Estimation of reproductive number of Influenza viruses using serological data

Hsiang-Yu Yuan; Steven Riley

May 7, 2015

Abstract

The effective reproductive numbers R_t is important parameter for understanding disease transmissibility but often difficult to obtain because of lack of true incidences and transmission history in a heterogeneous population. We present a method for estimation of the effective reproductive numbers in a population with aged mixing effects and constant changing immune responses, measured by serosurveillance.

We apply our method to 2009 pH1N1 data from the Hong Kong Serological Cohort Study. The effective reproductive number can be constructed in the course of infection and the antibody boosting shows significant different between different age groups while antibody protection is dependent on titres not age.

1 Introduction

The basic reproductive number R_0 ~~is one of the most important parameters to measure infectious diseases~~ allows us to measure the transmission potential for emerging diseases such as influenza, measles, and SARS, ~~etc. The number~~ and is defined as the average number of newly infected cases caused by one ~~typical~~ typically infected individual in ~~a completely nave~~ an otherwise naive population (**?**) ~~. However it is often not~~ XX change cite ot Heesterbeck 2001 acta

theoretica XX. However, models based only on the basic reproductive number are not able to reflect changes in behaviour, the impact of interventions nor variability in the distribution of immune states of individuals. Therefore, the concept of the ~~case because the depletion of susceptible individuals, presence of control interventions and heterogeneity in the population would affect the disease transmissibility~~the effective reproductive number is frequently used. The effective reproductive number R_t, ~~is therefore developed and~~ defined as the average number of secondary cases per primary case during the course of ~~the infection~~ an epidemic XX cite anderson and may here XX (**?**). A number of ~~methods are~~ analytical methods have been developed to measure ~~the~~ R_t ~~either from the initial growth rates or epidemic models (? , ? ,? ,?). All of which rely on the accurate measurement of the incidences. However, for influenza viruses, recent studies show that there is a significant proportion of asymptomatic cases(?), which would lead to an underestimate incidence using medicated data. On the contrary, Since human B cell immunity will be boosted after influenza infection, the course of infection history can be measured by serological surveillance.~~ using case incidence data, based on the idea that the number of cases at a given time is a function of the time series of previous cases, the serial interval distribution and the basic reproductive number ? . Although these approaches can be incredibly useful for measuring the degree to which transmissibility is above or below the self-sustaining threshold, they only describe a very superficial infection mechanism and do not explicitly describe the depletion of susceptible individuals nor other measures of herd immunity.

~~Whereas serological surveillance can be used to determine the disease outbreak, a recent study shows that low levels of antibody boosting and a higher level of baseline immunity would lead to a larger bias of incidence estimates (?) . Furthermore, because the inter-individual differences in physiological statuses,~~

~~the antibody boosting level varies among individuals, where some could obtain very high protection but some only partially protected. The partially protected individuals would still be susceptible to infection. Traditionally, 1:40 titres generally assumed to reach 50% protection against influenza infection, however, the study was originally done using adult volunteers against 1968 influenza strain (? ?) , but it is unclear whether different virus strains can lead to different outcome. Moreover, recent studies show that elderly do not have the same ability as younger subjects to mount an antibody response to influenza vaccination (zciteFagiolo1993). The individuals infection history also affect the baseline immunity. Together these heterogeneity in immune responses and virus strains could cause difficulties in determining the correct incidences under traditional SIR model approaches since a fully protection for recovered individual is usually assumed. An extended model to include differential susceptibility is essential to determine the incidence number~~Traditional serological assays such as the haemagglutinin inhibition (HI) assay XX cite XX and the microneutralization (MN) assay XX cite a definitive source XX allow us to measure the effective concentration of antibodies to specific strains of influenza in individual hosts. These titres have been mapped onto susceptibility for many years using the results of a small number of deliberate infection experiments XX cite hobson and hobson derivatives XX, with the assumption that susceptibility is not a binary state. Rather, susceptibility is inversely proportional to titre (on a $\log_2$ scale).

~~When age-specific structures are included, the estimate of R_t can reduce the levels of uncertainty (Nishiura 2010) given sufficient initial incidence data. Therefore, to~~ Age is also an important factor in the transmission of influenza. To be able to estimate an accurate reproduction number that reflects age effects, age mixing ~~effects along with~~ along with other age specific parameters are re-

quired. ~~Most~~ Many studies include age specific contact ~~pattern, this work for a~~ patterns layered on top of the traditional SIR model ~~assumption. However, to capture transmission dynamics in aged structure population age specific immune responses are needed to measure the effects on transmission.~~ assumptions.

In this study, we developed a novel approach including age specific antibody boosting and differential susceptibility in the transmission model to simultaneously estimate reproductive number and cumulative incidence using serological dynamics data.

2 Results

2.1 Patterns in the data

We obtained HI titers (Figure 1) for samples from 523 individuals from the baseline round (between 4 Jul 2009 and 28 Sep 2009), and from 465 individuals recruited during the first follow-up (between 11 Nov 2009 and 6 Feb 2010). In Hong Kong, the pandemic started in early May 2009 ~~where~~ with the first confirmed case announced on 1st of May. Between baseline and first follow-up, the proportion of the study population without detectable antibodies decreased from 90.1% to 80.0% XX confidence intervals needed here and throughout XX while the proportion of the population in all detectable titre classes increased. However, increases were much more substantial for titres between 1:80 and 1:320 (61%) and were far less substantial in the the population with titres less than or equal to 1:40.

2.2 Dynamics of the data

We defined a compartmental transmission model in which each age group and titre class was represented explicitly (see Methods) and estimated the param-

eters of the model using a Bayesian framework. Even without fitting to incidence data, the model was able to reproduce some key features of the incidence of influenza infection in Hong Kong during the initial waves of the 2009 pandemic. The peak of infection incidence in the best fitting model was 13 Sep 2009 (Figure2A) which is consistent with a peak of incidence of hospitalization of confirmed cases in the third and fourth weeks of September (**?**). Using the classic definition of seropositivity for incidence ($\geq 1:40$), the model estimated a cumulative infection attack rate between baseline and first follow-up of 16.4% [11 - 22%] for the study population. Again, this was consistent with the 16% [13-18%] non-age-standardized average rate of seroconversion from ~~the~~ a previous study (**?**). However, the duration of the best-fit model epidemic was considerably shorter that of the observed epidemic in Hong Kong. We note that the model structure did not contain any mechanism capable of reproducing a long tail (or shoulder) in the epidemic profile (see Discussion).

The underlying pattern of titre changes in the model is shown in Figure 2B. The black and red bands up to August reflect our assumption that pre-existing antibodies in each age group were uniform between 1:10 and 1:80 and that incidence did not reach a substantial proportion of the population until the end of the summer. Once incidence did increase, the 1:80 group had the highest prevalence of those with detectable antibodies. The pattern of antibody response is approximately constant after December once the best-fit model incidence fades out near the beginning of January.

2.3 Fitting model output to data

In order to compare the distribution of titres in the data and the best fit model, we examined the titres at the average time of sampling for each round, 11 Aug 2009 as T1 for the the baseline recruitment and 22 Dec 2009 as T2 for the

first-follow-up. Although there are some biases in the age specific comparison of model output and observed titres, given the small sample size for some age groups, the model fit is generally good. We compared the model-predicted distributions of titre to the observed titres for these time points for all age groups in Figure 3A and B. Model titres at T1 are very close to the assumed initial conditions but are not inconsistent with the observed tires. At T2 the proportions of all model titres (other than 1:10) overlap with the observed titres. The seroprevalence (proportion $\geq 1:40$) ~~increases~~ increased from 3.7% to 8.9% at T1 to 20.6% at T2 in the model output XX this not clear to me XX, which is also in good agreement with and increase from 5.1% to 19.8% in the observed sera (Table2). The geometric mean titres changed from 3.0 (T1) to 4.4 (T2), while the observed sera show the mean geometric titres changes from 3.8 to 4.8 (Table2B) ~~.~~ XX confidence intervals needed here and thoughout XX.

~~Although there are some systematic biases in the age specific comparison of model output and observed titres, given the small sample size for some age groups, the model fits are largely not inconsistent.~~ Among nave individuals, children show a highest proportion of seroconversion while older adults show a lowest (Figure3A). Among individuals with detectable antibodies, children shows a widely distributed antibody titres while mid-aged adults and elders, most individuals have low to medium titres. Seroprevalence appears to be highest in children group both from observed and model output (Table2A).

2.4 Boosting and protective effect of titres

We observed a significant differences in age specific estimates of antibody boosting following infection. The mean antibody boosting is among highest in children and decreases by age until mid age adults, whereas seniors show a higher boosting than mid aged adults (Table1). The largest increase of mean geometric

titres ~~occurs~~ occured XX check for tense XX in children group for 3.46 folds increase from 3.6 to 5.39, which agrees with sera data from children that changes for 3.68 folds increase from 3.92 to 5.80 (Table2B). The lowest increase of mean titres occurs in both elders and young adults from simulation whereas young adults show a relatively high increase from observed sera.

For adults aged 40 to 64, our results suggested that the low induced boosting may be offset by a higher level of protection per log unit antibody titre (Table1). For both children and young adult the protective titres are between 1:20-1:40. The high boosting levels with high protective titres together, as seen in children, could lead to high susceptibility before and strong antibody response after the infection. Despite slightly lower in mid age adults and a higher variability in seniors, overall, protective titres are not significant different by ages.

2.5 Reproduction number

To capture the heterogeneities of population immunity and contact behaviour, we estimated the effective reproductive number R_B in the presence of pre-existing antibody-derived protection. During the pandemic, R_B decreased from 1.25 to 0.83 and became stable after November (Figure4A) when disease incidence was low. The R_B declined slowly in the first three months of the outbreak and dropped rapidly during September and October, which is around the peak of incidences with frequent seroconversion occurred. Using SIR model, with the same age mixing effect, a higher R_t was present ~~at initial~~initially, which corresponds to R_0 and drops earlier and quicker than R_B (Figure4B). In the SIR model, the recovered individuals are fully protected to be reinfected therefore a stronger population immunity is gained after the infection ~~. The R_0~~ XX didn't understand this bit XX. The initial value of R_B using the parameters estimates from the serological model with the assumption that whole population is sus-

ceptible without pre-existing antibody is 1.88. If the 30% of protection exists in people without detectable antibody, the R_0 becomes 1.32, which is close to the SIR model.

2.6 ~~Sensitivity analysis~~

~~We compare the results to alternative models, one with less PUAb and one with no age specific boosting. All of the models can reproduce similar serological and disease dynamics patterns with only slightly differences. In order to check which model can be more realistic to reflect infectious processes, we compare the incidence rate from different models to paired sera sets. We acknowledge that comparing the model output to the paired sera would have some bias due to only part of samples are recruited for followed up. However, it still provide a way to measure the effects of different parameters. We calculate the proportion of incidences infected from naive and immune individuals in each age group. Among the three models, all model show a similar proportion of incidences from immune in children and young adults, which agrees with observed data (FigureS2). However, when age specific protection titres are not included, middle aged adults show a considerable high incidence from immune which contrast to the observed sera due to the higher protection titres than the age specific one for its age group. When PUAb drops, although middle aged adults show a considerable low incidence from immune individual , the overall incidences becomes too high which disagree with obserbed data. We also checked their impacts on the reproductive numbers. When PUAb decrease from 0.3 to 0.15, initial Rt slightly decreases (figS5). When there are no age specific protection titres, Rt is same initially but decreases slightly lower after the peak of outbreak.~~

3 Discussion

~~The accurate inference of effective reproductive number and incidence rate from temporal serological data depends on the correct estimates of age specific contact and immune response here. We assume the contact mixing is the same along the outbreak. The contact mixing will possibly change after disease was mitigated. We use a contact mixing estimated (which year? before or after the pandemic?) We also assume the antibody boosting is only age dependent but not related to the initial antibody levels . Although some assumptions are made, this is the first study to reconstruct the infection processes using more realistic immune boosting and protection. The strength of this approach is the study design would be less restricted and the number of estimates would be less biased. Because we can take temporal sera data~~

We have defined an SIR-like model for the transmission of influenza in which individual antibody titre levels are enumerated explicitly. This model structure allowed us to define the reproduction number R_B to be the transmissibility of an influenza strain in a population ~~, without the traditional paired serological surveillance and to predict the outcome of disease and serological dynamics. This provide a great flexibility in serological study design.~~

~~We observe a significant increase of mean antibody titresamong individuals showing detectable antibody (see Figure1 and Table2). Since the~~ that contains pre-existing ~~immunity of pH1N1 is low~~ partial immunity, as reflected by antibody titres. By estimated the parameters of the model using cross-sectional serological data from the 2009 pandemic in Hong Kong, ~~this increase could be partly due to repeated infection on individuals who have already been infected before round1. Because the infection on individuals who are weakly immune are still susceptible to infection, once the infection occurs, the immunity would be boosted to a higher level which increases the mean titres in the immune group.~~

~~Several facts support this possibility, first the antibody boosting levels are widely distributed from low to high boosting. Among seroconverted individuals, 19.2% of them would have equals to or less than 4 folds conversion (FigureS3). These individuals are weakly immune and are still highly susceptible. Second, we see there are significant proportion of individuals seroconverted from initially immune status. Overall, there are about 10% infection occurs for people having at least 1:10 detectable titres. The other possibility to cause the mean titres increase is that the new strain of viruses induce higher boosting after the initial outbreak. However, we don't have evidence that the pandemic H1N1 strain has changed too much on antigenicity nor other property that can trigger immune boosting.~~

~~The effect of recurrence of infection would make a fundamental difference on estimating incidence using seroprevalence data. Neither cross-sectional or paired seropravelance consider that same individual could be infected twice from the increase of titres. This could lead to underestimate the true disease incidences. Some of the possible causes of underestimation the real incidences using serological surveillance data have been described by ? where using 1:40 as a standard to define seroprevalence could be the source of bias. From our study, we suggest a lower immune boosting exists for age 40-64. A traditional sero-surveillance approach could also underestimate the true burden. The mean TP50 of young adults is 1:39.4 which is consistent to previous studies using seasonal H1 strains, whereas slightly lower values are found for children and mid aged adults but the difference is not significant. In our model, we incorporate a proportion of naive individuals that have~~ we were able to explain age-specific patterns of boosting and overall attack rates. Our analysis suggests that the 2009 pandemic strain had an underlying reproduction number of 1.9, which was reduced to only 1.3 by the presence of pre-existing ~~immunity to protect these~~

~~naive individuals to be infected. The reduced susceptibility among individuals lacking pre-existing antibodieshas been observed by ? . One of the possible explanation is the cellular immunity confer the protection of individuals without antibodies (?). Since if the individuals are protected, their antibodies won't be boosted into a detectable level. When the ratio decreases, the total incidences become higher which also lead to a higher reproductive number in early phase. Further studies would be needed to understand the proportion in the population have such protection but without any detectable antibodies.~~ or cross-reactive antibodies.

~~Impact for future public health control~~ The serological model in this study provides a framework to estimate the changes of reproductive number using time series serosurveillance data. It ~~allow~~ allows the study design to become easier to perform. In the future, once we have early estimates of antibody boosting and protective titres, we can infer the effective reproductive number and estimate the duration when the reproductive number can drop below than one, which is valuable for disease control and prediction.

4 Materials and Methods

4.1 Serological samples

We analyse new data from the baseline and 1st follow-up rounds of the Hong Kong influenza serological survey [?]. Here, we use hemagglutination inhibition assay results for the 2009 pandemic strain of H1N1 for all baseline samples and for all 1st round follow-up samples. Previously published analyses use microneutralization assay results only for the subset of individuals with paired samples. We also note that previous analyses was based only on fourfold rise or greater in paired samples and thus did not need to exclude individuals who reported

vaccination prior to the baseline visit. Here, because it was our objective to make inference on cross-sectional patterns of serology, for our primary analyses, we excluded all individuals who reported any influenza vaccination in the preceding years.

4.2 Transmission Model

The epidemiological and serological dynamics were simulated based on a disease transmission model with a immune response component. The serological model without age mixing effects was illustrated in the figureS1 by the extension of the basic SIR model to multiple levels, where each level represents a different antibody titre. Once individual becomes infected, viruses could transmit to any susceptible during the infectious periods. After the individual get infected, the antibody is boosted to a higher level according to a Poisson distribution. The transmission susceptibility for the titre given a contact is obtained based on the differential susceptibility. A higher antibody level will increase the protection of antibody and reduce the susceptibility of individuals. After recovery, infected individual becomes fully protected by both cellular and humoral immunity. At the same time, the antibody titres will be boosted to an elevated level . Individual later becomes susceptible again after the CTL immunity wanes.

When age mixing effects are considered, disease dynamics are described by the following equations. The model has not taken into account the demographics since the duration we consider here is less than one year. The birth and death rates would not generate significant differences.

$$\frac{dSi(a)}{dt} = -S_i \cdot \rho(i) \cdot \lambda(a) + \omega \cdot R_i(a) \tag{1}$$

$$\frac{dIi(a)}{dt} = S_i \cdot \rho(i) \cdot \lambda(a) + \frac{1}{T_g} \sum_{j=i}^{i_{max}} I_i(a) \cdot g_{ij} \tag{2}$$

$$\frac{dRi(a)}{dt} = \frac{1}{T_g} \sum_{j=i}^{i_{max}} I_i(a) \cdot g_{ij} \tag{3}$$

where a represents the age group for each individual, i is the disease susceptibility for susceptible individuals in the presence of titres level , λ is force of infection, is waning rate for CTL immunity. T_g is the duration of infection, g_{ij} is the probability of immune boosting from titres i to j. The force of infection on member of age class a who are completely naive within the population is

$$\lambda(a) = \beta \sum_{b=1}^{a_{max}} \left\{ m_{ab} \sum_{i=0}^{i_{max}} f(b) \cdot I_i(b) \right\} \tag{4}$$

where β is the basic unit of transmission rate, m_{ab} is the contact rate from the age class b to a, $f(b)$ is the virulence of viruses. We stratify sera samples into age groups, i.e. children and adolescent (2-19 y/o; for convenience, we define this group as children throughout the study), young adults (20-39 y/o), mid-age adults (40-64 y/o) and elders ($\geq$= 65 y/o). Contact mixing matrix of the four age groups is calculated from a community study in Hongkong (FigureS4).

4.3 Differential susceptibility

The susceptibility ρ is defined as $1 - \phi$ where ϕ is the proportion of individuals are protected from infection given a titre level. The protection is modelled in a two parameters logistic function.

$$\phi(i) = \frac{1}{1 + e^{I_\beta (i - I_\alpha)}} \tag{5}$$

where I_α determines the titres i^*, at which ϕ will drop 50% from the maximum value. The is equal to the antibody titre TP50. I_β determines the shape of the function.

4.4 Calculating effective reproductive number

The effective reproductive number was calculated using next generation approaches. Following the same notations in the study by **?**, the transmissions matrix T and the transitions $\sum$ can be obtained. Each element in T represents the newly 'transmitted' cases in age group a and titres i in a unit time per each infected individual in age group b and titres l, which can be calculated as $\frac{\lambda(a)\cdot S_i(a)}{I_l(b)} = \beta h_i(a) f M_{ab} S_i(a)$. $\sum$ represents the transitions of between cases in each age and titres group. Since the probability of boosting to all the other titres becomes one, each element in $\sum$ simply is the loss of infected cases due to recovery in a unit time $-\frac{1}{T_g}$ from our model.

4.5 MCMC approach

The posterior distributions of the parameters, including the transmission rate, age dependent antibody boosting, and two parameters for immune protection, are obtained after 10^6 steps in MCMC (Table1) to make sure ESS higher 100. The starting day of pandemic is predetermined as 1st May to match Honk Kong pandemic surveillance data while we assume 100 individuals are initially infected. Initial baseline antibodies that are larger than or equal to 1:40 are set to be 3.3% for age groups < 65 in population according to (ref) and for seniors we use 6.6% to represent a higher seroprevalence.

5 Figures

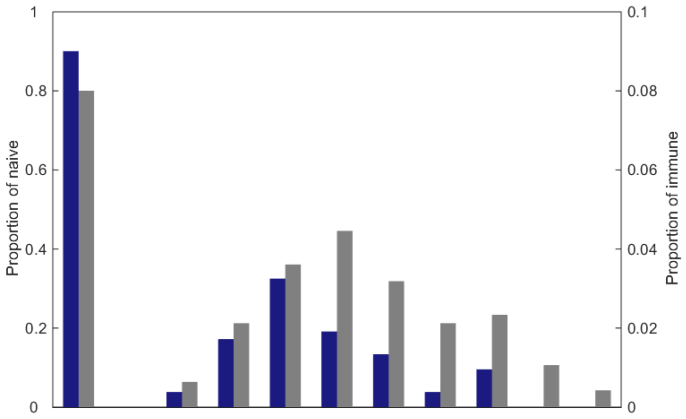

Figure 1: The serological profiles of nave and immune population at the initial and the followup rounds

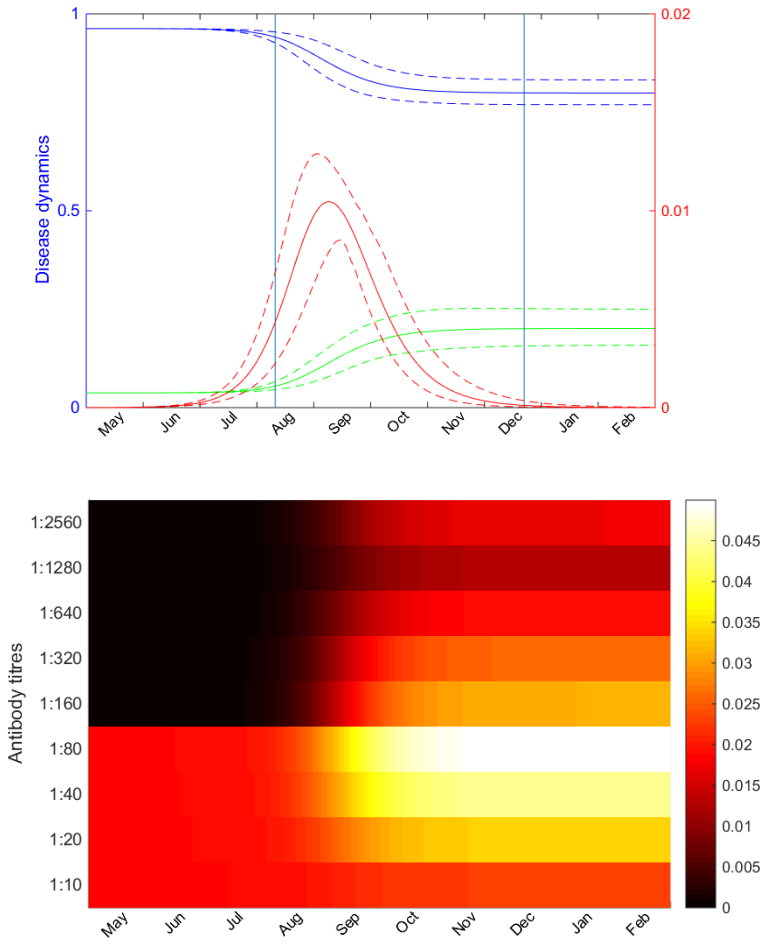

Figure 2: The sera and disease dynamics from model simulation. The dynamics are reconstructed using 100 samples from posterior distributions. a, The disease dynamics are calculated using the classic definition of seropositive (1:40). Susceptible individuals are defined as individuals who show titres less than 1:40 and is plotted in blue. Infecteds is plotted in red. Immune protected individuals are individuals who show titres greater or equal to 1:40. The number of immune protected individuals is plotted in green. Solid lines represent the mean of simulation from posterior distribution. Dashed lines represent 95% credible interval. b, The sera dynamics are simulated during the outbreak. Darker color represents a lower proportion and lighter represents a higher proportion in the population.

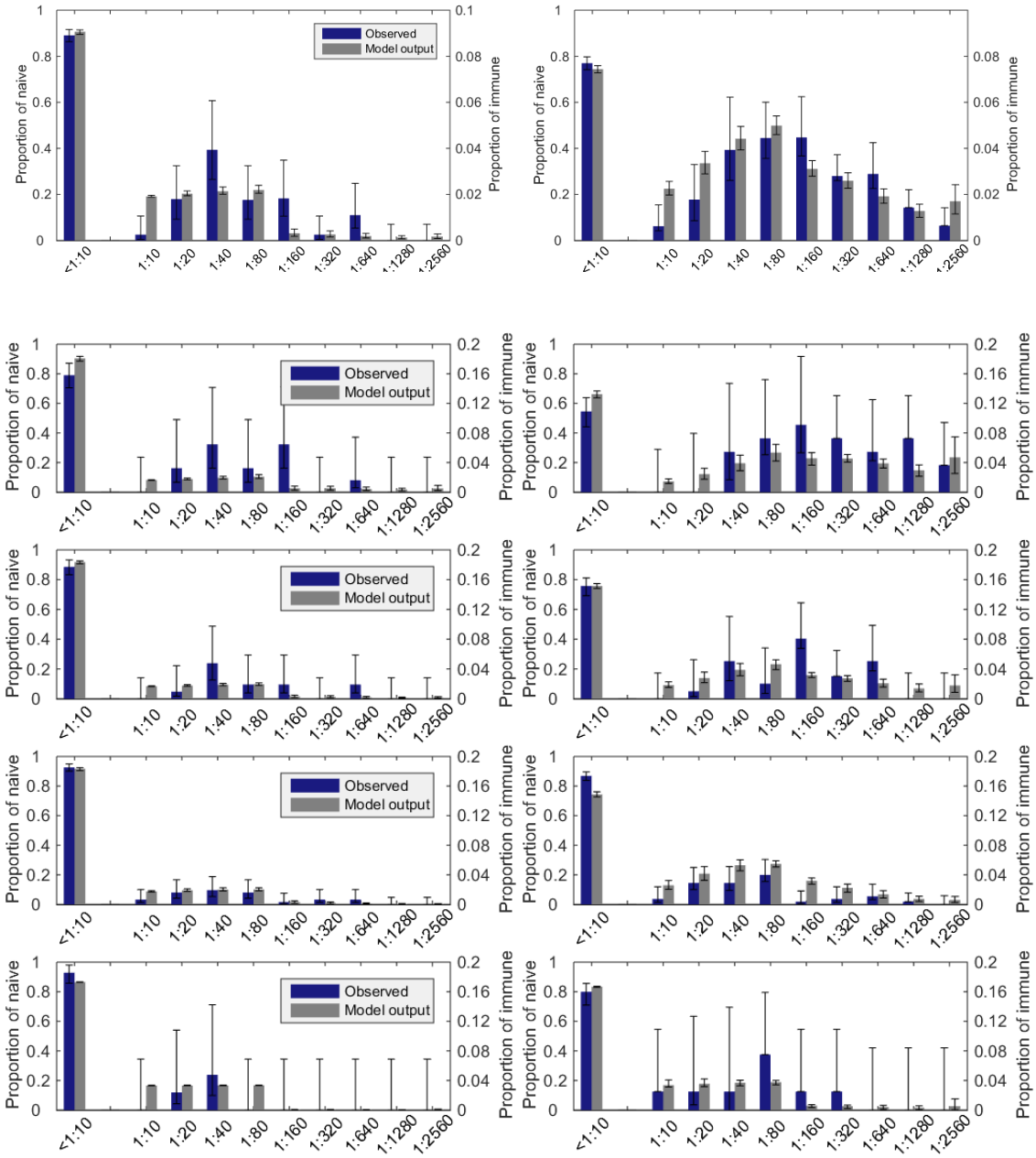

Figure 3: The simulated and the observed sera at the initial and the followup rounds. a,b sera data among overall population. c-j sera data among age groups.

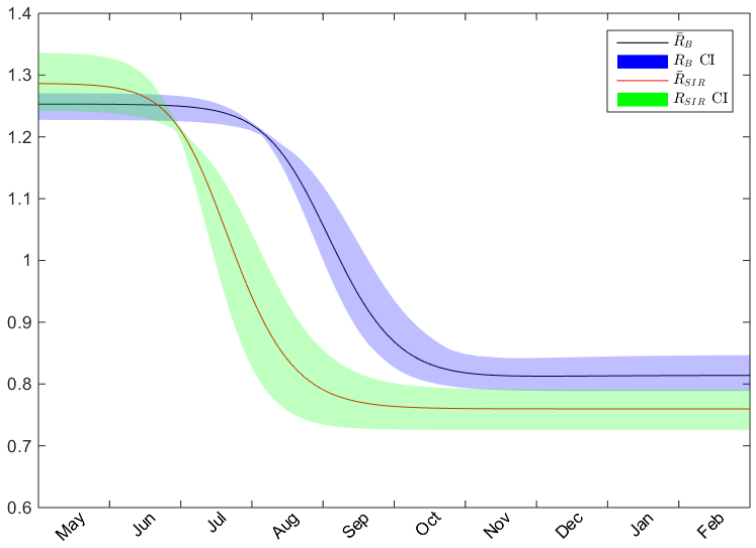

Figure 4: Figure4, Comparison of effective reproductive number with serological responses to the classical SIR model. Black line, the Reproductive number R_B estimated using serological model with age specific parameters. Red line, R_t estimated using the classical SIR model.

Table 1: The parameters estimated from the full age specific model and without age specific parameters. The minimum ESS is 53.

parameters	age specific	non age specific
β	0.0885 [0.0849-0.092]	0.09 [0.0864-0.093]
AbB_1	5.85 [4.96-6.87]	-
AbB_2	5.04 [4.12-6.03]	-
AbB_3	4.11 [3.36-4.96]	-
AbB_4	4.69 [1.90-7.54]	-
$TP50_1$	2.54 [0.74-4.67]	-
$TP50_2$	2.98 [0.67-8.25]	-
$TP50_3$	1.53 [0.57-3.37]	-
$TP50_4$	5.23 [0.81-9.65]	-
AbB	-	5.00 [4.47-5.56]
TP50	-	1.68 [0.62-3.06]
R_0	1.88 [1.80-1.95]	-

Table 2: The comparison of the observed sera and the model output. A, The seropravalene of initial and the followup rounds. B. The geometric mean titres (log scale) of initial and the followup rounds.

A				
age	initial	followup	X_{T1}	X_{T2}
Total	0.09 [0.07 0.12]	0.21 [0.17 0.24]	0.05 [0.046 0.064]	0.20 [0.184 0.214]
3-19	0.18 [0.15 0.21]	0.45 [0.41 0.50]	0.06 [0.050 0.083]	0.30 [0.274 0.327]
20-39	0.10 [0.08 0.13]	0.23 [0.20 0.27]	0.05 [0.042 0.058]	0.20 [0.178 0.215]
40-64	0.05 [0.04 0.08]	0.09 [0.07 0.12]	0.05 [0.041 0.058]	0.19 [0.171 0.209]
≥ 65	0.05 [0.03 0.07]	0.15 [0.12 0.18]	0.07 [0.067 0.070]	0.10 [0.085 0.105]
B				
Total	3.76 $\pm$1.53	4.80 $\pm$1.27	3.00 [2.801 3.237]	4.35 [4.100 4.624]
3-19	3.92 $\pm$1.38	5.80 $\pm$1.83	3.60 [3.152 4.086]	5.39 [4.794 5.964]
20-39	4.08 $\pm$1.55	4.92 $\pm$1.50	3.06 [2.813 3.339]	4.57 [4.081 5.091]
40-64	3.52 $\pm$1.69	3.67 $\pm$1.72	2.89 [2.712 3.110]	3.87 [3.479 4.321]
≥ 65	2.67 $\pm$0.47	3.63 $\pm$1.49	2.56 [2.511 2.636]	3.11 [2.621 3.678]

6 Supplementary

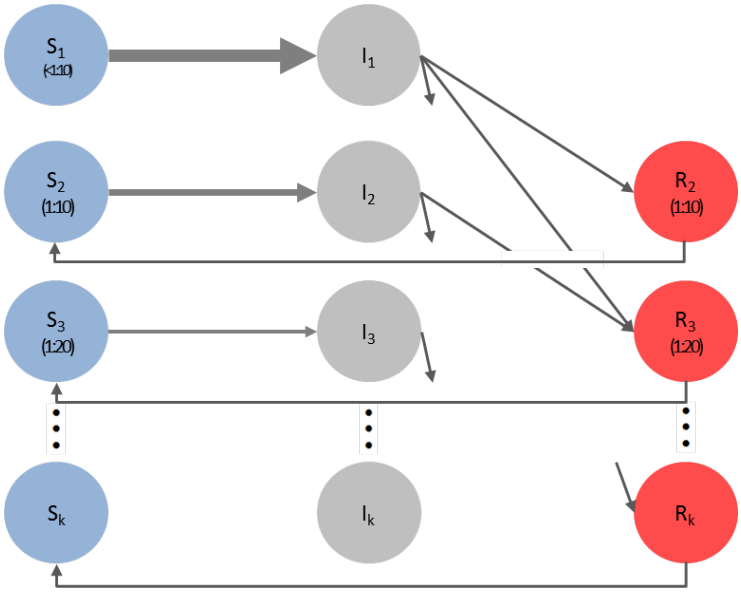

Figure S1: The illustrated schema of serological model without age mixing effects

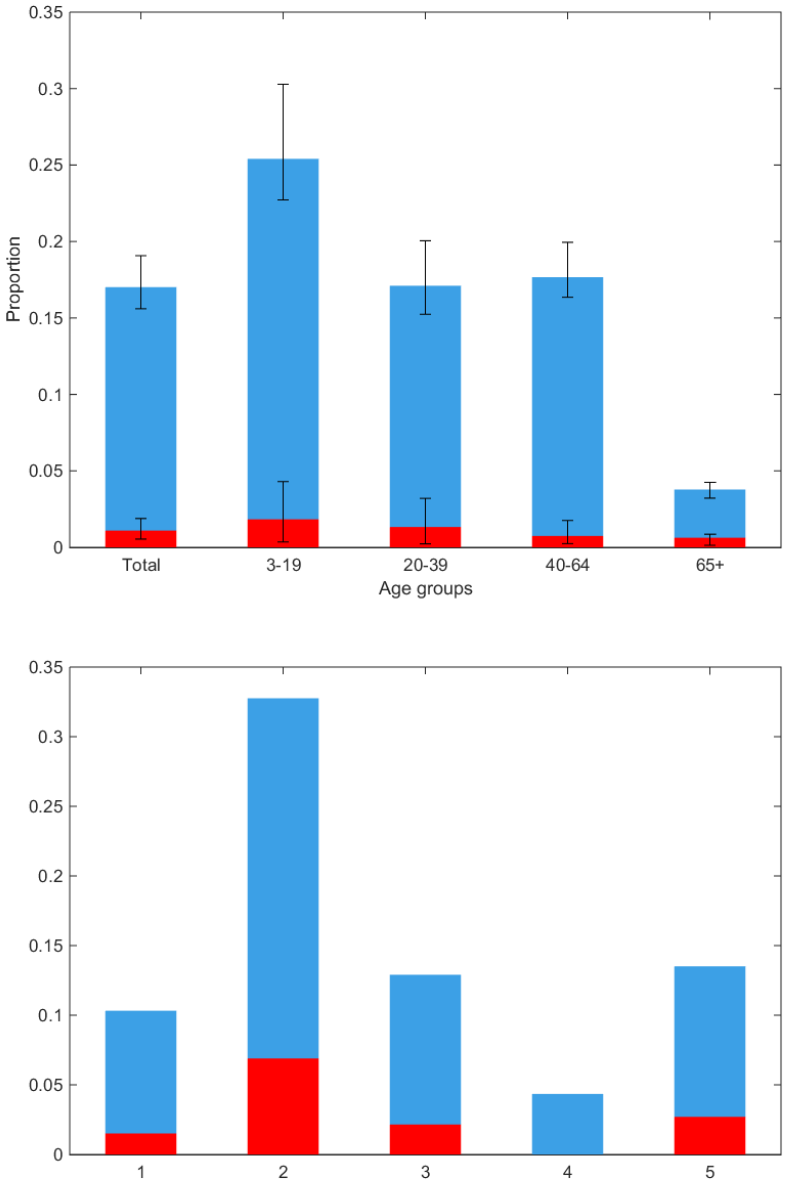
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Proportion
Total
3-19
20-39
40-64
65+
Age groups
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5

Figure S2

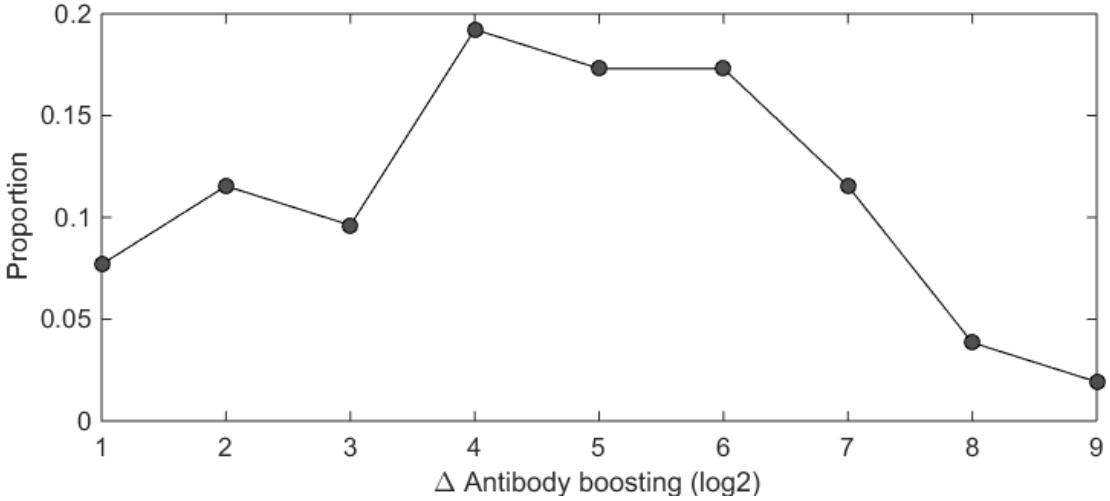

Figure S3: The distribution of antibody boosting from initial to the follow-up rounds

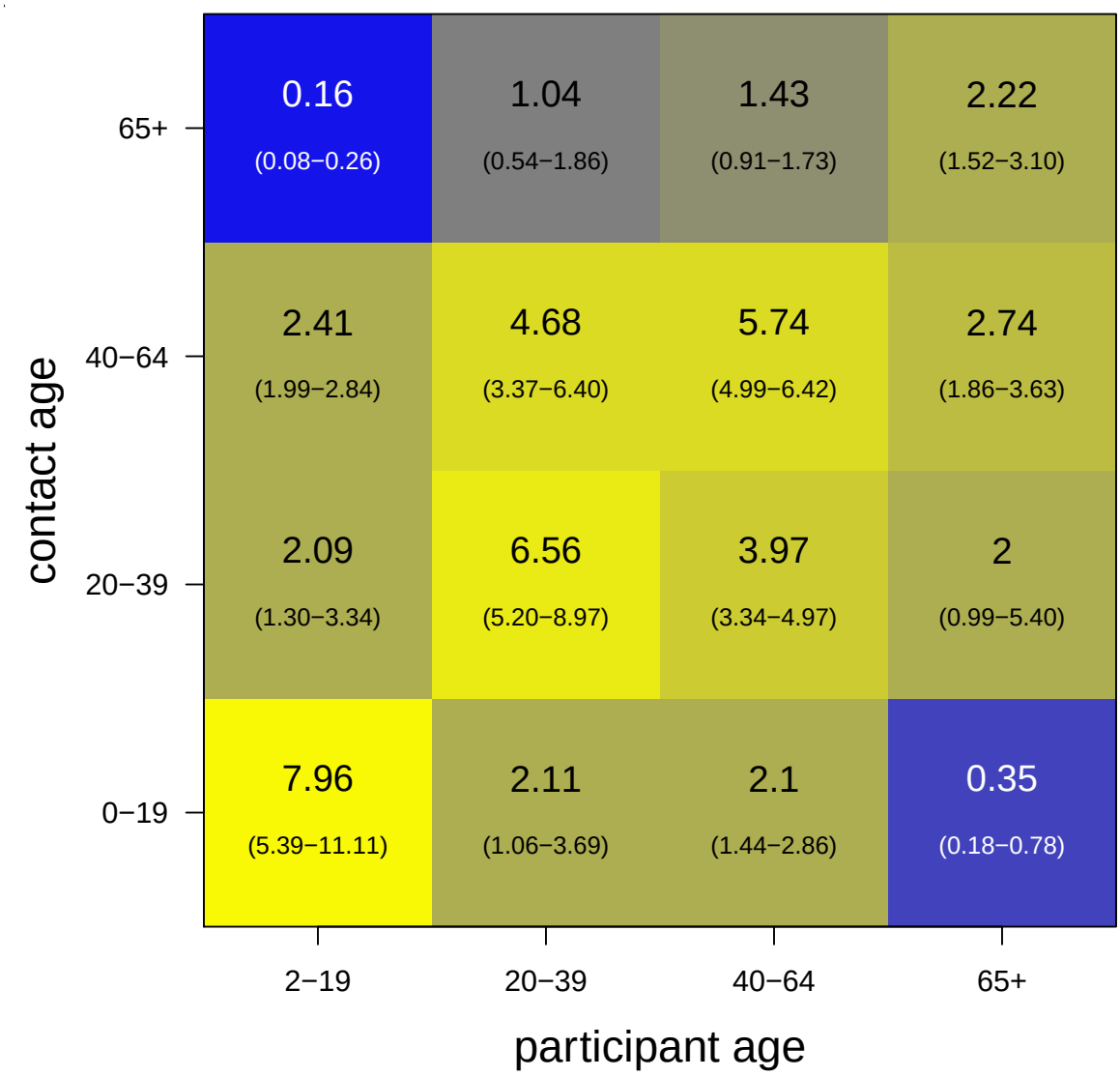

Figure S4: The contact mixing matrix among 4 age groups in HongKong (**?**)