterns in the Manicaland Province of Zimbabwe. *Plos One* **12**, e0170459, <https://doi.org/10.1371/journal.pone.0170459> (2017).
18. Dodd, P. J. *et al.* Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium tuberculosis Infection. *Am J Epidemiol* **183**, 156–166, <https://doi.org/10.1093/aje/kwv160> (2016).
19. Johnstone-Robertson, S. P. *et al.* Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. *Am J Epidemiol* **174**, 1246–1255, <https://doi.org/10.1093/aje/kwr251> (2011).
20. Kiti, M. C. *et al.* Quantifying Age-Related Rates of Social Contact Using Diaries in a Rural Coastal Population of Kenya. *Plos One* **9**, e104786, <https://doi.org/10.1371/journal.pone.0104786> (2014).
21. le Polain de Waroux, O. *et al.* Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: a survey in Southwest Uganda. *BMC Infectious Diseases* **18**, 172, <https://doi.org/10.1186/s12879-018-3073-1> (2018).
22. Ajelli, M. & Litvinova, M. Estimating contact patterns relevant to the spread of infectious diseases in Russia. *J Theor Biol* **419**, 1–7, <https://doi.org/10.1016/j.jtbi.2017.01.041> (2017).
23. Rolls, D. A. *et al.* Social encounter profiles of greater Melbourne residents, by location—a telephone survey. *BMC Infect Dis* **15**, 494, <https://doi.org/10.1186/s12879-015-1237-9> (2015).
24. McCaw, J. M. *et al.* Comparison of three methods for ascertainment of contact information relevant to respiratory pathogen transmission in encounter networks. *BMC Infect Dis* **10**, 166, <https://doi.org/10.1186/1471-2334-10-166> (2010).
25. Grijalva, C. G. *et al.* A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru. *PLoS One* **10**, e0118457, <https://doi.org/10.1371/journal.pone.0118457> (2015).
26. Ibuka, Y. *et al.* Social contacts, vaccination decisions and influenza in Japan. *Journal of epidemiology and community health* **70**, 162–167, <https://doi.org/10.1136/jech-2015-205777> (2016).
27. Fu, Y. C., Wang, D. W. & Chuang, J. H. Representative contact diaries for modeling the spread of infectious diseases in Taiwan. *PLoS One* **7**, e45113, <https://doi.org/10.1371/journal.pone.0045113> (2012).
28. Read, J. M. *et al.* Social mixing patterns in rural and urban areas of southern China. *Proceedings of the Royal Society B Biological Sciences* **281**, 20140268, <https://doi.org/10.1098/rspb.2014.0268> (2014).
29. Leung, K., Jit, M., Lau, E. H. Y. & Wu, J. T. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Sci Rep* **7**, 7974, <https://doi.org/10.1038/s41598-017-08241-1> (2017).
30. Kwok, K. O. *et al.* Social contacts and the locations in which they occur as risk factors for influenza infection. *Proceedings. Biological sciences* **281**, 20140709, <https://doi.org/10.1098/rspb.2014.0709> (2014).
31. Kwok, K. O., Cowling, B., Wei, V., Riley, S. & Read, J. M. Temporal variation of human encounters and the number of locations in which they occur: a longitudinal study of Hong Kong residents. *Journal of the Royal Society, Interface* **15**, <https://doi.org/10.1098/rsif.2017.0838> (2018).
32. *Data from published contact studies*, <http://www.socialcontactdata.org/>
33. Ozella, L. *et al.* Close encounters between infants and household members measured through wearable proximity sensors. *Plos One* **13**, e0198733, <https://doi.org/10.1371/journal.pone.0198733> (2018).
34. 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* **10**, e0136497, <https://doi.org/10.1371/journal.pone.0136497> (2015).
35. Kifle, Y. W. *et al.* Animal Ownership and Touching Enrich the Context of Social Contacts Relevant to the Spread of Human Infectious Diseases. *Plos One* **10**, e0133461, <https://doi.org/10.1371/journal.pone.0133461> (2015).
36. Qian, L. Shanghai Urban Master Plan (2016–2040). *Urban rail transit research*, 9–9 (2016).
37. Keeling, M. J. & Rohani, P. In *Modeling infectious diseases in humans and animals* Ch. 3, 69 (Princeton University Press, 2011).
38. Goyaerwaerts, N. *et al.* Estimating infectious disease parameters from data on social contacts and serological status. *Journal of the Royal Statistical Society Series C* **59**, 255–277, <https://doi.org/10.2307/40541685> (2010).
39. N. Wood, S. *Generalized Additive Models: An Introduction With R*. Vol. 66 (CRC Press, 2006).
40. J. Kucharski, A. *et al.* The Contribution of Social Behaviour to the Transmission of Influenza A in a Human Population. *Plos Pathogens* **10**, <https://doi.org/10.1371/journal.ppat.1004206> (2014).
41. Van Hoang, T. *et al.* A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections. *Epidemiology*, <https://doi.org/10.1097/ede.0000000000001047> (2019).
42. Gao, R. *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *The New England journal of medicine* **368**, 1888–1897, <https://doi.org/10.1056/NEJMoa1304459> (2013).
43. Guo, C. *et al.* Exposure history, post-exposure prophylaxis use, and clinical characteristics of human rabies cases in China, 2006–2012. *Sci Rep* **8**, 17188, <https://doi.org/10.1038/s41598-018-35158-0> (2018).
44. Zhou, H. *et al.* Human Rabies in China, 1960–2014: A Descriptive Epidemiological Study. *PLoS Negl Trop Dis* **10**, e0004874, <https://doi.org/10.1371/journal.pntd.0004874> (2016).
45. Mikolajczyk, R. T. & Kretzschmar, M. Collecting social contact data in the context of disease transmission: Prospective and retrospective study designs. *Social Networks* **30**, 127–135, <https://doi.org/10.1016/j.socnet.2007.09.002> (2008).

## Acknowledgements

We thank staff members of local Centers for Disease Control and prevention (Shanghai City, Huangpu District, Xuhui District, and Changning District), community health centers, and neighborhood committees for providing assistance with administration and data collection. This research was funded by the National Science Fund for Distinguished Young Scholars (81525023), the National Science and Technology Major Project of China (2018ZX10201001-010), and the Program of Shanghai Academic/Technology Research Leader (18XD1400300). P.K., A.R., S.F., and M.J. are funded by the National Institute of Health Research using Official Development Assistance (ODA) funding (16/137/109). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. S.F. is also supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (208812/Z/17/Z). S.L. is supported by the grants from the Ministry of Science and Technology of China (2016ZX10004222-009) and the National Natural Science Fund of China (81773498).

## Author contributions

H.Y. conceived, designed, and supervised the study. J.Z. wrote the drafts of the protocol. X.W., S.L., J.Y., H.L., S.F. and M.J. commented on drafts of the protocol. J.Z., M.L., Y.S., Q.W. and H.J. collected and checked data. J.Z. cleaned the data, analyzed the data, interpreted the data, and wrote the drafts of the manuscript. P.K., J.M.R., A.R., S.F. and M.J. commented on and revised drafts of the manuscript. All authors read and approved the final report.

### Competing interests

H.Y. has received investigator-initiated research funding from Sanofi Pasteur, GlaxoSmithKline, and Yichang HEC Changjiang Pharmaceutical Company. Other authors have declared that no competing interests exist.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-51609-8>.

**Correspondence** and requests for materials should be addressed to H.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019