Supplement to “Assessing the burden of Congenital Rubella Syndrome in China and evaluating mitigation strategies: a meta-population modeling study”

Qiru Su PhD,1,2 Professor Zhilan Feng PhD,3 Lixin Hao PhD,1 Chao Ma PhD,1 José E Hagan MD,4,5 Gavin B Grant MD,4 Ning Wen MD,1 Chunxiang Fan MD,1 Hong Yang MD,1 Lance E Rodewald MD,1,4,6 Huaqing Wang PhD,1,† John W Glasser PhD4

1Chinese Center for Disease Control and Prevention, 27 Nanwei Road, Xicheng District, Beijing 100050 PRC

2Institute of Pediatrics, Shenzhen Children's Hospital, 7019 Yitian Road, Shenzhen, Guangdong 518038 PRC

3Purdue University, 150 North University Street, West Lafayette, Indiana 47907 USA

4Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30333 USA

5Expanded Programme on Immunization, World Health Organization Regional Office for the Western Pacific, United Nations Avenue, Ermita, Manila 1000 Philippines

6Office of the World Health Organization Representative in China, 23 Dongzhimenwai Dajie, Chaoyang District, Beijing 100600 PRC

†Corresponding author:

Huaqing Wang PhD

National Immunization Program

Chinese Center for Disease Control and Prevention

27 Nanwei Road, Xicheng District

Beijing 100050 PRC

Voice: +86 10 63171724; facsimile: +86 10 83159534

hqwang@vip.sina.com

Documentation of Mathematica procedures: <https://reference.wolfram.com/language/>

Table of Contents

|  |  |  |
| --- | --- | --- |
| Section | Description | Pages |
| I | Overview of the Models and Parameter Sources  | 3-6 |
| II | Estimation of the Probabilities of Infection upon Contact with an Infectious Person  | 7-9 |
| III | The Next-Generation Approach and Age-Specific Results  | 9-10 |
| IV | Provincial Reproduction Numbers and their Ratios  | 10-12 |
| V | Estimation of Vaccination Rates | 12 |
| VI | Forcing the Regionally Stratified Simulation Model and Comparing its Predictions with Observations  | 12-14 |
| VII | Surveillance-Based Estimates of Rubella and CRS Incidence  | 14-16 |
| VIII | Estimation of Rubella and CRS Incidence via Demographically Realistic Modeling  | 16-19 |
| IX | Estimation of Reporting Efficiency  | 19-20 |
| X | Comments on the Methods Employed in a Recent Assessment of the Global Burden of CRS  | 20-21 |
| XI | Immune Profiles and Gradient by Province  | 21-32 |
| XII | References  | 32-33 |

**I. Overview of the Models and Parameter Sources**

Because the best model for any given purpose is as simple as possible (to facilitate evaluation), but not more so (as simplistic models can be misleading), we report analytical and simulation results from two age structured SEIR models of rubella in China. One is also spatially structured, and we analyze and simulate different versions of that model. Initial conditions and, insofar as possible, parameter values are the same (table S1). With one exception, we estimated them from observations in China. Here we describe calculations mentioned in the main text and figures, explain how we obtained the requisite information, or present results in greater detail.

We use the age- and location-stratified model (table S2, rows I and II) to 1) calculate the provincial and meta-population reproduction numbers (via the next-generation matrix); 2) identify the optimal strategy for accelerating rubella elimination via the gradient (multivariate partial derivative) of the 2014 effective reproduction number with respect to possible supplemental immunization rates; and 3) deduce times-to-elimination as a function of uptake among susceptible adolescents and young adults. The first two of these are analytical (static) and third is a simulation (dynamic) result. Nonetheless, they are consistent.

In section II, we describe how we obtain the age-specific probabilities of infection on contact used in all three models from published results of a pre-vaccination cross-sectional serological survey in China and study of face-to-face conversations in Europe. We explain why we do not use a more recent (post-vaccination) serological survey (whose results are illustrated by region in figure 1) for this purpose, but do use it for the initial conditions (table S1). We also explain why we use those contacts, rather than ones from a more recently published study in southern China.

In section III, we evaluate the basic reproduction number, age-specific contributions and prevalence for an age-stratified model (like that described in row III of table S2, but without demographic details) derived via the next-generation matrix approach by Glasser et al. (2010). The model with demographic details is analyzed by Feng et al. 2020. In our age- and location-stratified modeling (table S2, rows I and II), contacts decline exponentially with inter-location distances at age-specific rates. We ask readers interested in the derivation of expressions used in this calculation, for the reproduction numbers or gradient to consult Feng et al. (2017). For the rates that we use in our modeling of vaccine-preventable diseases in China, please consult Feng and Glasser (2019) or the supplement to Hao et al. (2019).

In section IV, we tabulate the provincial reproduction numbers derived from the model described in table S2, row I, and their ratios (illustrated in figure 3). Provincial stratification is not needed to evaluate the impact of supplemental immunization activities (SIAs) on time-to-elimination, a country-level phenomenon defined by the World Health Organization as absence of endemic disease for a one-year period (often approximated by reported annual incidence less than 1 per 106 people). Because China’s population is concentrated in the eastern and, to a lesser extent, central provinces, however, we a) stratify the version of our age- and location-specific model that we simulate (by evaluating the solution obtained via Mathematica’s NDSolve procedure) by region (table S2, row II).

In section V, we present the temporal relationship between rubella- and measles-containing vaccine uptake that enables us to estimate rubella vaccination rates from measles by location (Hao et al. 2019) for the effective reproduction number (figures 3) and gradient (figures 4) calculations and simulation modeling. Historical vaccination rates also enable us to estimate forces or hazard rates of infection from the 2014 serological survey, from which we could calculate contemporary probabilities of infection on contact with an infectious person (section I). However, as the difference is small and pre-vaccination probabilities (figure S3) are best suited for our analyses, we believe that the virtue of having a single set of parameters in the work reported here exceeds any possible inaccuracy in simulations due to our use of probabilities calculated from historical versus contemporary forces of infection.

Because reported incidence is seasonal, but systems of ordinary differential equations exhibit damped oscillations, we b) force the contact rates via a harmonic function (described in section VI) whose several coefficients we estimate within regions and age classes from surveillance reports (via Mathematica’s LinearModelFit procedure). While neither under- nor age-biased reporting (documented in section IX) should affect temporal patterns within those sub-populations, it precludes fitting the entire meta-population model to surveillance reports by adjusting any parameter(s). Rather, we c) scale the regional contact rates (i.e., adjust their magnitude without changing their age-distributions) so that predictions resemble observations summed over all ages (rightmost column of figure S7).

In section VII, we present surveillance-based estimates of rubella and CRS incidence both to illustrate these calculations, which are the same as the model-based ones with predictions replacing reports, but also because we compare them in the main text.

We also simulate an age-structured model that is not spatially stratified, but includes births, deaths, and ageing (table S2, row III), to 4) estimate the actual (versus reported) incidence of rubella in a relatively small but representative population, from which we 5) assess the accuracy of a) rubella reporting rates and b) CRS incidence. Stochastic simulations (via the discrete event/time approach of Renshaw 1991 programmed in C++ by Denis Taneri) also enable us to 6) estimate effective reproduction numbers for comparison with that deduced from our age- and province-stratified model. Again, despite those very different approaches, results are consistent. In section VIII, we describe the demographic parameters, simulation procedure and results in much greater detail than possible in the main text.

In section IX, we compare age-specific reported and simulated rubella infection rates and suggest possible explanations for the discrepancies. In section X, we comment on a study of the global burden of CRS whose results for China differ from ours. Finally, in section XI, we present the 2014 immune profiles and gradient by province-level jurisdiction (mainland China comprises 22 provinces, five municipalities and four autonomous administrative regions), results that are simply too numerous for the main text. Section XII contains the references to citations in this supplement.

Table S1. Common characteristics of the models.†

|  |  |  |
| --- | --- | --- |
| Characteristic | Details | Source(s) |
| Epidemiological structure  | Susceptible, Exposed (infected, but not yet infectious), Infectious, and Removed (by virtue of recovery from infection or immunization)  | 18-day latent and 5-day infectious periods are consistent with clinical observations (Heymann 2014) and yield ℜ0 in the range reported, respectively; immunity is lifelong, and mortality is the reciprocal of longevity, assumed to be 365.25×75 days.  |
| Age structure | <1, 1-4, 5-9, ..., 65+ years  | Abridged standard in demography |
| Forces of infection  | Calculated from a pre-vaccination cross-sectional serological survey (figure S1a)  | Su 1985  |
| Probabilities of infection on contact with an infectious person  | Calculated from forces of infection and contact rates (figure S3)  | Information about face-to-face conversations in Europe that Mossong et al. 2008 summarize  |
| Initial conditions  | Proportions susceptible and immune by age  | 2014 cross-sectional serological survey (Su et al. manuscript)  |
| Routine vaccination  | Provincial measles rates, relationship between rubella- and measles-containing vaccine uptake by dose (figure S5a)  | Hao et al. 2019  |

Table S2. Unique characteristics of the models.†

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Additional Features | Mixing | Approach | Results |
|  | Thirty-one locations (22 mainland provinces, five municipalities and four autonomous administrative regions)  | Empirical by age (single-year-of-age PolyMod observations aggregated into modeled classes) except for the gradient calculation (right), preferential by location (as described immediately below) | Obtained ℜ0 and ℜE via the next-generation matrix approach. Evaluated expressions derived by Feng et al. (2017) for ℜE and ∇ℜE (definitions) with respect to possible supplemental immunization rates. Those expressions assume proportional age-specific mixing  | Figures 2 and 3  |
|  | Three locations (eastern, central and western province-level jurisdictions), seasonal forcing, regional contact rates scaled to approximate surveillance reports  | Empirical by age, preferential by location (contacts decrease exponentially with inter-location distance at age-dependent rates) | Using infections during January 2018 (obtained as described in section VI), simulated (deterministically) SIAs among adolescents or young adults  | Figures 5 and 6 |
|  | Ageing, deaths by age, births by age of mother, both at 2014 rates, an otherwise unstratified population of 105 with the 2014 age- and gender-distributions  | Empirical (single-year-of-age PolyMod observations aggregated into modeled classes)  | Beginning with single newly infectious individuals aged <1, 1-4, 5-9, ..., 40-44 years, simulated (stochastically) until outbreaks ceased, calculated age-specific rubella and CRS incidence (section VIII) and under-reporting (section IX)  | Figures 4 |

†Section *x*, *x* = I, ..., X refers to this supplement. Figure and table numbers NOT preceded by the letter S are in the main text; others are here.

**II. Estimation of the Probabilities of Infection upon Contact with an Infectious Person**

In this section, we derive the age-specific probabilities of infection on contact with an infectious person that, together with rates of contact between members of the various age groups alone or also by spatial location, enable us to model rubella transmission. We begin by describing how we estimate the forces or hazard rates of infection and attack rates or proportions infectious. Feng and Glasser (2019, 2020) cover these topics in greater detail.

First, we define the hazard rate or force of infection among susceptible people in age group *i* as:



where *ai* is the *per capita* contact rate, *βi* is the probability of infection on contact with an infectious person, *cij* is the proportion of their contacts that members of group *i* have with those of group *j*, and *Ij* and *Nj* are the numbers infectious and total (i.e., sum over all epidemiological classes) in group *j*, respectively, whereupon their ratio is the probability that a randomly-contacted member of group *j* is infectious.

In transmission models, *λi* varies with time because *Ij* does. In our demographically realistic model (table S2, row 3; section VII), *Nj* also varies with time. Similarly, the number of new infections in any group is the product of the force of infection and number of susceptible people, both of which are dynamic. But, for present purposes, we assume that *λ*, *I* and *N* are at equilibrium.

Histories of infection and contact rates by age at any particular time enable us to estimate the unknown probabilities of infection on contact with an infectious person, *βi* (figure S3). If immunity following recovery from infection is lifelong, the probability of remaining susceptible at age *α* is:



where *λ*(*u*) is the force or hazard rate of infection at age *u*. One can estimate the *λ*(*u*) by fitting the cumulative probability of infection at age *α*, *PI*(*α*) = 1 – *PS*(*α*) to histories of infection, cross-sectional (i.e., including all ages) serological surveys (figure S1a) for example.

 

Figures S1. The a) probability of infection by age fitted (via Mathematica’s FindFit procedure) to proportions of 16 658 residents of 20 Chinese provinces with antibodies to rubella virus during the period 1979-80 (Su 1985) and b) estimated hazard rates or forces of rubella infection (the estimated coefficients are: *x* = 0.108, *y* = 0.25, and *z* = – 0.15) assuming that only 90% were symptomatic (and infectious).

We assume that the forces of infection have the functional form described by Farrington (1990), a linear increase followed by exponential decrease,  whereupon fitting involves estimating the coefficients *x*, *y*, and *z*. Given the forces or hazard rates of infection at ages *u* (figure S1b), one can calculate the attack rates  (figures S2). 

 

Figures S2. Age-specific a) attack rates and b) proportions infectious, obtained by averaging the continuous function on the left over the intervals illustrated on the right.

Empirical contact matrices have been described from proxies such as face-to-face conversations (Mossong et al. 2008) or periods sharing spaces (Zagheni 2008). Observed mean *per capita* contact rates *Cij*, quotients of the contacts by study participants aged *i* with people aged *j* during a period and numbers of such participants, are typically summed over all groups *j* to yield *ai* and then the *cij* are calculated by dividing the *Cij* by *ai*.

While contacts may be recorded by single year of age, they may be aggregated for publication into 5-year or larger groups. Read et al. (2014) recorded contacts from their study of mixing in southern China in age classes 0-4, 5-19, 20-64, and 65+ years, which are too broad for accurate transmission modeling (Hao et al. 2019, supplement). However, their contact numbers and durations by participant age are comparable to those from the PolyMod study (Read et al. 2014, figure S4). The synthetic Chinese matrix of Prem et al. (2017) resembles a composite of the 8 European countries involved in PolyMod, but we prefer the age-structure of abridged life tables (figure S2b), a standard in demography, because infants differ from other young children. Consequently, in our modeling of vaccine-preventable diseases in China, we aggregate the European observations, which we have by single year of age courtesy of Professor John Edmunds (London School of Hygiene and Tropical Medicine), into age classes 0, 1-4, 5-9, ..., 65+ years.

Given the *λi* and *Ij*/*Nj* calculated from the serological observations of Su (1985), illustrated in figures S1b and S2, which assume that 10% of survey participants with antibodies were asymptomatic (and not infectious), a conservative assumption (20-50% in the main text), and *ai* and *cij* from the PolyMod study, we solved the *n* = 15 equations for *λi* simultaneously for the unknown *βi* that are illustrated in figure S3.



Figure S3. Probabilities of infection on contact with an infectious person. The decline with age slows during adolescence, when about 20% of contacts result in infection.

Forces of infection can be derived from post-vaccination serological surveys if reliable historical vaccination records are available (Feng and Glasser 2020). While there are no records for the decade between licensure of an RCV and its inclusion in China’s EPI, we can estimate the unknown proportions immunized (section V, figure S5). Differences between the pre- and post-vaccination forces of infection are interesting, but using probabilities derived from one in our analyses and other in our simulations might lead to inconsistencies. So, we will compare quantities derived from these two surveys elsewhere.

**III. The Next-Generation Approach and Age-Specific Results**

In this section, we note that the reproduction numbers, equilibrium age-specific contributions and prevalence always may be calculated via the next-generation matrix approach of van den Driessche and Watmough (2001). Their terminology comes from demography, where reproduction numbers are average numbers of daughters per woman. Similarly, in matrix population models, next-generation matrices transform, by multiplication, vectors composed of the numbers by age in one generation into the corresponding vectors in subsequent generations (Caswell 2001).

In their supplement, Glasser et al. (2010) derive the reproduction number for an age structured SEIR model without demographic dynamics (i.e., births equal to deaths) via this approach. To emphasize the importance of mixing in transmission of the pathogens causing infectious diseases among the members of host populations, we write next-generation matrices K as products of diagonal matrices whose elements are the sub-population reproduction numbers and contact matrices whose elements are the *cij* described above. That is,



The largest eigenvalue of the next-generation matrix is the average factor by which successive vectors differ in magnitude, or average number of secondary infections per infectious person. Because heterogeneity increases reproduction numbers (Feng et al. 2015, Glasser et al. 2016), the ℜ0 from this age-stratified model without demographic dynamics, 7⋅15, is slightly less than that calculated from our age- and province-stratified model (and reported in the main text).

The dominant eigenvalue has two associated nonzero vectors whose dot products with this matrix or its transpose equal their products with the eigenvalue. They also have biological interpretations: the right eigenvector is prevalence of infection and left is contributions to the basic reproduction number, both by sub-population (figures S4).



Figures S4. Age-specific a) contributions to rubella’s ℜ0 and b) equilibrium prevalence calculated using the serological observations of Su (1985) and contact observations summarized by Mossong et al. (2008).

**IV. Provincial Reproduction Numbers and their Ratios**

In this section, we tabulate the provincial reproduction numbers (table S3) reported in the main text (figures 2), which were derived from the age- and location-stratified model of Feng et al. (2017) via the next-generation matrix approach described in the previous section. In that model, age-specific contacts decline exponentially with inter-location distance at age-dependent rates that are reported in the supplement to Hao et al. (2019).

Table S3. Provincial basic and effective reproduction numbers, ℜ0*i* and ℜE*i* respectively, and proportions of provincial populations that are susceptible, ℜE*i*/ℜ0*i*.

|  |  |  |  |
| --- | --- | --- | --- |
| Province | ℜ0*i* | ℜE*i* | Ratio |
| Anhui | 9·19601 | 1·007601 | 0·109569 |
| Beijing | 8·41163 | 0·322716 | 0·038365 |
| Chongqing | 8·737076 | 1·628638 | 0·186405 |
| Fujian | 7·510264 | 1·110029 | 0·147802 |
| Gansu | 7·872194 | 0·890822 | 0·113161 |
| Guangdong | 7·224998 | 1·268451 | 0·175564 |
| Guangxi | 7·346261 | 1·238845 | 0·168636 |
| Guizhou | 7·925165 | 0·848905 | 0·107115 |
| Hainan | 5·88089 | 0·929079 | 0·157983 |
| Hebei | 8·865302 | 1·445784 | 0·163083 |
| Heilongjiang | 4·953663 | 0·426083 | 0·086014 |
| Henan | 9·451852 | 1·268419 | 0·134198 |
| Hubei | 9·323048 | 1·145185 | 0·122834 |
| Hunan | 8·727226 | 1·179521 | 0·135154 |
| Inner Mongolia | 7·327767 | 1·27207 | 0·173596 |
| Jiangsu | 8·711712 | 1·1239 | 0·129011 |
| Jiangxi | 8·610814 | 1·203475 | 0·139763 |
| Jilin | 5·933146 | 0·693436 | 0·116875 |
| Liaoning | 7·087443 | 0·899893 | 0·12697 |
| Ningxia | 8·072227 | 0·617542 | 0·076502 |
| Qinghai | 6·776343 | 0·759676 | 0·112107 |
| Shaanxi | 8·967647 | 0·957373 | 0·106759 |
| Shandong | 8·921229 | 1·163524 | 0·130422 |
| Shanghai | 8·027371 | 0·511318 | 0·063697 |
| Shanxi | 9·00634 | 1·204338 | 0·133721 |
| Sichuan | 8·058667 | 0·891483 | 0·110624 |
| Tianjin | 8·581049 | 0·998412 | 0·116351 |
| Tibet | 4·444372 | 0·634328 | 0·142726 |
| Xinjiang | 3·147807 | 0·380385 | 0·120841 |
| Yunnan | 6·435573 | 1·174423 | 0·182489 |
| Zhejiang | 8·038145 | 0·458849 | 0·057084 |

Because ℜE*i* ≈ ℜ0*i* × (1 – *pi*), where *pi* is the proportion of sub-population *i* that is immune, the ratio ℜE*i*/ℜ0*i* is interpretable as proportion of sub-population *i* that is susceptible (e.g., less than 4% of Beijing’s population). While one can become immune naturally as well as by vaccination, to some extent these proportions reflect the quality of provincial immunization programs.

We derived the reproduction numbers described in this supplement and illustrated in the main text via next-generation matrices, but our approach for identifying optimal strategies for reducing effective numbers requires explicit expressions. In section 3.6, Feng et al. (2015) proved that the gradient (vector of partial derivatives of ℜE with respect to parameters representing vaccination) quantifies the relative importance of sub-populations in vaccine allocation. Feng et al. (2017) derived an expression for ℜE from our age- and location-stratified model (table S2, rows I and II) assuming that mixing is proportionate with respect to age, whereupon ℜE is the trace of the next-generation matrix. We evaluated its gradient with respect to supplemental immunization rates (via Mathematica’s Grad procedure).

**V. Estimation of Vaccination Rates**

In this section, we describe how we estimated the rubella immunization rates for the 2014 effective reproduction number and gradient calculations and simulation models from annual increases in proportions with antibodies to measles between serological surveys during 2006 and 2014 that were not due to infection (Hao et al. 2019).

We multiply provincial proportions immunized against measles by logistic functions that we fit to annual ratios of the uptake of rubella- and measles-containing vaccine from 2008 to 2013 (Su et al. 2018), the curves and symbols, respectively, on figure S5 left. We calculate the immunization rates, where *p* is the proportion immunized during period *t*.

We use the 2014 provincial rates in our modeling, but superimpose the uptake of RCV reported by Su et al. (2018) by dose (symbols) for China on weighted averages of our provincial estimates from 2005-15 on figure S5 right.



Figure S5. Annual a) ratios of first and second dose uptake of rubella- and measles-containing vaccines (RCV and MCV, respectively) in China from 2008 to 2013 and b) administrative estimates of RCV uptake (symbols) reported by Su et al. (2018) superimposed on our estimated proportions immunized against rubella from 2005-15 in China (blue curve). The equations for the fitted curves (panel a) are 

**VI. Forcing the Regionally Stratified Simulation Model and Comparing its Predictions with Observations**

In this section, we describe the version of our age- and location-stratified model that we simulated (table S2, row II) to estimate times-to-elimination (reported annual incidence < 1 per 106 people) as a function of SIA uptake. First, we calculated regional contact rates as population-density-weighted averages of their constituent provincial ones.

Table S4. Daily contact rates for the age- and region-stratified model (table S2, row II) obtained as weighted averages of their constituent provincial population densities.

|  |  |  |  |
| --- | --- | --- | --- |
| Age | East | Central | West |
| <1 | 9.178 | 9.534 | 8.837 |
| 1-4 | 11.013 | 11.389 | 10.721 |
| 5-9 | 15.333 | 15.682 | 15.147 |
| 10-14 | 18.692 | 18.982 | 18.574 |
| 15-19 | 18.202 | 18.502 | 18.075 |
| 20-24 | 14.024 | 14.389 | 13.807 |
| 25-29 | 14.252 | 14.614 | 14.04 |
| 30-34 | 14.382 | 14.742 | 14.173 |
| 35-39 | 14.876 | 15.231 | 14.679 |
| 40-44 | 14.838 | 15.193 | 14.64 |
| 45-49 | 13.399 | 13.768 | 13.166 |
| 50-54 | 13.061 | 13.433 | 12.82 |
| 55-59 | 11.853 | 12.23 | 11.582 |
| 60-64 | 9.955 | 10.323 | 9.635 |
| 65+ | 7.792 | 8.113 | 7.414 |

Then we fit the function



where *t* is time in months and *a-h* are coefficients, to surveillance results from 2005 to 2016 by age group and region via Mathematica’s LinearModelFit procedure (figure S6, left). The polynomial terms account for secular trends and harmonic ones for seasonal fluctuations. After scaling as needed to ensure positive values, we multiplied the harmonic portion of these functions by the contact rates (figure S6, right).



Figure S6. Fit of the function *Fi* to a) infant reports in the eastern provinces of China from 2005 through 2016 and b) product of the infant contact rate (table S4) and harmonic portion for a two-year period. Narrow peaks occur during summer for ages ≤ 9 and ≥ 40 (as illustrated) and broader ones during winter (i.e., spanning the entire school year) for ages 10-39 years.

Next, using the initial conditions and parameter values (described in tables S1 and S2, row II), we simulated this age- and region-stratified model (i.e., evaluated the solution obtained via Mathematica’s NDSolve procedure) from 2014 through 2018, adjusting the regional contact rates in table S4 by constant factors so that predictions matched reported incidence over all age classes (figure S7, rightmost column).



Figure S7. Fit of the age- and location-stratified model (table S2, line II) to reports (dots) for the first two groups, aged < 1 and 1-4 years (left and middle, respectively) and overall (right), in the eastern, central and western provinces of China (top, middle, and bottom rows, respectively).

The final values of *S*(*t*), *E*(*t*), *I*(*t*), and *R*(*t*) were then used as new initial conditions to assess the impact of SIAs among adolescents or young adults on time-to-elimination (figures 6). Proportions of susceptible students entering middle and high school (the first years of age groups 4 and 5) were vaccinated during the first month of the school year (September) for 3 years beginning in 2018. Alternatively, proportions of susceptible young adults (i.e., one fifth of age groups 6 and 7) were vaccinated at the same time.

**VII. Surveillance-Based Estimates of Rubella and CRS Incidence**

In this section, we estimate the incidence of infection among pregnant women (infants with CRS if none chose to terminate their pregnancies) from disease surveillance (table S5). We multiply age-specific reported or simulated infections by proportions female, birth rates by age of mother and – because the risk of CRS varies with gestational age, approaching one during the first trimester and zero afterwards – the fraction of a pregnancy at risk.

Table S5. Estimating the burden of CRS from disease surveillance in China during 2014. We multiply the sum over all ages of products of infections reported, proportions female, and birth rates by 0·287, the proportion of each pregnancy at risk.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Age Group (years) | Rubella Reports | Population | Reported Incidence (per 106) | Proportions Female | Birth Rates (per year) | Estimated Infants w/CRS |
| 0 | 1 140 | 13 840 633 | 82·366 | 0⋅460244088 | 0 | 0 |
| 1-4 | 1 621 | 64 006 083 | 25·326 | 0⋅461103297 | 0 | 0 |
| 5-9 | 1 432 | 76 802 920 | 18·645 | 0⋅456290688 | 0 | 0 |
| 10-14 | 1 725 | 70 908 759 | 24·327 | 0⋅457580593 | 0 | 0 |
| 15-19 | 1 768 | 78 733 577 | 22·455 | 0⋅465636588 | 0·011194384 | 2·641 |
| 20-24 | 1 175 | 110 444 039 | 10·639 | 0⋅483493969 | 0·079773466 | 12·988 |
| 25-29 | 1 200 | 120 249 392 | 9·979 | 0⋅496170772 | 0·093621378 | 15·975 |
| 30-34 | 679 | 100 420 925 | 6·762 | 0⋅493888182 | 0·049027563 | 4·712 |
| 35-39 | 430 | 99 503 650 | 4·321 | 0⋅489430636 | 0·017040403 | 1·028 |
| 40-44 | 228 | 124 037 713 | 1·838 | 0⋅489147599 | 0·003963693 | 0·127 |
| 45-49 | 134 | 120 740 876 | 1·11 | 0⋅49163728 | 0·001073870 | 0·02 |
| 50-54 | 77 | 94 779 805 | 0.812 | 0⋅493383306 | 0 | 0 |
| 55-59 | 49 | 80 789 538 | 0·607 | 0⋅491318948 | 0 | 0 |
| 60-64 | 30 | 74 948 905 | 0·47 | 0⋅500365218 | 0 | 0 |
| 65+ | 78 | 137 675 182 | 0·567 | 0⋅521096767 | 0 | 0 |
| Total | 11 766 | 1 367 881 995 | 8·6 |  |  | 37·5 |

We estimate this fraction from information reported by Miller et al. (1982), who followed more than a thousand pregnant women with laboratory confirmed rubella infections to determine the outcome. The integral of the function fitted to probabilities of CRS given maternal infection in figure S8 is 11·46 weeks. Thus, the fraction of a pregnancy at risk is 11·46/40 = 0⋅287.

In 2014, 11 766 rubella infections were reported in China (table S5), from which we estimate that roughly 38 women could have borne children with CRS. However, such calculations underestimate the burden of CRS to the extent that rubella is under-reported among reproductive-aged adults. We can estimate the burden of rubella and CRS more accurately by demographically realistic modeling (section VIII).



Figure S8. Product of logistic functions fitted to the probabilities of fetal infection given maternal infection,  and CRS given fetal infection,  where *w* is gestational age in weeks (Miller et al. 1982).

**VIII. Estimation of Rubella and CRS Incidence via Demographically Realistic Modeling**

In this section, we estimate the incidence of rubella and CRS by simulating a demographically realistic SEIR model (table S2, row III) of 105 people having the 2014 age and gender distributions, death rates by age and birth rates by age of mother, vaccination rates and immunity profile of China (table S6).

In this model, individuals are born and die at observed rates, *fi* and *μi*, and age at rates *θI* (figure S9). In simulations, routine vaccination occurs during infancy rather than at birth, when the fraction *σ* is vaccinated in the analyses of Feng et al. (2020). In both models, all age classes may receive supplementary vaccination at rates *χi*.



Figure S9. Transition diagram for the demographically realistic SEIR model. Each epidemiological class has *n* = 15 sub-groups (horizontal flows) with transition rates *θi* due to aging.

As demographic parameters were available in the age groups of abridged life tables (table S6), we used them, obtaining initial conditions (i.e., age-specific proportions immune and their complements) by averaging functions over the appropriate intervals (e.g., figure S2).

Table S6. Demographic parameters from the 2014 China Statistical Yearbook (Xing and Ye 2014). Proportions immune are predictions of the logistic regression equation fitted to observations from the 2014 serological survey (illustrated by region in figures 1).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Age (years) | Proportion | Pr(female) | Pr(immune) | Birth Rates (per day) | Death Rates (per day) |
| 0 | 0·010118294 | 0·460244088 | 0·952987896 |  | 0·000012611 |
| 1-4 | 0·046792108 | 0·461103297 | 0·927679783 |  | 0·000001368 |
| 5-9 | 0·056147328 | 0·456290688 | 0·860239891 |  | 0·000000591 |
| 10-14 | 0·05183836 | 0·457580593 | 0·807664692 |  | 0·000000715 |
| 15-19 | 0·057558749 | 0·465636588 | 0·809641899 | 0·000030649 | 0·000000983 |
| 20-24 | 0·080740911 | 0·483493969 | 0·857213625 | 0·000218408 | 0·000001032 |
| 25-29 | 0·087909185 | 0·496170772 | 0·917252062 | 0·000256321 | 0·000001717 |
| 30-34 | 0·073413441 | 0·493888182 | 0·961732705 | 0·000134230 | 0·000002333 |
| 35-39 | 0·072742861 | 0·489430636 | 0·985002042 | 0·000046654 | 0·000002851 |
| 40-44 | 0·090678665 | 0·489147599 | 0·994643471 | 0·000010852 | 0·000004776 |
| 45-49 | 0·088268488 | 0·49163728 | 0·998128003 | 0·000002940 | 0·000006298 |
| 50-54 | 0·06928946 | 0·493383306 | 0·999315318 |  | 0·000011967 |
| 55-59 | 0·059061774 | 0·491318948 | 0·999720372 |  | 0·000015263 |
| 60-64 | 0·054791938 | 0·500365218 | 0·999864079 |  | 0·000027780 |
| 65+ | 0·100648435 | 0·521096767 | 0·999930172 |  | 0·000124867 |

The mixing matrix was derived as described in section II from face-to-face conversations, a proxy for the inter-personal contacts by which respiratory pathogens may be transmitted, recorded by participants in 8 European countries (Mossong et al. 2008) because they could be aggregated into those intervals. As noted above, otherwise comparable Chinese observations (Read et al. 2014, figure S4) are only available for 0-4, 5-19, 20-64, and 65+ years (Jonathan M. Read, personal communication), which are too crude for accurate transmission modeling.

We simulated this model via the discrete event/time method (Renshaw 1991) programmed in C++. In such simulations, time intervals vary inversely with the sum of all state-transition rates. There are as many *per capita* rates as age groups, not only of ageing, but of dying, giving birth (if female), being vaccinated or infected (if susceptible), becoming infectious (if infected) and recovering (if infectious). While those are fixed, the transition rates – products of the *per capita* rates and numbers in relevant states – vary (e.g., the rates of infection also depend on the numbers and ages of susceptible and infectious people).

We begin simulations by making single individuals newly infectious. Even if they do not die or recover before contacting a susceptible person, an outbreak may not occur (figure S3). At each time, the event occurs whose normalized rate corresponds to a random number drawn from a multinomial distribution. Then possible transition rates are (re)calculated and summed to determine the next time interval, rates are normalized, a random number in the unit interval is chosen, and the event with the corresponding normalized rate occurs until there are no new infections. No simulations lasted more than 3 years and most ended within 2 years.

The age distribution of simulated infections does not vary with the age of newly infectious individuals (figure S10b) and, except for the youngest group, resembles that of the 2014 reports (table S5, figure S10a). But, because model individuals are not only heterogeneous with respect to age, and hence contact rates, but mix non-randomly, the age of newly infectious individuals does affect other results.



Figures S10. Age distributions of a) 2014 reported and b) simulated rubella infections, the latter as a function of the newly infectious person’s age.

In particular, newly infectious infants differ from others. A greater proportion lead to outbreaks, implying a higher effective reproduction number and incidence than for older individuals. This almost certainly reflects preferential contacts between young children and parents together with the relatively high probability of susceptible infants being infected (figure S3). In other words, infants are super-spreaders. Because the other reproduction numbers are consistent with estimates from our age- and province-stratified model and infants are unlikely to initiate rubella outbreaks, the estimates reported in the main text are for infectious individuals aged 1-44 years.

Also, because we must extrapolate to the population of China, whose size is more than 104 times the model population, spatially heterogeneous, and mobile, our estimates in the main text are based on simulations with final size > 1. In other words, we assume that outbreaks are inevitable in such a large, spatially heterogeneous, and mobile population. To include all *m* = 200 simulations with initial newly infectious individuals aged < 1, 1-4, … years, sum products of the incidence estimates and quotients of outbreaks and simulations in table S7.

Table S7. Results from 2-year simulations of our demographically realistic model of rubella in China. From left to right, columns are ages of initial newly infectious individuals, proportions of simulations with final size > 1, the corresponding effective reproduction numbers (in an un-stratified population model without demographic dynamics), and estimates (based on simulations with final size > 1) of annual rubella and potentially CRS incidence per 105 people.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Initial Newly Infectious Person’s Age | Quotient of Outbreaks and Simulations | Effective ℜE | Rubella Incidence  | Potential CRS Incidence  |
| < 1 | 0·82 | 5·56 | 43·73 | 0·207 |
| 1-4 | 0·45 | 1·82 | 13·81 | 0·075 |
| 5-9 | 0·365 | 1·57 | 7·16 | 0·037 |
| 10-14 | 0·455 | 1·83 | 10·49 | 0·056 |
| 15-19 | 0·365 | 1·57 | 13·05 | 0·071 |
| 20-24 | 0·305 | 1·83 | 8·13 | 0·04 |
| 25-29 | 0·255 | 1·34 | 10·89 | 0·052 |
| 30-34 | 0·285 | 1·4 | 5·25 | 0·024 |
| 35-39 | 0·34 | 1·52 | 8·61 | 0·058 |
| 40-44 | 0·265 | 1·36 | 8·24 | 0·048 |

**IX. Estimation of Reporting Efficiency**

In this section, we compare simulated rubella incidence rates (table S7) with those reported (table S5), concluding that fewer than 10% of infections were reported. Under-reporting is expected for mild diseases, but its magnitude is difficult to assess. Comparing age-specific reported and simulated incidences, all save infant and post-reproductive adult infections appear to have been under-reported (figure S11). As these discrepancies do not affect the reproductive ages, at least not directly, we do not believe that they affect our estimate of CRS incidence.



Figure S11. Ratios of age-specific reported and simulated incidence indicate that infections among people aged 1-49 years were under-reported in 2014.

Regarding the discrepancy among the elderly, we stabilized the polynomial function fitted to results from the 2014 serological survey, which extended only through age 29 years, by adding a few people aged 65+ years with the immunity predicted from our estimate of ℜ0 (i.e., 1 – 1/ℜ0). This may have over-estimated immunity among the elderly in 2014.

Regarding the discrepancy among infants, the 2014 serological survey indicates that < 5% were susceptible (table S6). We assumed that routine vaccination occurred throughout infancy, albeit at a lower rate than it may actually have during the last 4 months. We considered sub-dividing the infant group into ages < and ≥ 8 months, but estimating the requisite demographic parameters would have introduced other inaccuracies. Moreover, as the serological survey indicates that most young adults were immune, we reasoned that most infants < 8 months of age would have maternal antibodies, which we did not model.

In the analyses reported by Feng et al. (2020), it was convenient to have vaccination occur at birth (figure S9). Neither their assumption nor ours is correct, as routine vaccination in China comprises doses of MMR at 8 and MR at 15-18 months.

As mentioned in the main text, the incidence of rubella has since increased dramatically in China, so an alternative explanation for the apparent under-reporting among people aged 1-49 years is that we predicted the ongoing outbreak. In either event, we believe that our estimates of the incidence of rubella and CRS are as reliable as possible at present.

**X. Comments on the Methods Employed in a Recent Assessment of the Global Burden of CRS**

In this section, we comment on the global CRS burden assessment of Vynnycky et al. (2016), who approximate even where more detailed calculations (e.g., China) could have been performed. Their stated purpose was to reduce computational burden, but this may also have increased comparability. Nonetheless, we are puzzled by their use of Schenzle’s (1984) model, whose age groups are single years, to assess the burden in countries that vaccinate against rubella. This was unnecessary and may be misleading. Also, their estimate of the force of infection among adults seems too low.

First, Vynnycky et al. (2016) assert that wider age ranges introduce inaccuracies associated with ageing. Insofar as ageing occurs at constant *per capita* rates, sojourns are exponentially distributed, whereupon most members spend less than the mean time in such states and a few more. Given those authors’ lengthily simulations to obtain quasi-stable age-distributions, such inaccuracies would indeed be a problem. (Our simulations last at most three years, so this is inconsequential.) However, Hethcote (2000) explained how to calculate stable age-distributions in the continuous case and Feng et al. (2020) provide the corresponding expression when age is discrete. They also reformulate in terms of mean sojourns and probabilities of ageing from one group to the next.

We use the actual 2014 age-distribution of China in our models. But, to illustrate this point, we compare the stable age-distribution given the 2014 birth and death rates with those for the same population growth rate, but types I and II mortality, in which people die at their expectation of life at birth or at constant rates, respectively, in figures S12.



Figures S12. For the actual 2014 mortality in China and types I and II, a) stable age distributions and b) corresponding survival curves.

Second, Vynnycky et al. (2016) estimated only two forces of infection and effective contact rates, essentially for children and adults. Based on the pre-vaccination serological survey of Su (1985), the most reasonable of their several estimates are 295 and two per 1 000 susceptible people aged < 13 and ≥ 13 years, respectively. Averaging the function illustrated in figure S1b (which assumes that 10% of survey participants with antibodies were asymptomatic and not infectious) over those intervals and multiplying by 1 000 for comparison, we obtain 224·7 and 15·3. Thus, Vynnycky et al. (2016) seem to have under-estimated incidence among adults.

In our model 2014 population, an average of 1 011 births occur per year. To obtain CRS incidence per live birth, divide our estimates (for 105 people) by 1 011 births. Our estimate in the main text, 0⋅051 per 105 people, becomes approximately 5·1 per 105 live births, which is roughly five times the estimate of Vynnycky et al. (2016).

**XI. Immune Profiles and Gradient by Province**

The main text includes the rubella immune profiles for Eastern, Central and Western China (figures 1) and provincial gradients in 2014 grouped by region (figures 4). In this section, we describe the immune profiles and gradient separately by province-level jurisdiction (figures S13).































































Figures S13. Rubella a) immune profiles and b) gradient in 2014 by province. The lines on the left are weighted logistic regressions with single years of age modeled as a cubic polynomial. On the right, they are the multivariate partial derivatives of the 2014 meta-population effective reproduction number with respect to possible supplemental vaccination rates.

**XII. References**

Caswell H. *Matrix Population Models: Construction, Analysis and Interpretation*, 2nd Ed. Sunderland, MA: Sinauer Associates, 2001

Farrington CP. Modeling risks of infection for measles, mumps and rubella. Stat Med 1990; **9**:953–67

Feng Z, Hill AN, Smith PJ, Glasser JW. An elaboration of theory about preventing outbreaks in homogeneous populations to include heterogeneity or non-random mixing. J Theor Biol 2015; **386**:177–87

Feng Z, Glasser JW. Mixing in meta-population models. Chapter 3 in *The Dynamics of Biological Systems*, A Bianchi, T Hillen, M Lewis, and Y Yi, eds. Springer, 2019 (in press)

Feng Z, Glasser JW. Estimating age-specific hazard rates of infection from cross-sectional observations. Revista de Matemática: Teoría y Aplicaciones 2020; **27**:123–40

Feng Z, Feng Y, Glasser JW. Influence of demographically realistic mortality schedules on vaccination strategies in age-structured models. Theor Popul Biol 2020 **132**:24–32

Glasser JW, Taneri D, Feng Z, et al. Evaluation of targeted influenza vaccination strategies via population modeling. PLoS ONE 2010; **5**(9):e12777 (doi:10.1371/journal.pone.0012777)

Glasser JW, Feng Z, Omer SB, Smith PJ, Rodewald LE. The effect of heterogeneity in uptake of the measles, mumps, and rubella vaccine on the potential for outbreaks of measles: a modelling study. Lancet Infect Dis 2016; **16**:599–605

Hao L, Glasser JW, Su Q, et al. Evaluating vaccination policies to accelerate measles elimination in China: a meta-population modeling study. Int J Epidemiol 2019; **48**:1240–51

Hethcote HW. The mathematics of infectious diseases, SIAM Review 2000; **42**:599–653

Heymann DL. Control of Communicable Diseases Handbook, 20th Ed. Washington, DC: American Public Health Association Press, 2014

Miller E, Cradock-Watson JE, Pollock TM. Consequences of confirmed maternal rubella at successive stages of pregnancy. Lancet 1982; **320**:781–84

Mossong J, Hens N, Jit M, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 2008; **5**:381–91

Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Comput Biol 2017; **13**:e1005697 (https://doi.org/10.1371/journal.pcbi.1005697)

Read JM, Lessler J, Riley S, et al. Social mixing patterns in rural and urban areas of southern China. Proc R Soc B 2014; **281**:20140268

Renshaw E. Modeling Biological Populations in Space and Time. Cambridge, England: Cambridge Univ Press, 1991

Schenzle D. An age-structured model of pre- and post-vaccination measles transmission. IMA J Math Appl Med Biol 1984; **1**:169–91

Su Q, Ma C, Wen N, et al. Epidemiological profile and progress toward rubella elimination in China 10 years after nationwide introduction of rubella vaccine. Vaccine 2018; **36**:2079–85

Su Q, Zhu Z, Hao L, et al. The first population-based rubella serologic survey of 1-29 years old in China: are we close to rubella elimination? Manuscript (in preparation)

Su W. Rubella in the People's Republic of China. Rev Infect Dis 1985; **7**:S72

van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci 2002; **180**:29–48

Vynnycky E, Adams EJ, Cutts FT, et al. Using seroprevalence and immunisation coverage data to estimate the global burden of congenital rubella syndrome, 1996-2010: A systematic review. PLoS ONE 2016; **11**(3): e0149160 (doi:10.1371/journal.pone.0149160)

Xing Z, Ye Z, Editors-in-Chief. China Statistical Yearbook – 2014. China Statistics Press, Beijing, PRC (<http://www.stats.gov.cn/tjsj/ndsj/2014/indexeh.htm>)

Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong J, Edmunds WJ. Using time-use data to parameterize models for the spread of close-contact infectious diseases. Am J Epidemiol 2008; **168**:1082–90