Published in final edited form as:

Stat Methods Med Res. 2018 July ; 27(7): 2050–2059. doi:10.1177/0962280216673676.

Estimating age-specific reproductive numbers—A comparison of methods

Carlee B Moser^{1,2} and Laura F White²

¹Center for Biostatistics in AIDS Research, Harvard T. H. Chan School of Public Health, Boston, MA, USA

²Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

Abstract

Large outbreaks, such as those caused by influenza, put a strain on resources necessary for their control. In particular, children have been shown to play a key role in influenza transmission during recent outbreaks, and targeted interventions, such as school closures, could positively impact the course of emerging epidemics. As an outbreak is unfolding, it is important to be able to estimate reproductive numbers that incorporate this heterogeneity and to use surveillance data that is routinely collected to more effectively target interventions and obtain an accurate understanding of transmission dynamics. There are a growing number of methods that estimate age-group specific reproductive numbers with limited data that build on methods assuming a homogeneously mixing population. In this article, we introduce a new approach that is flexible and improves on many aspects of existing methods. We apply this method to influenza data from two outbreaks, the 2009 H1N1 outbreaks in South Africa and Japan, to estimate age-group specific reproductive numbers and compare it to three other methods that also use existing data from social mixing surveys to quantify contact rates among different age groups. In this exercise, all estimates of the reproductive numbers for children exceeded the critical threshold of one and in most cases exceeded those of adults. We introduce a flexible new method to estimate reproductive numbers that describe heterogeneity in the population.

Keywords

reproductive number; heterogeneity; social mixing; H1N1; transmission models; influenza

1 Introduction

Transmission parameters that are of interest during an infectious disease outbreak depend on contact patterns between individuals. Although many methods assume homogenous mixing, ignoring heterogeneity in contact rates when estimating transmission parameters, such as the

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav

Corresponding author: Carlee B Moser, Harvard University, 651 Huntington Ave, FXB 513, Boston, MA 02144, USA. cmoser@sdac.harvard.edu.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

basic reproductive number (R_0), can give misleading results,^{1–5} which can be particularly problematic as an outbreak is unfolding and resources for intervening are limited. There is a growing body of work that incorporates information on social mixing when estimating transmission parameters and uses social mixing survey data to quantify the contact rates among different age groups.^{1,6–8} Such surveys have also shown that individuals of similar ages tend to mix together.^{9–12}

This conclusion is of particular importance given that many studies have found that children are often the driving force of an outbreak and that interventions targeted to this group, such as school closures, could positively impact the course of an epidemic.^{2,5,13–20} Accounting for heterogeneity in transmission and also estimating age-group specific reproductive numbers could provide insight into differences in transmission potential by age and interventions can be more effectively targeted to reduce the size of an outbreak, even as an outbreak is unfolding.

Some methods have been developed to provide estimates of age-group specific reproductive numbers. Nishiura et al.^{16,21} introduced an approach to estimate age-group specific reproductive numbers using a next-generation matrix (NGM) that incorporated the overall R_0 , the susceptibility of the age groups, and the frequency of contacts between age groups. The frequency of contacts was assumed to be known and was based on data from a social mixing survey.⁹ At first, only minors and adults were examined, but later this was expanded to include six age groups; they found that those aged 13–19 could sustain transmission of pandemic H1N1 in Japan.

Glass et al.²² expanded on this work and introduced an approach that modifies two existing homogenous transmission models (the methods of Wallinga and Teunis and White and Pagano)^{23,24} to account for age heterogeneity in the estimation of R_0 with the NGM. They also used social mixing survey data from the POLYMOD study¹⁰ to calibrate their model. Overall, the reproductive number estimates were slightly higher than Nishiura et al.,^{16,25} but they observed a similar trend of high child-to-child transmission in Japan. A limitation of this method, however, is that the population can only be divided only into two groups (i.e., adults and children).

White et al.²⁶ adapted the Wallinga and Teunis model²⁴ to account for the likelihood of transmission between multiple age groups. Social mixing survey data was used as a surrogate for the likelihood of transmission and allowed for multiple reproductive numbers to be estimated. In modeling the H1N1 outbreak data from South Africa, they considered multiple age categories stratified into 5-year intervals up to 45+ years of age. The largest average R_t was observed for 15–19 and 20–24 year olds, and the overall average R_t was similar to the homogenous estimate obtained with the original Wallinga and Teunis approach.

In this article, we address how social mixing data can be used to better account for heterogeneity in the estimation of the reproduction number as an outbreak is unfolding. Specifically, we introduce a new approach to estimate age-specific reproductive numbers that readily incorporates available data as an outbreak is unfolding, with the goal of better

informing interventions to limit the outbreak. Our method builds on the approach introduced by White and Pagano²³ and uses a Bayesian framework to incorporate details on contact rates among groups with social mixing data. With our approach we can estimate the serial interval and provide more intuitive estimates of the variability in the estimates of R_0 . We illustrate this method and compare it to the three previously described methods using data from the 2009 H1N1 pandemic influenza outbreaks in South Africa and Japan.

2 Methods

2.1 South Africa data

The National Institute for Communicable Diseases (NICD) of the National Health Laboratory (NHLS) in South Africa maintained a database of laboratory-confirmed cases of H1N1 during the 2009 pandemic. Data were collected throughout the country beginning in April and lasting until October, and include basic demographic information and spatial and temporal data for each case.^{27–29} Data from the first 2 weeks of the outbreak are not included in the analysis due to large gaps between cases, likely due to imported cases and lack of sustained transmission. The total outbreak included 12,340 cases over a 91-day period. In the initial epidemic growth phase, there were 2387 cases with nonmissing age from days 19–53 of the outbreak. Each case was classified as a child or adult, with children defined as those <15 years of age and adults as those ≥15 years of age.^{30,31} Figure 1(a) displays the total epidemic curve, distinguishing between child and adult cases, and the white vertical line displays the cut-point of the epidemic growth phase. Figure 2 shows contact tracing data on a subset of the confirmed and probable cases, which is used to quantify the serial interval distribution in the methods outlined below.

The age distributions of children and adults during the initial epidemic growth period and total outbreak period are shown in Table 1. Children ranged in age from <1 to 14.9 years, and adults from 15 to 90 years. Children were on average 8–9 years old and adults were 25–28 years old.

2.2 Japan data

Nishiura et al.¹⁶ provide a detailed summary of the 2009 H1N1 outbreak in Japan. Briefly, the Ministry of Health, Labour and Welfare in Japan reported 361 confirmed native cases of H1N1 by 1 June 2009, of which, 302 (84%) were less than 20 years-of-age (children) and 59 (16%) were adults (≥20 years of age); Figure 1(b) displays the total epidemic curve. The initial reported outbreak (as of 1 June 2009) of confirmed native cases occurred between 9 and 29 May, with the initial epidemic growth phase occurring during the first 9 days. During the initial epidemic growth phase (days 1–9), indicated by a vertical white line in Figure 1(b), there were 194 reported child cases and 24 reported adult cases. Children and adult cases are classified as <20 years of age and ≥20 years of age, respectively, to be consistent with Nishiura et al.;¹⁶ this differs from the age cut off used for the South Africa analysis. Because there is no contact tracing data available from this outbreak, two samples obtained from other H1N1 outbreaks will be considered to quantify the serial distribution for the analysis of this data. The primary analysis will use the contact tracing sample collected in South Africa (Figure 2); a sensitivity analysis will use contact tracing data from an

Australian outbreak, shown in McBryde et al.³² This contact tracing sample has a slightly larger mean, but has the same maximum length of 6 days.

2.3 Social mixing data

Mossong et al.¹⁰ conducted a population-based survey (POLYMOD) across eight European countries including Belgium, Germany, Finland, Great Britain, Italy, Luxembourg, The Netherlands, and Poland. Participants were asked to describe their physical and nonphysical contacts in diary entries, with 7290 people reporting 97,904 contacts. Summaries of contacts for different age group categories ranging from 0 to 70+, as well as other characteristics, were reported. We aggregate the age groups and estimate that children (<15 years) contact other children 50% of the time while adults (≥ 15 years) contact other adults 75% of the time. As a sensitivity analysis, we include data from a second social mixing survey that was conducted by Johnstone-Roberston et al.¹² in a South African Township. A total of 571 participants completed diary entries about their daily physical and nonphysical contacts and reported a total of 29,125 contacts. We aggregate the contacts in a similar manner and estimate that children contact other children 60% of the time and adults contact other adults 80% of the time.

2.4 Statistical methods

2.4.1 Homogenous estimation of R_0 —We first review two common approaches for estimating homogenous reproductive numbers that can be implemented in an ongoing outbreak and use only basic surveillance data. These methods are the foundations for the methods examined in this article, which estimate heterogeneous reproductive numbers.

The first method, introduced by Wallinga and Teunis,²⁴ requires an estimate of the serial interval and data from a complete epidemic curve. With this information, the authors propose an elegant estimator of the effective reproductive number, given by

$$P_{ij} = \frac{w(t_i - t_j)}{\sum_{i \neq k} w(t_i - t_k)}, \text{ where } p_{ij} \text{ represents the relative likelihood that case } i \text{ has been infected}$$

by case j , given their difference in time of symptom onset is $t_i - t_j$ which is expressed in terms of the distribution of the serial interval (or generation interval), $w(\tau)$. The likelihood that case i has been infected by case j is normalized by the likelihood case i has been infected by any other case k . The effective reproductive number, R_e is then determined by summing the p_{ij} over all cases i . The Walling and Teunis method has also been modified to provide estimates of R_t as an outbreak is unfolding.³³

White and Pagano²³ introduced an approach that also uses the data from an epidemic curve and is designed to estimate the basic reproductive number (and serial interval) during the epidemic phase of the outbreak. By assuming that the number of cases an individual infects is Poisson distributed and that the serial interval can be defined by a multinomial distribution, the following likelihood function can be constructed and maximized to obtain

$$\text{estimates of the reproductive number and serial interval } L(R_0, \mathbf{p}) = \prod_{t=1}^T \frac{e^{-\mu_t} \mu_t^{N_t}}{N_t!}, \text{ where}$$

$$\mu_t = R_0 \sum_{j=1}^{\min(k, t)} N_{t-j} p_j.$$

In this model, p_i describes the probability of a serial interval that is i days long, R_0 is the basic reproductive number, and N_t is the number of new secondary cases at a given time t . Alternatively, if a reliable estimate of the serial interval exists, one can obtain an estimate of

$$R_0 \text{ using the following formula: } \hat{R}_0 = \frac{\sum_{t=1}^T N_t}{\sum_{t=1}^T \sum_{j=1}^{\min(k,t)} p_j N_{t-j}}.$$

2.4.2 Heterogeneous estimation of R_0 —Modifications of both of these approaches have been proposed to account for heterogeneity; we highlight three previously developed methods and introduce our approach. We define R_C and R_A to be the average number of secondary cases a primary child or adult case will infect.

The first two heterogeneous approaches were developed by Glass et al.²² and use the White and Pagano (WP) and Wallinga and Teunis (WT) methods^{23,24} to estimate reproductive numbers for children and adults by constraining the NGM to take the following forms: separable, high child-to-child, contact frequency, and proportional mixing. We focus on the contact frequency matrix, which is calibrated with contact rates from social mixing data. The

contact frequency matrix form is $\begin{bmatrix} \rho_C a^2 & (1 - \rho_A)ab \\ (1 - \rho_C)ab & \rho_A b^2 \end{bmatrix}$, where ρ_A and ρ_C are the fraction of

adult contacts that are with other adults and the fraction of child contacts that are with other children, respectively. The contact rates are calibrated using data from the POLYMOD study, and a and b are derived from the likelihood functions (WP or WT). The WP-based approach (Glass WP) stratifies the WP likelihood function by the age groups, such that C_t is the number of child cases on day t and A_t is the number of adult cases on day t . Glass further

defines $\hat{C} = \sum_{t=1}^T C_t$, $\hat{A} = \sum_{t=1}^T A_t$, $\tilde{C} = \sum_{t=1}^T \sum_{j=1}^{\min(k,t)} C_{t-j} w_j$, and

$\tilde{A} = \sum_{t=1}^T \sum_{j=1}^{\min(k,t)} A_{t-j} w_j$, where w_j represents the serial interval. Implementing the

NGM constraints and using the contact frequency estimator, the reproductive numbers are

estimated by $R_C = \rho_C a^2 + (1 - \rho_C)ab$ and $R_A = (1 - \rho_A)ab + \rho_A b^2$, where $b = \frac{\hat{C} - \rho_C a^2 \tilde{C}}{(1 - \rho_A)a\tilde{A}}$ and

a is the solution of

$$\rho_C \tilde{C}^2 (\rho_A + \rho_C - 1)a^4 + [(1 - \rho_C - \rho_A - \rho_A \rho_C)\tilde{C}\hat{C} - (1 - \rho_A)^2 \tilde{A}\hat{A}]a^2 + \tilde{C}^2 \rho_A = 0.$$

A similar approach is taken for the Glass WT models, which are written in terms of the overall homogenous reproductive number (R) and the value f , the fraction of total cases that are children, which is calculated for each day of the outbreak and is set as 0.30 for days without child cases. The reproductive numbers are then determined by $R_C = \rho_C a^2 + (1 - \rho_C)ab$ and $R_A = (1 - \rho_A)ab + \rho_A b^2$, where $q_C = \rho_C a^2$, $q_A = \rho_A b^2$,

$q_A = 1 - (1 - q_C) \frac{f^2(1 - \rho_C)}{(1 - f)^2(1 - \rho_A)}$, and q_C is the solution of

$$(\rho_A + \rho_C - 1)q_C^2 + (1 - \rho_C - \rho_A - \rho_A \rho_C - \frac{(1 - \rho_A)^2(1 - f)^2}{f^2})q_C + \rho_A \rho_C = 0.$$

The third approach we consider is based on work by White et al.,²⁶ which modifies the Wallinga and Teunis model²⁴ (White WT) to estimate multiple reproductive numbers using social mixing survey data. In their article, the probability that an infection occurred between two cases is adjusted to include an additional component $w_{g_j g_i}$, which is the likelihood of transmission between two individuals in age group g_i and g_j . Then, the probability of an infectious contact between two individuals, i and j , is $P(t'_j \rightarrow t_i) = P_{[t_i - t'_j]} w_{g_j g_i}$, where t_i denotes the i th individual with symptom onset on day t , t'_j denotes the j th individual with symptom onset on day t' , and $P_{[t_i - t'_j]}$ is the probability of infection between two individuals on days t_i and t'_j using the serial interval only. The estimator for R_t is then defined as $R_{t'_j} = \sum_{s=t'+1}^{\min(T, t'+k)} n_s q_{s, t'_j}$, where n_s denotes the number with symptom onset on day s and $q_{s, t}$ denotes the relative probability that case s was infected by case t . The $w_{g_j g_i}$ are determined from contact rates from the POLYMOD social mixing survey.

The novel method, we introduce in this article (Moser WP), utilizes a Bayesian framework to estimate group-specific reproductive numbers and extends a previous modification to the White and Pagano model.³⁴ This approach not only allows estimates for the reproductive numbers, but also the serial interval, and provides credible intervals for each estimated parameter. This method assumes that the population is stratified into two groups, adults and children; because children and adults infect each other, the group reproductive numbers can be viewed as a combination of reproductive numbers as follows: $R_C = R_{CC} + R_{CA}$ and $R_A = R_{AA} + R_{AC}$. For example, R_{CA} is the average number of secondary adult cases a child will infect. This formulation is similar to the Glass formulation, in which they define a reproductive number matrix for two types of individuals.

The White and Pagano likelihood function is modified to allow for estimation of multiple

reproductive numbers, $L(\mathbf{R}_0, \mathbf{p}) = \prod_{t=1}^T \prod_{g=1}^2 \frac{e^{-\mu_{tg}} \mu_{tg}^{N_{tg}}}{N_{tg}!}$, where

$\mu_{tg} = \sum_{h=1}^2 \mathbf{R}_{hg} \sum_{j=1}^{\min(k, t)} N_{t-j, h} p_j$ and g indexes the age groups (assumed here to be just adults and children, but expandable to more groups).

N_{tC} and N_{tA} represent the number of new child and adult cases, respectively, at given time t . The $\mathbf{p}$ represents the SI distribution and is assumed to be multinomial. In this model, we assume the same SI for children and adults, although this assumption could be relaxed by further modifying the likelihood.

In order to include the social mixing survey data in our model, we reparametrize the reproductive numbers to be functions of the proportion of contacts between and within age groups: $R_{CC} = \rho_{CC} R_C$, $R_{AC} = \rho_{AC} R_A$, $R_{CA} = \rho_{CA} R_C$, and $R_{AA} = \rho_{AA} R_A$, where, for example, ρ_{CA} is the proportion of contacts children have with adults.

Each parameter in this model is assigned a prior distribution. The SI prior distribution can be set to be a constant, parameterized with a uniform prior, or can be informed with contact tracing data.³⁴ Contact tracing samples were collected during the H1N1 outbreak in South Africa (see Figure 2) and can be used to inform the SI prior; in this case we use a Dirichlet distribution with $\alpha = (9, 5, 5, 7, 2, 1)$. As a sensitivity analysis, a second SI distribution was considered for the analysis of the Japan data, obtained from McBryde et al.,³² with the SI prior following a similar Dirichlet distribution with $\alpha = (5, 10, 9, 7, 4, 1)$. R_C and R_A each are set to have log-normal prior distributions, such that $\log(R_C)$ and $\log(R_A)$ are normally distributed with mean = 0 and variance = 1000, which are noninformative prior distributions. Because the Japan outbreak has a limited number of cases, the reproductive number prior distributions are adjusted to be slightly less noninformative with mean = 0.10 and variance = 5, but still have a large range. The contact proportions, ρ_C and ρ_A , are given Dirichlet priors with hyperparameters set to be the average number of contacts between and within age groups obtained from the social mixing data; although the hyperparameters could be adjusted accordingly.

In this analysis, the contact rates for adults and children are obtained from the POLYMOD social mixing data by aggregating the contacts across age groups: 50% of child contacts are with other children and 75% of adult contacts are with other adults. The contact rates are used to inform the prior distributions such that the total prior counts are equal to ~1% of the total outbreak size and are proportional to the contact rates.

3 Results

The South Africa and Japan H1N1 data were analyzed using the four methods outlined above. Basic reproductive numbers were estimated for the White and Pagano based methods (Moser WP and Glass WP) and daily effective reproductive numbers were estimated for the Wallinga and Teunis based models (Glass WT and White WT). Daily effective reproductive numbers were averaged over the initial epidemic growth phase of each outbreak (days 19–53 for South Africa and days 1–9 for Japan), to be comparable to the basic reproductive numbers.

Estimates of the age-specific reproductive numbers for each of the four methods are displayed in Figure 3 and Table 2. In South Africa, all estimates exceed the critical threshold of one and are consistent with previously published estimates for this epidemic.^{26,28,34} The WP-based methods estimated $R_C > R_A$, indicating that children, on average, infected more people than adults (Glass WP: $R_C = 1.6$, $R_A = 1.4$; Moser WP: $R_C = 1.5$, $R_A = 1.3$). Adults still contributed to the growth of the outbreak with super-critical reproductive numbers, but to a lesser degree than children. The WT-based methods produced estimates for adults and children that are likely within the bounds of estimation error, with adults producing more secondary cases than children on average ($R_C = 1.3$, $R_A = 1.4$).

The results from the Japan analysis were similar to those previously reported;^{16,22} children were the driving force of the outbreak with $R_C = 2.1$ – 2.9 . For all methods, the reproductive number estimates for adults were below 1. Although the estimates of the reproductive numbers resulted in similar conclusions, regardless of analytic method, slight difference

were still observed depending upon the underlying method. WP-based methods again had higher estimates for R_C compared to the WT-based methods, but an inconsistent pattern for R_A .

Unlike the other approaches, the Moser WP approach provides an estimate of the 95% credible interval, as well as estimates for the SI. The mean and 95% CI for the mean of the serial interval are 2.17 (1.78–2.62) days in South Africa, which are comparable to other reported estimates from this outbreak.^{28,29,34} In Japan, a slightly higher mean SI estimate was obtained, 2.78 (2.26–3.33) days.

Daily effective reproductive number estimates tended to exhibit substantial variation over the first 2 weeks in the South Africa outbreak (Figure 4(a)). After this initial timeframe, the White WT estimates for children and adults converged and had little variability over time; however, the Glass WT estimates continued to show substantial stochasticity. The results for the Japan outbreak are plotted in a similar fashion (Figure 4(b)), but lack estimates for some days due to limited adult cases. The Glass WT and White WT methods do not estimate reproductive numbers on days without adult or child cases. We observe that the estimates for child reproductive numbers are initially high and quickly decline to less than the critical value 1 after the initial epidemic phase. The White WT method initially estimates adult reproductive numbers to be greater than 1, which explains the slightly higher estimate for R_A for this method.

To further examine the observed differences between the WT- and WP-based South Africa results, a moving-average plot of the effective reproductive numbers is shown in Figure 5. For adults, both WT-based methods have similar moving-average estimates over the course of the outbreak, and were similar to the WP-based estimates at around day 60, which is within a few days of the initial epidemic growth cut point. For children, the Glass WT estimates were almost always larger than those from White WT, and both were consistently lower than the WP-based estimates.

We repeated the analysis using the Johnstone-Robertson social mixing data as a sensitivity analysis and found that although the contact matrix was more assortative, the results were virtually identical. In addition, we considered using the SI distribution from McBryde et al.³² for the Japan analysis. Consistent with other findings using this SI distribution, the estimates for the reproductive numbers were similar, but slightly higher than the results obtained using the South African SI distribution, which is attributed to using a SI distribution with a larger mean.²²

4 Discussion

In this article, we introduce a novel and flexible method for estimating heterogeneous reproductive numbers that is based on the White and Pagano estimator of the reproductive number.²³ We compare this method to three other methods that use the White and Pagano or Wallinga and Teunis likelihoods and estimate child and adult reproductive numbers for the 2009 H1N1 outbreaks in South Africa and Japan. All methods resulted in reasonable estimates for the reproductive numbers in South Africa (estimates ranging from 1.2 to 1.5);

however, the results varied depending on the underlying transmission model. The WP-based estimates for adults were consistently lower than those for children. The WT-based estimates for adults and children were very similar during the initial epidemic growth phase, with slightly larger estimates for adults. This was consistent with the findings in White et al.,²⁶ which saw that those aged 15–24 had the largest reproductive numbers.

All methods similarly estimated high reproductive numbers for children in Japan (estimates ranging from 2.1 to 2.9) and all estimated adult reproductive numbers to be less than 1, which was also consistent with previously reported findings that children were the drivers of this outbreak.^{16,21,22} Differences in estimates were observed between the approaches depending upon the underlying transmission model, with WP-based methods estimating higher child reproductive numbers than the WT-based methods.

As a sensitivity analysis, we also considered the proportional matrix from Glass et al., which is another asymmetric matrix type and is recommended when contact patterns are not symmetric between age groups.^{22,35} This matrix is calibrated with the proportion of children in the population (estimated to be 25% for this analysis). When we analyzed the South Africa outbreak using the proportional matrix with Glass WT the estimates were closer to the WP-based methods ($R_C = 1.56$ and $R_A = 1.28$). In contrast, when analyzing the Japan data using a proportional matrix the estimates were similar ($R_C = 2.24$ and $R_A = 0.24$). In fact, the results for Japan were the essentially same regardless of matrix choice (separable: $R_C = 2.22$ and $R_A = 0.27$; high child-to-child: $R_C = 2.20$ and $R_A = 0.36$), which was consistent with what Glass et al. noted. We speculate that this is likely due to the large number of child cases.

The Wallinga and Teunis model can be sensitive to the underlying population dynamics. In South Africa, the near equivalence of adult and child reproductive numbers could be attributable to this, since adults comprise the majority of reported cases (62%) and could explain some of the observed differences between the estimates. Including additional age-group categories that can further distinguish younger and older children and younger and older adults may provide further insights since simply dichotomizing the age groups may not be sufficient. Our proposed method allows for this extension.

Our method carries some of the same limitations as the existing comparable methods. Specifically, if the reporting rate is different for children and adults then all methods will produce biased results.^{22,26,36} We also assume that the contact rates from the social mixing data are appropriate surrogates for the transmission probability, which is an untested hypothesis. In the analysis of the Japan H1N1 data, the prior distributions for the reproductive numbers were modified to be slightly less noninformative, which is sometimes necessary with small outbreaks.³⁴ The homogenous Bayesian model, which this approach is derived from, can have issues estimating reproductive numbers when the outbreak is small, especially relative to the maximum SI length.³⁴ In this outbreak, the initial epidemic growth phase is the first 9 days (approximately 3–4 serial interval lengths) of the outbreak and during which only 24 adult cases were reported (~11% of total cases during this time period). Adjusting the prior distribution to slightly constrain reproductive numbers estimates

within a reasonable range of values offers a reasonable solution for small and/or ongoing outbreaks.

In the results that we have shown, we limit our analysis to two groups: adults and children. We note that this might be too crude of a stratification of the data and suggest that when possible finer stratification of age be incorporated, as shown in the methods section. This is possible with the White WT method and the method that we have introduced. However, the Glass et al. estimators currently only allow for two groups. Extension to further age groups requires adequate social mixing data for the age groups, such as is provided by the Mossong et al. surveys.¹⁰ The computation time also increases with the number of groups.

We have introduced a novel and flexible method for estimating age-specific reproductive numbers. We show that the results that we obtain are comparable to other methods with similar assumptions, which supports the validity of this novel approach. Our method allows for estimates of variability of the reproductive numbers and also allows us to use contact tracing data as prior information to obtain an estimate of the serial interval, which are not features of currently available methods. We can also extend this method to incorporate multiple groups with a simple modification to the likelihood function. Our approach is more computationally intensive than those proposed by Glass et al.,²² but requires far fewer a priori assumptions and imputation of missing data. As the tools exist to perform these types of analyses are improving, we advocate for their routine implementation in the analysis of future and ongoing outbreaks of infectious disease.

Acknowledgments

The authors thank Brett Archer for providing the South African H1N1 outbreak data and Helen Jenkins for helpful feedback on the article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institute of General Medical Sciences (grant number U545GM088558).

References

1. Melegaro A, Jit M, Gay N, et al. What types of contacts are important for the spread of infections? Using contact survey data to explore European mixing patterns. *Epidemics*. 2011; 3:143–151. [PubMed: 22094337]
2. Hens N, Ayele GM, Goeyvaerts N, et al. Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. *BMC Infect Dis*. 2009; 9:187. [PubMed: 19943919]
3. Farrington CP, Unkel S, Anaya-Izquierdo K. Estimation of basic reproduction numbers: individual heterogeneity and robustness to perturbation of the contact function. *Biostatistics*. 2013; 14:528–540. [PubMed: 23266419]
4. Farrington CP, Whitaker HJ, Unkel S, et al. Correlated infections: quantifying individual heterogeneity in the spread of infectious diseases. *Am J Epidemiol*. 2013; 177:474–486. [PubMed: 23403987]
5. Eames KT, Tilston NL, Brooks-Pollock E, et al. Measured dynamic social contact patterns explain the spread of H1N1v influenza. *PLoS Comput Biol*. 2012; 8:e1002425. [PubMed: 22412366]

6. Goeyvaerts N, Hens N, Ogunjimi B, et al. Estimating infectious disease parameters from data on social contacts and serological status. *J R Stat Soc Ser C (Appl Stat)*. 2010; 59:255–277.
7. Apolloni A, Poletto C, Colizza V. Age-specific contacts and travel patterns in the spatial spread of 2009 H1N1 influenza pandemic. *BMC Infect Dis*. 2013; 13:176. [PubMed: 23587010]
8. Hens N, Calatayud L, Kurkela S, et al. Robust reconstruction and analysis of outbreak data: influenza A (H1N1) v transmission in a school-based population. *Am J Epidemiol*. 2012; 176:196–203. [PubMed: 22791742]
9. 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; 164:936–944. [PubMed: 16968863]
10. Mossong J, Hens N, Jit M, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*. 2008; 5:e74. [PubMed: 18366252]
11. Hens N, Goeyvaerts N, Aerts M, et al. Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infect Dis*. 2009; 9:5. [PubMed: 19154612]
12. Johnstone-Robertson SP, Mark D, Morrow C, et al. Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. *Am J Epidemiol*. 2011; 174:1246–1255. [PubMed: 22071585]
13. Chowell G, Viboud C, Munayco CV, et al. Spatial and temporal characteristics of the 2009 A/ H1N1 influenza pandemic in Peru. *PLoS One*. 2011; 6:e21287. [PubMed: 21712984]
14. Gambhir M, Sverdlow DL, Finelli L, et al. Multiple contributory factors to the age distribution of disease cases: a modeling study in the context of influenza A(H3N2v). *Clin Infect Dis*. 2013; 57(Suppl 1):S23–S27. [PubMed: 23794728]
15. Eames KT, Tilston NL, Edmunds WJ. The impact of school holidays on the social mixing patterns of school children. *Epidemics*. 2011; 3:103–108. [PubMed: 21624781]
16. Nishiura H, Castillo-Chavez C, Safan M, et al. Transmission potential of the new influenza A (H1N1) virus and its age-specificity in Japan. *Euro Surveill*. 2009; 14:19227. [PubMed: 19497256]
17. Nishiura H, Ejima K, Mizumoto K, et al. Cost-effective length and timing of school closure during an influenza pandemic depend on the severity. *Theor Biol Med Model*. 2014; 11:5. [PubMed: 24447310]
18. Paine S, Mercer G, Kelly P, et al. Transmissibility of 2009 pandemic influenza A (H1N1) in New Zealand: effective reproduction number and influence of age, ethnicity and importations. *Euro Surveill*. 2010; 15 pii: 19591.
19. Schanzer D, Vachon J, Pelletier L. Age-specific differences in influenza A epidemic curves: do children drive the spread of influenza epidemics? *Am J Epidemiol*. 2011; 174:109–117. [PubMed: 21602300]
20. Worby CJ, Chaves SS, Wallinga J, et al. On the relative role of different age groups in influenza epidemics. *Epidemics*. 2015; 13:10–16. [PubMed: 26097505]
21. Nishiura H, Chowell G, Safan M, et al. Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009. *Theor Biol Med Model*. 2010; 7:1. [PubMed: 20056004]
22. Glass K, Mercer G, Nishiura H, et al. Estimating reproduction numbers for adults and children from case data. *J R Soc Interface*. 2011
23. White LF, Pagano M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Stat Med*. 2008; 27:2999–3016. [PubMed: 18058829]
24. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol*. 2004; 160:509–516. [PubMed: 15353409]
25. Nishiura H, Chowell G, Heesterbeek H, et al. The ideal reporting interval for an epidemic to objectively interpret the epidemiological time course. *J R Soc Interface/R Soc*. 2010; 7:297–307.
26. White LF, Archer B, Pagano M. Determining the dynamics of influenza transmission by age. *Emerging Themes Epidemiol*. 2014; 11:4.
27. Archer BN, Tempia S, White LF, et al. Reproductive number and serial interval of the First Wave of Influenza A(H1N1)pdm09 Virus in South Africa. *PLoS ONE*. 2012; 7:e49482.doi: 10.1371/journal.pone.0049482 [PubMed: 23166682]

28. Archer BN, Tempia S, White LF, et al. Reproductive number and serial interval of the first wave of influenza A (H1N1) pdm09 virus in South Africa. 2012
29. Archer BN, Timothy GA, Cohen C, et al. Introduction of 2009 pandemic influenza A virus subtype H1N1 into South Africa: clinical presentation, epidemiology, and transmissibility of the first 100 cases. *J Infect Dis.* 2012; 206:S148–S153. [PubMed: 23169962]
30. Anderson, BA., Phillips, H. Adult mortality (age 15–64): based on death notification data in South Africa, 1997–2004. South Africa: Statistics South Africa; 2006.
31. Hosegood V, Vanneste A-M, Timæus IM. Levels and causes of adult mortality in rural South Africa: the impact of AIDS. *AIDS.* 2004; 18:663–671. [PubMed: 15090772]
32. McBryde E, Bergeri I, van Gemert C, et al. Early transmission characteristics of influenza A (H1N1) v in Australia: Victorian state, 16 May–3 June 2009. *Euro Surveill.* 2008; 14:19356.
33. Cauchemez S, Boëlle P-Y, Donnelly CA, et al. Real-time estimates in early detection of SARS. *Emerg Infect Dis.* 2006; 12:110. [PubMed: 16494726]
34. Moser CB, Gupta M, Archer BN, et al. The impact of prior information on estimates of disease transmissibility using Bayesian tools. *PLoS One.* 2015; 10:e0118762.doi: 10.1371/journal.pone.0118762 [PubMed: 25793993]
35. Glass K, Kelly H, Mercer GN. Pandemic influenza H1N1: reconciling serosurvey data with estimates of the reproduction number. *Epidemiology.* 2012; 23:86–94. [PubMed: 22089631]
36. White LF, Pagano M. Reporting errors in infectious disease outbreaks, with an application to Pandemic Influenza A/H1N1. *Epidemiologic Perspectives & Innovations.* 2010; 7:12. [PubMed: 21159178]

Figure 1.
Epidemic curves for children and adults in a) South Africa and b) Japan.

Figure 2.
Contact tracing sample distribution from South Africa.

Figure 3.

Estimates of reproductive numbers in a) South Africa and b) Japan for children and adults for each method (M-WP: Moser WP, G-WP: Glass WP, G-WT: Glass WT, W-WT: White WT).

Figure 4.
Daily effective reproductive number estimates for WT-based methods in a) South Africa and
b) Japan.

Figure 5.
Moving average daily effective reproductive numbers compared to basic reproductive numbers in South Africa.

Table 1

Age distribution for children and adults in South Africa.

Days	N (%)	Mean (SD)	Median (range)
Children			
19–53	901 (38)	9.8 (3.6)	10.8 (0–14.9)
19–109	5650 (46)	8.2 (4.1)	8.7 (0–14.9)
Adults			
19–53	1486 (62)	24.7 (11.2)	20 (15–80)
19–109	6690 (54)	28.5 (12.6)	24.6 (15–90)

Table 2

Reproductive number estimates for children and adults in the 2009 H1N1 influenza outbreaks in South Africa and Japan.

Methods	South Africa		Japan	
	R_C (95% CI)	R_A (95% CI)	R_C (95% CI)	R_A (95% CI)
Moser WP	1.54 (1.12–1.97)	1.28 (1.02–1.56)	2.86 (2.21–3.65)	0.34 (0.24–0.57)
Glass WP	1.63	1.41	2.87	0.21
Glass WT	1.36	1.40	2.10	0.27
White WT	1.25	1.42	2.15	0.79

CI: credible interval.