Reconstructing the household transmission of influenza in the suburbs of Tokyo based on clinical cases

Masaya Masayoshi Saito (saitohm@sun.ac.jp)
The Institute of Statistical Mathematics https://orcid.org/0000-0002-4840-6515
Nobuo Hirotsu
Hirotsu Clinic
Hiroka Hamada
The Institute of Statistical Mathematics
Mio Takei
The Institute of Statistical Mathematics
Keisuke Honda
The Institute of Statistical Mathematics
Takamichi Baba
Shionogi Co. & Ltd.
Takahiro Hasegawa
Shinogi Co. & Ltd.
Yoshitake Kitanishi
Shionogi CO. & Ltd.

Research

Keywords: Influenza, Household transmission, Mathematical model, Stochastic simulation, Infectious period, Secondary attack ratio

Posted Date: January 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-37881/v2

Version of Record: A version of this preprint was published on February 10th, 2021. See the published version at https://doi.org/10.1186/s12976-021-00138-x.

Reconstructing the household transmission of influenza in the suburbs of Tokyo based on clinical cases

Masaya M. Saito[1, 2], Nobuo Hirotsu[3], Hiroka Hamada[2], Mio Takei[2], Keisuke Honda[2], Takamichi Baba[4], Takahiro Hasegawa[4], Yoshitake Kitanishi[4]

[1] University of Nagasaki Siebold, Nagasaki, Japan

[2] The Institute of Statistical Mathematics, Tokyo, Japan

[3] Hirotsu Clinic, Kawasaki, Japan

[4] Shionogi & Co., Ltd., Osaka, Japan

Abstract

Background

Influenza is a public health issue that needs to be addressed strategically. The assessment of detailed infectious profiles is an important part of this effort. Household transmission data play a key role in estimating such profiles. We used diagnostic and questionnaire-based data on influenza patients at a Japanese clinic to estimate the detailed infectious period (as well as incubation period, symptomatic and infectious periods, and extended infectious period after recovery) and the secondary attack ratio (SAR) of influenza for households of various sizes based on a modified Cauchemez-type model.

Results

The data were from enrolled patients with confirmed influenza who were treated at the Hirotsu Clinic (Kawasaki, Japan) with a neuraminidase inhibitor (NAI) during six northern hemisphere influenza seasons between 2010 and 2016. A total of 2,342 outpatients, representing 1,807 households, were included. For influenza type A, the average incubation period was 1.43 days (95% probability interval: 0.03-5.32 days). The estimated average symptomatic and infective period was 1.76 days (0.33-4.62 days); the extended infective period after recovery was 0.25 days. The estimated SAR rose from 20% to 32% as household size increased from 3 to 5. For influenza type B, the average incubation period, average symptomatic and infective period, and extended infective period were estimated as 1.66 days (0.21-4.61), 2.62 days (0.54-5.75) and 1.00 days, respectively. The SAR increased from 12% to 21% as household size increased from 3 to 5.

Conclusion

All estimated periods of influenza type B were longer than the corresponding periods for type A. However, the SAR for type B was less than that for type A. These results may reflect Japanese demographics and treatment policy. Understanding the infectious profiles of influenza is necessary for assessing public health measures.

Keywords

Influenza, Household transmission, Mathematical model, Stochastic simulation, Infectious period,

Secondary attack ratio

Background

Simulation-based studies [1, 2] have been effectively used to assess the burden of influenza on society and the reaction of public health, as well as to help understand the dynamical nature of an epidemic [3]. The household is considered a particularly useful artificial experimental environment in which all members of the family are expected to experience intense contact [4, 5]. The risk of infection attributed to infective individuals is directly reflected in the data as a series of infection events in a particular household, which is largely independent of other households based on the intense contact. Accordingly, household data have played a key role in addressing estimation problems and have received particular attention in the analysis of influenza infection. In a number of previous studies, a Reed & Frost-type model has been used to estimate the probability that one susceptible subject will experience at least one contact with one infected household subject per unit time [6, 7] using a chain binomial model. In this type of model, the parameters are essentially estimated based on the number of infected subjects at each point in time. As an alternative, Longini et al. [8] proposed a constructive way to estimate the probability of being infected by an infective household member or community from the final numbers at the end of the epidemic. Carrat et al. [9] conducted a longitudinal study of the household transmission of influenza in the 1999/2000 season that included the start and end times of the illness for 946 households. Cauchemez et al. [4] applied a Bayesian Markov chain Monte Carlo (MCMC) to the outcome of the study to estimate the force of infection and the distribution of the infectious period simultaneously.

Although the infective period is one of the most clinically important parameters to describe natural history, it alone is not sufficient to describe influenza epidemics. There have been numerous attempts to identify other parameters, both in Japan and elsewhere. Frequency of social contacts and the secondary attack ratio (SAR), which is the probability of any member (of $n - 1$) being infected by the primary source, in household contacts are also essential factors for describing epidemics, in addition to the infectious and latent periods. Wallinga et al. [10] quantified the concept of social contacts in an age-specific contact matrix, which was then applied by Mossong et al. [11] in an investigation in

Europe. Carcione et al. [12] estimated the SAR in households during the first circulation of the pandemic influenza A(H1N1) 2009. In Japan, Uchida et al. [13, 14, 15] conducted questionnaire-based studies of school outbreaks. Takeuchi et al. [16] and Ibuka et al. [17] investigated social contacts in a village in Miyazaki prefecture and among age-stratified responders recruited online, respectively. Nishiura & Oshitani [5] estimated the SAR in households for the pandemic influenza A(H1N1) 2009.

Hirotsu et al. [6] used data from the Hirotsu Clinic in Kawasaki City, a major city in the greater Tokyo area, to conduct a single-center, prospective, observational study (UMIN‐CTR: UMIN000024650) involving the transmission of influenza during six influenza seasons (2010–2016). The data were taken from the records of 2,342 outpatients, representing 1,807 households, who were diagnosed with influenza A or B. Each household record consisted of the diagnosis as well as the infection history of other household members who were tracked via a questionnaire provided to the outpatients. These records serve as the basis for the current study.

Given the underlying information available in the diagnostic records of the Hirotsu Clinic, we sought to estimate the latent and infectious periods and to reconstruct household transmission in a pair of simulations. The latent period of influenza is hardly identifiable via routinely corrected epidemic data except when a small outbreak occurs, one induced by clearly identified primary cases [6]. Moreover, asymptomatic agents may have a non-negligible influence on the epidemic [18, 19].

We employ a modified Cauchemez-type household transmission model. After conducting two simulation trials to produce a detailed infection profile, we explore the information extracted from the Hirotsu Clinic's influenza diagnosis records. In the first step, estimates of the infectious period are produced by combining the available records and the simulation model. In the second step, we simulated inter-household transmission assuming a particular force of infection between households and calibrate the assumed value so that the number of simulated infected households for each household size agrees with the reality represented in the data. It is expected that the new and more detailed infectious profile developed here for influenza will contribute to public health globally.

The remainder of the paper is organized as follows: In the Methods section, the dataset is introduced and the methods applied to the dataset are described mathematically. In the Results section, parameter

estimates related to the natural history of influenza are presented. In the Discussion section, we summarize the outcomes of our trials and discuss the study's strengths and limitations. Finally, we offer concluding remarks and indicate future research directions in the Conclusion section.

Methods

Data source

The data were derived from enrolled patients with confirmed influenza who were treated at the Hirotsu Clinic (Kawasaki, Japan) with a neuraminidase inhibitor (NAI) during the six northern hemisphere influenza seasons between 2010 and 2016. A total of 2,342 outpatients, representing 1,807 households, were included (Table 1). A majority of the households have 3-5 members (parents and their children), as summarized in Table 1b. This is reflected in the age-specific distribution of the total infected cases over the six-year period shown in Fig. 1, where the 0-11 age group (preschool and elementary school children) and the 30-49 age group are predominant. The dataset includes individuals based on infection, either diagnosed or indicated by the family. For this reason, the corresponding 'divisor' necessary to obtain a crude estimate of the infection probability is not clearly defined. The divisors are calculated by counting all members of the household once a year if the household experiences the disease. The jagged lines in Fig. 1 show the crude infection probability for influenza A or B . While a general downward trend (see Ref [22]) is apparent in the case of type B, there is no clear trend for type A. Patients of any age who were diagnosed with influenza A or B using rapid influenza diagnostic tests (RIDTs) were eligible for inclusion in the study. ImmunoAce® Flu (Tauns Laboratories, Inc., Shizuoka, Japan) was used for the differential diagnosis of influenza A and B. In order to share the data among members of the research team, the original Hirotsu Clinic data were first anonymized by keeping only assigned identifiers, along with the times of infection-related events and the relationship between identifiers in order to reconstruct households. This anonymization did not affect the analysis. The infection events consisted of the onset (based on questionnaire responses), diagnosis at the clinic, and recovery (occasionally N/A was entered to indicate the disappearance of

132 constitutional symptoms, specifically antipyresis < 37.5°C). The dosage and administration of the
133 NAIs were as per the package insert for each product. Secondary infection patients were defined as
134 household members who were diagnosed with the same influenza type as the index patient within 8
135 days after the onset of symptoms in the index patient. Occasionally, two or more members of a
136 household may be simultaneously (or nearly so) infected outside the household and be introduced into
137 the household as primaries. Figure 1 shows that the serial interval frequency decreases
138 monotonically up to 0.5 days for influenza type A and up to 1.25 days for influenza type B. This
139 downward trend may be at least partially attributable to the introduction of multiple primaries, as
140 discussed in more detail later.
141

142 Transmission model in a household

143 Cauchemez et al. [4] conducted a longitudinal investigation of 334 households over a 15-day period
144 in winter 1999-2000 and applied their novel household transmission model to estimate parameters
145 describing the natural history of (seasonal) influenza infection. The approach we propose is similar to
146 this in concept. In the model, the time variation of the infectivity attributed to an individual is modeled
147 as a piecewise constant function (Figure 3) that includes four period parameters: pre-symptomatic and
148 non-infective (a), pre-symptomatic and infective (b), symptomatic and infective (d), and extended
149 infective after recovery (e). By definition, infectivity sustains for period $b + d + e$. Period d was
150 observed from the data; period $b + e$, on the other hand, was estimated via maximum likelihood
151 estimation. Because the individual values of b and e could not be broken out from the combined $b + e$
152 value, we set $b = 0$ only for conceptual completeness.
153 We employ a parametric model (which will be explained in the next subsection) to describe a and
154 d as random variables and obtain their point estimates via maximum likelihood estimation. The
155 subscripted version (e.g., a_i for a) denotes the values for an individual i. We assume that a given
156 individual may be infected per unit time according to the probability of the sum of the infectivity
157 attributed to the rest of the family members. Specifically, if family member i acquires infectivity at

time $t_i - b_i$ (t_i is the time of the illness onset informed by the data) and loses it at $t_i' + e_i$ (t_i' is the time of the disappearance of constitutional symptoms, which may be unavailable in the data), then the infection probability per unit time (i.e., the force of infection, FOI) that member j is infected at time t is given by

$$\lambda_j(t) = \frac{\lambda_0}{(n-1)^\alpha} \sum_{i \neq j} \mathbb{1}\,(t_i - b_i \leq t \leq t_i' + e_i), \tag{1}$$

where n is the number of members in the household (i.e., the household size), λ_0 is a constant controlling the FOI, and $\mathbb{1}(\cdot)$ is an indicator function: $\mathbb{1}\,(\mathrm{True}) = 1$ and $\mathbb{1}\,(\mathrm{False}) = 0$. Cauchemez et al. [4] employed this type of power-law risk (assuming $\lambda_j(t) \propto n^{-a}$) to check the effect of the density of infectives and obtained an estimate of α = 0.84 (95% CrI: 0.46-1.21). Later, Ferguson et al. used α = 0.8, a value close to the mean determined by Cauchemez et al., in their pandemic simulation study [1]. Division by the number of other family members (the case of $\alpha = 1$) assumes that infective contacts occur in an exclusive time-sharing manner; the absence of the division ($\alpha = 0$) implies that the FOIs exert influence equally on the population of concern, irrespective of family size. The former setting is appropriate for diseases that require close contact for infection, including influenza, while the latter well matches, for example, diseases where the infection is induced by polluted agents [20]. However, as we will see, because the SAR is non-negligibly large in large families, setting $\alpha = 1$ appears to over-reduce the FOI: an infective agent may have a conversation with two or more family members. For this reason, we introduced an empirically determined power for scaling.

Estimation of parameters

We employ a rather descriptive statistical approach to estimating a, d, and e. Setting b = 0, the symptomatic period $d^{(h,i)}$ of the i-th infected member in household h is informed by data as $t'^{(i)} - t^{(i)}$; summing such realizations over all households and members therein, we have the empirical distribution (*i.e.*, histogram) of period d, along with its mean $\mathrm{E}[d]$. Similarly, in principle, collecting the serial interval instance $t_{\mathrm{int}}^{(h)} \coloneqq t^{(h,2)} - t^{(h,1)}$ over households h, we have the empirical

184 distribution for t_{int}. As mentioned in the data source subsection, however, cases in which two
185 family members who were simultaneously infected (or nearly so) outside the home may occasionally
186 appear. In the expert opinion of one of the authors, an infected agent barely develops sufficient
187 infectiousness within 24 hours. In our study, a serial interval within 24 hours was observed in 54 of
188 290 influenza A pairs (18%) and in 14 of 135 influenza B pairs (18%). It may be that these
189 simultaneous pairs increased the proportion of short serial intervals in the distribution. Figure 2
190 shows the left ends of the histograms of the crude distributions of t_{int} for influenza A and B,
191 respectively. As is apparent in figure, for influenza B, there is a persistent downward trend up to 1.25
192 days, followed by an upward trend that forms the left side of an approximately bell-shaped distribution
193 (the full histogram is shown in Figure 4). Though there is a possibility that a relatively large
194 infectious incubation period of a primary can yield a small (occasionally negative) serial interval, our
195 dataset is not sufficient to deal separately with the infectious incubation period (as noted above). As
196 a consequence, we empirically adjusted the contribution from simultaneous primary cases in the crude
197 serial interval distribution under the assumption that this downward trend is mainly attributed to
198 simultaneous infections, in particular where t_{int} is close to zero. Given that this trend is almost
199 linear, we assumed that the count between t_{int} and $t_{\mathrm{int}} + \Delta t$ is attributable to true household
200 transmission with probability $\propto t_{\mathrm{int}} \Delta t$ and discarded stochastically the data of $t_{\mathrm{int}} < 1.25$ days
201 accordingly. The same procedure was applied to influenza A, with a different cut-off of $t_{\mathrm{int}} < 0.5$
202 days, although the downward trend here is not as apparent as in the case of influenza B. After these
203 adjustments, the "observed" serial interval t_{int} can be modeled as the summation of the interval τ_{int}
204 between the infection times of the primary and secondary infections (i.e., the "intrinsic" serial interval)
205 and the incubation period *a* of the secondary subject. For ease of computation of the *a* distribution, we
206 introduce two simplifications. First, the interval of the two cases is assumed to follow a uniform
207 distribution truncated at the mean infective period τ_{ifv}: $p(\tau_{\mathrm{int}}) = \mathbb{1}\,(0 \leq \tau_{\mathrm{int}} \leq \tau_{\mathrm{ifv}})/\tau_{\mathrm{ifv}}$. Second,
208 the incubation period follows a gamma distribution: $a \sim \mathrm{Gam}\,(\mathrm{shape} = k_a, \mathrm{scale} = \theta_a)$. The
209 distribution form of t_{int} is then written as

$$p(t_{\text{int}}|\, k_a, \theta_a, \tau_{\text{ifv}}) = \frac{d}{dt_{\text{int}}} \int_0^{\min(t_{\text{int}}, \tau_{\text{ifv}})} \frac{1}{\tau_{\text{ifv}}} \int_{\tau_{\text{int}}}^{t_{\text{int}}} \text{Gam}(t'_{\text{int}} - \tau_{\text{int}}|\, k_a, \theta_a) dt'_{\text{int}}\, d\tau_{\text{int}}$$

$$= \frac{1}{\tau_{\text{ifv}}} \int_0^{\min(t_{\text{int}}, \tau_{\text{ifv}})} \text{Gam}(t_{\text{int}} - \tau|\, k_a, \theta_a) d\tau.$$

By maximizing the likelihood $\prod_h p\left(t_{\text{int}} = t_{\text{int}}^{(h)}|\, k_a, \theta_a, \tau_{\text{ifv}}\right)$, we have the distribution of incubation period a. Technically, the simultaneous optimization of $(k_a, \theta_a, \tau_{\text{ifv}})$ is sensitive to the initial conditions. Hence, given τ_{ifv} in a certain range, we optimize for (k_a, θ_a). Since a point estimate of the infective period τ_{ifv} can be equated to $\text{E}[d + e]$, $\tau_{\text{ifv}} - \text{E}[d]$ serves as a point estimate of e. Scaling power α is determined as follows. The SAR is $1 - e^{-\lambda\tau_{\text{ifv}}} \approx \lambda\tau_{\text{ifv}}$ with $\lambda = (n-1) \cdot \lambda_0/(n-1)^\alpha$, from Eq. (1). Our dataset allows us to compute the SAR for household sizes n = 3, 4, and 5 which together comprise approximately 90% of all households (a small number of households were of size 6 and 7). Therefore, the value of α with $\lambda_0\tau_{\text{ifv}}$ is determined by the regression

$$\log \text{SAR}_n = \log \lambda_0 \tau_{\text{ifv}} + (1 - \alpha) \log(n - 1).$$

The value of λ_0 is determined as an MLE via a comparison of the number of secondary cases in the data and in the simulation. Suppose that i secondary cases appear in a simulated household of size n with probability $p_{i/n}$ and that $\left(p_{i/n}\right)_{i=0}^{n-1}$ is obtained through multiple simulation runs with different seeds. Then the likelihood for the comparison is $L(\lambda_0) = \prod_{n=3,4,5} \prod_{i=0}^{n-1} \left(p_{i/n}\right)^{m_{i/n}}$, given $m_{i/n}$ real households yielded i secondary cases. In other words, λ_0 is chosen so that the KL divergence is minimized.

Results

Estimation of duration parameters

To estimate the duration parameters, we first summarized the dataset in the form of histograms for the symptomatic period and the serial interval. The dataset histograms for influenza A, along with the related estimation results, are shown in Figure 4a. We identified 1,389 cases in which the patient

234 exhibited symptoms and 290 transmissions from the primary to the secondary subject. The
235 symptomatic period distribution is well approximated by a gamma distribution with shape = 2.97 and
236 scale = 0.59 days; that is, period *d* is 1.76 days on average, with a 95% probability interval (hereafter
237 95% I) of 0.33-4.62. The point estimate of the infective period as the MLE is 2.01 (left-panel in Figure
238 4a). The difference between the infective and symptomatic periods is the extended infective period
239 after recovery, *e*; here, $e = 2.01 - 1.76 = 0.25$ days. The incubation period is extracted as a gamma
240 distribution with shape = 0.99 and scale = 1.46 days (mean: 1.43 days; 95% I: 0.03-5.32 days).
241 The results for influenza B are shown in Figure 4b. For the analysis here, we identified 760 cases
242 where the patient exhibited symptoms and 135 transmissions from the primary to the secondary subject.
243 A gamma distribution with shape = 3.05 and scale = 0.86 days (mean: 2.62 days; 95% I: 0.54-5.75
244 days) was fitted to the symptomatic period data; notably, the period here is longer than in the case of
245 influenza A. The serial interval was also longer and cut off at approximately 6 days, while the infective
246 period was estimated to be 3.62 days, which yields an incubation period of 1.66 days (95% I: 0.21-
247 4.61 days), as well as a relatively long extended infective period after recovery, $e = 3.62 - 2.62 = 1.00$
248 days.
249

250 # Estimation of FOI

251 We then produced a point estimate of the FOI coefficient λ_0 after fixing scaling power α. In the case
252 of influenza A, the SAR increased from 20% to 32% as household size increased from 3 to 5, for which
253 $\alpha = 0.32$ is optimal. This is a much smaller value than that used in a previous study, where Ferguson
254 et al. [1], who carried out agent simulations and employed a Cauchemez-type model as an internal
255 process, used $\alpha = 0.8$. With $\alpha = 0.32$ and an infective period of 3.73 days, we ran 1,024 simulations
256 to obtain the ML estimate of λ_0. The likelihood as a function of λ_0 is shown in Figure 5a; the optimal
257 value was found to be $\lambda_0 = 0.065$. The corresponding SAR distribution is shown in Figure 6a. As
258 shown, the simulated distribution well reproduces the data.
259 The same procedure was applied to influenza B. Compared to type A, type B exhibited a smaller
260 SAR (range: 12%-21%) and a much longer infective period (5.73 days). The resulting estimates are

also different: $\alpha = 0.17$ and $\lambda_0 = 0.015$. The likelihood function is shown in Figure 5b. The SAR distribution in Figure 6b is similarly reproduced, and two or more secondary cases appear less frequently than for type A.

Discussion

We estimated parameters to describe an influenza natural history, the incubation and infective periods, and the FOI coefficient, using both diagnostic and questionnaire-based data obtained in a clinic located in the suburbs of Tokyo. While the study produced several useful insights, it is not without limitations. Although the estimated incubation period is rather inescapably obscure because it is hidden by the different symptomatic periods for individuals and was reflected in a quite skewed distribution with shape less than unity, it was estimated to be roughly 1.5 days, both for type A and type B. For the latter, a uniform infective period distribution with bi-level infectivity (non-infective or infective) does not fully capture the nature of the influenza (e.g., virus titer over time). The difference between the infective period and the symptomatic period, for which we produced point estimates, was 0.25 days for influenza type A and 1.00 days for type B. While the estimate for type B would seem to be consistent with what is commonly known about this disease, the estimate for type A appears excessively short. According to the Enforcement Regulations for School Health and Safety Act in Japan [21], patients with an influenza virus infection are banned from attending school for 2 days (or 3 days for infant children) after the resolution of fever. Our estimates of the mean of the extended infective period after the resolution of fever were less than 2 or 3 days and were thus consistent with the enforcement regulation. In this respect, our constructed models would appear to be realistic. However, it should be noted that the choice of cut-off points to split simultaneous infection from true household transmission is influential in the estimation and, although the choice in the case of type B is rather clear, it is less obvious for type A.

The scaling power was estimated to be quite small ($\alpha = 0.2$ to 0.3) relative to that reported by Cauchemez et al. However, it should be noted that Cauchemez's α = 0.84 estimate involved substantial uncertainty (95% CrI: 0.46-1.21), and that the best fit power is different for $1/n^{\alpha}$ versus

289 $1/(n-1)^{\alpha}$ (for example, $1/(n-1)^{0.3}$ well fits to $1/n^{0.42}$). A straightforward explanation for
290 this might be that the dataset covered households in which family members tended to spend most of
291 their time with one another. However, two (or more) primary cases introduced almost simultaneously
292 can elevate the apparent SAR in large families. Our analysis assumed that the first reported case
293 infected the second, and so on. The fraction of the ignored tail (> 8 days) should be available to model
294 the probability that multiple members were infected simultaneously, referring to epidemic surveillance
295 in Tokyo. It should also be noted that the SAR may be underestimated due to vaccination and the
296 basic strategy of early diagnosis followed by early treatment. While our dataset does not inform the
297 vaccine effect separately since vaccination records are available only for infected household members,
298 we can compare the SAR values between all households and those with a late-diagnosed primary case.
299 Here, we identify a primary to be late-diagnosed if the waiting time c from the onset to the diagnosis
300 is greater than 1 day, which is the case for 25% of households with type A and 33% of those with type
301 B. Figure 7 shows the number of secondary cases according to household size. The uncertainties
302 at the various points on the distributions are 95% CIs based on binomial sampling. Considering only
303 the mean (crude proportion), the weights shift to a larger number of secondary cases, particularly for
304 type A in four-member households. Late diagnosis and reduced awareness may raise the probability
305 of household outbreaks. However, the uncertainties of the stratified and unstratified distributions
306 overlap one another and we did not find a significant difference.
307

308 # Conclusion

309 As this study was conducted in Japan, the results are likely to reflect Japanese demographics and
310 treatment policy. Although it is important to assess infectious profiles in various countries, the
311 number of studies that have estimated the infectious duration and FOI of influenza in Japan is quite
312 limited. In general, it is useful to understand the infectious profiles of influenza for examining public
313 health measures.
314 We expect that our estimates of the infective period based on the present situation in an urban area
315 in Japan will be informative to school health officials and helpful in determining the span of school

closures and attendance suspensions. However, more research will be needed to improve the accuracy and applicability of our results.

Declarations

Ethics approval and consent to participate

The protocol was approved by the ethics committee of Shionogi & Co., Ltd. Informed consent was not required, as the anonymized data were shared with the Institute of Statistical Mathematics only, as required by the Act on the Protection of Personal Information.

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

TB, TH and YK are employees of Shionogi & Co., Ltd. NH has received research funding and has served as a consultant, advisory board member, and/or speaker for Shionogi & Co., Ltd.

Funding

Not applicable

Authors' contributions

All authors participated in the analytical process and the interpretation of analysis results, and in the drafting, critical revision, and approval of the final version of the manuscript. NH was responsible for data collection and anonymization. MS and HH conducted the statistical analysis. MT developed a program for anonymization and provided it to NH.

Acknowledgements

Anonymization of the original dataset was carried out at the Hirotsu Clinic using a dedicated code

provided by Professor Kazuhiro Minami at the Institute of Statistical Mathematics, Japan. We would like to express our deepest gratitude for his contribution.

References

1. Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, Meeyai A, Iamsirithaworn S, Burke DS. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 2005; 437(7056):209–214.
2. Longini IM Jr, Nizam A, Xu S, Ungchusak K, Hanshaoworakul W, Cummings DAT, Halloran ME. Containing pandemic influenza at the source. *Science* 2005; 309(5737):1083–1087.
3. Colizza V, Barrat A, Barthélemy M, Vespignani A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci U S A* 2006; 103(7):2015–2020.
4. Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boëlle PY. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data, *Stat Med* 2004; 23(22):3469–3487.
5. Nishiura H, Oshitani H. Household transmission of influenza (H1N1-2009) in Japan: age-specificity and reduction of household transmission risk by Zanamivir treatment. *J Int Med Res* 2011; 39(2):619–628.
6. Hirotsu N, Saisho Y, Hasegawa T. The effect of neuraminidase inhibitors on household transmission in Japanese patients with influenza A and B infection: a prospective, observational study. *Influenza Other Respir Viruses* 2019; 13(2):123–132.

7. Abbey H. An examination of the Reed-Frost theory of epidemics. *Hum Biol* 1952; 24(3):201–233.
8. Longini IM, Koopman JS. Household and community transmission parameters from final distributions of infections in households. *Biometrics* 1982; 38(1):115–126.
9. Carrat F, Sahler C, Rogez S, Leruez-Ville M, Freymuth F, Gales CL, Bungener M, Housset B, Nicolas M, Rouzioux C. Influenza burden of illness: estimates from a national prospective survey of household contacts in France. *Arch Intern Med* 2002; 162(16):1842–1848.
10. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am J Epidemiol* 2006; 164(10):936–944.
11. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 2008; 5(3):e74.
12. Carcione D, Giele CM, Goggin LS, Kwan KS, Smith DW, Dowse GK, Mak DB, Effler P. Secondary attack rate of pandemic influenza A(H1N1) 2009 in western Australian households, 29 May-7 August 2009. *Euro Surveill* 2011; 16(3):19765.
13. Uchida M, Tsukahara T, Kaneko M, Washizuka S, Kawa S. Swine-origin influenza a outbreak 2009 at Shinshu university, Japan. *BMC Public Health* 2011; 11:79.

14. Uchida M, Kaneko M, Hidaka Y, Yamamoto H, Honda T, Takeuchi S, Saito M, Kawa S. Effectiveness of vaccination and wearing masks on seasonal influenza in Matsumoto city, Japan, in 2014/2015 season: an observational study among all elementary schoolchildren. *Prev Med Rep* 2016; 5:86–91.
15. Uchida M, Takeuchi S, Saito MM, Koyama H. Effects of influenza vaccination on seasonal influenza symptoms: a prospective observational study in elementary schoolchildren in Japan. *Heliyon* 2020; 6(2):e03385.
16. Takeuchi S, Yamauchi T, Kuroda Y. Increased fertility in an aging society will affect the spread of infectious diseases: a case study in Miyazaki prefecture. *Jpn J Health & Human Ecology* 2014; 80(1):17–22 (in Japanese).
17. Ibuka Y, Ohkusa Y, Sugawara T, Chapman GB, Yamin D, Atkins KE, Taniguchi K, Okabe N, Galvani AP. Social contacts, vaccination decisions and influenza in Japan. *J Epidemiol Community Health* 2016; 70(2):162–167.
18. Brauer F. Modeling Influenza: Pandemics and Seasonal Epidemics. In: Brauer F, van den Driessche P, Wu J (eds) *Mathematical Epidemiology. Lecture notes in mathematics, vol 1945*. Springer: Berlin, Heidelberg, 2008.
19. Somes MP, Turner RM, Dwyer LJ, Newall AT. Estimating the annual attack rate of seasonal influenza among unvaccinated individuals: a systematic review and meta-analysis. *Vaccine* 2018; 36(23):3199–3207.
20. Vynnycky E, White RG. *An Introduction to Infectious Disease Modelling*. Oxford University Press: New York, 2010.
21. The Website (in Japanese). (Available at: https://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=333M50000080018#137) [accessed on June 9, 2020].

22. Kawado M, Hashimoto S, Ohta A, Oba MS, Taniguchi K, Sunagawa T, Matsui T, Nagai M, Murakami Y. Improvement of influenza incidence estimation using auxiliary information in sentinel surveillance in Japan. The Open Infectious Diseases Journal 2018; 10: 29-36.

Tables

Table 1

Summary of diagnosed influenza cases covered by our questionnaire investigation.

a. Number of cases in each year

Year	Type	Cases	Households	Total Cases
1	A	243	189	
	B	104	86	
2	A	263	204	
	B	222	177	
3	A	259	189	
	B	11	8	270
4	A	175	147	
	B	289	227	464
5	A	339	249	
	B	18	13	357
6	A	237	168	
	B	182	150	419
	Total	2342	1807	

b. Number of households of each household size

Household size	1	2	3	4	5	6	7	Total
A	37	62	318	544	149	30	6	1146
B	16	15	182	340	85	18	5	661
Total	53	77	500	884	234	48	11	1807

Figures

Figure 1. Age-specific distributions of the total infected cases over six years (black portion of the histograms) and the corresponding age-specific population (black+white) for influenza A and B. The lines show a crude estimate of the infection probability, calculated as the number of infectious individuals divided by the population.

Figure 2. Partial serial interval histograms of the crude number of secondary cases. Only the left side of the histograms (≤ 2 days) is shown. In the type A histogram, there is a downward trend up to 0.5 days; in the type B histogram, there is a downward trend up to 1.25 days (type B). These may be attributable to simultaneous infections outside the household.

Figure 3. Natural history of infective people and the variation of infectivity. In an actual situation, a person may be infected at some unknown point in time (Infection) and the infectivity to other people gradually increases up to its maximum around the time when the illness is well developed and recognized (Illness onset). It then is drained as the process of recovery from infection proceeds, which may be clinically observed by antipyresis (Antipyresis), though weak infectivity may remain. For simplicity, such time variation of infectivity is modeled using a piecewise constant function that takes a non-zero constant value λ_0 only from one point in time (labeled as Infectious) to another point near Antipyresis. The modeled infectivity function is temporally controlled by four period parameters: *a* (pre-symptomatic and non-infectious), *b* (pre-symptomatic and infectious), *d* (symptomatic and infectious) and *e* (extended infective after recovery**).**

Figure 4. Summary of household infection data and estimates of the parameters for influenza types. a. Type A and b. type B (bottom row). The left column shows the histogram of the symptomatic period and its fitting to a gamma distribution. The middle column shows the histogram of the serial interval, its fitting to the distribution constructed as a convoluted gamma distribution (blue curve), and the extracted infectious period (annotation) and incubation period (red curve) via the fitting (see the Methods section for detail). The right column shows the negative likelihood against the assumed infective period.

Figure 5. Likelihood curves for the two influenza types. a. Type A and b. type B are shown with respect to the number of secondary cases in a household as a function of the force of infection (FOI) coefficient λ_0. The likelihood values are determined by simulation runs with different values of λ_0 and

457 $\Delta\lambda_0 = 0.005$; optimal values of λ_0 are 0.065 for A and 0.015 for B.

458

459 **Figure 6.** Comparison of the number of secondary cases in a household. Households have 3, 4, or 5
460 members in the actual data (white) and the simulation (black) for the two influenza types: a. type A
461 and b. type B. The simulation run is mainly controlled by the force of infection (FOI) coefficient λ_0
462 and the household size scaling power a; their ML estimates are shown.

463

464 **Figure 7.** Comparison of the number of secondary cases in households with $c_1 \leq 1$ (*early*-
465 diagnosed primary) and those with $c_1 > 1$ (*late*-diagnosed primary): a. type A and b. type B.

Figures

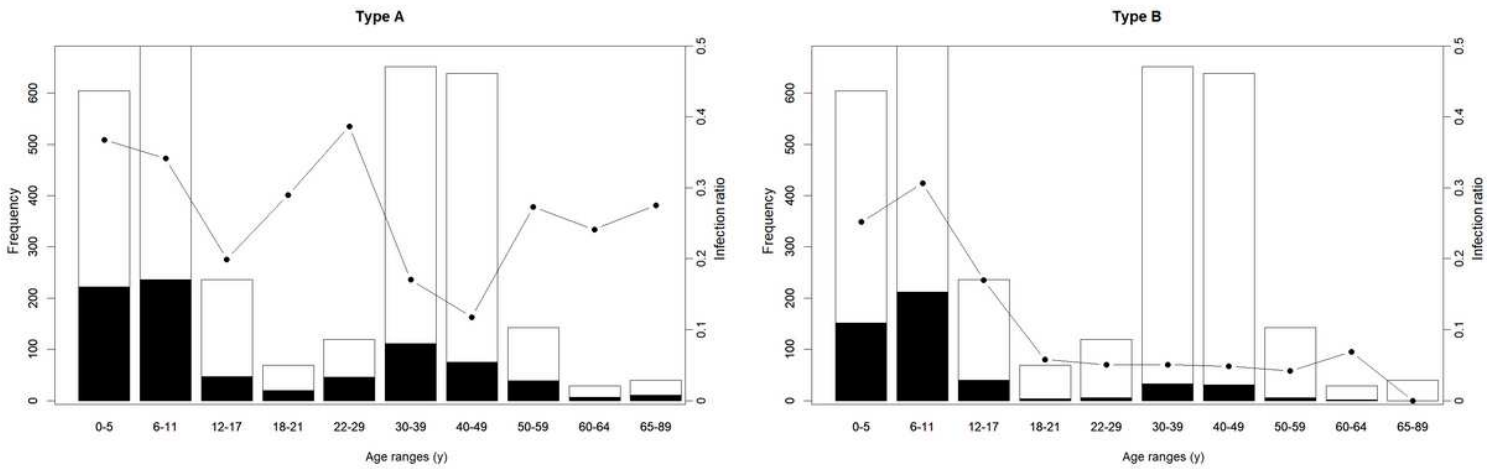

Figure 1

Age-specific distributions of the total infected cases over six years (black portion of the histograms) and the corresponding age-specific population (black+white) for influenza A and B. The lines show a crude estimate of the infection probability, calculated as the number of infectious individuals divided by the population.

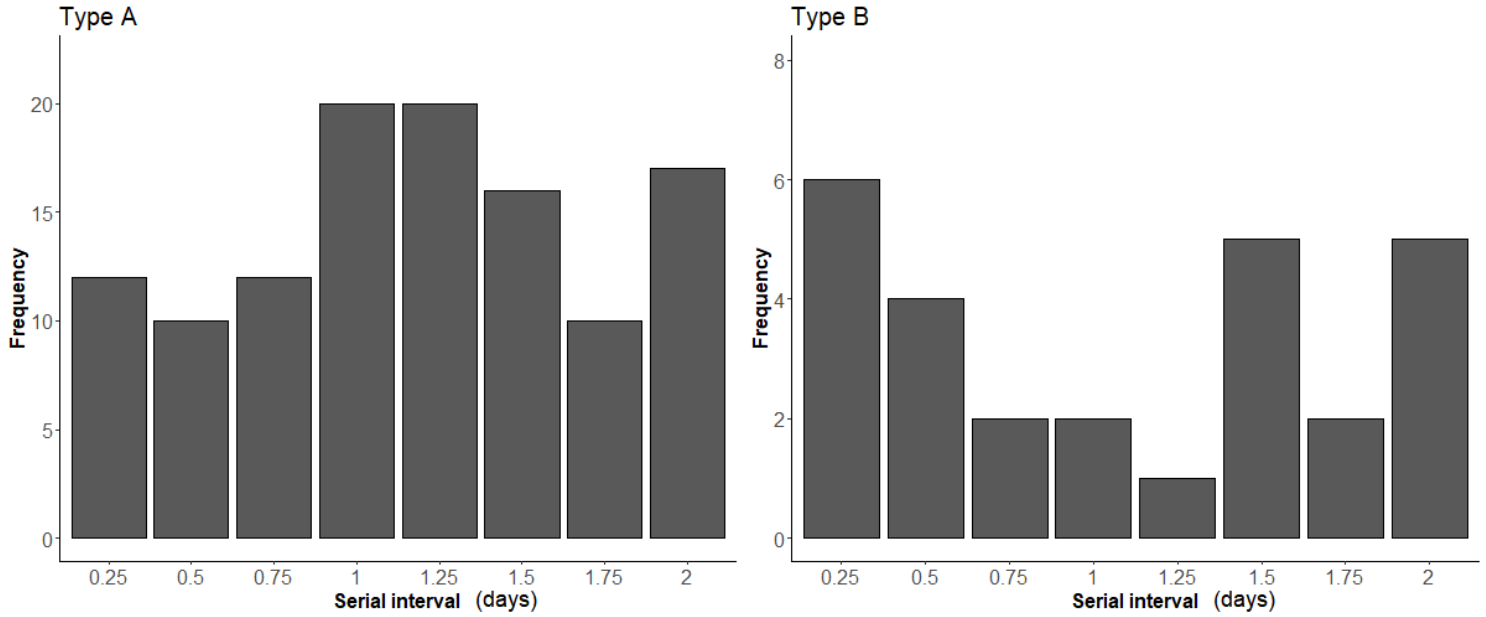

Figure 2

Partial serial interval histograms of the crude number of secondary cases. Only the left side of the histograms (≤ 2 days) is shown. In the type A histogram, there is a downward trend up to 0.5 days; in the type B histogram, there is a downward trend up to 1.25 days (type B). These may be attributable to simultaneous infections outside the household.

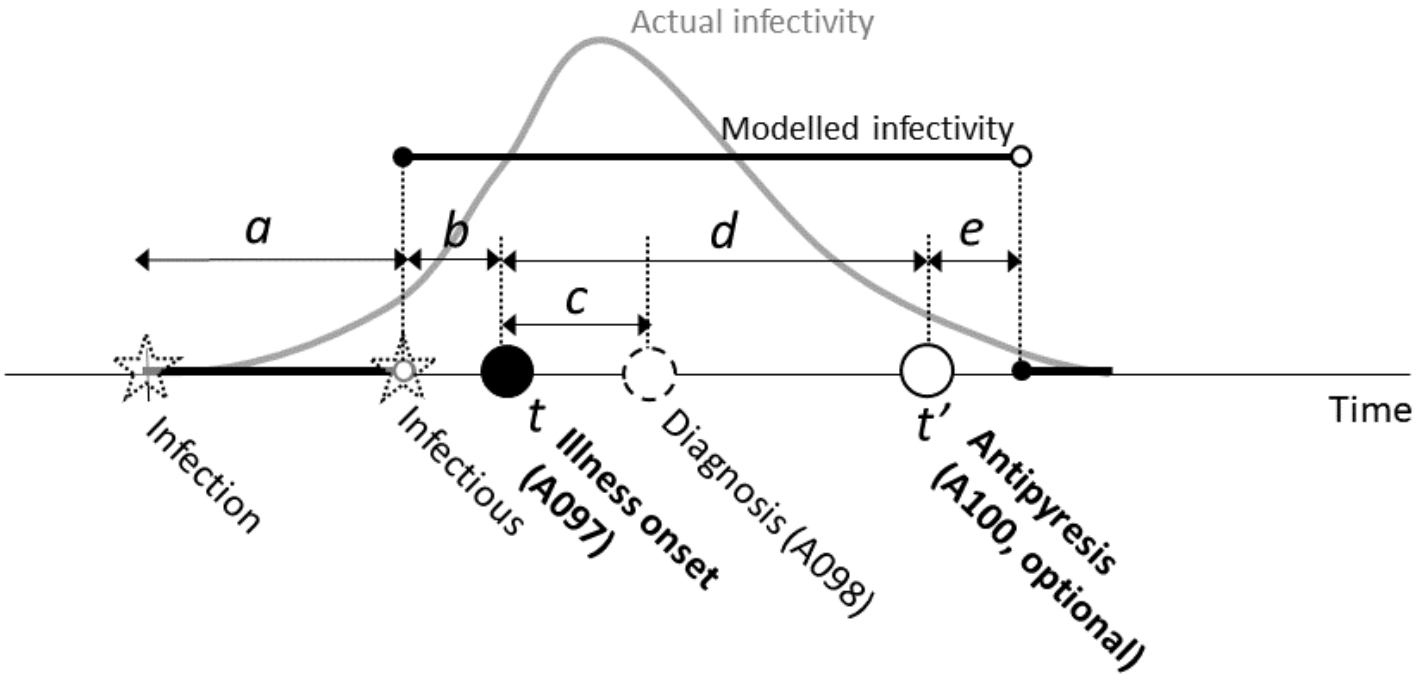

Figure 3

Natural history of infective people and the variation of infectivity. In an actual situation, a person may be infected at some unknown point in time (Infection) and the infectivity to other people gradually increases up to its maximum around the time when the illness is well developed and recognized (Illness onset). It then is drained as the process of recovery from infection proceeds, which may be clinically observed by antipyresis (Antipyresis), though weak infectivity may remain. For simplicity, such time variation of infectivity is modeled using a piecewise constant function that takes a non-zero constant value λ0 only from one point in time (labeled as Infectious) to another point near Antipyresis. The modeled infectivity function is temporally controlled by four period parameters: a (pre-symptomatic and non-infectious), b (pre-symptomatic and infectious), d (symptomatic and infectious) and e (extended infective after recovery).

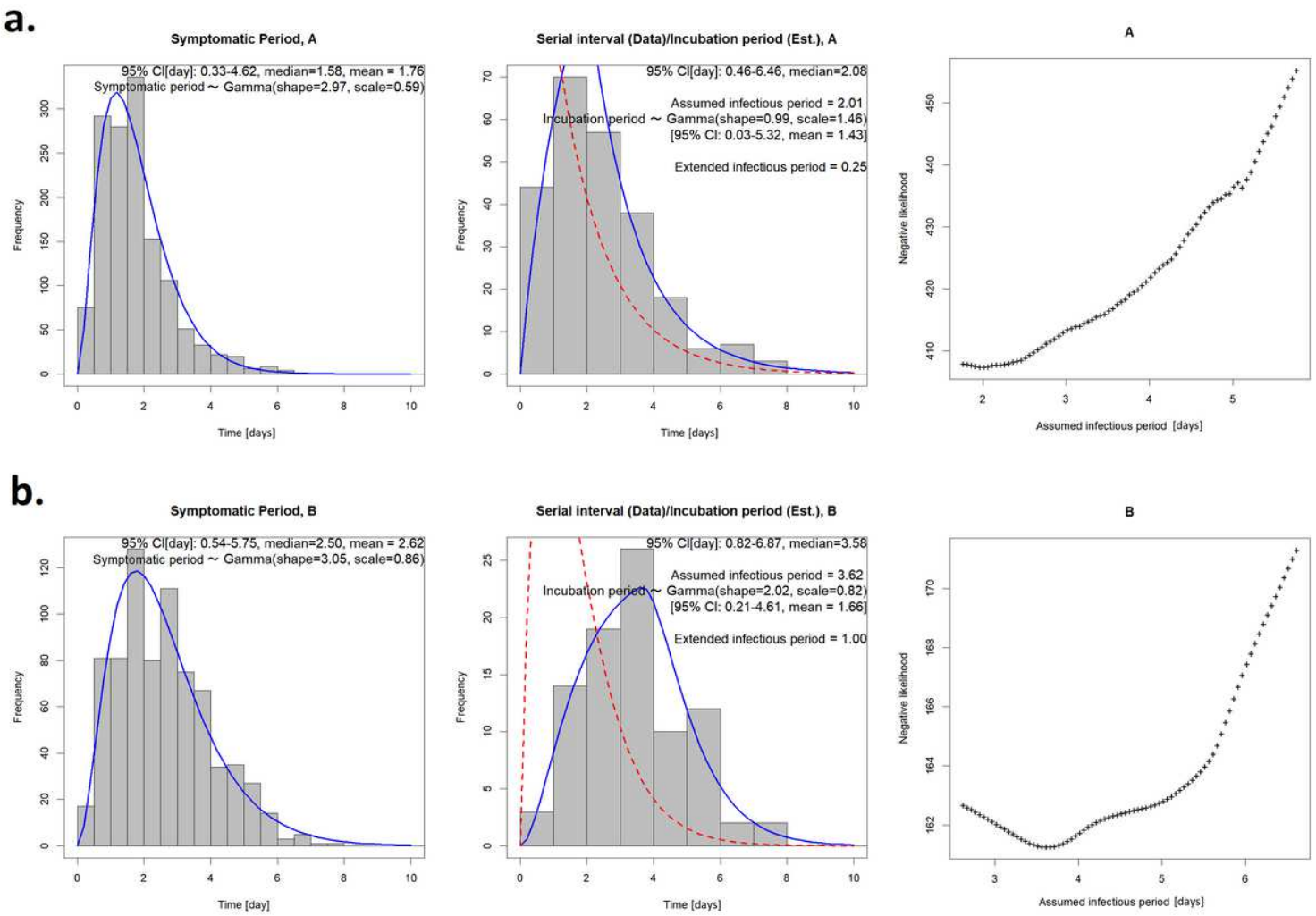

Figure 4

Summary of household infection data and estimates of the parameters for influenza types. a. Type A and b. type B (bottom row). The left column shows the histogram of the symptomatic period and its fitting to a gamma distribution. The middle column shows the histogram of the serial interval, its fitting to the distribution constructed as a convoluted gamma distribution (blue curve), and the extracted infectious period (annotation) and incubation period (red curve) via the fitting (see the Methods section for detail). The right column shows the negative likelihood against the assumed infective period.

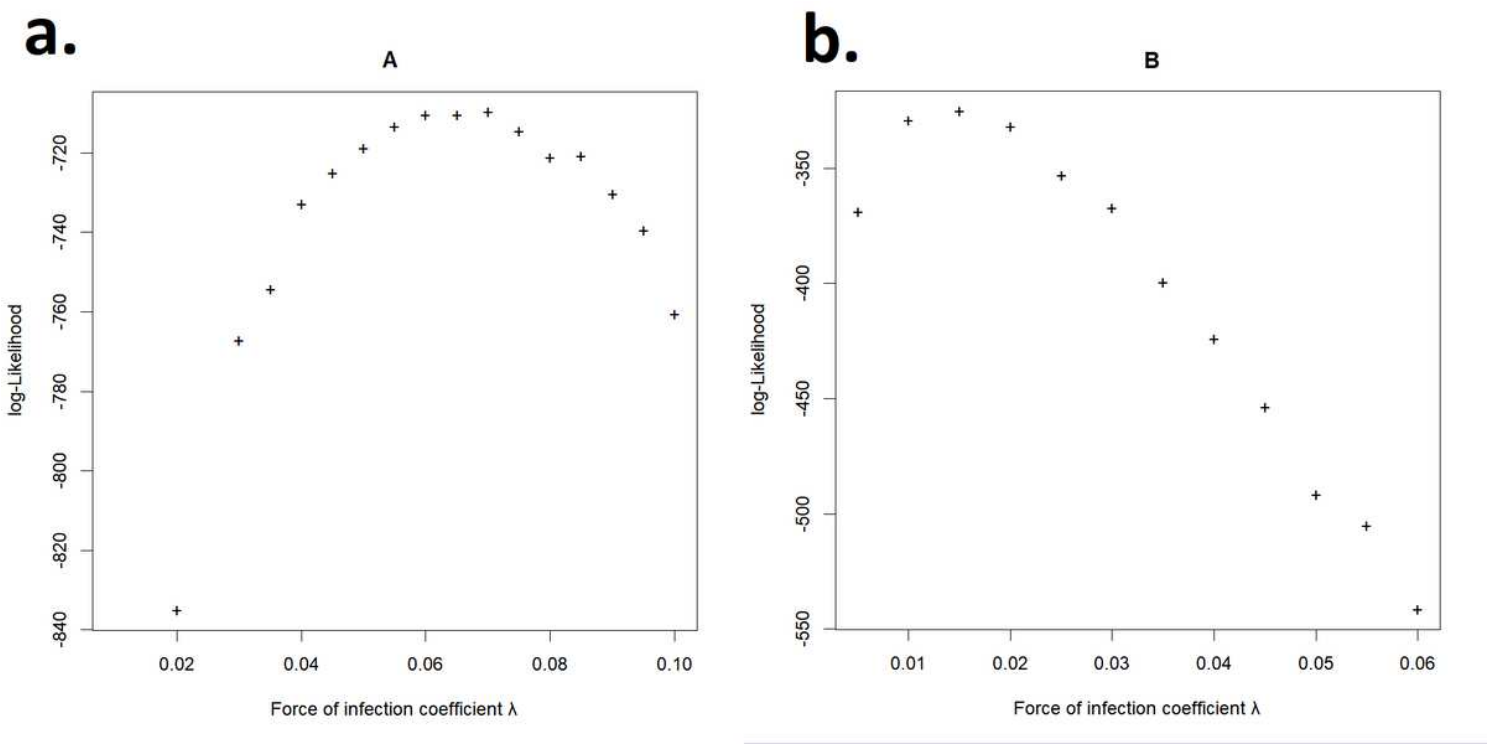

Figure 5

Likelihood curves for the two influenza types. a. Type A and b. type B are shown with respect to the number of secondary cases in a household as a function of the force of infection (FOI) coefficient λ0. The likelihood values are determined by simulation runs with different values of λ0 and Δλ0 = 0.005; optimal values of λ0 are 0.065 for A and 0.015 for B.

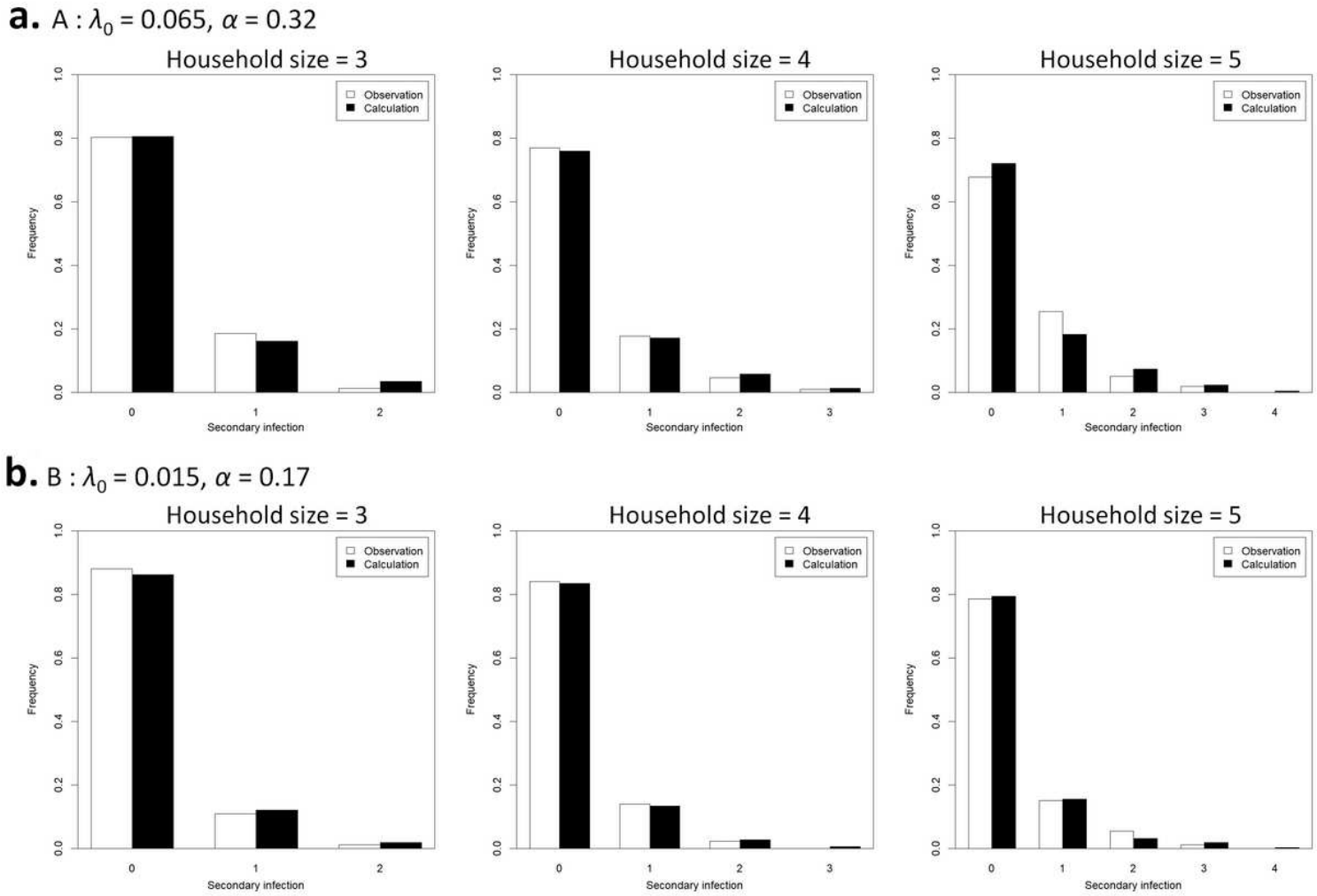

Figure 6

Comparison of the number of secondary cases in a household. Households have 3, 4, or 5 members in the actual data (white) and the simulation (black) for the two influenza types: a. type A and b. type B. The simulation run is mainly controlled by the force of infection (FOI) coefficient λ0 and the household size scaling power a; their ML estimates are shown.

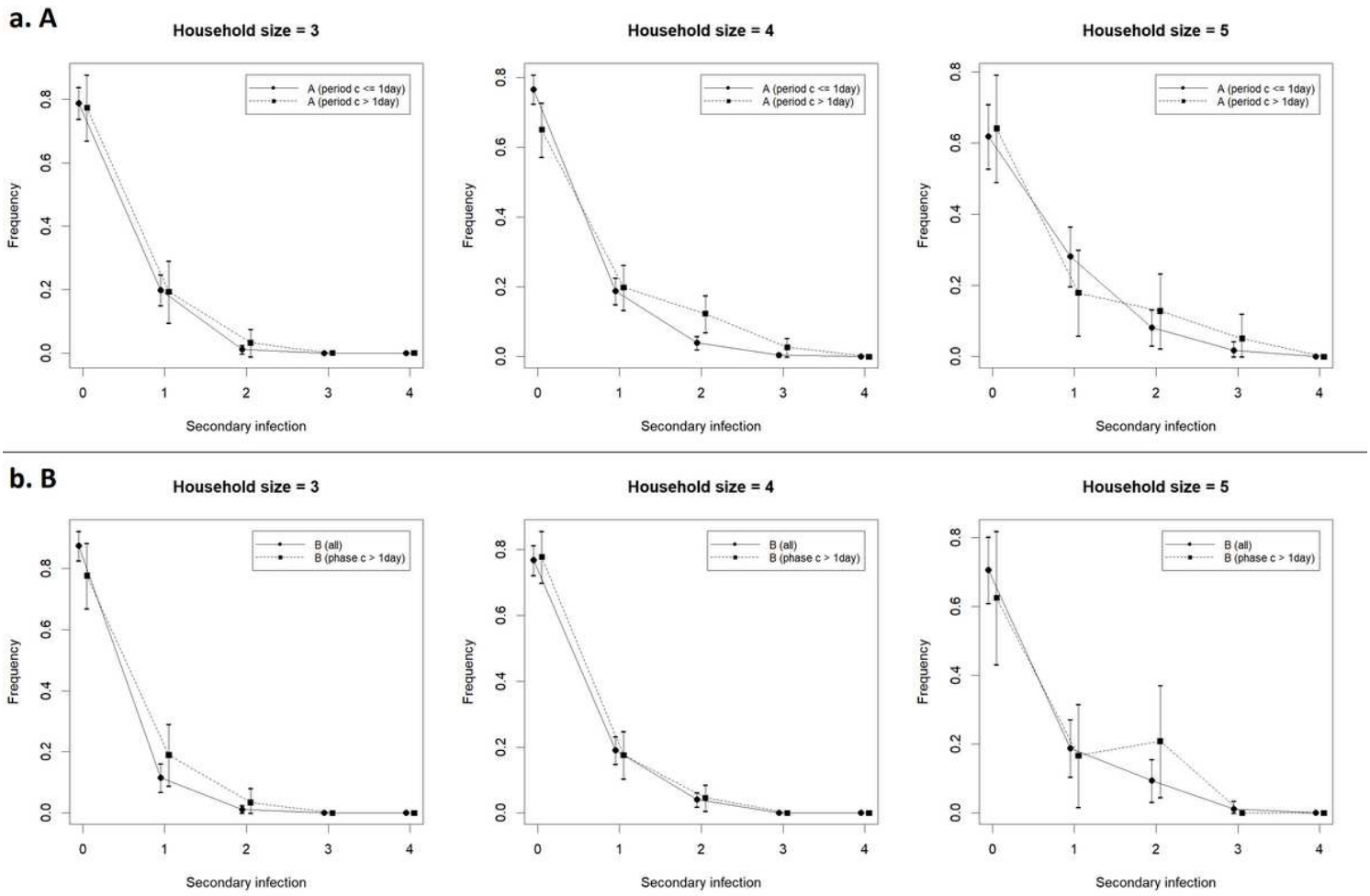

Figure 7

Comparison of the number of secondary cases in households with c_1≤ 1 (early-diagnosed primary) and those with c_1> 1 (late-diagnosed primary): a. type A and b. type B.