eAppendix to Transmissibility of seasonal and pandemic influenza in a cohort of households in Hong Kong in 2009

1 Statistical Model

We extend the model first proposed in a general sense by Ludwig[1] and as a statistical model by Longini and Koopman[2] and later discussed in a Bayesian context by O'Neill et al.[3] The Longini-Koopman model assumes a Reed-Frost type chain-binomial infection process within households but modifies the basic model to allow for the possibility of infection from the outside community. Consider that the probability an individual escapes infection from the community is q_c and the probability an individual escapes infection from the household given that another household member is infected is q_h. Then the probability an individual escapes infection is q_c given no other family members is infected, $q_c q_h$ given one family member has been infected and $q_c q_h^2$ given two family members have been infected.

Now consider a family of s susceptible individuals of which j are infected. Denote w_{js} as a family where j out of s susceptible individuals were infected. Define $\alpha_{js} = Pr(W_{js} = w_{js}|q_c, q_h)$ and $\alpha_{00} = 1$. Then for all $j \neq s$

$$\alpha_{js} = \binom{s}{j} \alpha_{jj} (q_c q_h^j)^{s-j}$$

and consequently by the law of total probability

$$\alpha_{ss} = 1 - \sum_{j \neq s} a_{js}.$$

We extend the basic combinatorics argument to include four parameters. Let C_c and C_h be the community and home escape probabilities respectively for children and A_c and A_h be the community and home escape probabilities respectively for adults. Let w_{ijst} be a family where i out of s susceptible children and j out of t susceptible adults are infected. Let $\alpha_{0000} = 1$ then for $j \neq t$

$$\alpha_{ijst} = \alpha_{ijij} \binom{s}{i} (C_c C_h^{i+j})^{s-i} \binom{t}{j} (A_c A_h^{i+j})^{t-j}$$

and again by the law of total probability

$$\alpha_{stst} = 1 - \sum_{i,j \neq t} \alpha_{ijst}.$$

Suppose that in the mth family, $m \in \{1, \ldots, 117\}$, we observe $w_m = w_{ijst}$ specifically that i out of s susceptible children are infected and j out of t susceptible adults are infected. Let α_m be the probability be defined as $Pr(W_{ijst} = w_{ijst} | C_c, A_c, C_h, A_h)$. The likelihood is then

$$L(C_c, A_c, C_h, A_h | W) = \prod_m \alpha_m$$

and the log likelihood is

$$l(C_c, A_c, C_h, A_h | W) = \sum_m \log \alpha_m.$$

2 Community Probability of Infection and Secondary Attack Proportion

Following the convention of Longini et al[4] we define the community probability of infection for children and adults respectively as $1 - C_c$ and $1 - A_c$ and the secondary attack proportion respectively as $1 - C_h$ and $1 - A_h$. The secondary attack proportion can be interpreted as the probability that a susceptible individual will be infected by another individual in the same household who has already been infected. The community probability of infection can be interpreted as the probability of acquiring infection from the community during the period of study.

3 Priors

Bases on our review of the literature we used semi-informative priors beta[1.5, 6] for the secondary attack proportion and beta[1.2, 6] for the community probability of infection in pH1N1 and sH3N2 and beta[1.5, 30] for the community probability of infection in sH1N1. We expect that the secondary attack proportions for pH1N1 and sH3N2 would fall in the range 5%-50% given our previous studies in Hong Kong[5],[6] and studies elsewhere.[7] Therefore we used a wide beta[1.5, 6] with mean 20% prior for the secondary attack proportion.

Pandemic A(H1N1) and seasonal A(H3N2) were most prevalent during the study period and it is believed that in a typical season around 16% of the population are infected, therefore we used a wide beta[1.2, 6] with mean 16% prior for the community probability of infection with pH1N1 and sH3N2.[4] However seasonal A(H1N1) did not circulate as widely as pH1N1 and sH3N2 during our study period, and we chose a wide beta[1.5, 30] prior with a lower mean of 5% for that parameter.[8] We did not use different priors for children and adults due to the absence of explicit information available in the literature of the relative differences in risk of infection from the community or in households by age.

The plots of priors versus posteriors in eFigures 1-3 show that the data have a strong influence on most posterior estimates. The data did not contain much information on the secondary attack proportion of sH1N1 among children with antibody titers $\leq$ 1:40, and the posterior distribution appeared to be similar to the prior (eFigure 3). Because the sample size of children with sH1N1 was relatively small, further studies are needed to clarify the secondary attack proportion associated with this virus in our setting.

4 MCMC

We employed a random walk Metropolis-Hastings algorithm to generate posterior distributions for the parameters.[9] The proposal density for all parameters was Uniform[0, .10] centered around the current parameter value. A separate accept-reject "Metropolis within Gibbs" step was performed for each parameter conditional on the remaining three parameters.

The MCMC was run for 15000 iterations. The first 5000 iterations were considered burn-in and not used in calculation of the posterior distributions. The last 10000

iterations were summarized as draws from the posterior distribution. MCMC chains were plotted and checked to ensure that they were mixing properly. Raftery-Lewis diagnostics were calculated to ensure the number of iterations for burn-in were sufficient.[10]

5 Data Imputation

Individuals with missing paired serology were considered to have missing infection status. Data imputation was used in the MCMC algorithm to account for their missing infection status.[11] Consider a child with 2 other infected family members. Then that child's escape probability is $C_c C_h^2$ and his probability of infection is $1 - C_c C_h^2$. For each adult and child with missing infection status at each iteration i of the MCMC algorithm we imputed their infection status by drawing from a random bernouli distribution with probability $1 - A_c^{(i)}(A_h^{(i)})^z$ and $1 - C_c^{(i)}(C_h^{(i)})^z$ respectively where z is the number of other infected household members. The imputed infection status of individuals with missing data was then used to calculate the log likelihood at iteration $i + 1$.

6 Chi-square goodness-of-fit tests

Chi-square goodness-of-fit tests were performed for all models to ensure no model exhibited inadequate fit.[2,12]

References

[1] Ludwig D. Final size distribution for epidemics. *Math Biosci* 1975;23(1-2):33-46.

[2] Longini IM, Jr., Koopman JS. Household and Community Transmission Parameters from Final Distributions of Infections in Households. *Biometrics* 1982;38(1):115-26.

[3] O'Neill PD, Balding DJ, N.G. B, et al. Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *J Royal Stat Soc Ser C* 2000;49(4):517-42.

[4] Longini IM, Jr., Koopman JS, Monto AS, et al. Estimating household and community transmission parameters for influenza. *Am J Epidemiol* 1982;115(5):736-51.

[5] Cowling BJ, Chan KH, Fang VJ, et al. Facemasks and hand hygiene to prevent influenza transmission in households: a cluster randomized trial. *Ann Intern Med* 2009;151(7):437-46.

[6] Cowling BJ, Chan KH, Fang VJ, et al. Comparative epidemiology of pandemic and seasonal influenza A in households. *N Engl J Med* 2010;362(23):2175-84.

[7] Yang Y, Sugimoto JD, Halloran ME, et al. The transmissibility and control of pandemic influenza A (H1N1) virus. *Science* 2009;326(5953):729-33.

[8] Cowling BJ, Ng S, Ma ES, et al. Protective efficacy of seasonal influenza vaccination against seasonal and pandemic influenza virus infection during 2009 in Hong Kong. *Clin Infect Dis* 2010;51(12):1370-9.

[9] Robert CP, Casella G. *Monte Carlo statistical methods. 2nd ed.* New York: Springer; 2004.

[10] Raftery AE, Lewis SM. The number of iterations, convergence diagnostics and generic Metropolis algorithms. In: Gilks C, Spiegelhalter DJ, Richardson S, eds. *Practical Markov Chain Monte Carlo.* London: Chapman and Hall, 1995;115-130.

[11] O'Neill PD. A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. *Math Biosci* 2002;180:103-14.

[12] Haber M, Longini IM, Jr., Cotsonis GA. Models for the Statistical Analysis of Infectious Disease Data. *Biometrics* 1988;44(1):163-73.

eTable 1: Twelve study subjects with a fourfold or greater rise to more than one influenza subtype during the follow-up period and justification for final subtype classification.

ID	sH1N1 baseline/ convalescent titer (ratio)	sH3N2 baseline/ convalescent titer (ratio)	pH1N1 baseline/ convalescent titer (ratio)	Confirmation by RT-PCR	Other infected family members	ARI date	ILI date	Final classification and justification for decision
9101/0	160/640 (4)	160/320 (2)	5/80 (16)	n/a	1 had sH3N2	n/a	n/a	pH1N1 (titers)
9110/2	5/40 (8)	5/40 (8)	(n/a)/5	n/a	3 had sH3N2	n/a	n/a	sH3N2 (based on family members)
9122/4	80/160 (2)	320/2560 (8)	20/160 (8)	pH1N1 6/7/2010	2 had pH1N1, 1 had sH3N2	n/a	n/a	pH1N1 (RT-PCR)
9128/3	5/40 (8)	5/80 (16)	(n/a)/20	n/a	None	13/9/09	n/a	sH3N2 (titers)
9161/1	5/80 (16)	10/2560(256)	(n/a)/5	n/a	None	n/a	n/a	sH3N2 (titers)
9167/3	10/160 (16)	5/40 (8)	(n/a)/40	n/a	2 had pH1N1	1/9/09	n/a	pH1N1 (other family members and ARI date)
9195/3	20/80 (4)	5/5(1)	5/160 (32)	n/a	1 had ILI 9/910	13/5/09	13/5/09	pH1N1 (titers and family member)
9203/3	5/80 (4)	640/640 (1)	5/320 (64)	n/a	2 had sH3N2	11/5/09	11/5/09	pH1N1 (titers)
9206/3	5/5 (1)	5/640 (128)	5/160 (32)	pH1N1 15/9/2010	1 had pH1N1	13/9/09	13/9/09	pH1N1 (RT-PCR)
9208/1	5/5 (1)	5/40 (8)	10/40 (4)	n/a	None	n/a	n/a	sH3N2 (titers)

9211/0	5/80 (16)	5/5 (1)	10/40 (4)	n/a	1 had pH1N1	18/9/09	18/9/09	pH1N1 (due to family members and ILI date)
9214/3	640/640 (1)	5/160 (32)	10/40 (4)	n/a	2 had pH1N1	25/9/09	25/9/09	pH1N1 (due to family members and ILI date)

* pH1N1 = 2009 pandemic A(H1N1); sH1N1 = seasonal A(H1N1); sH3N2 = seasonal A(H3N2). n/a = not measured. The initial dilution was 1:10, responses undetectable at this level (i.e. <1:10) were coded as 5 for each strain.

eTable 2: Sample sizes and proportions infected in subgroup analyses excluding individuals with higher baseline seasonal A antibody titers.

	N (% of total)	pH1N1 N (%)	sH1N1 N (%)	sH3N2 N (%)
All individuals				
Children	161	35 (22%)	4 (2%)	13 (8%)
Adults	222	19 (9%)	4 (2%)	14 (6%)
sH1N1 ≤1:160				
Children	99 (61%)	17 (17%)	4 (4%)	12 (12%)
Adults	221 (100%)	19 (9%)	4 (2%)	13 (6%)
sH3N2 ≤1:160				
Children	98 (61%)	23 (23%)	3 (3%)	12 (12%)
Adults	216 (97%)	19 (9%)	4 (2%)	13 (6%)
sH1N1 ≤1:40				
Children	58 (36%)	11 (19%)	4 (7%)	8 (24%)
Adults	204 (92%)	16 (8%)	4 (2%)	11 (5%)
sH3N2 ≤1:40				
Children	60 (37%)	13 (22%)	2 (3%)	11 (18%)
Adults	198 (89%)	17 9%)	3 (2%)	13 (7%)

* pH1N1 = 2009 pandemic A(H1N1); sH1N1 = seasonal A(H1N1); sH3N2 = seasonal A(H3N2). Proportions calculated based on all individuals with paired antibody titers available.

eTable 3: Estimates of the Community Probability of Infection and the Secondary Attack Proportion (SAP) for children and adults who did not receive seasonal vaccination.

Individuals assumed to be susceptible	Influenza A subtype*	Cumulative community probability of infection (per season)				Secondary attack proportion			
		Children	(95% CI)	Adults	(95% CI)	Children	(95% CI)	Adults	(95% CI)
Individuals with antibody ≤1:160	pH1N1	0.19	(0.12- 0.27)	0.07	(0.04- 0.12)	0.12	(0.01- 0.28)	0.10	(0.02- 0.21)
	sH1N1	0.05	(0.02- 0.11)	0.02	(0.01- 0.04)	0.10	(0.00- 0.29)	0.08	(0.00- 0.24)
	sH3N2	0.15	(0.08- 0.25)	0.05	(0.03- 0.09)	0.24	(0.04- 0.54)	0.11	(0.03- 0.23)
Individuals with antibody ≤1:40	pH1N1	0.19	(0.12- 0.27)	0.07	(0.04- 0.11)	0.12	(0.01- 0.29)	0.10	(0.02- 0.20)
	sH1N1	0.07	(0.02- 0.13)	0.02	(0.01- 0.05)	0.17	(0.01- 0.42)	0.08	(0.00- 0.22)
	sH3N2	0.15	(0.08- 0.24)	0.05	(0.03- 0.09)	0.24	(0.05- 0.49)	0.12	(0.03- 0.26)

Note: CI = credibility interval;

A(H3N2).

eTable 4: Estimates of the Community Probability of Infection and the Secondary Attack Proportion (SAP) for children and adults for pH1N1 excluding all individuals who had been infected with either seasonal A(H1N1) or A(H3N2)..

Individuals assumed to be susceptible	Influenza A subtype*	Cumulative community probability of infection (per season)				Secondary attack proportion			
		Children	(95% CI)	Adults	(95% CI)	Children	(95% CI)	Adults	(95% CI)
Individuals with antibody ≤1:160	pH1N1	0.21	(0.14- 0.29)	0.06	(0.02- 0.10)	0.19	(0.05- 0.34)	0.09	(0.01- 0.17)
Individuals with antibody ≤1:40	pH1N1	0.21	(0.14- 0.29)	0.06	(0.02- 0.10)	0.18	(0.05- 0.33)	0.09	(0.02- 0.17)

Note: CI = credibility interval. * pH1N1 = 2009 pandemic A(H1N1).

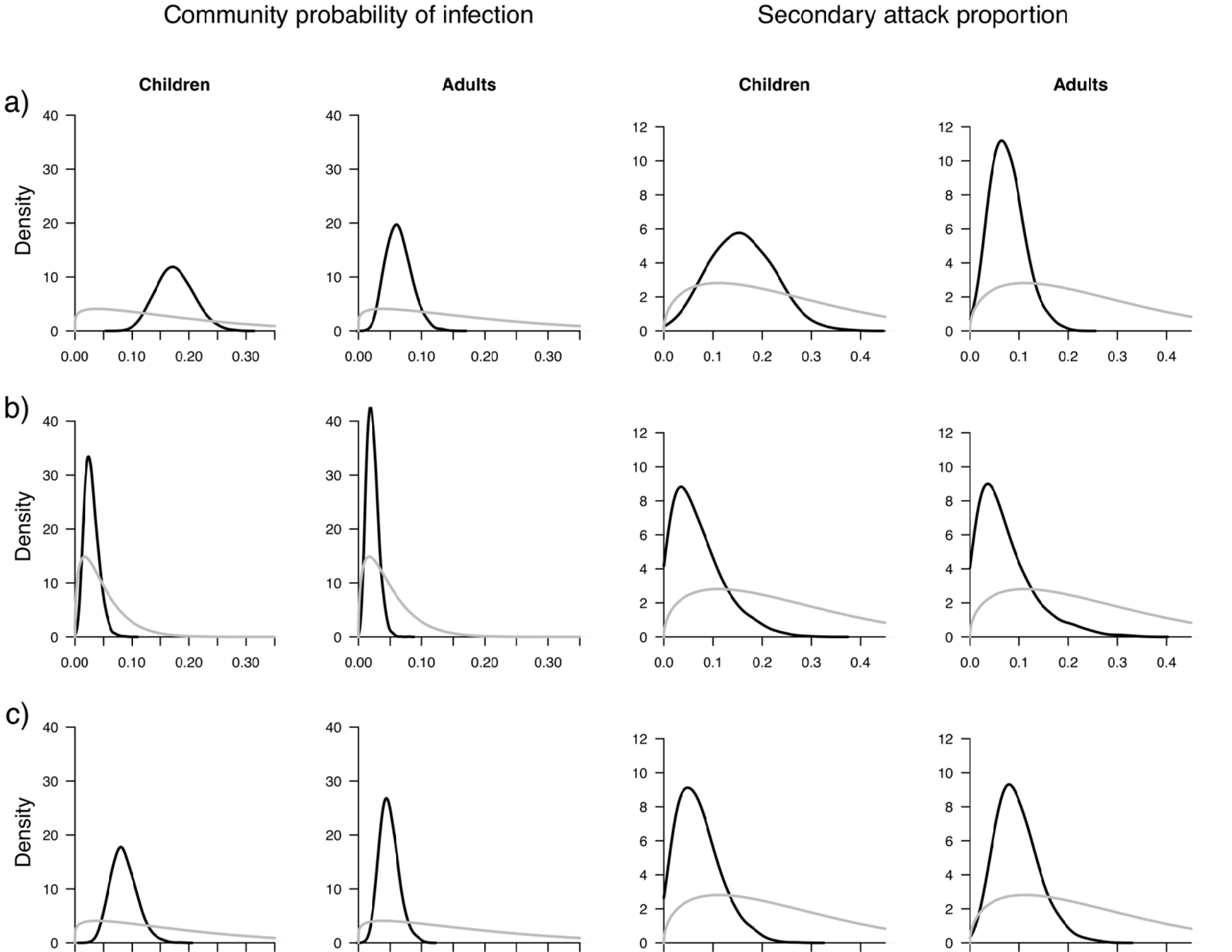
eTable 5: Estimates of the Community Probability of Infection and the Secondary Attack Proportion (SAP) for children and adults allowing individuals to be infected with more than one strain.

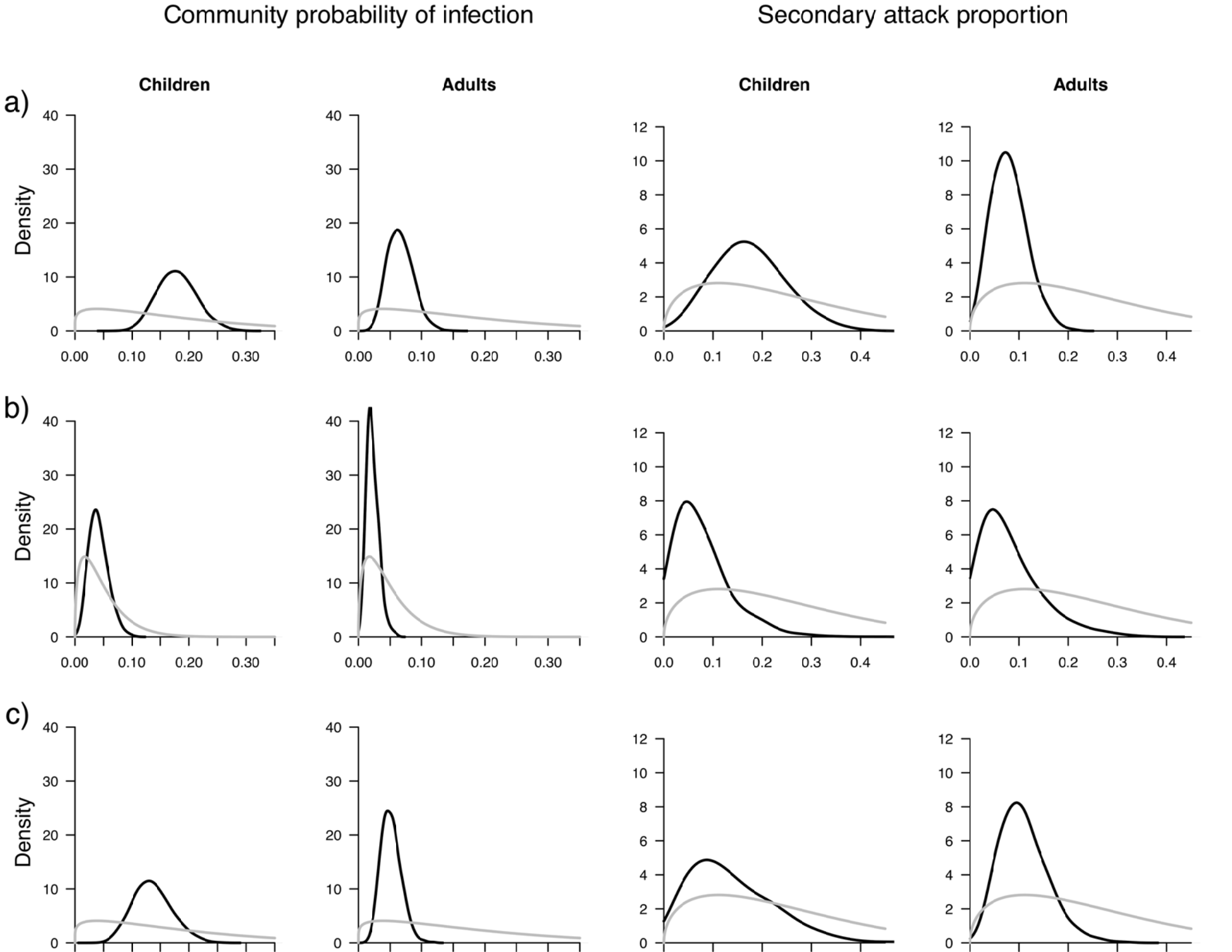
Individuals assumed to be susceptible	Influenza A subtype*	Cumulative community probability of infection (per season)				Secondary attack proportion			
		Children	(95% CI)	Adults	(95% CI)	Children	(95% CI)	Adults	(95% CI)
All individuals	pH1N1	0.18	(0.12- 0.25)	0.07	(0.03- 0.12)	0.15	(0.03- 0.28)	0.07	(0.01- 0.15)
	sH1N1	0.06	(0.03- 0.09)	0.04	(0.02- 0.07)	0.05	(0.00- 0.14)	0.04	(0.00- 0.11)
	sH3N2	0.11	(0.07- 0.16)	0.05	(0.02- 0.09)	0.06	(0.00- 0.16)	0.08	(0.02- 0.17)
Individuals with antibody ≤1:160	pH1N1	0.18	(0.11- 0.25)	0.07	(0.04- 0.12)	0.18	(0.05- 0.33)	0.07	(0.02- 0.15)
	sH1N1	0.08	(0.04- 0.13)	0.04	(0.02- 0.07)	0.08	(0.00- 0.24)	0.04	(0.00- 0.13)
	sH3N2	0.15	(0.09- 0.23)	0.06	(0.03- 0.09)	0.13	(0.01- 0.31)	0.09	(0.02- 0.19)
Individuals with	pH1N1	0.18	(0.11- 0.24)	0.08	(0.04- 0.12)	0.18	(0.05- 0.31)	0.08	(0.02- 0.15)

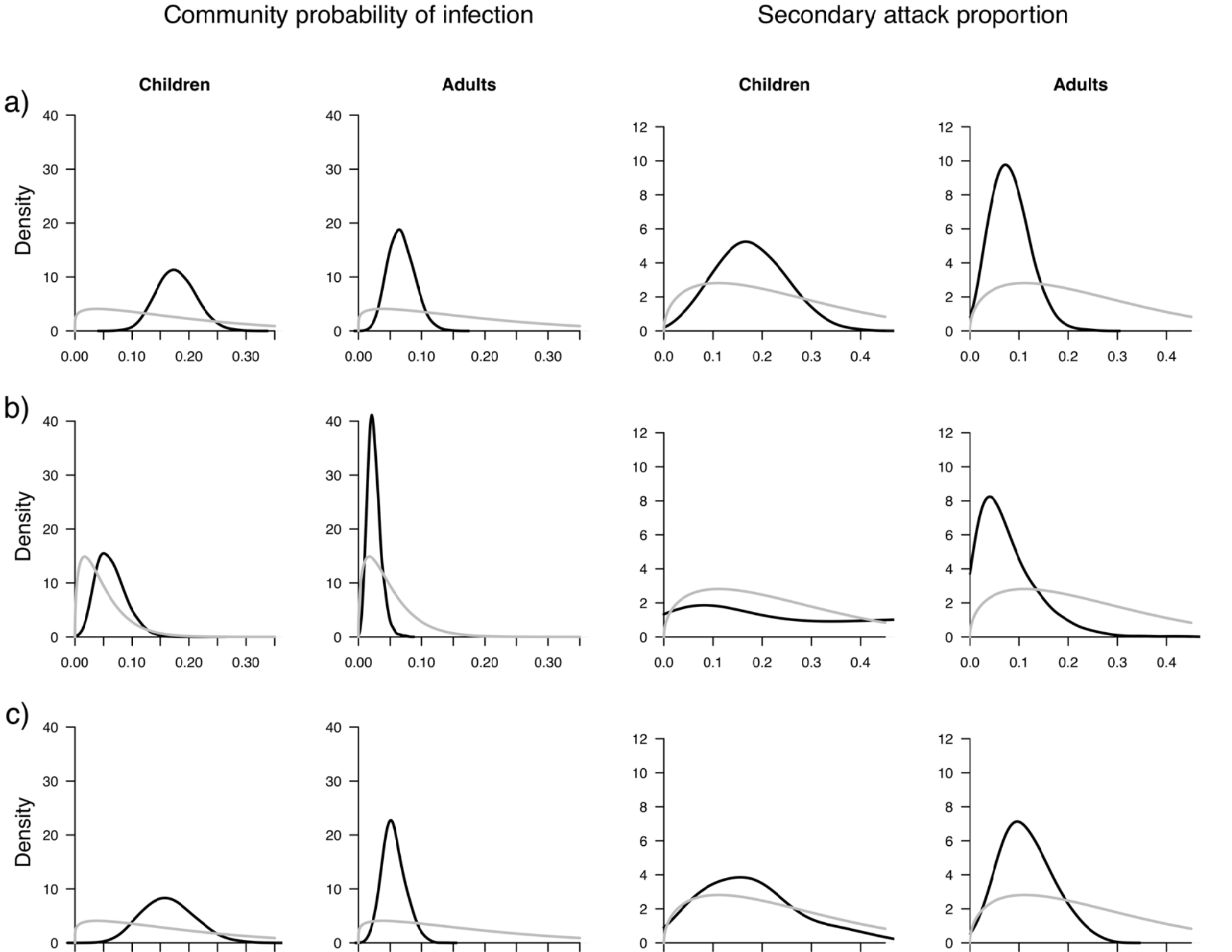
antibody ≤1:40	sH1N1	0.11	(0.05- 0.19)	0.04	(0.02- 0.07)	0.12	(0.00- 0.49)	0.03	(0.00- 0.12)
	sH3N2	0.20	(0.11- 0.30)	0.05	(0.02- 0.10)	0.17	(0.02- 0.45)	0.10	(0.03- 0.21)

Note: CI = credibility interval;

* pH1N1 = 2009 pandemic A(H1N1); sH1N1 = seasonal A(H1N1); sH3N2 = seasonal A(H3N2).


Community probability of infection
Secondary attack proportion
Children
Adults
Children
Adults
a)
b)
c)
Density
Density
Density
0.00
0.10
0.20
0.30
0.0
0.1
0.2
0.3
0.4


Community probability of infection
Secondary attack proportion
Children
Adults
Children
Adults
a)
b)
c)
Density
Density
Density
0.00
0.10
0.20
0.30
0.0
0.1
0.2
0.3
0.4


Community probability of infection
Secondary attack proportion
Children
Adults
Children
Adults
a)
b)
c)
Density
Density
Density
0.00
0.10
0.20
0.30
0.0
0.1
0.2
0.3
0.4