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 and assessment of detailed infectious profiles is important for that. Household transmission data have played a key role for estimating these profiles. From one clinic's influenza diagnostic and questionnaire-based data, we aimed to estimate the detailed infectious period (incubation, symptomatic and infective, and extended infective after recovery) and secondary attack ratio (SAR) for each household size of influenza by using modified Cauchemez-type model.

Results

The data source was derived from enrolled patients with confirmed influenza who were treated at the Hirotsu Clinic (Kawasaki, Japan) with a neuraminidase inhibitor (NAI) during the 6 Northern Hemisphere influenza seasons between 2010 and 2016. A total of 2,342 outpatients representing 1,807 households were included. For influenza type A, average incubation period and its 95% probability interval was 1.43 (0.03-5.32) days. Estimated average symptomatic and infective period was 1.76 (0.33-4.62) days and point estimated extended infective period after recovery was 0.25 days. Estimated SAR elevated from 20% to 32% as household size increases from 3 to 5. For influenza type B, average incubation period, average symptomatic and infective period, and extended infective period were estimated 1.66 (0.21-4.61) days, 2.62 (0.54-5.75) days and 1.00 days, respectively. SAR was increased 12% to 21% as household size increases from 3 to 5.

Conclusion

All estimated periods of influenza type B were longer than those of influenza type A. On the other hands, SAR of type B was less than that of type A. These results may reflect Japanese demographics and treatment policy. It is useful to understand the infectious profiles of influenza for examining 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 for 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 experience intense contact [4, 5]. Accordingly, household data have played a key role in solving estimation problems and have received particular attention in influenza infection. In many of these studies, a Reed & Frost-type model was 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. In 2004, Cauchemez et al. [4] applied a new methodology to estimate the force of infection and the distribution of the infectious period simultaneously using a Bayesian Markov chain Monte Carlo (MCMC) approach. In this research, the data from Carrat et al. [9], which included the start and end time of illness for 946 households, were used as characteristic data. As a result, the authors were able to provide detailed information on infection duration. It is this method that serves as the focus of the present paper.

Although the infective period is one of the most clinically important parameters to describe natural history, it alone is not enough 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**Error! Reference source not found.**] 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.

For a longitudinal study of household transmission, a single-center, prospective, observational study (UMIN‐CTR: UMIN000024650) at the Hirotsu Clinic in Kawasaki City, a major city in the greater Tokyo Area, enrolled 2,342 Hirotsu Clinic outpatients and 1,807 households with confirmed influenza during six influenza seasons (2010–2016). In addition to the diagnostic records of individual patients, the accumulated dataset includes the relationships between family members obtained from responses to a questionnaire administered to study participants. 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 propose a new household transmission model using a modified Cauchemez-type approach and 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 infectious duration are produced by combining the available records and the simulation model. The second step involves a simulation of inter-household transmission assuming a particular force of infection between households and calibrates 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. Then, in the "Discussion" section, we summarize the outcomes of our trials and discuss the study's strengths and limitations. Finally, we state conclusion remark and future direction in the "Conclusion" section.

Methods

Data source

The data source was derived from enrolled patients with confirmed influenza who were treated at the Hirotsu Clinic (Kawasaki, Japan) with a neuraminidase inhibitor (NAI) during the 6 Northern Hemisphere influenza seasons between 2010 and 2016. A total of 2,342 outpatients representing 1,807 households (Table 1) were included. Patients of any age who were diagnosed with influenza A or B using rapid influenza diagnostic tests (RIDTs) were eligible. 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 constitutional symptoms, specifically antipyresis < 37.5℃). The dosage and administration of the NAIs were as per the package insert for each product. Secondary infection patients were defined as household members who were diagnosed with the same influenza type as the index patient within 8 days after the onset of symptoms in the index patient. It is expected that secondary patients had simultaneous infection with the index patient and secondary infection from the index patient. Figure 1 shows that the number of secondary patients was monotonically decreased to 0.5 day in influenza type A and to 1.25 day in influenza type B. It may show that simultaneous infection with the index patient have occurred.

Transmission model in a household

Cauchemez et al. [4] conducted a longitudinal investigation of 334 households over a 15-day period and applied their novel household transmission model to estimate parameters describing the natural history of influenza infection. The approach we propose is similar to this in concept. In the model, the time variation of the infectivity attributed to an individual is modeled as a piecewise constant function (Figure 2) that includes four period parameters: asymptomatic and non-infective a, asymptomatic and infective b, symptomatic and infective d, and extended infective after recovery e. By definition, infectivity sustains for period $b + d + e$. Period d was observed from the data; period $b + e$, on the other hand, was estimated via maximum likelihood estimation. Because the individual values of b and e could not be broken out from the combined $b + e$ value, we set $b = 0$ only for conceptual completeness.

We employ a parametric model (which will be explained in the next subsection) to describe a and d as random variables and obtain their point estimates via maximum likelihood estimation. The subscripted version (e.g., a_i for a) denotes the values for an individual i. We assume that a given individual may be infected per unit time according to the probability of the sum of the infectivity 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 illness onset informed by the data) and loses it at $t'_i + e_i$ (optionally informed by 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: $\mathbb{1}\,(\text{True}) = 1$ and $\mathbb{1}\,(\text{False}) = 0$. We introduce the division factor for scaling in a similar manner to Ferguson's 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

158 well matches, for example, diseases where the infection is induced by polluted agents [20]. However,
159 as we will see, because the SAR is non-negligibly large in large families, setting $\alpha = 1$ appears to
160 over-reduce the FOI: an infective agent may have a conversation with two or more family members.
161 For this reason, we introduced an empirically determined power for scaling.
162

163 # Estimation of parameters

164 We employ a rather descriptive statistical approach to estimating a, d, and e. Setting $b = 0$, the
165 symptomatic period $d^{(h,i)}$ of the i-th infected member in household h is informed by data as $t'^{(i)} -$
166 $t^{(i)}$; summing such realizations over all households and members therein, we have the empirical
167 distribution (*i.e.*, histogram) of period d, along with its mean $\mathrm{E}[d]$. Similarly in principle, collecting
168 the serial interval instance $t_{\mathrm{int}}^{(h)} \coloneqq t^{(h,2)} - t^{(h,1)}$ over households h, we have the empirical
169 distribution for t_{int}. However, as is mentioned in Data source subsection, two infected family
170 members occasionally appear almost simultaneously, and these should be discarded from the
171 construction of the serial interval distribution. Figure 1 shows a part of histograms of the crude
172 distribution of t_{int} for influenza A and B, respectively. For influenza B, there is a downward trend
173 up to 1.25 days, followed by an upward trend as a part of bell-shaped distribution (see the entire
174 histogram is shown Figure 3b). We consider this downward trend is mainly attributed to
175 simultaneous infections, in particular t_{int} close to zero. Considering this trend is almost linear, we
176 assumed that the count between t_{int} and $t_{\mathrm{int}} + \Delta t$ is attributed to true household transmission in
177 probability $\propto t_{\mathrm{int}} \Delta t$ and discarded stochastically the data of $t_{\mathrm{int}} < 1.25$ days accordingly. The
178 same procedure is applied to influenza A with a different cut-off, $t_{\mathrm{int}} < 0.5$, though the downward
179 trend is not so clear as influenza B. After such adjustments, the "observed" serial interval t_{int} can
180 be modeled as the summation of the interval τ_{int} between the infection times of the primary and
181 secondary infections (i.e., the "intrinsic" serial interval) and the incubation period a of the secondary
182 subject. For ease of computation of the a distribution, we introduce two simplifications. First, the
183 interval of the two cases is assumed to follow a uniform distribution truncated at the mean infective

period τ_{ifv}: $p(\tau_{\text{int}}) = \mathbb{1}(0 \le \tau_{\text{int}} \le \tau_{\text{ifv}})/\tau_{\text{ifv}}$. Second, the incubation period follows a gamma distribution: $a \sim \text{Gam}(\text{shape} = k_a, \text{scale} = \theta_a)$. The 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. 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 3a. We identified 1,389 cases in which the patient exhibited symptoms and 290 transmissions from the primary to the secondary subject. The symptomatic period distribution is well approximated by a gamma distribution with shape = 2.97 and scale = 0.59 days; that is, period *d* is 1.76 days on average, with a 95% probability interval (hereafter 95% I) of 0.33-4.62. The point estimate of the infective period as the MLE is 2.01 (left-panel in Figure 3a). The difference between the infective and symptomatic periods is the extended infective period after recovery, *e*; here, $e = 2.01 - 1.76 = 0.25$ days. The incubation period is extracted as a gamma distribution with shape = 0.99 and scale = 1.46 days (mean: 1.43 days; 95% I: 0.03-5.32 days).

The results for influenza B are shown in Figure 3b. For the analysis here, we identified 760 cases where the patient exhibited symptoms and 135 transmissions from the primary to the secondary subject. A gamma distribution with shape = 3.05 and scale = 0.86 days (mean: 2.62 days; 95% I: 0.54-5.75 days) was fitted to the symptomatic period data; notably, the period here is longer than in the case of influenza A. The serial interval was also longer and cut off at approximately 6 days, while the infective period was estimated be 3.62 days, which yields the incubation period of 1.66 days (95% I: 0.21-4.61 days), as well as a relatively long extended infective period after recovery, $e = 3.62 - 2.62 = 1.00$ days.

Estimation of FOI

We then produced a point estimate of the FOI coefficient λ_0 after fixing scaling power α. In the case of influenza A, the SAR elevated from 20% to 32% as household size increases from 3 to 5, for which $\alpha = 0.32$ is optimal. This is a much smaller value than that used in a previous study, where Ferguson et al. [1], who carried out agent simulations and employed a Cauchemez-type model as an internal process, used $\alpha = 0.8$. With $\alpha = 0.32$ and an infective period of 3.73 days. We ran 1,024 simulations to obtain the ML estimate of λ_0. The likelihood as a function of λ_0 is shown in Figure 4a; the optimal value was found to be $\lambda_0 = 0.065$. The corresponding SAR distribution is shown in Figure 5a. As shown, the simulated distribution well reproduces the data.

The same procedure was applied to influenza B. Compared to type A, type B exhibited a smaller 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 4b. The SAR distribution in Figure 5b 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 day both for type A and 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 correspond to general medical intuition, the estimate for type A appears excessively short. According to Enforcement Regulations for School Health and Safety Act in Japan [21], patients with influenza virus infection are banned from attending school within 2 days (or 3 days for infant children) after the resolution of fever. These estimates of the mean of the extended infective period after the resolution of fever were lower than 2 or 3 days and corresponded to the description of this enforcement regulation in Japan. Therefore, our constructed models would make sense in this regard. But it should be noted that the choice of cut-off points which splits simultaneous infection from the true household transmission is in fact influential in the estimation and it is rather obscure in type A. The scaling power was estimated to be quite small ($\alpha = 0.2$ to 0.3). A straightforward explanation for this might be that the dataset covered households in which family members tended to spend most of their time with one another. However, two (or more) primary cases introduced almost simultaneously can elevate the apparent SAR in large families. Our analysis assumed that the first reported case infected the second, and so on. The fraction

of the ignored tail (> 8 days) should be available to model the probability that multiple members were infected simultaneously, referring to epidemic surveillance in Tokyo. It should also be noted that the SAR may be underestimated due to the basic strategy of early diagnosis followed by early treatment.

Conclusion

All estimated periods of influenza type B were longer than those of influenza type A. On the other hands, SAR of type B was less than that of type A. This study was conducted in Japan; therefore, these results may reflect Japanese demographics and treatment policy. Although it is important to assess these infectious profiles in each country, the number of studies that estimated infectious duration and FOI of influenza in Japan is very limited. It is useful to understand the infectious profiles of influenza for examining public health measures. More research may be performed to improve the accuracy of the results which can potentially be used for better public health.

Declarations

Ethics approval and consent to participate

The protocol was approved by the ethics committee of Shionogi & Co., Ltd. Informed consent was not required, because the anonymized data was shared with the Institute of Statistical Mathematics only as required by law, 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 analysis method and 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 provide it to NH.

Acknowledgements

Anonymization of the original dataset was carried out at 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].

Tables

Table 1

Summary of diagnosed influenza cases covered by our questionnaire investigation.

Year	Type	Cases	Households	Total Cases
1	A	243	189	
	B	104	86	347
2	A	263	204	
	B	222	177	485
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	

Figures

Figure 1. Histogram of the crude number of secondary cases for each serial interval (the part < 2 days). In both the types, there is a downward trend up to 0.5 days (type A) and 1.25 days (type B), which may be attributed to simultaneous infections outside the household.

Figure 2. 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 Infective) to another point near Antipyresis. The modeled infectivity function is temporally controlled by four period parameters: *a* (asymptomatic and non-infective), *b* (asymptomatic and infective), *d* (symptomatic and infective) and *e* (extended infective after recovery**).**

Figure 3. 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 infective period (annotation) and incubation period (red curve) via the fitting (see methods section for detail). The right column shows the negative likelihood against the assumed infective period.

Figure 4. 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

399 $\Delta\lambda_0 = 0.005$; optimal values of λ_0 are 0.065 for A and 0.015 for B.

400

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