Web Material

Risk of Transmission and Viral Shedding from the Time of Infection for Respiratory Syncytial Virus in Households

AUTHORS

Hirono Otomaru, Johanna Beulah T. Sornillo, Taro Kamigaki, Samantha Louise P. Bado, Michiko Okamoto, Mariko Saito-Obata, Marianette T. Inobaya, Edelwisa Segubre-Mercado, Portia P. Alday, Mayuko Saito, Veronica L. Tallo, Beatriz P. Quiambao, Hitoshi Oshitani, and Alex R Cook.

CONTENTS

Web Appendix 1

An example of the data collection at the household with a follow-up period is described in Web Figure 1. Once an RSV outbreak started, the enrolled households were visited by nurses every three to four days (Web Figure 1 (A) 'Regular household visit'). The purpose of these visits was to record if any symptoms of cough, coryza or difficulty of breathing were present, and if so, two nasal swabs were collected from the household members. One of the swabs was tested for RSV by a rapid test and immediately discarded, and the other swab was kept for PCR regardless of the result of the rapid test.

If the rapid test was positive for RSV, one nasal swab was collected from the rest of members of the household regardless of symptoms and this day became Day 0 of follow-up (Web Figure 1, 'Date of HH member Rapid test positive'). Following three visits during the follow-up period (Day 2, 5, and 8 in Web Figure 1), one nasal swab for laboratory confirmation was collected from the household members till Day 10. Thus, the specimen collection for PCR in the follow-up period was conducted on four occasions over the follow-up period to Day 10 (Web Figure 1, 'Visit during Follow-up period'). The schedule of visit was flexible in the case that the household members were not available for the sample collection. Specifically, the 1^{st} , 2^{nd}, 3^{rd}, and 4^{th} visits in the follow-up period were conducted during days 0–1, 2–4, 5–7, and 8–10, respectively.

If multiple household members presenting symptoms were found, the younger children were tested by a rapid test first followed by other household members. For instance, if both a child and her mother show symptoms, we tested the child first, and if the result was negative, the mother was tested next. If the result of the child was positive, the mother would not be tested by the rapid test. Regardless of the results of the rapid tests, another nasal swab was collected

for PCR from both of them. If the both rapid tests were negative, the follow-up period did not begin, however, the collected nasal swabs were kept for PCR.

To increase data of the follow-up data collection, we invited children and household members to participate in follow-up data collection if they were found to be RSV positive by rapid test screening at the primary health care facility, even if they were not part of the current enhanced cohort. Once informed consent was obtained, the follow-up data collection launched (Web Figure 1 (B)). After the follow-up period, the regular visits were also performed for these households.

Web Appendix 2

The collected specimens were stored at 4°C in a refrigerator and transported with ice packs to the Research Institute for Tropical Medicine in Manila twice a week. Viral RNA was extracted from the specimens using Viral QIAamp MinElute Virus Spin kit (QIAGEN, Hilden, Germany), and extracted RNA with Random Primer (Invitrogen Carlsbad, CA, USA) was reverse transcribed to complementary DNA using Moloney Murine Leukemia Virus Reverse Transcriptase (Takara, Shiga, Japan). Quantitative reverse transcription Polymerase Chain Reaction (PCR) was performed to obtain the CT values by using TaqMan Fast Advanced Master Mix together with optimized standards (Applied Biosystems, Foster City, USA) by using RSV specific primers and probe listed in Web Table 1. Based on the obtained CT values of standards, estimated viral loads of specimens were calculated. The assay was conducted in duplicate in single run to get a mean estimated viral load. Nucleotide sequencing targeting the G gene was conducted for specimens tested RSV positive by Big Dye Terminator version 1.1 (Applied Biosystems) to determine RSV subgroup. The detailed procedure is described elsewhere (1).

Web Appendix 3

Informative prior of RSV incubation period

The difference between the date of onset based on symptom records and the estimated time of infection was defined as the incubation period. An incubation period τ_i was considered to follow lognormal distribution: $\tau_i \sim \log N\,(\alpha, \beta)$ and an informative prior was applied to τ_i. To obtain the informative prior of α, β, we used a previous report about the incubation period of RSV (2), which reported that the estimated incubation period with 95% CI of RSV were 3.1 (2.5–3.8), 3.8 (3.3–4.4), 4.4 (3.9–4.9), 5.1 (4.5–5.7), 6.3 (5.2–7.3) days for 5th, 25th, 50th, 75th, and 95th percentile with dispersion 1.24. We estimated a mean and standard deviation of α and β through the simulation and fitting and obtained $\alpha \sim N(1.48, 0.0042^2)$ and $\beta \sim N(0.22, 0.056^2)$.

Log-Likelihood function

A model-based viral load was determined by following equation.

$$\mu(t) = cb^a t^{a-1} \exp(-bt)/\Gamma(a)$$

Parameters $a,\ b,\ c$ were further re-parameterized as θ_1, θ_2 and θ_3 to aid convergence. Specifically, we defined θ_1, θ_2 and θ_3 to be

$$\theta_1 = (a-1)/b$$

$$\theta_2 = cb(a-1)^{a-1} \exp(1-a)/\Gamma(a)$$

$$\theta_3 = c.$$

In which case, $c = \theta_3$, $b = (a-1)/\theta_1$ and a is the solution to

$$\theta_1\theta_2/\theta_3 = (a-1)^a \exp(1-a)\,/\Gamma(a).$$

We defined a likelihood function as $f(\mu(t); a, b, c)$ and $\mu_{(t)}$ was assumed to follow a one-sided truncated normal distribution. Censoring was assumed to occur at $x_{in} < 1$. This

censoring was observed in the analysis of the estimation of viral loads. The maximum of cycle value in our runs of the laboratory confirmation corresponded to $0 < x_{in} < 1$.

Next, the incubation period was defined as the following.

$$\tau_i = T_i^{inf} - Date\ of\ onset$$

For asymptomatic cases, the period from T_i^{inf} to the date of collection of the first positive sample was used as the initial value, but that was not included in the likelihood calculation. The incubation period τ_i was assumed to follow a log-normal distribution and $g\ (\tau_i; \alpha, \beta)$ was defined as a likelihood function.

The log likelihood was obtained by summing the log density for the viral load for each uncensored observation, the log cumulative density function evaluated at the assumed detection limit for each censored observation, and the log density of the incubation periods:

$$\log \mathrm{L}\ (x_t|\Theta)\ = \sum_{t=1}^{Tmax} \ln f(\mu(t); a, b, c) \sum_{i}^{N} \ln g(\tau_i; \alpha, \beta)$$

where N is the number of participants and $Tmax$ was the maximum of n_i. The description of parameters was summarized in Web Table 2. The codes used can be found at: https://github.com/hiro-oto/RSV_viralload.

MCMC run

The parameter vector was thus $\Theta = \left(\sigma, \theta_1, \theta_2, \theta_3, T_i^{inf}, \tau_i, \alpha, \beta\right)$, which was estimated with Markov Chain Monte Carlo (MCMC). We employed the Metropolis-Hastings algorithm and using thinning of every 100 iterations from 1 000 000 MCMC steps for each parameter, obtained 10 000 simulated posterior observations. The simulation was preceded by 3 000 burn in iterations, which were discarded from the analysis. Proposals were tuned on pilot runs.

The posterior estimates of the parameter values

We obtained μ as the modelled viral loads by the posterior of a, b and c. The mean of those were estimated to be a = 8.8, $b = 1.3$, and $c = 24.7$.

The covariance matrix M was described as follows:

$$\mathrm{M} = \begin{pmatrix} 2.3 & 0.3 & -1.7 \\ 0.3 & 0 & -0.3 \\ -1.7 & -0.3 & 2.7 \end{pmatrix}$$

Geweke's convergence diagnostic

MCMC convergence was tested by a Geweke's convergence diagnostic. The absolute value of the Z-scores were confirmed to be less than 1.96 for $\theta_1, \theta_2, \theta_3, \alpha$, and β. Only three of T_i^{inf} had an absolute value of the Z-score of 1.96 or higher.

Web Appendix 4

To identify associations between model-based viral loads and potential risk factors, we tested their correlation by using Generalized Estimating Equations (GEE) (3). We calculated residuals as the difference between the model-based mean viral load to the estimated viral load at each time point of specimen collection and used them as a dependent variable. An exchangeable correlation structure was assumed.

In the GEE, our input data is the following.

$$\mathbb{E}(y_i) = \varepsilon_i$$

$$g(\varepsilon_i) = X_i^T \beta$$

where $i = 1 \ldots N$, X is the $n_i \times 3$ design matrix, and $T = t_{in}$. The covariance vector (the presence of symptom, age and if the index case or not) was denoted by $\beta = (\beta_1, \beta_2, \beta_3)$. The dependent variable was y_j, the difference of model-based viral loads on T_{in} and viral loads estimated through PCR.

The GEE model was defined as follows.

$$U(\beta) = \sum_{i=1}^{N} D_i^T V_i^{-1} S_i = 0$$

where

$$D_i^T = \frac{\partial \varepsilon_i}{\partial \beta}$$

$$V_i = A^{1/2} R_i A^{1/2} \Theta$$

$$A_j = diag\{v(\varepsilon_i)\}$$

$$S_i = y_i - \varepsilon_i$$

and R_i is a working correlation matrix and Θ is a scale parameter.

An extended Cox proportional hazard model was used to evaluate risks of the household transmission as a function of viral load of the index case.

The Cox model is expressed by the hazard function denoted by $h(\mathbb{t})$.

We defined that h_0 is a baseline hazard and $\mathbb{t}$ denotes a time elapsed from the household exposure. $h(\mathbb{t})$ is the hazard function determined through explanatory variables. We used the presence of cough in the index case and the age of the at-risk household members as the covariates. The age group of the at-risk household members was x_1. The presence of cough in the index case varies with time and was represented as $x_2(\mathbb{t}) = (x_2(1) \ldots x_2(T_i))$, where T_i denotes the time when a secondary case was infected or when 21 days elapsed after the household exposure.

Thus, the hazard function is:

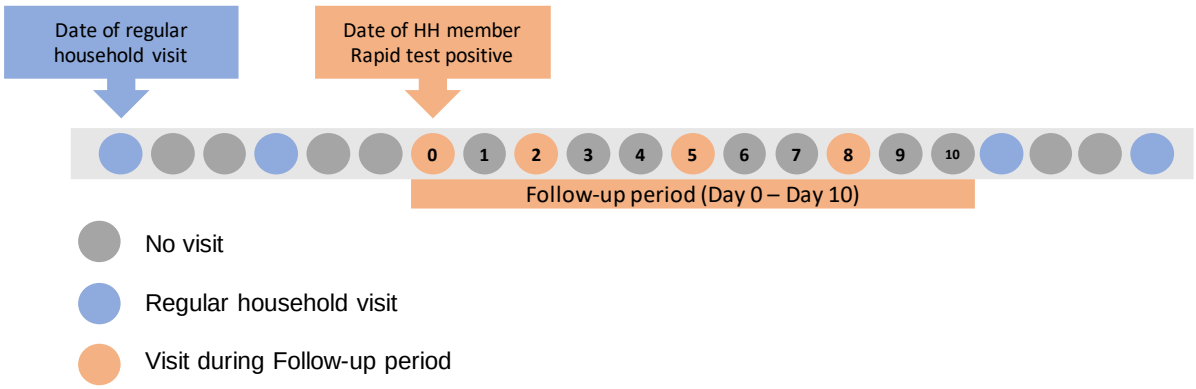
$$h(\mathbb{t}, x(t), \beta) = h_0(\mathbb{t}) \cdot \exp(\beta_1 x_1 + \beta_2 x_2(\mathbb{t}))$$

The hazard ratio for age group is:

$$HR\big(\mathbb{t}, x_1 = 0, x_1 = 1; x_2(\mathbb{t})\big) = \exp(\beta_1)$$

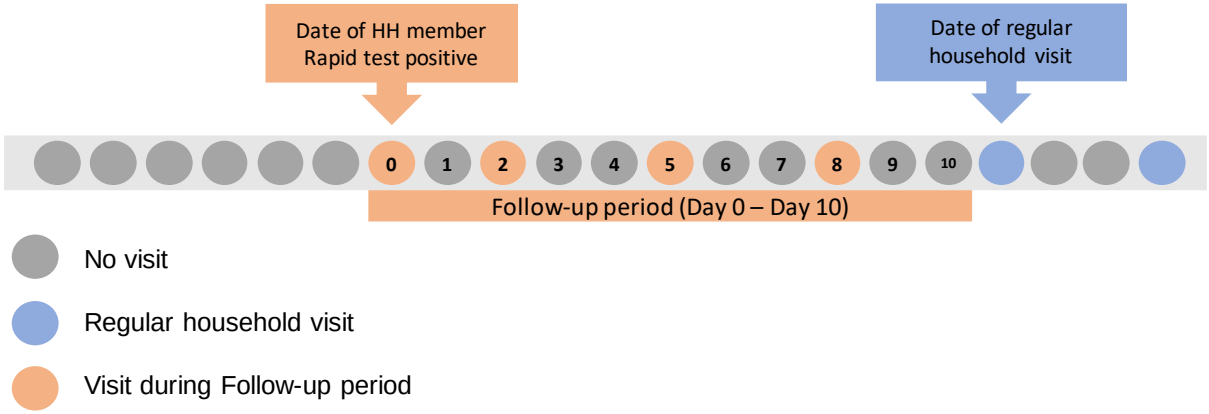
Web Appendix 5

Web Figure 2 shows when specimen collections have been conducted and its PCR results including before and after the follow-up period for members of households analysed in the viral load model. Because of the difference in sensitivity between rapid tests and PCR or other technical reasons, there were three cases where the PCR results of specimens were positive even though the rapid test conducted at the same time results were negative. Therefore, specimens collected up to 7 days prior to the start of the follow-up period that were PCR positive were used in the analysis of the viral load model.



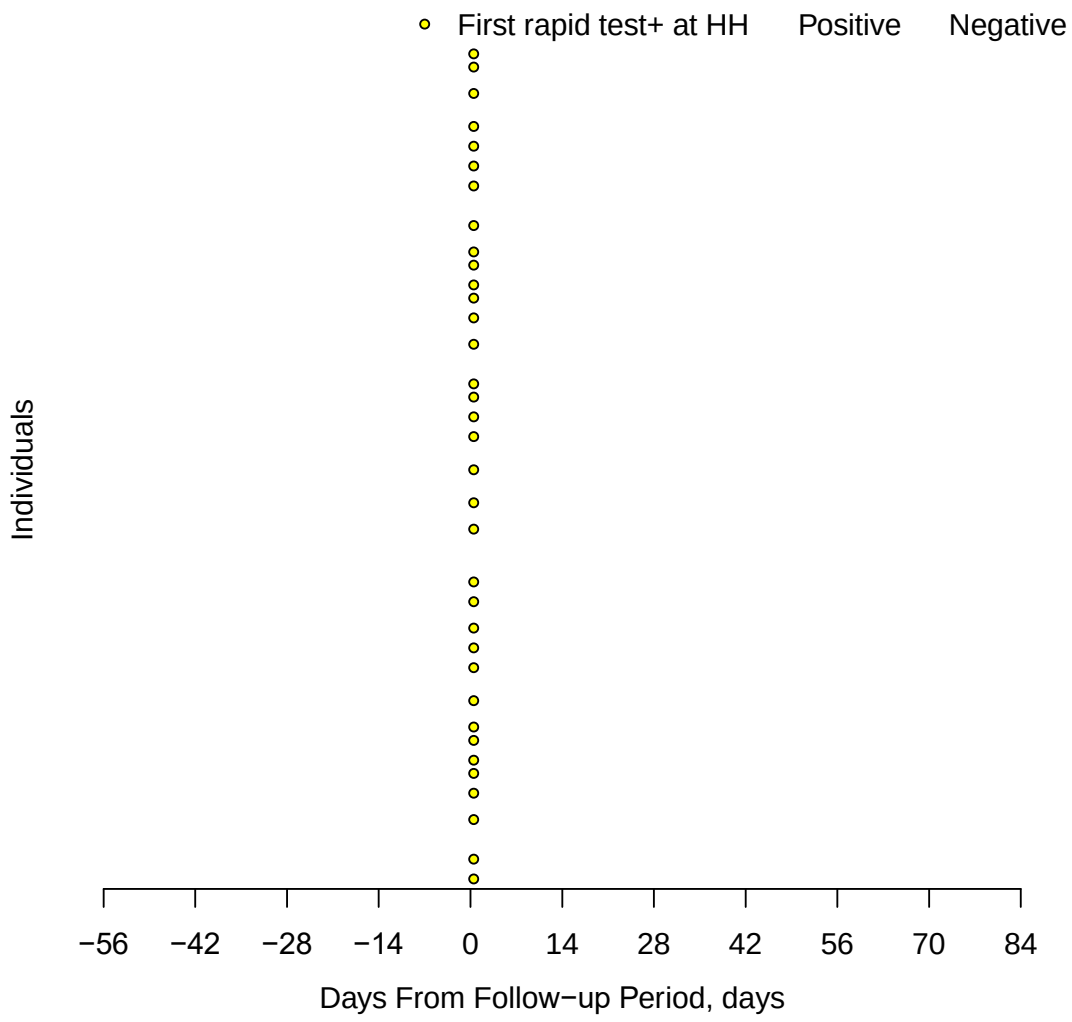
Web Figure 1 (A). An example of the data collection.

The color of the circle indicates the presence and type of household visit. Gray indicates no visit, blue indicates regular visit (not a follow-up period visit), and orange indicates follow-up period visit.



Web Figure 1 (B). An example of the data collection for those who joined in the study through the rapid kit screening at the primary health care facility.

The color of the circle indicates the presence and type of household visit. Gray indicates no visit, blue indicates regular visit (not a follow-up period visit), and orange indicates follow-up period visit.



Web Figure 2. List of sample collection dates and their test results

One row represents an individual and one column represents a day. Blue squares indicate PCR-negative specimens, and red squares indicate PCR-positive specimens. Yellow circles indicate the rapid test positive that triggered the follow-up period. The blue line separates the households. HH: household.

Web Figure 3. The fitting of the measured viral load to the estimate from the inferred time of infection for each case.

Solid lines represent the mean value of the model-based viral load, and blue shading represents the 95 percent credible interval for the mean. The black circles describe the estimated viral load for the positive, and the white circles the negative samples at the assumed limit of detection.

Web Table 1. Sequence information of primers and probe used to detect RSV

Primer/Probe	Sequence (5′ – 3′)
Forward primer	GCTCTTAGCAAAGTCAAGTTRAATGATACA
Reverse primer	GTTTYTGCACATCATAATTRGGAGT
Probe	(5′-VIC)-CTRTCATCCAGCAAATA-(MGB-3′)

Web Table 2. Summary of parameters and data used in the model

Notation	Description	Source
t_{in}	Difference between the time point T_i^{inf} and T_{in}^S	Data
T_{in}^S	The nth sample for individual i, and in total S specimens were collected from i	Data
T_i^{inf}	Unknown time of infection	Estimated by model
x_{in}	The viral load estimated through PCR	Data
$\mu(t_{in})$	Mean of x_{in} in normal distribution	Estimated by model
σ	Standard deviation of x_{in}	Estimated by model
a, b, c	Parameters to define $\mu(t)$	Estimated by model
$\theta_1, \theta_2, \theta_3$	Parameters to define a, b, c	Estimated by model
τ_i	Log-scale incubation period of individual i. Incubation period is a time between the time of infection (T_i^{inf}) and the date of onset.	Estimated by model
α	Log-scale mean of log-normal distribution	Estimated by model
β	Log-scale variance of log-normal distribution	Estimated by model

1. References

1. Malasao R, Okamoto M, Chaimongkol N, Imamura T, Tohma K, Dapat I, et al. Molecular characterization of human respiratory syncytial virus in the Philippines, 2012-2013. PLoS ONE. 2015;10(11):2012–3.
2. Capelozza ALA, Dezotti MSG, Alvares LC, Fleury RN, Ana ES. Incubation periods of acute respiratory viral infections: a: Systematic review : 2005;34(1):32–5.
3. Zeger SL, Liang K-Y, Albert PS. Models for Longitudinal Data: A Generalized Estimating Equation Approach. Biometrics. 1988;44(4):1049–60.