Fine-scale family structure shapes influenza transmission risk in households: insights from a study of primary school students in Matsumoto city, 2014/15.

Akira Endo[1*], Mitsuo Uchida[2], Adam J. Kucharski[1,3], Sebastian Funk[1,3]

[1] Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom

[2] Department of Public Health, Graduate School of Medicine, Gunma University

[3] Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

Abstract

Background: Households are important settings for the transmission of seasonal influenza. Previous studies found that the per-person risk of within-household transmission decreases with household size. However, more detailed heterogeneities driven by household composition and contact patterns have not been studied.

Methods: We employed a mathematical model which accounts for infections both from outside and within the household. The model was applied to citywide primary school

* Correspondence to: akira.endo@lshtm.ac.uk
Keppel St., Bloomsbury, London, WC1E 7HT, UK.

surveillance data of seasonal influenza in 2014/15 season in Matsumoto city, Japan. We compared a range of models to estimate the structure of household transmission.

Results: Familial relationship and household composition strongly influenced the transmission patterns of seasonal influenza in households. Children had substantially high risk of infection from outside the household (up to 20%) compared with adults (1-3%). Intense transmission was observed within-generation (between children/parents/grandparents) and also between mother and child, with transmission risks typically ranging around 5-20% depending on the pair and household composition.

Conclusions: We characterised heterogeneity in household transmission patterns of influenza. Children were identified as the largest source of secondary transmission, with family structure influencing infection risk. This suggests that vaccinating children would have stronger secondary effects on transmission than would be assumed without taking into account transmission patterns within the household.

Abbreviations: CPI, community probability of infection; RDK, rapid diagnostic kit; SITP, susceptible-infectious transmission probability; MCMC, Markov-chain Monte Carlo; WBIC, widely-applicable Bayesian information criterion; CrI, credible interval.

Introduction

Respiratory infectious diseases transmitted by droplets, exemplified by influenza, are known to spread over social contact networks (1,2). Social settings which involve frequent contacts play important roles in the transmission dynamics (3,4). Households are considered as one of the main layers of transmission, as individuals come in close contact with each other both conversationally and physically on a daily basis (5–7). Many epidemiological studies have used household data to investigate the transmission dynamics of influenza within households (8,9), particularly in terms of the secondary attack rate (the number of household secondary cases divided by the number of household members at risk). However, this assumes that an index case (the first case in a household, who is considered to be infected outside the household) is responsible for all subsequent household cases, and that all the other household members are equally at the risk of secondary infection.

The possibility of co-primary infections and tertiary transmissions are neglected under such assumptions (8); potentially heterogeneous transmission patterns between household members are also radically simplified. The former limitation can be addressed by mathematical models which separately estimate the risk of infection from outside the household (community probability of infection; CPI) and the within-

54 household transmission risk (10). Many household studies have employed the Longini-
55 Koopman model and other related models to study within-household transmission
56 dynamics of influenza (11–17).

57 On the other hand, potentially-heterogeneous transmission patterns have not
58 been fully studied with empirical data. Multiple household modelling studies
59 incorporated factors including age, vaccination status and antibody titres (14,16,18–20)
60 to account for heterogeneity, but these are merely individual risk factors that determine
61 relative susceptibility of individuals. Considering typical behaviours within the family,
62 it is natural to expect rich heterogeneity in household contact patterns related to familial
63 relationships and household compositions, on top of those individual factors (6).
64 However, to our best knowledge, household size is the only covariate which has been
65 used to characterise contact behaviours in household models (13,14,17,18,21). Besides,
66 due to the limited sample size of households in these studies, a rationale on the
67 quantitative effect of household size in transmission has not been established. Familial
68 roles/relationships have been paid far less attention to in household studies; we found
69 only one field study on influenza that included familial roles as a covariate, a
70 descriptive study that did not quantify the risk by familial roles (22).

Households serve as important units in intervention policies (23,24). Tailored quantification of the transmission risks from outside and inside the household will help prioritising and promoting household-level prevention strategies including vaccination. If specific compositions of households have a higher risk of outbreak than others, intervention policies may be optimised by particularly targeting such households. Moreover, as vaccine uptake is shown to be influenced by perceived risk of infection and vaccine effectiveness (25,26), identifying the household-specific risk of infection and the possible reduction by vaccines may support highlight the individual benefit of vaccination.

In the present study, we applied a highly flexible household transmission model that accounts for heterogeneity to a large dataset to investigate the within-household transmission dynamics of seasonal influenza. The dataset included more than 10,000 primary school students with the infection status not only of students but also of their household members, which was expected to provide broader understanding on the within-household transmission dynamics. Particularly laying our focus on the effect of familial roles and household compositions, we compared multiple models with different levels of complexity to find the best model to describe the transmission patterns.

Methods

89 **Data source**

90 We used data from a citywide primary school influenza survey. At the end of
91 the 2014/15 season (early March), parents of students at all 29 public primary schools in
92 Matsumoto city, Nagano prefecture, Japan, were asked to respond to a questionnaire
93 consisting of a variety of questions including whether the students had influenza during
94 the season, onset date and observed symptoms, vaccination history, family composition
95 and who in the same household had influenza episodes during the season. The data was
96 originally collected for an observational study on the effect of prevention measures
97 against seasonal influenza (Uchida et al., 2017) (27). In the present study, we only
98 considered data on influenza episodes in students, their household composition and
99 influenza episodes in the household members. Participants reported the number of
100 siblings in the household, and also ticked the type of family members (such as “father”,
101 “younger sister” or “uncle”) with whom they live, as well as whether they acquired
102 influenza in the 2014/15 season. Among 13,217 students eligible, 11,390 (86%)
103 responded to the survey. After removing those with missing values, 10,486 surveys
104 were used in the present study. Characteristics of the population and frequent household
105 compositions are shown in Tables 1 and 2. Further details of the data collection can be

found in the original study (27). The analysis was approved by the ethics committee at London School of Hygiene & Tropical Medicine (approval number: 2715).

In the survey, all students who reported acquiring influenza also reported that they were diagnosed at a medical institution. For other household members, clinical diagnosis was not clearly required on the question sheet. In Japan, rapid diagnostic kits (RDKs) are usually used for suspected patients. International systematic reviews estimated that the sensitivity and specificity of RDKs are 50-70% and 98-99%, respectively (28,29). However, the sensitivity for studies conducted in Japan included in these reviews was relatively high (range: 72.9-96.4%), consistent with other earlier studies conducted in Japan (30–32). Considering that many Japanese primary schools encourage students presenting influenza-like symptoms to consult medical institutions so that they are granted absence, we believe that the reported influenza episodes in the dataset were sufficiently inclusive for our analysis. We also performed sensitivity analysis to address possible underreporting in the survey (described later).

Heterogeneous chain binomial model

We employed the chain-binomial model presented in (33) which allows for heterogeneous transmission. Let $\boldsymbol{N}$ be a vector representing the number of family

124 members stratified by individual type (e.g., father, mother, child, etc.) in a household.
125 The probability that a certain combination of individuals (represented by a vector $\boldsymbol{n}$) in
126 the household are infected by the end of the season is given by the following recursive
127 equations.

$$\pi(\boldsymbol{n};\boldsymbol{N},\boldsymbol{\varepsilon},H)=\pi(\boldsymbol{n};\boldsymbol{n},\boldsymbol{\varepsilon},H)\prod_{k}\binom{N_k}{n_k}S_k(\boldsymbol{n},\boldsymbol{\varepsilon},H)^{N_k-n_k},$$
$$\pi(\boldsymbol{n};\boldsymbol{n},\boldsymbol{\varepsilon},H)=1-\sum_{\boldsymbol{\nu}<\boldsymbol{n}}\pi(\boldsymbol{\nu};\boldsymbol{n},\boldsymbol{\varepsilon},H). \quad (1)$$

128 where N_k and n_k are the k-th component of $\boldsymbol{N}$ and $\boldsymbol{n}$, respectively ($1 \le k \le K$). The sum
129 $\sum_{\boldsymbol{\nu}<\boldsymbol{n}}$ is taken for all vector $\boldsymbol{\nu}$ satisfying $0 \le \nu_k \le n_k$ $(\forall k)$ and $\boldsymbol{\nu} \ne \boldsymbol{n}$. $\boldsymbol{\varepsilon}$ is the risk of
130 external infection for each type of individual (a heterogeneous version of CPI; we avoid
131 the term CPI as our model assumes household members experiences infection from
132 different sources outside the household and not from a single "general community").
133 The susceptible-infectious transmission probability (SITP) ρ_{kl} is the probability of
134 within-household transmission for a specific infectious-susceptible pair (17) and has
135 been used to quantify within-household transmission. However, it is more convenient to
136 use the effective household contact matrix $H=(\eta_{kl})$ in the model; η_{kl} is defined to
137 satisfy $\rho_{kl}=1-\exp(-\eta_{kl})$, and is interpreted as the amount of contact that leads to
138 within-household transmission (effective contact) from type l to k. That is, η_{kl} denotes
139 the amount of exposure that an individual k experiences when another individual of type

140 l in the same household is infectious. $S_k(\boldsymbol{n}, \boldsymbol{\varepsilon})$, the probability that a type k individual

141 escapes infection from both outside and inside the household, is given as

$$S_k(\boldsymbol{n}, \boldsymbol{\varepsilon}, H) = (1 - \varepsilon_k) \exp\left(-\sum_l \eta_{kl} n_l\right). \quad (2)$$

142 $(1 - \varepsilon_k)$ is the probability that the individual is not infected outside the household, and

143 $\exp(-\sum_l \eta_{kl} n_l)$ is the probability that the individual is not infected from any of the

144 household infectives. When a dataset $\{\boldsymbol{N_i}, \boldsymbol{n_i}\}$ contains the family composition and

145 infection status in each household i, the likelihood function is given as

$$L(\boldsymbol{\varepsilon}, H; \{\boldsymbol{N_i}, \boldsymbol{n_i}\}) = \prod_i \pi(\boldsymbol{n_i}; \boldsymbol{N_i}, \boldsymbol{\varepsilon}, H). \quad (3)$$

146 The likelihood $\pi(\boldsymbol{n_i}; \boldsymbol{N_i}, \boldsymbol{\varepsilon}, H)$ is computed by recursively applying Equation (1) starting

147 with $\pi(\mathbf{0}; \mathbf{0}, \boldsymbol{\varepsilon}, H) = 1$.

148 In the present study, we classified each individual in households as one the

149 following type: “father”, “mother”, “student”, “sibling”, or “other”. “Students” are

150 participants of the survey (i.e., students of primary schools in Matsumoto city), and

151 “siblings” are their elder/younger siblings, who may have also been recruited in the

152 survey if they are primary school students (however, they are not linked in the data and

153 thus unidentifiable as participants). The parameters for “students” and “siblings” were

154 differentiated because “siblings” are not necessarily primary school students, therefore

155 their characteristics may be different from “student”. “Father” and “mother” were

156 labelled as "single-parent" if they are only one parent in the family; models were
157 considered in model selection where their parameter values were differentiated from
158 cohabiting parents (details described in "model selection"). Most individuals classified
159 as "other" were grandparents (90.1%). Uncles/aunts accounted for 6.7%, and the
160 remaining 3.2% was "none of the above categories".
161
162 **Transmission risk in households**
163 We modelled the possible heterogeneity in household transmission by
164 parameterising the effective household contact matrix $H = (\eta_{kl})$. Our basic
165 assumptions are: (i) each pairs of individuals have a specific "intensity of contact"; (ii)
166 the relative importance of each household contact may be reduced if an individual
167 experiences a large amount of household contacts in total; (iii) the contact intensity
168 adjusted by the total amount of contact is proportional to the force of infection. That is,
169 we modelled η_{kl} as

$$\eta_{kl} = \beta \frac{c_{kl}}{C_k^{\gamma}}. \quad (4)$$

170 C_k represents the total number of household contacts experienced by an individual of
171 type k, which we introduced to investigate how η_{kl} differs in households of different

sizes and compositions. Noting that the number of individuals in contact is $N_k - 1$ if k=l, we get

$$C_k = \sum_l c_{kl} (N_l - \delta_{kl}), \tag{5}$$

where δ_{kl} is the Kronecker delta. The value of the exponent parameter γ determines how strongly η_{kl} is scaled by C_k, which associates our model with density-dependent vs. frequency-dependent mixing assumptions (34). γ=0 corresponds to the density dependent mixing assumption, where the force of infection is proportional to the total number of contacts (weighted by intensity) with infectives, whereas γ=1 corresponds to the frequency dependent mixing assumption, where it is the proportion of infectious contacts among total contacts that matters. In addition to γ=0 and γ=1, γ was also allowed to be estimated as a free parameter in the model selection, representing a mixture of density-dependent and frequency-dependent mixing.

The contact intensity matrix (c_{kl}) is interpreted as the per-individual version of the contact matrix ($c_{kl} = b_{kl}/N_l$ where b_{kl} is the contact matrix). c_{kl} is generally a $K \times K$ matrix and contains too many parameters to estimate. We therefore reduced the number of parameters by categorising contacts into the following 5 pairs first:

$$c_{kl} = \begin{cases} c_{\mathrm{CC}}\ (\mathrm{Child} - \mathrm{Child}) \\ c_{\mathrm{FC}}\ (\mathrm{Father} - \mathrm{Child}) \\ c_{\mathrm{MC}}\ (\mathrm{Mother} - \mathrm{Child}) \\ c_{\mathrm{OC}}\ (\mathrm{Other} - \mathrm{Child}) \\ c_{\mathrm{AA}}(\mathrm{Adult} - \mathrm{Adult}) \end{cases} \tag{6}$$

187 Child included both "student" and "sibling", and adult included "father", "mother" and
188 "other". (In models where "single-parent" is a separate type, another parameter
189 c_{SC} (Single parent − Child) was added.) The matrix was assumed to be symmetric, i.e,
190 $c_{kl} = c_{lk}$. Since we did not have a measurement for the intensity of household contacts
191 in our dataset, we used relative values of c_{kl} in our analysis where c_{AA} was assumed to
192 be 1. β is approximately equal to the probability of transmission in a (hypothetical)
193 household composed of only father and mother (since $\frac{c_{kl}}{C_k^{\gamma}} = 1$ regardless of γ).
194
195 **Statistical analysis and model selection**
196 We sampled parameter values from a posterior distribution yielded from the
197 likelihood function (3) and priors in Table 3 using the Markov-chain Monte Carlo
198 (MCMC) method. An optimal variance-covariance matrix for proposal was explored by
199 Adaptive-Metropolis algorithm and then Random-walk Metropolis algorithm was used
200 to obtain final samples. All MCMC sampling was performed using the R package
201 {LaplacesDemon}. The scripts to produce MCMC samples for the main results is
202 reposited on GitHub (https://github.com/akira-endo/HHstudy_FluMatsumoto2014-15).
203 First, we tested various possible combinations of assumptions on the effective
204 contact matrix and the risk of external infection (shown in Table 3), and compared their

goodness of fit by Widely-applicable Bayesian Information Criterion (WBIC) (35). Model variants included (i) homogeneous or heterogeneous mixing in households (c_{kl}), (ii) uniform or heterogeneous risk of external infection (ε_k), (iii) the value of the exponent parameter (γ), and (iv) whether the parameter values for a single parent is differentiated from those of cohabiting parents. Characteristics of compared models are documented in the supplementary materials, Section 1. WBIC for each model was computed from 80,000 MCMC samples which were thinned from 125,000 samples × 8 chains, so that the chains had ESS ~40,000.

We then used the models selected by WBIC to estimate the parameters. As final samples, 10,000 thinned samples were recorded from 40,000 pre-thinned MCMC samples. It was ensured that the effective sample size (ESS) was at least 500 for each parameter.

Using the estimated parameters, we computed the source-stratified risk of infection and the risk attributable to the introduction into the household (see the supplementary materials Section 2 for further details).

Further model development

When the parameters were estimated with the best model selected, we found that the estimates for c_{FC} and c_{OC} were very similar, which suggested that we might be able to equate these two parameters and further stratify the contacts between adults (c_{AA}) with the degree of freedom earned. We tested some other contact intensity matrices, including

$$c_{kl} = \begin{cases} c_{\mathrm{CC}}\ (\mathrm{Child} - \mathrm{Child}) \\ c_{\mathrm{MC}}\ (\mathrm{Mother} - \mathrm{Child}) \\ c_{\mathrm{FM}}\ (\mathrm{Father} - \mathrm{Mother}) \\ c_{\mathrm{OO}}\ (\mathrm{Other} - \mathrm{Other}) \\ c_{\mathrm{X}}\ (\mathrm{Cross\ generational}) \end{cases} \tag{7}$$

which gave the best performance in the end. Explored candidate models and selection results are detailed in the supplementary materials Section 2.

Sensitivity analysis

We performed sensitivity analysis to address potential biases in our dataset. We considered in our sensitivity analysis (i) ascertainment bias, (ii) different susceptibility in children, (iii) multiple counting of households and (iv) censoring of sibling cases.

The first two points are related to the assumptions in our models. Influenza can have a low reporting rate due to mild clinical presentation (including asymptomatic infections), and therefore some infectious individuals may not have been included in our dataset. The reporting rate of influenza is considered to be very high in primary school

238 students in Japan, who are often required to report influenza to their schools. On the
239 other hand, the reporting rate of adults can be lower, as they may be less likely to seek
240 medical treatment than children. A serosurvey conducted in Japan after the 2009/10
241 H1N1 influenza pandemic suggested that while influenza in children were almost fully
242 reported, the reporting rate of adults were relatively low (30-50%) (36).
243 Another possible difference between adults and children is susceptibility:
244 adults may be less likely to be infected by the same amount of exposure due to the
245 previous history of infections or stronger immune systems than children. Conversely,
246 children may exhibit lower susceptibility if the vaccine uptake for them is higher than
247 adults. The majority of household transmission studies from a systematic review (8)
248 reported significant association between susceptibility and age (although this becomes
249 the minority when limited to the studies with PCR-confirmed cases). Our baseline
250 model assumes that transmissibility β is identical between individuals, but in reality
251 transmissibility might depend on the age of the susceptibles.
252 The remaining points explored in sensitivity analysis are inherent limitations in
253 our dataset. One of the limitations is that, because students in the same household
254 responded to the questionnaire separately, households with multiple siblings may have
255 been counted more than once. As this was an anonymous questionnaire, data obtained

256 from different students were not linked with each other even if they were from the same
257 household. If there was more than one child in a household who was eligible for the
258 study, the same household transmissions can appear multiple times in the dataset, which
259 could modify the results. Lastly, because of the design of the questionnaire, the number
260 of influenza cases in siblings may have been underreported. The questionnaire asked
261 whether each type of individual in the same household had influenza during the season,
262 and the respondents ticked if at least one individual of that type was infected since it
263 was a yes-no question. Therefore, even if there was more than one case in the same type
264 of individuals, the number was not reported and treated as a single case; that is, if a
265 respondent has two older brothers, he/she only reports that "older brother had
266 influenza", and there was no distinction on the dataset whether it was only one or both
267 of them.
268 Each potential source of bias was addressed by incorporating the data-generating
269 process causing the bias into the model. Technical details of the sensitivity analysis can
270 be found in the supplementary materials Section 3.
271

272 **Results**

We found that considerable heterogeneity existed in both the risk of external infection and the risk of within-household transmission (Table 3 and Figure 1). The best performing mathematical model suggested that children had a comparatively high risk of infection outside the household: 20% in the primary school students and 16% in their siblings, compared to only 1-3% in adults. Within-household contact patterns showed strong generational clustering. High contact intensities were observed within the same generation (between siblings, parents and grandparents), and the intensity of cross-generational contacts was less than half the intensity within the same generation. Contact between mothers and children was an exception to this, showing a higher intensity than between parents. The estimated contact intensity relative to that between parents (father-mother) was highest between other-other (1.97; CrI: 1.10-3.24), most of whom were grandparents in our data, followed by mother-child (1.16; CrI: 1.00-1.32) and child-child (1.04; 0.88-1.23). The model did not support a significant difference between parameter estimates for single and cohabiting parents.

The inferred networks of household transmission suggest that various contact patterns between household members exist in different household compositions. The contact intensity between individuals are shown in network graphs (Figures 3A-3C) for three selected characteristic household composition models, “nuclear family”: FM-2

(see Table 2 for the notation), (b) "many-siblings family": FM-4, and (c) "three-generation family": FM-2-2. Mothers served to bridge between the generations of children and parents; clusters of grandparents were relatively independent of other household members.

Overall risk of infection and the breakdown of infection source presented in Figures 3D–3F suggests that risk of infection in children was mostly from outside the household, whereas larger proportion of risk in adults was attributed to within-household transmission. Risk of within-household infection increased when more children were in the household (Figure 3E); however, the influence of additional members categorised as "others" (grandparents in most cases) was minimal, probably due to their low risk of external infection and contact intensity (Figure 3F). On the other hand, for grandparents in a typical three-generation household, the risk of infection from inside the household was twice the risk from outside.

Once influenza was brought into a household by a student, the conditional risk of infection in other members of the household became substantially higher; the implication of disease introduction into households can be seen in the simulated risk of infection after introduction (Figures 3G–3I). In "nuclear family" and "three-generation

family" models, the risk in adults increased by a factor of 2-3 if a primary school student in the family was infected.

The effective household contacts that each type of individual experiences are displayed in Figure 4, indicating the substantial variation in household contact patterns between individuals and between households. SITP typically ranged around 5-20%, depending on the contact pair and household composition. Reflecting the estimated value of γ=0.5 (CrI: 0.3-0.7), the total amount of effective household contacts was greater in larger households, but the weight of each single contact (the effective contact corresponding to a contact with one individual in the household) decreased with household size. This is because the effective household contact η_{kl} that one experiences followed an "inverse square root law", i.e., η_{kl} is inversely proportional to the square root of the total amount of contact C_k ($\eta_{kl} \propto C_k^{0.5}$; see Equation 4).

While Figure 4 summarises the heterogeneous within-household transmission patterns, one must note that the secondary transmission is conditional to infection in the primary case. When the contacts were weighted by the risk of external infection to visualise the source of primary and secondary infections for each individual, it can be seen that the children were responsible for the most of secondary transmissions within households (Figures 5): as children were more than five times likely to acquire

326 influenza from outside the household than adults, they were the most likely source of
327 secondary transmission. As a consequence, the individual risk of infection was mostly
328 determined the number of children in the household.

329 The sensitivity analysis suggested that the effective household contacts
330 between children may have been lower than the baseline estimates under some
331 assumptions (Figure S1). However, the overall trend did not change substantially. The
332 importance of children introducing influenza into household remained unchanged
333 throughout the sensitivity analysis.

334 The predicted and observed frequency of data compared in Figure S2 illustrate
335 the goodness of fit of our model. The model prediction was highly consistent with the
336 observed outcome patterns, suggesting our model successfully described the
337 heterogeneous transmission patterns of influenza in households.

338

339

340 **Discussion**

341 We applied a household-based mathematical model to a large-scale influenza
342 survey data including 10,000 primary school students and their families in Matsumoto
343 city, Japan, 2014-15. With the dataset of an extensive sample size on morbidity and

344 familial roles of household members, the model captured heterogeneous transmission
345 patterns in households in greater detail than previous household studies.

346 Our results are supportive of the common perception that influenza is brought
347 into households by schoolchildren (37). With their high probability of contracting
348 influenza outside the household, they were responsible for most secondary
349 transmissions within households. Once they brought virus from outside the household,
350 their mother and other siblings were exposed to a higher risk of within-household
351 secondary transmission. The estimated breakdown of infection source showed that
352 within-household transmission accounted for a large proportion of the overall risk in
353 adults. The relative importance of within-household transmission was especially
354 highlighted in grandparents in “three-generation” households. In a typical three-
355 generation family composed of two children, two parents and two grandparents, the risk
356 of infection in grandparents was tripled by within-household transmission. Besides, it
357 must be noted that an infection of a grandparent is likely to be followed by that of
358 another due to a high transmission risk between grandparents. These emphasise the
359 importance of controlling school epidemic and household contagion, as the symptoms
360 of influenza tends to be more severe in the elderly (37–39).

The results of the present study could have implications for household-level control measures. There are two steps in a household outbreak: introduction and within-household transmission. Due to the different risk patterns between the two steps, the focus of prevention measures should also change accordingly. At the pre-introduction stage when no one in the household is yet infected with influenza, the primary target is to prevent the first infection in the household from happening. Children, with the risk of external infection up to 20%, are most likely to be the first case in the household and thus should be prioritised at this stage. As the high risk of external infection is probably from schools (3), household members are advised to monitor the trend of school outbreaks and guide children to comply with daily precautions (40,41). Our results suggest that vaccinating children is an effective strategy not only because their risk of infection is high but also because they are responsible for a substantial fraction of within-household secondary infections. Especially for adults living with many children, protecting children from infection is as important as (or even more important in some cases) protecting themselves. If one of the household members contracts influenza despite the pre-introduction control effort, the primary target shifts to preventing further transmissions within the household. Household members are now exposed to an infectious person within the same household, which substantially elevates their risk. At

this post-introduction stage, preventing subsequent transmissions is important because every additional infection further increases the exposure. Our findings about household transmission patterns can be used to identify key individuals in the household network. For example, if the primary case is a child, the most probable secondary case is either the mother or another sibling. If the mother gets infected, that may be followed by a transmission to either the father or another child. Direct transmissions between children and father/grandparent may be relatively rare. Grandparents are suggested to be at comparatively low risk from other household members. However, their contacts with each other are closer than any other pair of household members, which warrants attention provided the high disease burden of influenza in the elderly.

To our best knowledge, the present study first reported a parametric relationship between within-household influenza transmission and household composition with high precision. With a detailed dataset consisting of up to 10,000 households, the present study was able to employ a highly flexible modelling framework to explore previously used modelling assumptions in great detail. A decrease of the per-person risk of within-household infections with household size has been observed in previous studies (8); our model selection supported that this reduced effect of household contact is better characterised as a function of the total amount of contact

397 experienced by an individual (C_k) rather than the household size (N), and that the
398 relationship follows an inverse square root law. Previous modelling studies used
399 different frameworks to study the relationship between SITP and household
400 composition. Cauchemez et al. (2014) (14) selected the frequency-dependent mixing
401 assumption (SITP inversely proportional to N) over the density-dependent mixing (SITP
402 independent of N). Many similar studies were also supportive of the frequency-
403 dependent mixing assumption (13,18,21), while Azman et al. (2013) reported an
404 increased transmission rate in larger household (SITP proportional to $N^{0.7}$; although not
405 conclusive due to the limited sample size). One of the strengths of our results is that not
406 only did we propose a better alternative measure to scale SITP than household size, we
407 also differentiated the model from both density- and frequency dependent models with a
408 sufficient support. The best model suggested that within-household transmission
409 patterns lies half-way between the two extremes of density- and frequency-dependent
410 models (we call this the semi-density-dependent model as the total effective contact
411 experienced by an individual is proportional to the total contact intensity to the power of
412 0.5). Although a similar approach (without incorporating heterogeneous contact
413 patterns) was employed in (18), where the authors estimated the STIP proportional to
414 $N^{1.2}$, their CrI was too wide (0.13-2.3) to be conclusive. The large-scale dataset enabled

415 us to obtain a narrower CrI (0.30-0.72) that distinguished the model with significance
416 from the density- and frequency-dependent models. In the semi-density-dependent
417 model, the total amount of effective contact increases in larger household despite the
418 reduced importance of each contact (Figure 4). Therefore, if the risk of external
419 infection is similar between household members, having many household members is a
420 risk factor (which is not usually the case in the frequency-dependent model) because the
421 effect of reduced SITP is outweighed by the increased number of household members
422 who potentially bring infection into the household. Although such effect was not clearly
423 visible in the present study due to the almost exclusive primary infections in children
424 (Figure 5), more distinct characteristics may be seen in other epidemic settings with the
425 semi-density dependent model.

426 Multiple limitations in the present study must be acknowledged. Firstly, the
427 case definition in the dataset was not very strict. The data was collected by self-written
428 questionnaires and it was impossible to validate their response. In the dataset, all student
429 cases were reported to be with a clinical diagnosis, and more than 95% of diagnoses
430 were based on RDKs (42). Considering that primary school students in Japan are highly
431 motivated to visit medical institutions to obtain a leave of absence from school, we
432 believe that our data was able to capture influenza incidence in primary schools at high

accuracy. However, it is not clear if the same applies to their household members; diagnosis was not explicitly required for household members on the question sheet, although the term "influenza" rather than "influenza-like illness" was used. Moreover, subclinical infections may have been present both in children and adults. Because of this, we considered underreporting in the sensitivity analysis, leaving the main conclusions unaltered. Secondly, our model formulation is only one possible candidate for parameterising within-household transmission patterns. "Contact" in our model was merely a hypothetical quantity and may not be directly related to actual physical or social contacts. We also had to use a relatively simple contact pattern matrix for successful parameter estimation. Although our model successfully explained the current data incorporating in an interpretable manner, further development may be sought in the future, including empirical characterisations of household contact patterns which is currently lacking. A recent study have suggested the possible age-dependency in the contact frequency between siblings (6), but the age of household members were not available in the current dataset. More informative dataset and understanding of age-dependent household contact patterns will yield further clarification on this point. Furthermore, one must be aware that our analysis based on a unique study population, i.e., households with at least one primary school student in Matsumoto city, may not be

451 overgeneralized. Extrapolating our household transmission model to household
452 compositions not included in the dataset, e.g., household with no children, may be
453 unreliable. Thirdly, the present study radically simplified the risk factors of individuals.
454 Covariates other than familial roles and household compositions, e.g., comorbidities,
455 vaccination history, previous exposures or habits of personal hygiene, were not
456 considered. The risk of external infection in children was estimated as a single value,
457 which may potentially vary between classes, grades and schools. Overdispersion in
458 infectiousness as addressed in (13,43,44) was also assumed to be negligible.
459 Nonetheless, it is of note that the model had a fairly good performance despite
460 considerable simplification.

461 Although more follow-up studies that supplement our findings are to be
462 awaited, we believe that the present study has presented useful insights on the
463 household-level dynamics of influenza. Understanding of the household-specific contact
464 patterns will help us illustrate how influenza spreads across multiple social settings and
465 facilitate individual and political decisions on disease control accounting for household-
466 specific characteristics.

467

468 **Author contributions**

469 AE: conceptualisation, methodology, software, formal analysis, visualisation, writing

470 (original draft preparation)

471 MU: data curation, writing (review & editing)

472 AK: methodology, supervision, writing (review & editing)

473 SF: methodology, supervision, writing (review & editing)

474

475 **Tables**

476 Table 1. The number of individuals and influenza cases in each type

Individual type		Counts	Cases	Attack ratio (%)
Student	Overall	10,410	2,137	20.5
	Male	5,311	1,132	21.3
	Female	5,099	1,005	19.7
	Grade 1	1,831	406	22.2
	2	1,773	363	20.5
	3	1,731	342	19.8
	4	1,717	375	21.8
	5	1,674	322	19.2
	6	1,684	329	19.5
Father		9,201	629	6.8
Mother		10,260	866	8.4
Sibling		12,632	2,320	18.4
Other		4,356	191	4.4

477 * The number of respondents and cases for “Father”, “Mother”, “Sibling” and “Other”

478 is obtained from the response to the questionnaire and may be redundant due to the

479 inclusion of multiple students from the same household.

480

481 Table 2. Frequency distribution table for compositions of households included in the

482 retrospective data

Order	Composition	# of households	Order	Composition	# of households
1	FM-2	3,915	11	M-3	160
2	FM-3	1,971	12	FM-1-2	134
3	FM-1	899	13	FM-1-1	97
4	FM-2-2	606	14	M-1-2	86
5	M-2	429	15	M-2-2	80
6	FM-2-1	415	16	FM-2-3	70
7	FM-3-2	297	17	FM-3-3	57
8	FM-4	250	18	FM-4-2	55
9	FM-3-1	232	19	M-1-1	43
10	M-1	205	20	M-2-1	39
				Subtotal	10,040 (95.7%)

483 Only 20 most frequent compositions are shown, accounting for 95.7% of the total

484 10,486 responses. Household compositions are denoted in the following manner.

485 FM: households with both father and mother; M: households with only mother; The first
486 number: the total number of siblings in the household; The second number (where
487 applicable): the number of other members (mostly grandparents) in the household.

488

489 Table 3. Parameter estimates by the best model.

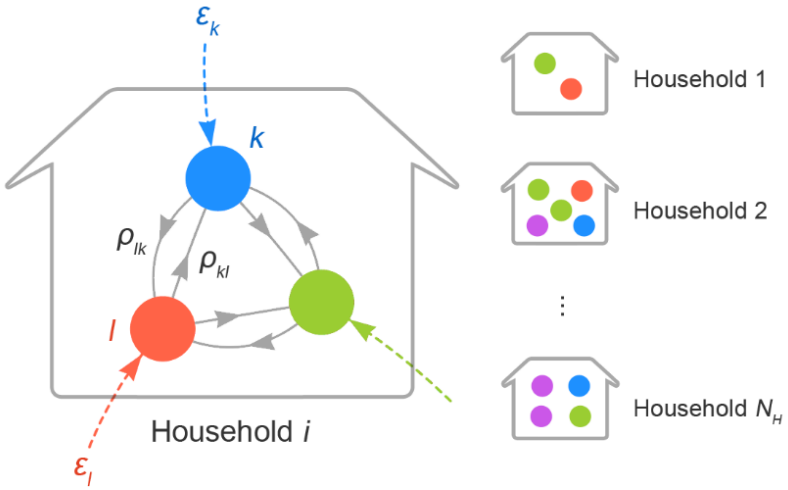
Parameter		Prior	Estimate (95% CrI)
External risk (ε_k)	Student	1-LogUnif(0,1)*	0.197 (0.188-0.207)
	Sibling		0.161 (0.153-0.169)
	Mother		0.035 (0.030-0.040)
	Father		0.038 (0.033-0.043)
	Other		0.013 (0.009-0.017)
Contact intensity (c_{kl})	Child-Child	Unif$(0,\infty)$	1.04 (0.88-1.23)
	Mother-Child		1.16 (1.00-1.32)
	Father-Mother		1 (0.748-1.282)
	Other-Other		1.97 (1.10-3.24)
	Cross generational		0.43 (0.35-0.52)
Transmissibility (β)		(not sampled by MCMC)	0.20 (0.16-0.24)
Exponent parameter (γ)		Uniform$(-\infty,\infty)$	0.51 (0.33-0.69)

490 * Cumulative force of infection is uniformly distributed.

491

492 **Figures**

493



494

495 Figure 1. A schematic illustration of household chain-binomial model.

496 Nodes in different colours corresponds to different types of individuals (e.g., father,

497 sibling, etc.). Transmission patterns are illustrated taking household *i* as an example.

498 Coloured dotted edges represent the risk of external infection $\boldsymbol{\varepsilon}$ to each individual. Solid

499 grey edges denote person-to-person transmission risk (PTR) from one type of person to

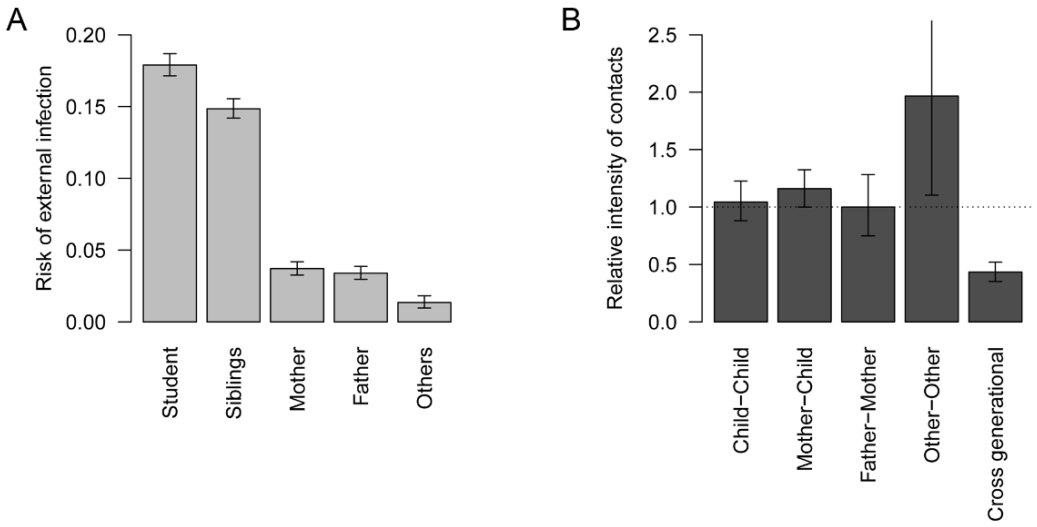
500 another. PTR from type *l* to *k* is given as ρ_{kl}, which refers to the risk of transmission

501 given that the individual of type *l* is infectious. Households have different compositions

502 and ρ_{kl} may also vary according to the composition. On the other hand, $\boldsymbol{\varepsilon}$ is the risk from

503 outside the household and thus assumed to be identical across households.

504



505

506 Figure 2. Estimated risk of external infection and relative intensity of within-household

507 contact. (A) Estimated risk of external infection for each type of individual. (B)

508 Estimated relative intensity of within-household contact. Values are scaled so that the

509 median of contact intensity between adults is 1 (horizontal dotted line). Whiskers

510 indicate 95% credible intervals (CrI).

511

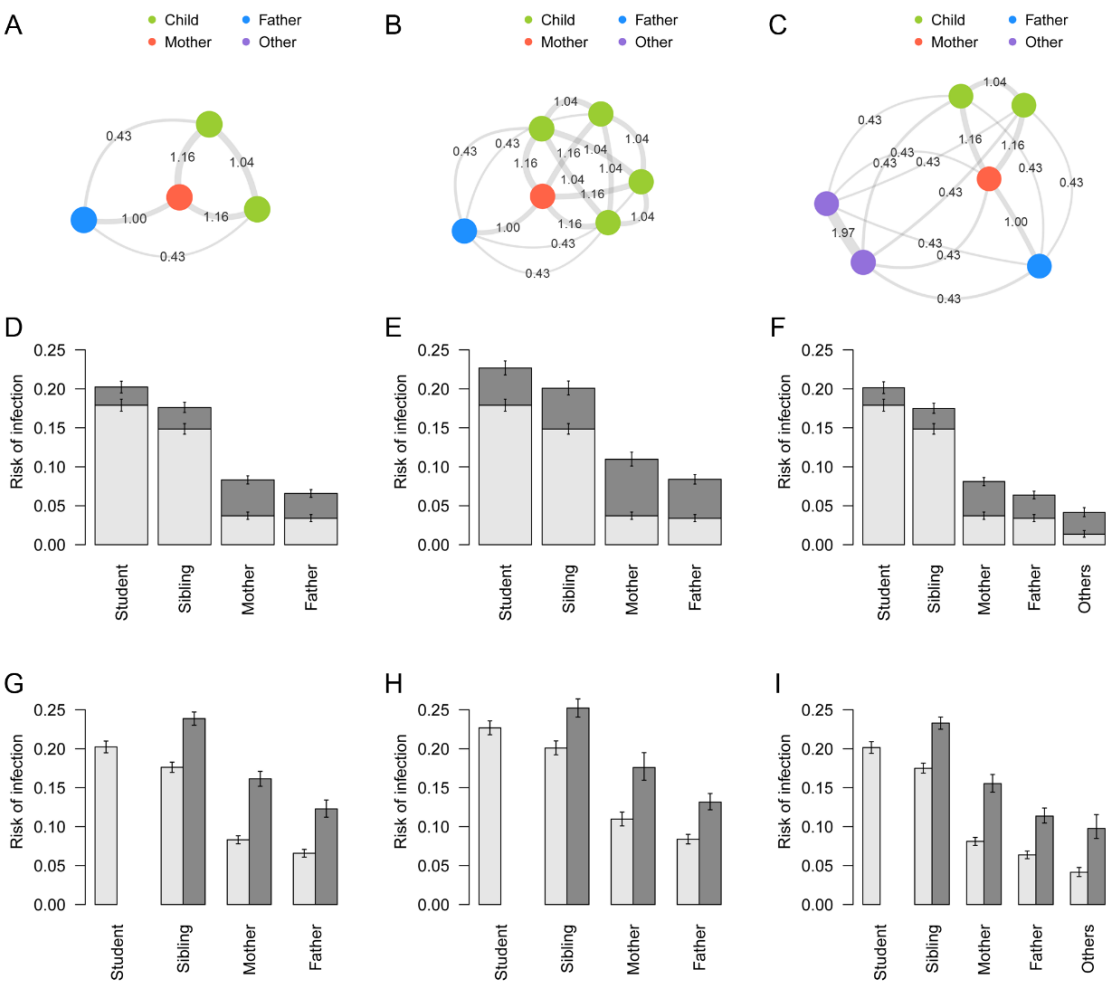


Figure 3. Contact patterns and risk of infection in specific household compositions.

(A)-(C) Network graphs showing contact intensity between individuals for different household compositions: (A) "nuclear family", (B) "many-siblings family", (C) "three-generation family". Node colours represent the type of individuals. Edges denote the relative intensity of contact (c_{kl}) between individuals.

(D)-(F) Risk of infection in households of different compositions stratified by source. Light grey: risk of infection from outside the household; dark grey: risk of infection

520 from within the household. Whiskers indicate the 95% CrI.

521 (G)-(I) Unconditional risk of infection and conditional risk given introduction of

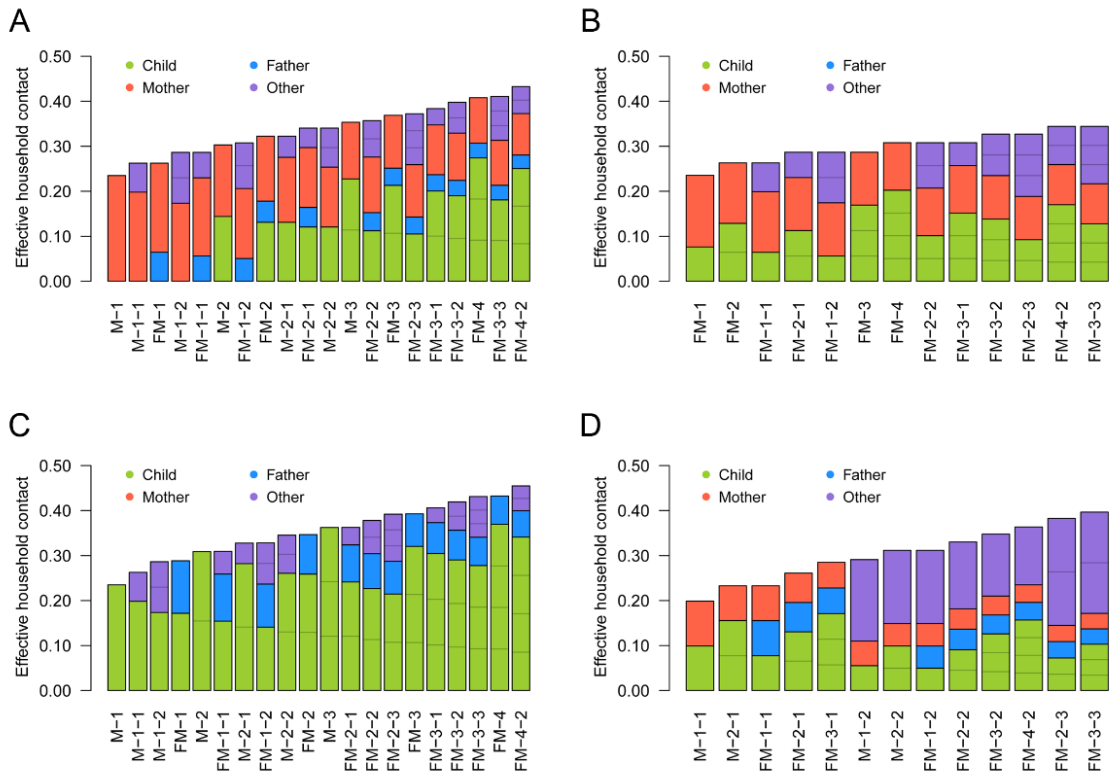
522 infection into a household. Light grey: overall risk of infection for each individual in the

523 household; dark grey: risk of overall infection conditional that a student is infected

524 outside and introduces infection into the household. Infection of the student is given and

525 thus the conditional risk for the student is not shown. Whiskers indicate the 95% CrI.

526



527

528 Figure 4. The effective amount of contacts experienced by individuals (R_{kl}) in different

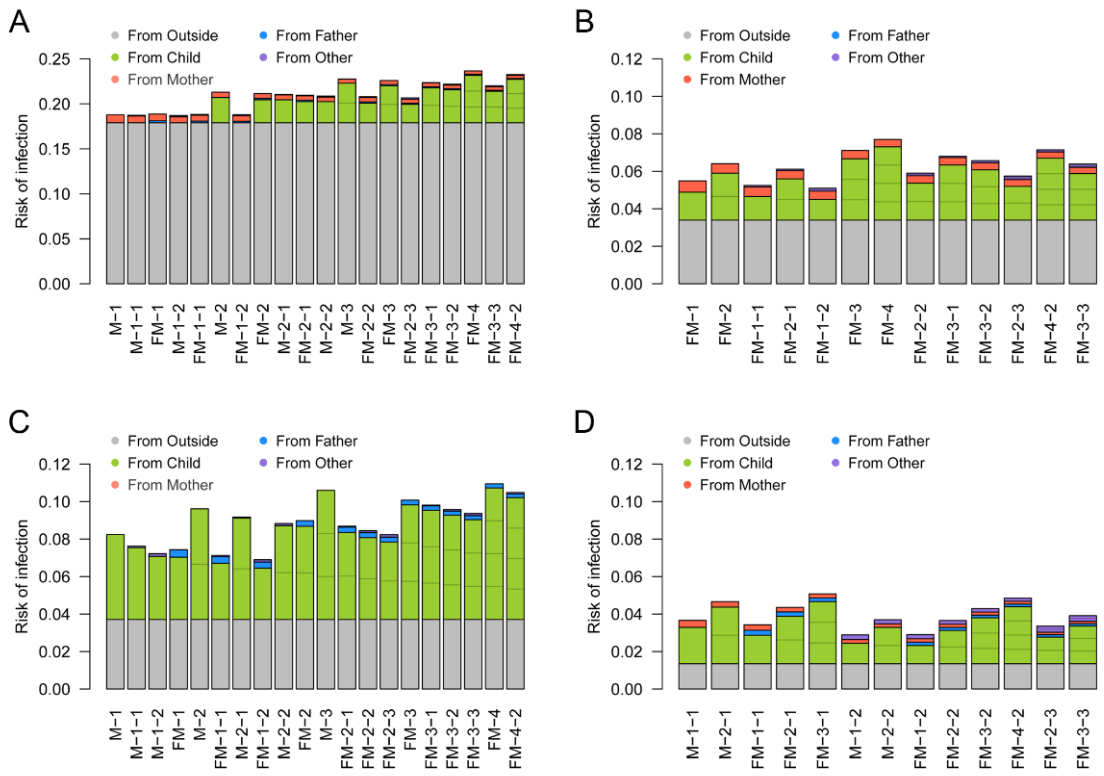
529 household compositions.

530 (A) Child; (B) Father; (C) Mother; (D) Other. The coloured compartments denote the

531 breakdown of effective contacts allocated to each individual in the household, which

532 corresponds to SITP given that individual is infectious.

533



534

535 Figure 5. The risk of primary/secondary infection to individuals in different household

536 compositions and its source.

537 (A) Child; (B) Father; (C) Mother; (D) Other. The coloured compartments denote the

538 breakdown of sources. Household compositions are displayed in the same order as

539 Figure 4. The risk of primary infection in children was set to be 16.4%, the average

between those of "students" and "siblings". Note that the scale of the y axis in (E) is different from the other 3 panels.

References

1. le Polain de Waroux O, Flasche S, Kucharski AJ, Langendorf C, Ndazima D, Mwanga-Amumpaire J, et al. Identifying human encounters that shape the transmission of Streptococcus pneumoniae and other acute respiratory infections. Epidemics. 2018;

2. Christakis NA, Fowler JH. Social network sensors for early detection of contagious outbreaks. PLoS One. 2010;

3. Ackerman E, Elveback LR. Simulation of infectious disease epidemics. Springfield, Ill.: C. C. Thomas; 1984.

4. Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature. 2006;

5. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008;

555 6. Goeyvaerts N, Santermans E, Potter G, Torneri A, Kerckhove K Van, Willem L, et al. Household
556 members do not contact each other at random: implications for infectious disease modelling. Proc
557 R Soc B Biol Sci. 2018;285(1893):20182201.

558 7. Ibuka Y, Ohkusa Y, Sugawara T, Chapman GB, Yamin D, Atkins KE, et al. Social contacts,
559 vaccination decisions and influenza in Japan. J Epidemiol Community Health. 2016;

560 8. Tsang TK, Lau LLH, Cauchemez S, Cowling BJ. Household Transmission of Influenza Virus.
561 Trends Microbiol. 2016;24(2):123–33.

562 9. Lau LLH, Nishiura H, Kelly H, Ip DKM, Leung GM, Cowling BJ. Household transmission of
563 2009 pandemic influenza A (H1N1): a systematic review and meta-analysis. Epidemiology. 2012;

564 10. Longini IM, Koopman JS. Household and community transmission parameters from final
565 distributions of infections in households. Biometrics. 1982;38(1):115–26.

566 11. Becker NG, Britton T. Statistical studies of infectious disease incidence. J R Stat Soc B.
567 1999;61(2):287–307.

568 12. Ball F, Neal P. A general model for stochastic SIR epidemics with two levels of mixing. Math
569 Biosci. 2002;180:73–102.

13. House T, Inglis N, Ross J V., Wilson F, Suleman S, Edeghere O, et al. Estimation of outbreak severity and transmissibility: Influenza A(H1N1)pdm09 in households. BMC Med. 2012;10(1):117.

14. Cauchemez S, Ferguson NM, Fox A, Mai LQ, Thanh LT, Thai PQ, et al. Determinants of Influenza Transmission in South East Asia: Insights from a Household Cohort Study in Vietnam. PLoS Pathog. 2014;10(8):2–9.

15. O'Neill PD, Balding DJ, Becker NG, Eerola M, Mollison D. Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. J R Stat Soc Ser C-Applied Stat. 2000;49(4):517–42.

16. Wardell R, Prem K, Cowling BJ, Cook AR. The role of symptomatic presentation in influenza A transmission risk. Epidemiol Infect. 2017;

17. Azman AS, Stark JH, Althouse BM, Vukotich CJ, Stebbins S, Burke DS, et al. Household transmission of influenza A and B in a school-based study of non-pharmaceutical interventions. Epidemics. 2013;

18. Cauchemez S, Bhattarai A, Marchbanks TL, Fagan RP, Ostroff S, Ferguson NM, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. Proc Natl Acad Sci. 2011;108(7):2825–30.

587 19. Cauchemez S, Donnelly CA, Reed C, Ghani AC, Fraser C, Kent CK, et al. Household
588 transmission of 2009 pandemic influenza A (H1N1) virus in the United States. N Engl J Med.
589 2009;

590 20. Tsang TK, Cauchemez S, Perera RAPM, Freeman G, Fang VJ, Ip DKM, et al. Association
591 between antibody titers and protection against influenza virus infection within households. J
592 Infect Dis. 2014;

593 21. Cauchemez S, Ferguson NM, Wachtel C, Tegnell A, Saour G, Duncan B, et al. Closure of schools
594 during an influenza pandemic. The Lancet Infectious Diseases. 2009.

595 22. Thai PQ, Mai LQ, Welkers MRA, Hang NLK, Thanh LT, Dung VTV, et al. Pandemic H1N1
596 virus transmission and shedding dynamics in index case households of a prospective Vietnamese
597 cohort. J Infect. 2014;

598 23. Wu JT, Riley S, Fraser C, Leung GM. Reducing the impact of the next influenza pandemic using
599 household-based public health interventions. PLoS Med. 2006;

600 24. Budge PJ, Griffin MR, Edwards KM, Williams J V., Verastegui H, Hartinger SM, et al. Impact of
601 home environment interventions on the risk of influenza-associated ARI in Andean Children:
602 Observations from a prospective household-based cohort study. PLoS One. 2014;

603 25. Yeung MPS, Lam FLY, Coker R. Factors associated with the uptake of seasonal influenza
604 vaccination in adults: A systematic review. Journal of Public Health (United Kingdom). 2016.

605 26. Wu S, Su J, Yang P, Zhang H, Li H, Chu Y, et al. Factors associated with the uptake of seasonal
606 influenza vaccination in older and younger adults: A large, population-based survey in Beijing,
607 China. BMJ Open. 2017.

608 27. Uchida M, Kaneko M, Hidaka Y, Yamamoto H, Honda T, Takeuchi S, et al. Effectiveness of
609 vaccination and wearing masks on seasonal influenza in Matsumoto City, Japan, in the 2014/2015
610 season: An observational study among all elementary schoolchildren. Prev Med Reports.
611 2017;5:86–91.

612 28. Chartrand C, Leeflang MMG, Minion J, Brewer T, Pai M. Accuracy of rapid influenza diagnostic
613 tests: A meta-analysis. Annals of Internal Medicine. 2012.

614 29. Bruning AHL, Leeflang MMG, Vos JMBW, Spijker R, de Jong MD, Wolthers KC, et al. Rapid
615 Tests for Influenza, Respiratory Syncytial Virus, and Other Respiratory Viruses: A Systematic
616 Review and Meta-analysis. Clin Infect Dis. 2017;

617 30. Yamazaki M, Mitamura K, Ichikawa M, Kimura K, Komiyama O, Shimizu H, et al. [Evaluation
618 of flow-through immunoassay for rapid detection of influenza A and B viruses]. Kansenshogaku
619 Zasshi. 2004;

620 31. Hara M, Takao S, Fukuda S, Shimazu Y, Miyazaki K. [Comparison of three rapid diagnostic kits
621 using immunochromatography for detection of influenza A virsuses]. Kansenshogaku Zasshi.
622 2004;

623 32. Hara M, Sadamasu K, Takao S, Shinkai T, Kai A, Fukuda S, et al. [Evaluation of
624 immunochromatography test for rapid detection of influenza A and B viruses using real-time
625 PCR]. Kansenshogaku Zasshi. 2006;

626 33. Longini IM, Koopman JS, Haber M, Cotsonis GA. Statistical inference for infectious diseases:
627 Risk-specific household and community transmission parameters. Am J Epidemiol. 1988;

628 34. Begon M, Bennett M, Bowers RG, French NP, Hazel SM, Turner J. A clarification of
629 transmission terms in host-microparasite models: Numbers, densities and areas. Epidemiol Infect.
630 2002;

631 35. Watanabe S. A Widely Applicable Bayesian Information Criterion. 2013;14:867–97.

632 36. Mizumoto K, Yamamoto T, Nishiura H. Age-dependent estimates of the epidemiological impact
633 of pandemic influenza (H1N1-2009) in Japan. Comput Math Methods Med. 2013;

634 37. Glezen WP. Emerging infections: Pandemic influenza. Epidemiologic Reviews. 1996.

635 38. Thompson WW, Shay DK, Weintraub E, Brammer L, Bridges CB, Cox NJ, et al. Influenza-
636 associated hospitalizations in the United States. J Am Med Assoc. 2004;

637 39. Schanzer DL, Tam TWS, Langley JM, Winchester BT. Influenza-attributable deaths, Canada
638 1990-1999. Epidemiol Infect. 2007;

639 40. Jefferson T, Foxlee R, Del Mar C, Dooley L, Ferroni E, Hewak B, et al. Physical interventions to
640 interrupt or reduce the spread of respiratory viruses: Systematic review. BMJ. 2008;

641 41. Aiello AE, Murray GF, Perez V, Coulborn RM, Davis BM, Uddin M, et al. Mask Use, Hand
642 Hygiene, and Seasonal Influenza‐Like Illness among Young Adults: A Randomized
643 Intervention Trial. J Infect Dis. 2010;

644 42. Uchida M, Kaneko M, Hidaka Y, Yamamoto H, Honda T, Takeuchi S, et al. Prospective
645 epidemiological evaluation of seasonal influenza in all elementary schoolchildren in Matsumoto
646 city, Japan, in 2014/2015. Jpn J Infect Dis. 2017;

647 43. Ball F. A unified approach to the distribution of total size and total area under the trajectory of
648 infectives in epidemic models. Adv Appl Probab. 1986;

649 44. Van Boven M, Koopmans M, Van Beest Holle MDR, Meijer A, Klinkenberg D, Donnelly CA, et
650 al. Detecting emerging transmissibility of avian influenza virus in human households. PLoS
651 Comput Biol. 2007;3(7):1394–402.

652