

Supplementary appendix to:

Estimation of SARS-CoV-2 age-specific mortality during the early stages of an epidemic: a modelling study in Hubei, China and northern Italy

Anthony Hauser^a, Michel J. Counotte^a, Charles C. Margossian^b, Garyfallos Konstantinoudis^c, Nicola Low^a, Christian L. Althaus^a, and Julien Riou^{a,*}

^aInstitute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

^bDepartment of Statistics, Columbia University, New York, NY

^cMRC Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

*Corresponding author (julien.riou@ispm.unibe.ch)

March 22, 2020

Contents

1	Data	2
1.1	General data	2
1.2	Data from Hubei, China	2
1.3	Data from Northern Italy	2
2	Model	6
2.1	Transmission model structure	6
2.2	Initial conditions and simulation settings	7
2.3	Dummy compartments and outputs	7
2.4	Delayed mortality	7
2.5	Inference	8
2.6	Estimates of mortality	9
2.7	Implementation and diagnostics	9
3	External validation	11
4	Additional results	12
4.1	Additional results in Hubei, China	12
4.2	Additional results in northern Italy	13
5	Sensitivity analyses	14
5.1	Ascertainment in the highest age class	14
5.2	Reduced time between disease onset and death	16
6	Stan code	17

1 Data

1.1 General data

Some data and parameter values are not specific to the area considered. The mean incubation period was fixed to 5.9 days, following a study of cases and their contacts in Shenzhen, China [1]. The mean time from disease onset to isolation was fixed to 2.4 days, following the same Shenzhen study [1]. Other works have reported very similar values for both periods [2, 3, 4]. These durations were transformed into rates by taking the inverse and used in the system of ordinary differential equations (section 2). As such, this results in exponential distributions of incubation times and time from disease onset to isolation, which is acceptable considering the aim of the SEIR model in this context: obtain a good description of case incidence by date of disease onset in each age class.

Conversely, one of the main objectives of this work was to account for bias due to the delay between disease onset and death. For the time from disease onset to death, we used a log-normal distribution with mean 20.2 and standard deviation 11.6 estimated from patient data and accounting from right-censoring [3]. We considered other mean durations in sensitivity analyses (section 5). This continuous probability distribution was discretized by day in order to be used in the model (Figure S1).

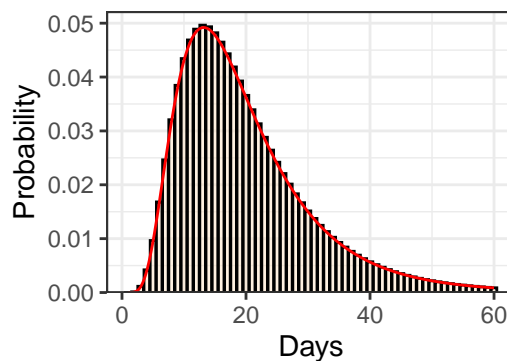


Figure S1: Probability density function of the time from disease onset to death (red) and corresponding discretization by day between 1 and 60 days (columns).

1.2 Data from Hubei, China

We model the dynamics of the COVID-19 epidemic in Hubei from January 1st to February 11th, 2020, considering the introduction of strong control measures from January 20th [5]. Our analysis was based on a report published by the Chinese CDC describing all Chinese COVID-19 cases reported up to February 11th, 2020 [6]. From these reports, we extracted the consolidated number of confirmed cases of COVID-19 *by day of disease onset*, the number of deaths per day, and the age distributions of cases and deaths on 11 February 2020 (Figure S2). We also used data about the age distribution of the Chinese population, and about the daily number of potentially infectious contacts by age group in Shanghai [7], which we assumed were similar to the population of Hubei.

We also extracted data on the proportion of comorbidities among reported COVID-19 cases and deaths [6]. Age-specific prevalences of diabetes, chronic respiratory diseases, cardiovascular diseases and hypertension in China were extracted from the IHME website [8] (Figure S3).

1.3 Data from Northern Italy

We model the dynamics of the COVID-19 epidemic in Northern Italy from 8 February to 3 March, 2020. Our analysis was based on reports from the Istituto Superiore di Sanita and the Dipartimento della Protezione Civile describing all Italian COVID-19 cases reported up to 3 February, 2020 [9, 10]. From these reports, we extracted the consolidated number of confirmed cases of COVID-19 *by day of disease onset*, the number of deaths per day, and the age distributions of cases and deaths at 15 March 2020 (Figure S4). We also collected data about the age distribution of the Italian population [11], as well as its contact patterns by age group using results from the POLYMOD study [12].

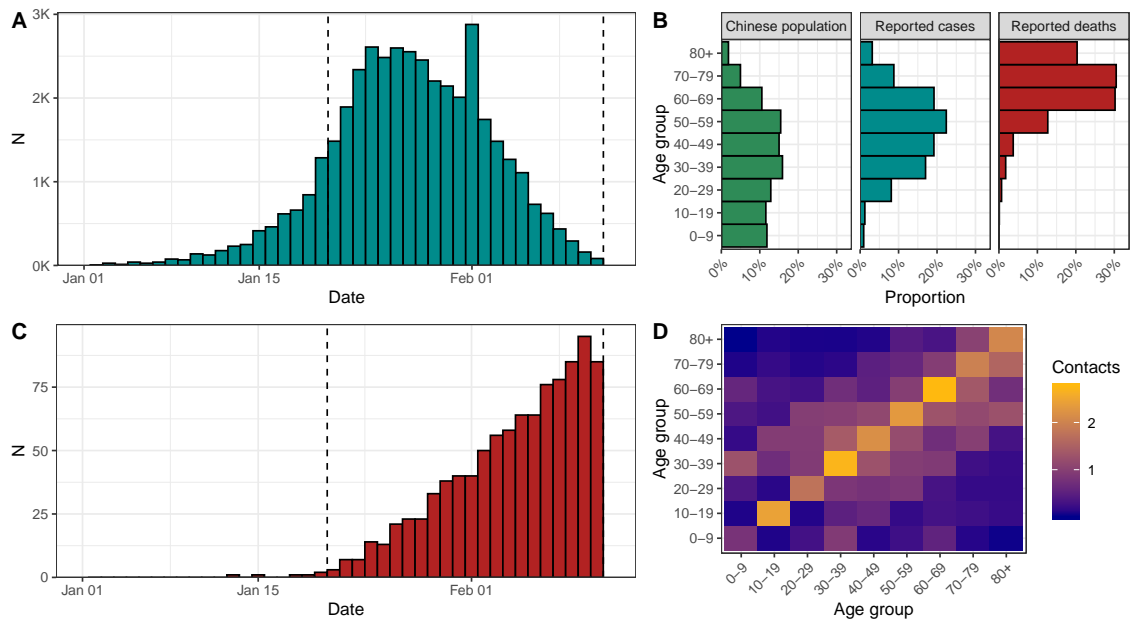


Figure S2: Data used to fit the model in Hubei, China. (A) Reported confirmed cases of COVID-19 in Hubei by date of disease onset until 11 February, 2020. (B) Age distribution of the Chinese population compared to that of confirmed cases and of deaths due to COVID-19. (C) Reported deaths until 11 February, 2020. (D) Matrix representing the average number of daily contacts between each age class in Shanghai (source: Zhang et al., 2019 [7]).

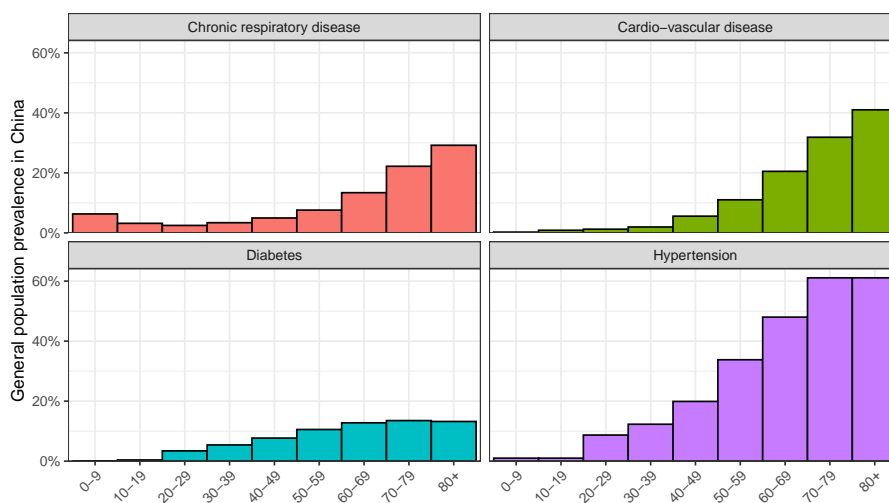


Figure S3: General population prevalence of four comorbidities by age class in China (source: IHME).

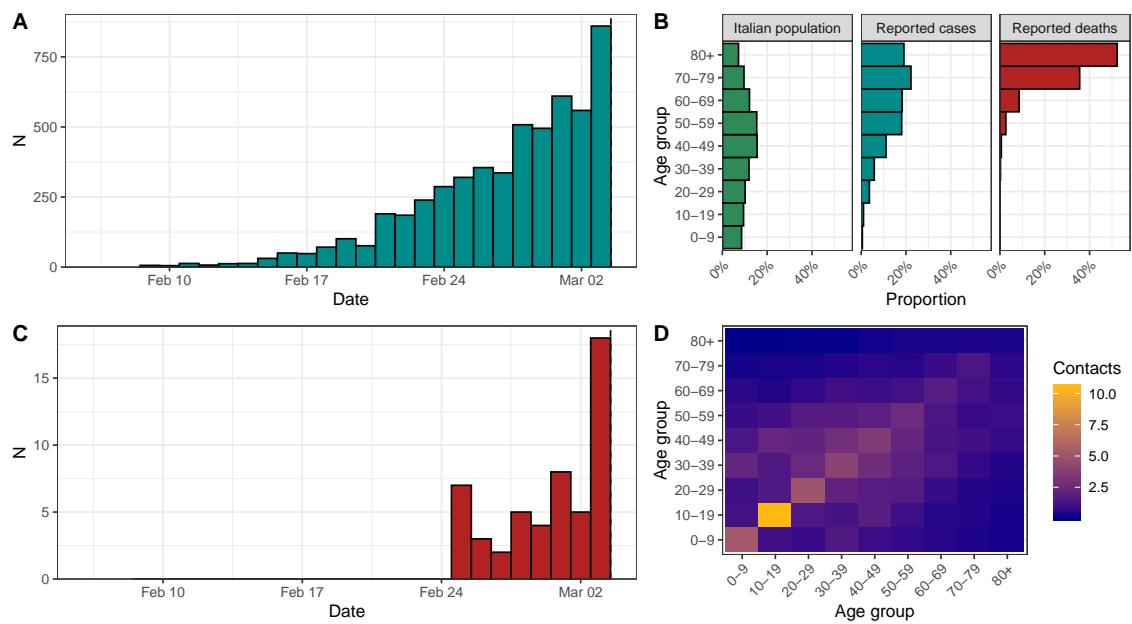


Figure S4: Data used to fit the model. (A) Reported confirmed cases of COVID-19 in Italy by date of disease onset until 11 February, 2020 (blue), (B) Age distribution of the Italian population compared to that of confirmed cases and of deaths due to COVID-19. (C) Reported deaths in Italy until 11 February, 2020 (red). (D) Matrix representing the number of daily contacts between each age class.

Symbol	Comment	Value	Unit	Source
A	Reported cases of COVID-19 in Hubei from 1 January to 11 February 2020 and in Italy from 8 February to 3 March	see Fig S2A & S4A	-	Hubei: [6], Italy: [9, 10]
B	Age distribution of all confirmed cases in China up to 11 February 2020* and in Northern Italy up to 15 March 2020	see Fig S2B & S4B	-	Hubei: [6], Italy: [9, 10]
C	Deaths among patients with confirmed COVID-19 infection by day of report in China from 1 January to 11 February 2020 † and in Northern Italy from 8 February to 3 March	see Fig S2C & S4C	-	Hubei: [6], Italy: [9, 10]
D	Age distribution of the deaths among patients with confirmed COVID-19 infection that occurred in China up to 11 February 2020 and in Northern Italy up to 15 March ‡	see Fig S2B & S4B	-	Hubei: [6], Italy: [9, 10]
E	Age distribution of the population in China and Northern Italy	see Fig S2B & S4B	-	[11, 13]
F	9×9 matrix with contacts between each age class in Shanghai (used for China) and in Europe (POLYMOD, used for Italy)	see FigS2D & S4D	per day	[14, 12]
G	Distribution of time from disease onset to death	$\log\mathcal{N}(20.2, 11.6)$	days	[15]
$1/\tau$	Incubation period	5.9	days	[1]
$1/\mu$	Time from disease onset to isolation	2.4	days	[1]
γ	Time from disease onset to death (mean - standard deviation)	20.2 - 11.6	days	[16]

*As 85% of these are in Hubei, we assume that the age distribution is the same in overall China. † As 95% of these are in Hubei, the numbers were scaled down by 5%. ‡ We assume that the age distribution is the same in overall China.

Table 1: Summary of data sources and fixed parameters.

2 Model

We developed an approach aimed at estimating the true mortality associated with COVID-19. With regards to a specific epidemic, the model relies upon data on cases by date of disease onset (which is difficult to obtain), mortality by date of death, and the age distribution of both cases and deaths. In a first step, we use an age-stratified susceptible-exposed-infected-removed (SEIR) compartmental model to describe the dynamics of transmission of the virus in the population of Hubei, including the decrease in incidence after the introduction of control measures on 20 January. In a second step, we estimate the mortality of infected individuals with regards to the probability of dying upon infection and the delay between symptom onset and death. We fit the model to data on the time dynamics and the age distribution of both reported cases and deaths. We implemented the model in a Bayesian framework using Stan. All code and data are available from https://github.com/jriou/covid_adjusted_cfr/. We applied this approach to COVID-19 epidemics in two different settings where appropriate data is available: Hubei (China) and Northern Italy.

2.1 Transmission model structure

We use an age-stratified susceptible-exposed-infected-removed (SEIR) compartmental model using ordinary differential equations (ODEs), with a distinction between asymptomatic and symptomatic infections. We stratify the population in nine ten-year age groups ($0-9, 10-19, \dots, 70-79, 80+$). The population of each age group k is divided into five compartments: susceptible (S_k), incubating or exposed (E_k), infected with symptoms (I_k), infected and asymptomatic (A_k), and removed (R_k). Note that the number of individuals in each compartment is scaled by the total population of the area (59,020,000 for Hubei, 25,276,000 for Lombardia, Emilia Romagna, Veneto, Marche and Piemonte), so that the sum of all compartments is 1. Since we focus on the first wave of infection in Hubei and the early stage of the epidemic in northern Italy, the total size of the population has a very small impact on the transmission dynamics. Susceptible individuals may become infected through contacts with infected individuals with symptoms (assuming that incubating or asymptomatic individuals are not infectious) from any other age group. The force of infection λ_k depends on the added number of infectious individuals in each age class, weighted by a contact matrix \mathbb{F} and the proportion of infection upon contact with an infectious individual β .

In addition, we include a time-dependent forcing function to account for the reduction in transmissibility following the introduction of control measures in Hubei on 20 January 2020 in Hubei. In northern Italy, this effect is ignored as strict control measures were implemented on 9 March 2020, after the period of interest. We model the progressive introduction of control measures using the following logistic function:

$$f(t, \eta, \nu, \xi) = \eta + \frac{1 - \eta}{1 + e^{\xi(t - t_c - \nu)}} \quad (1)$$

where η is the relative reduction in transmission after control measures are fully implemented, ξ is the slope of the logistic function modelling the implementation of the control measures, and ν is the delay (in days after t_c , corresponding to 20 January 2020) until the control measures are 50% effective.

The force of infection λ_k can then be expressed as:

$$\lambda_k(t) = f(t, \eta, \nu, \xi) \beta S_k(t) \sum_{l=1}^9 \frac{I_l(t)}{E_l} \mathbb{F}_{k,l} \quad (2)$$

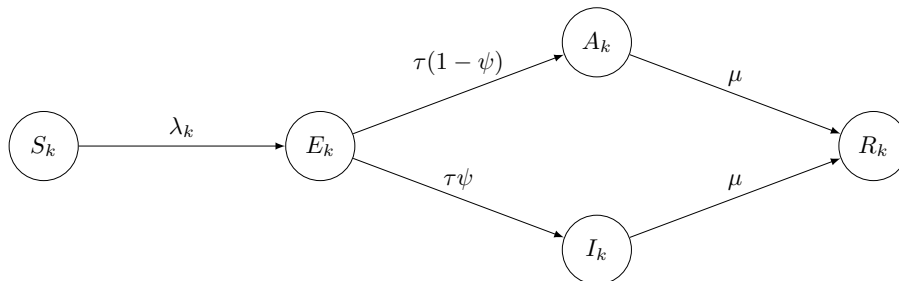


Figure S5: Schematic description of the COVID-19 transmission model. The dashed circles represent the dummy compartments used to record cumulative incidences in asymptomatic and symptomatic people.

After the incubation period $1/\tau$, a proportion ψ of infected people develop symptoms and become infectious while the remaining remain asymptomatic and do not transmit the disease further. After a period $1/\mu$, individuals with symptoms are identified and isolated, and thus removed from being infectious. The infection dynamics are thus controlled by 4 parameters: $\{\beta, \tau, \psi, \mu\}$.

2.2 Initial conditions and simulation settings

The simulation start ($t = 0$) is set on the day before the first data point (31 December 2019 for Hubei, 7 February 2020 for northern Italy). At this date, most individuals are assumed to be susceptible and are distributed across age groups according to the population age distribution \mathbb{E} . The number of people in the exposed compartments at $t = 0$ is controlled by a parameter π , and also distributed according to \mathbb{E} . Other compartments are set to 0.

2.3 Dummy compartments and outputs

In addition to the five compartments per age group, we add two dummy compartments to record the cumulative incidence of new symptomatic infections by day of disease onset, C_k :

$$\frac{dC_k^I}{dt} = \tau\psi E_k \quad (3)$$

and the cumulative incidence of new asymptomatic infections, D_k :

$$\frac{dC_k^A}{dt} = \tau(1 - \psi)E_k \quad (4)$$

Moreover, we need the following outputs in order to fit the model:

- the number of new symptomatic infections per day and by age group:

$$\Delta C_{k,t}^I = C_k^I(t) - C_k^I(t-1) \quad (5)$$

- the total number of new infections per day:

$$\Delta C_t^I = \sum_k^9 \Delta C_{k,t}^I \quad (6)$$

- the total number of new reported infections per day, introducing the age-specific ascertainment proportion parameter ρ_k :

$$R_t = \sum_k^9 \rho_k \Delta C_{k,t}^I \quad (7)$$

- the age distribution of all reported cases up to $t_{\max} = 11$ February 2020 in Hubei and up to $t_{\max} = 3$ March in northern Italy:

$$D_k^{\text{cases}} = \frac{\rho_k C_k^I(t_{\max})}{\sum_k^9 \rho_k C_k^I(t_{\max})} \quad (8)$$

where $t_{\max++}$ represent the index of the last simulated day.

To allow identifiability of the parameters, we remove a degree of freedom by fixing the ascertainment proportion of the highest age group ρ_9 to 100%. This assumes that all individuals aged 80 and more that are infected with symptoms seek medical care, are identified and reported to the authorities. The ascertainment proportions for other groups $\rho_{1,\dots,8}$ are estimated from data.

2.4 Delayed mortality

Mortality is considered outside of the system of ODEs, using an age-specific mortality parameter ϵ_k . In accordance to data, we only consider the mortality *among cases with a date of disease onset during the period of interest*, up to $t_{\max} = 11$ February 2020 for Hubei and to $t_{\max} = 8$ March 2020 for Italy. We calculated the mortality until 60 days after the last date of disease onset for each country. To do so, we generated a discretized distribution of time from disease onset to death \mathbb{G} of length 60 (Figure S1).

To fit the model and to estimate ϵ_k , we need the following mortality-related outputs:

- the number of deaths in age group k among people infected before t_{\max} and 60 days after t_{\max} , to account for the delay between disease onset and deaths:

$$M_{k,t} = \epsilon_k \sum_d^{60} \Delta C_{k,t-d}^I \mathbb{G}_d \quad (9)$$

- the daily number of deaths summed over age groups, assuming that all deaths are reported:

$$M_t = \sum_k^9 M_{k,t} \quad (10)$$

- the age distribution of all deaths occurring up to t_{\max} :

$$D_k^{\text{deaths}} = \frac{\sum_{t=1}^{t_{\max}} M_{k,t}}{\sum_{t=1}^{t_{\max}} M_t} \quad (11)$$

2.5 Inference

The objective of this step is to estimate the following parameters: $\theta = \{\beta, \eta, \xi, \nu, \psi, \pi, \rho_k, \epsilon_k, \phi_1, \phi_2\}$. We set weakly-informative priors for most of these parameters, which are summarized in Table 2. For the proportion of symptomatic infections ϕ , we used an informative prior distribution $\psi \sim \text{Beta}(522, 114)$ corresponding to an estimate of 82.1% (95% central range: 79.8-84.5) [17]. Implementing this information as a beta distribution allowed the propagation of uncertainty on ψ into the final estimates. For the initial proportion of infected, we used an informative prior distribution $\pi \sim \text{Beta}(1, 999)$ corresponding to a 95% central range of 0-0.0037 of the total population. This range is actually quite large and weakly-informative as it corresponds to a number of initial infections at $t = 0$ in the range $[0 - 640]$.

Symbol	Comment	Support	Prior
β	Probability of transmission upon contact	$[0 - 1]$	Beta(1, 1)
η	Reduction in transmission after control measures	$[0 - 1]$	Beta(1, 1)
ξ	Slope of logistic function for control measures implementation	$[0.5 - 1.5]$	Beta(1, 1) + 0.5
ν	Shift of logistic function for control measures implementation	$[0 - \infty[$	Exp(1)
ψ	Proportion of symptomatics of 82% on the Diamond Princess [18]	$[0-1]$	Beta(522, 114)
ϵ_k	Proportion of deaths among symptomatic cases (by age group)	$[0-1]$	Beta(1, 1)
ρ_k	Reporting rate of symptomatic cases (by age group)	$[0-1]$	Beta(1, 1)
π	Initial proportion of exposed (at $t = 0$)	$[0-1]$	Beta(1, 999)
ϕ_1	Overdispersion for cases	$[0 - \infty[$	Exp(1/100)
ϕ_2	Overdispersion for deaths	$[0 - \infty[$	Exp(1/100)

Table 2: Summary of parameters.

We simultaneously fitted our model to four data sets:

1. the number of reported cases of COVID-19 by day of disease onset:

$$\Pr(\theta|\mathbb{A}) = \text{Neg-Bin}(\mathbb{A}|R_t, \phi_1) \quad (12)$$

2. the age distribution of all confirmed cases up at t_{\max} :

$$\Pr(\theta|\mathbb{B}) = \text{Multinomial}(\mathbb{B}|D_k^{\text{cases}}) \quad (13)$$

3. the number of deaths among patients with confirmed COVID-19 infection by day of report:

$$\Pr(\theta|\mathbb{C}) = \text{Neg-Bin}(\mathbb{C}|M_t, \phi_2) \quad (14)$$

4. the age distribution of the deaths among patients with confirmed COVID-19 infection that occurred before t_{\max} :

$$\Pr(\theta|\mathbb{D}) = \text{Multinomial}(\mathbb{D}|D_k^{\text{deaths}}) \quad (15)$$

This leads to the following likelihood:

$$\Pr(\theta|\mathbb{A}, \mathbb{B}, \mathbb{C}, \mathbb{D}) = \Pr(\theta|\mathbb{A}) \cdot \Pr(\theta|\mathbb{B}) \cdot \Pr(\theta|\mathbb{C}) \cdot \Pr(\theta|\mathbb{D}) \quad (16)$$

2.6 Estimates of mortality

Table 3 summarizes the different estimates of mortality according to whether we adjust for preferential ascertainment of severe cases and/or right-censoring of deaths.

		Adjusting for preferential ascertainment	
		No	Yes
Adjusting for right-censoring of deaths	No	$\frac{M_{t_{\max}}}{R_{t_{\max}}}$	$\frac{M_{t_{\max}}}{C_{t_{\max}}^I}$
	Yes	$\frac{M_{t_{\max}+60}}{R_{t_{\max}+60}}$	$\frac{M_{t_{\max}+60}}{C_{t_{\max}+60}^I}$

Table 3: Mortality among symptomatic infections.

We also estimate the adjusted mortality in the overall infected (symptomatic or asymptomatic) population:

$$\frac{M_{t_{\max}+60}}{C_{t_{\max}+60}^I + C_{t_{\max}+60}^A} \quad (17)$$

2.7 Implementation and diagnostics

We implemented the model in *Stan* [19], that uses Hamiltonian Monte Carlo sampling. We estimated the posterior distributions by sampling 500 iterations after 500 iterations for warm-up in four independent chains. We assessed convergence visually and by using the Gelman-Rubin convergence diagnostic \hat{R} . We also checked for divergent transitions.

Mixing of chains was good in both Hubei and northern Italy (Figures S6 and S7), with a maximal \hat{R} of 1.011 in Hubei and 1.006 in northern Italy, more than 500 effective sample size and no divergent transition.

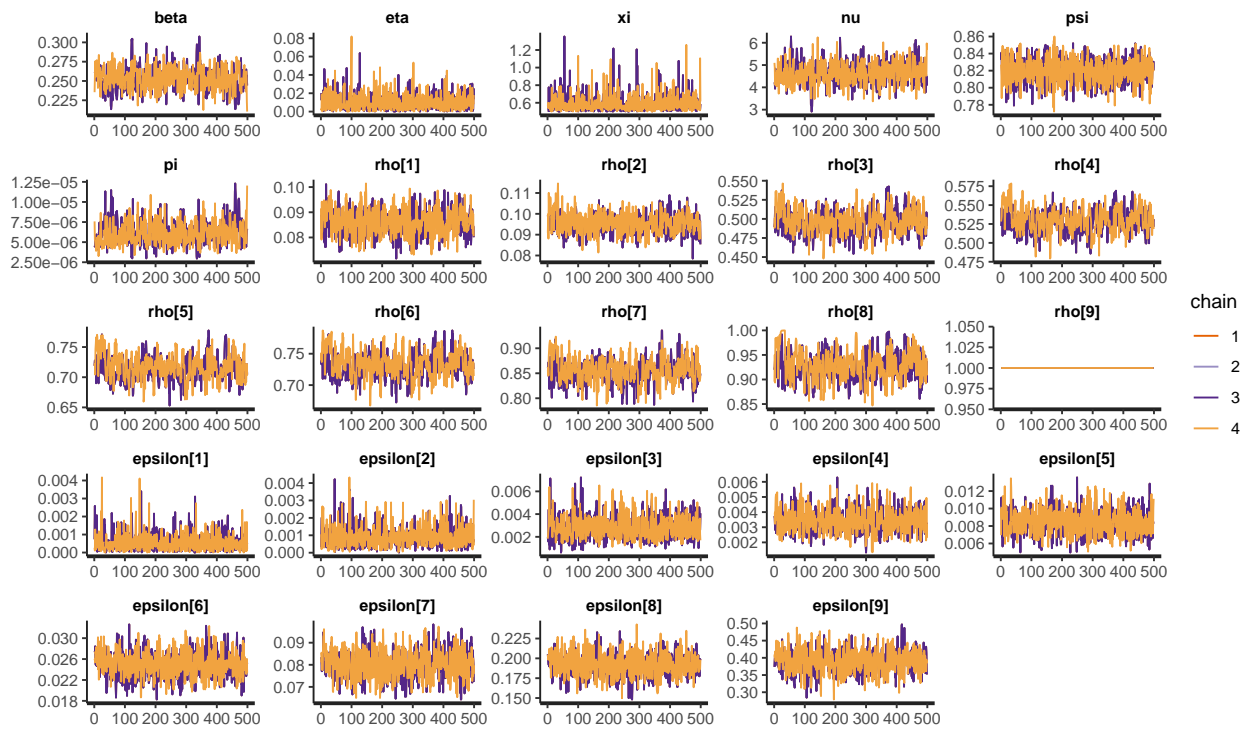


Figure S6: Trace plot of the model parameters in Hubei, China.

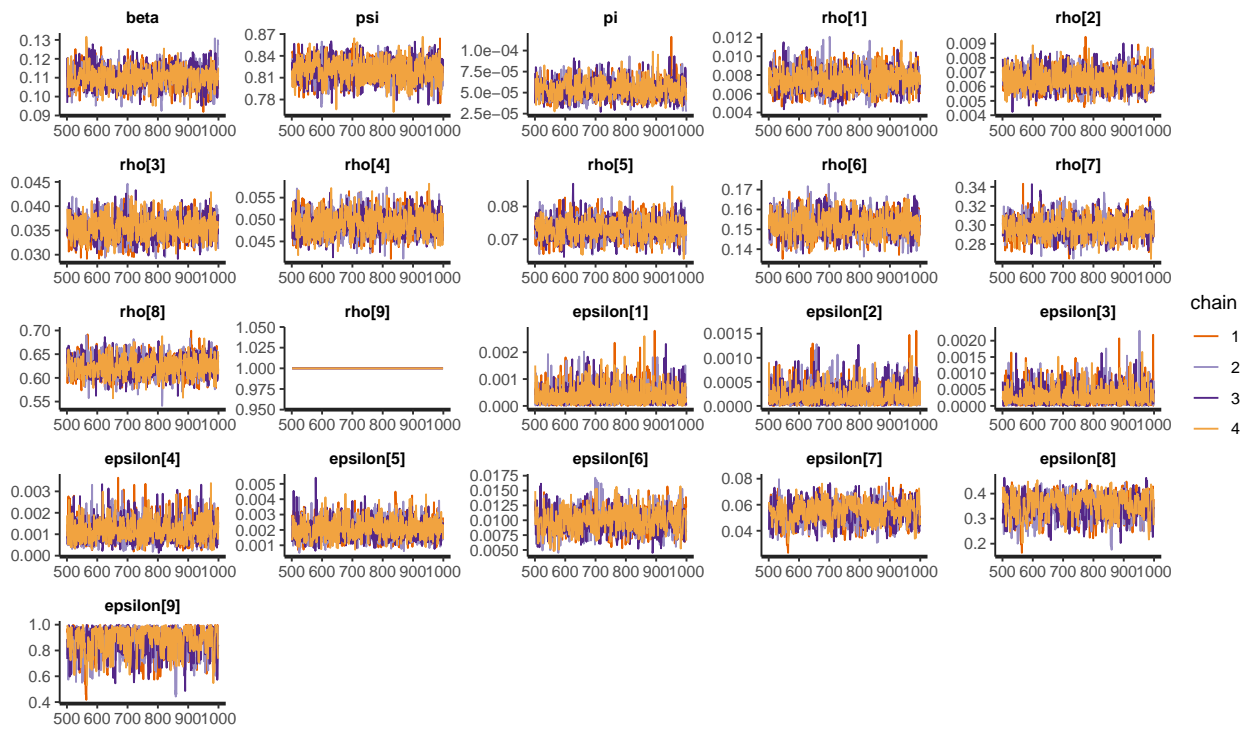


Figure S7: Trace plot of the model parameters in northern Italy.

3 External validation

Some level of external validation of our approach for correcting right-censoring in deaths could be conducted using later reports of deaths in Hubei, China (Figure S8). Visual comparison shows that prediction of deaths align well with later reports (triangles), especially if we average with zero deaths with days with a large number of reported deaths. An important caveat of this comparison is that model prediction only concerns deaths among people infected before 11 February, while later reported deaths may also concern people infected after this date. In Hubei, as the model captured the epidemic wave almost until the end, this effect is less important, but this is why external validation was not conducted for northern Italy. In any case, the good alignment of predictions and later death reports may serve as an external indication of the adequacy of the model, especially regarding the time between disease onset and deaths (this can be compared with a sensitivity analysis using a shorter duration from disease onset to death, see section 5).

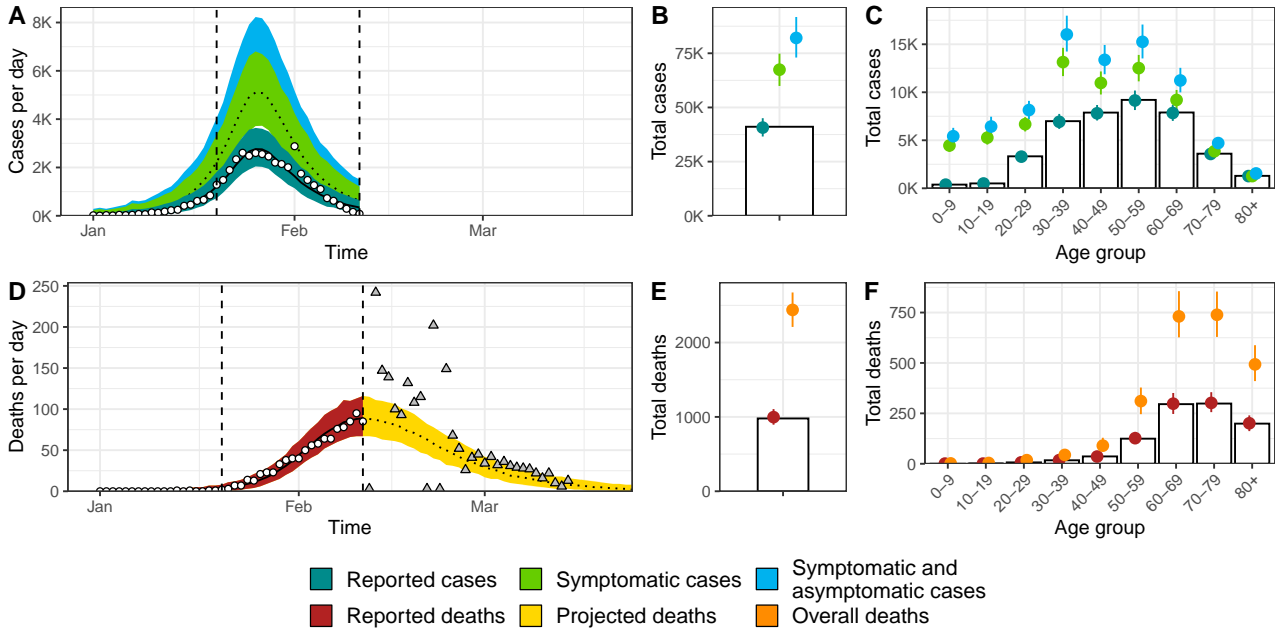


Figure S8: Model fit for Hubei, China of (A) incident cases of SARS-CoV-2 infection by date of disease onset, (B) total cases, (C) age distribution of cases, (D) incidence of deaths, (E) total deaths and (F) age distribution of deaths. White circles and bars represent data. Lines and shaded areas or points and ranges show the posterior median and 95% credible intervals for five types of model output: reported cases, symptomatic cases, overall cases (i.e. symptomatic and asymptomatic cases), reported deaths until 11 February 2020, and overall deaths including these that will occur after this date.

4 Additional results

4.1 Additional results in Hubei, China

Table 4: Posterior distributions of the parameters in Hubei, China.

Parameter	Interpretation	Posterior (median and 95%CrI)
β	Probability of transmission upon contact	0.25 [0.22-0.28]
ψ	Proportion of symptomatic infections	0.82 [0.79-0.85]
π	Initial proportion of infected in the population	0.000006 [0.000004- 0.000010]
η	Reduction of transmission due to control measures	0.008 [0-0.031]
ν	Delay of implementation of control measures (days)	4.6 [3.7-5.7]
ξ	Slope of implementation of control measures	0.56 [0.50-0.87]

Table 5: Posterior distributions of the age-specific parameters in Hubei, China: the proportion of ascertainment ρ_k and the proportion of deaths among symptomatics ϵ_k

Age group	ρ_k	ϵ_k
0-9	0.09 [0.08-0.1]	0 [0-0.002]
10-19	0.09 [0.09-0.1]	0.001 [0-0.002]
20-29	0.49 [0.47-0.53]	0.003 [0.001-0.005]
30-39	0.53 [0.5-0.56]	0.003 [0.002-0.005]
40-49	0.71 [0.68-0.76]	0.008 [0.006-0.011]
50-59	0.73 [0.69-0.77]	0.025 [0.02-0.03]
60-69	0.85 [0.8-0.9]	0.08 [0.068-0.093]
70-79	0.93 [0.87-0.99]	0.191 [0.164-0.219]
80+	1	0.386 [0.325-0.454]

Table 6: Age-specific mortality associated with infection by SARS-CoV-2 in Hubei up to 11 February 2020.

Age group	Crude	Among symptomatics	Among all infected
0-9	0.002 [0-0.011]	0 [0-0.002]	0 [0-0.001]
10-19	0.002 [0-0.014]	0.001 [0-0.002]	0.001 [0-0.002]
20-29	0.002 [0.001-0.005]	0.003 [0.001-0.005]	0.002 [0.001-0.004]
30-39	0.003 [0.001-0.005]	0.003 [0.002-0.005]	0.003 [0.002-0.004]
40-49	0.005 [0.003-0.007]	0.008 [0.006-0.011]	0.007 [0.005-0.009]
50-59	0.014 [0.01-0.018]	0.025 [0.02-0.03]	0.02 [0.017-0.025]
60-69	0.038 [0.031-0.046]	0.08 [0.068-0.093]	0.065 [0.055-0.076]
70-79	0.084 [0.069-0.102]	0.191 [0.164-0.219]	0.157 [0.133-0.181]
80+	0.159 [0.123-0.199]	0.386 [0.325-0.454]	0.316 [0.264-0.377]

4.2 Additional results in northern Italy

Table 7: Posterior distributions of the parameters in northern Italy.

Parameter	Interpretation	Posterior (median and 95%CrI)
β	Probability of transmission upon contact	0.11 [0.10-0.12]
ψ	Proportion of symptomatic infections	0.82 [0.79-0.85]
π	Initial proportion of infected in the population	5.1×10^{-5} [3.4×10^{-5} - 7.6×10^{-5}]

Table 8: Posterior distributions of the age-specific parameters in northern Italy: the proportion of ascertainment ρ_k and the proportion of deaths among symptomatics ϵ_k

Age group	ρ_k	ϵ_k
0-9	0.01 [0.01-0.01]	0 [0-0.001]
10-19	0.01 [0.01-0.01]	0 [0-0.001]
20-29	0.04 [0.03-0.04]	0 [0-0.001]
30-39	0.05 [0.04-0.05]	0.001 [0-0.002]
40-49	0.07 [0.07-0.08]	0.002 [0.001-0.004]
50-59	0.15 [0.14-0.16]	0.01 [0.006-0.014]
60-69	0.3 [0.28-0.32]	0.054 [0.034-0.069]
70-79	0.62 [0.58-0.67]	0.357 [0.224-0.427]
80+	1 [1-1]	0.89 [0.562-0.996]

Table 9: Age-specific mortality associated with infection by SARS-CoV-2 in northern Italy up to 3 March 2020.

Age group	Crude	Among symptomatics	Among all infected
0-9	0 [0-0.038]	0 [0-0.001]	0 [0-0.001]
10-19	0 [0-0.02]	0 [0-0.001]	0 [0-0.001]
20-29	0 [0-0.005]	0 [0-0.001]	0 [0-0.001]
30-39	0 [0-0.009]	0.001 [0-0.002]	0.001 [0-0.002]
40-49	0.001 [0-0.007]	0.002 [0.001-0.004]	0.002 [0.001-0.003]
50-59	0.004 [0-0.012]	0.01 [0.006-0.014]	0.008 [0.005-0.011]
60-69	0.013 [0.003-0.034]	0.054 [0.034-0.069]	0.044 [0.028-0.057]
70-79	0.044 [0.014-0.109]	0.357 [0.224-0.427]	0.293 [0.182-0.353]
80+	0.077 [0.024-0.181]	0.89 [0.562-0.996]	0.729 [0.461-0.826]

5 Sensitivity analyses

5.1 Ascertainment in the highest age class

In the main analysis, our results depended on the assumption that older individuals that have more severe symptoms are very likely to be identified (ascertainment proportion of 100% for people aged 80 and more with symptoms). In the absence of an outside reference point, the reporting rate cannot be estimated from surveillance data only, and an assumption has to be made. In this sensitivity analysis, we make the assumption that only 50% of all people aged 80 years and more that were infected with SARS-CoV-2 in Hubei in that period and had symptoms were identified and reported as cases. This modification leads to a linear scaling of the results by a factor 2:

- estimated total number of symptomatic cases: 134,700 (95%CrI: 120,200-150,700) total symptomatic cases instead of 67,400;
- estimated total number of symptomatic and asymptomatic cases: 164,400 (95%CrI: 145,000-185,200) instead of 82,300;
- adjusted mortality among infected individuals with symptoms: 1.8% (95%CrI: 1.6-2.1) instead of 3.6% (95%CrI: 3.1-4.1);
- adjusted mortality among all infected individuals: 1.5% (1.3-1.7) instead of 3.0% (95%CrI: 2.6-3.4).

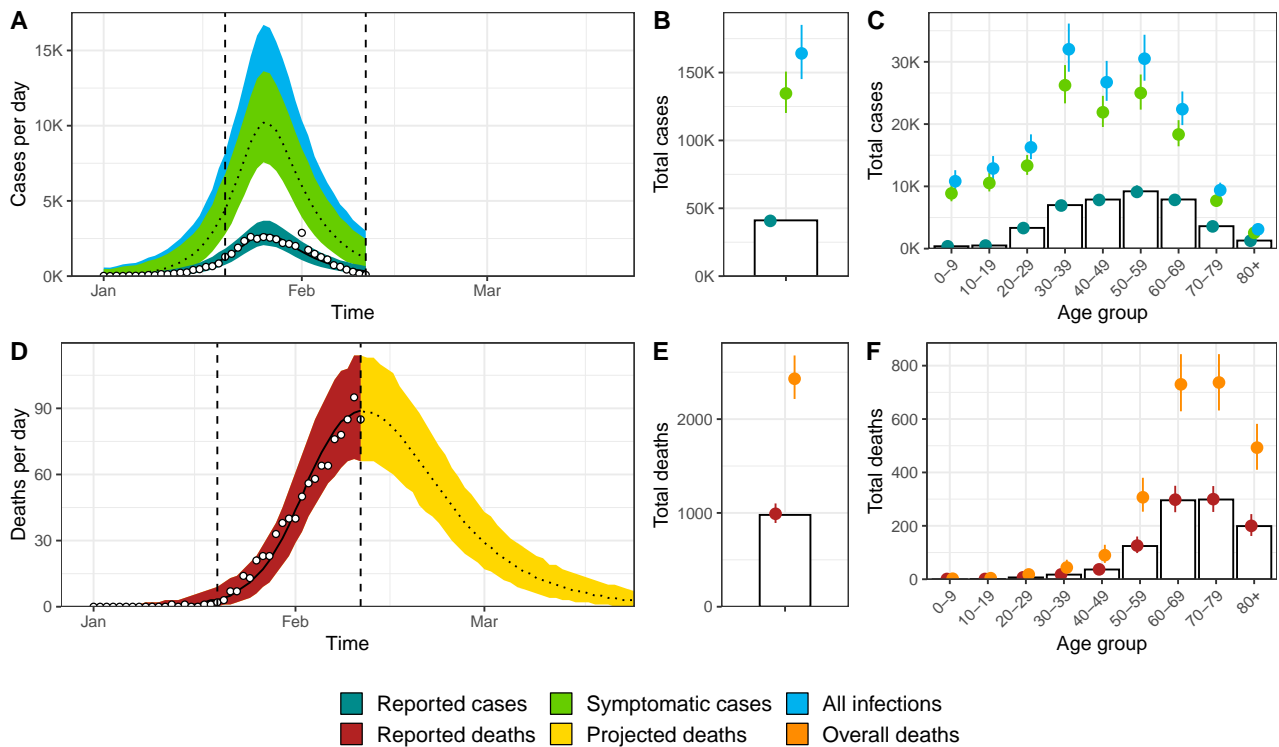


Figure S9: Model fit for Hubei, China **with a fixed ascertainment proportion of older symptomatic patients of 50% instead of 100%**. (A) Incident cases of SARS-CoV-2 infection by date of disease onset, (B) total cases, (C) age distribution of cases, (D) incidence of deaths, (E) total deaths and (F) age distribution of deaths. White circles and bars represent data. Lines and shaded areas or points and ranges show the posterior median and 95% credible intervals for six types of model output: reported cases, symptomatic cases, overall cases (i.e. symptomatic and asymptomatic cases), reported deaths until 11 February 2020, projected deaths after 11 February 2020 and overall deaths.

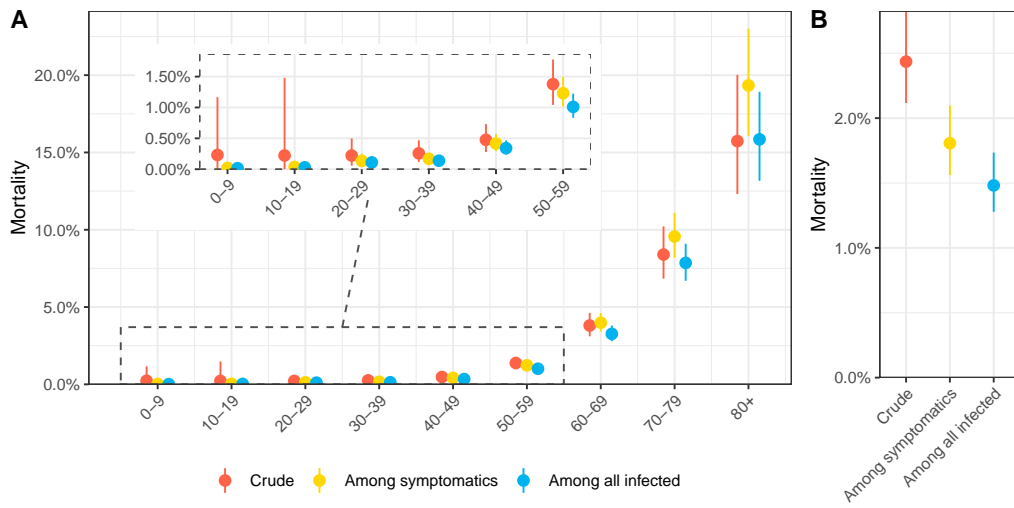


Figure S10: (A-B) Estimates of mortality during the first wave of the SARS-CoV-2 epidemic in Hubei, China by age group and overall **with a fixed ascertainment proportion of older symptomatic patients of 50% instead of 100%**. (C) Observed prevalence four comorbidities among deaths associated with SARS-CoV-2 infection in Hubei (purple) compared to the expected prevalence given the age distribution of deaths and the age-specific prevalence of each comorbidity in the Chinese population (white bar). Points and ranges show the posterior median and 95% credible intervals.

5.2 Reduced time between disease onset and death

In the main analysis, the delay between disease onset and death was implemented using a log-normal distribution with mean 20.2 days and standard deviation 11.6. In this sensitivity analysis, we ran the analysis in Hubei using instead a log-normal distribution with mean 15 days and standard deviation 10. This resulted in a worse fit to data and less convincing external validation (Figure S11D).

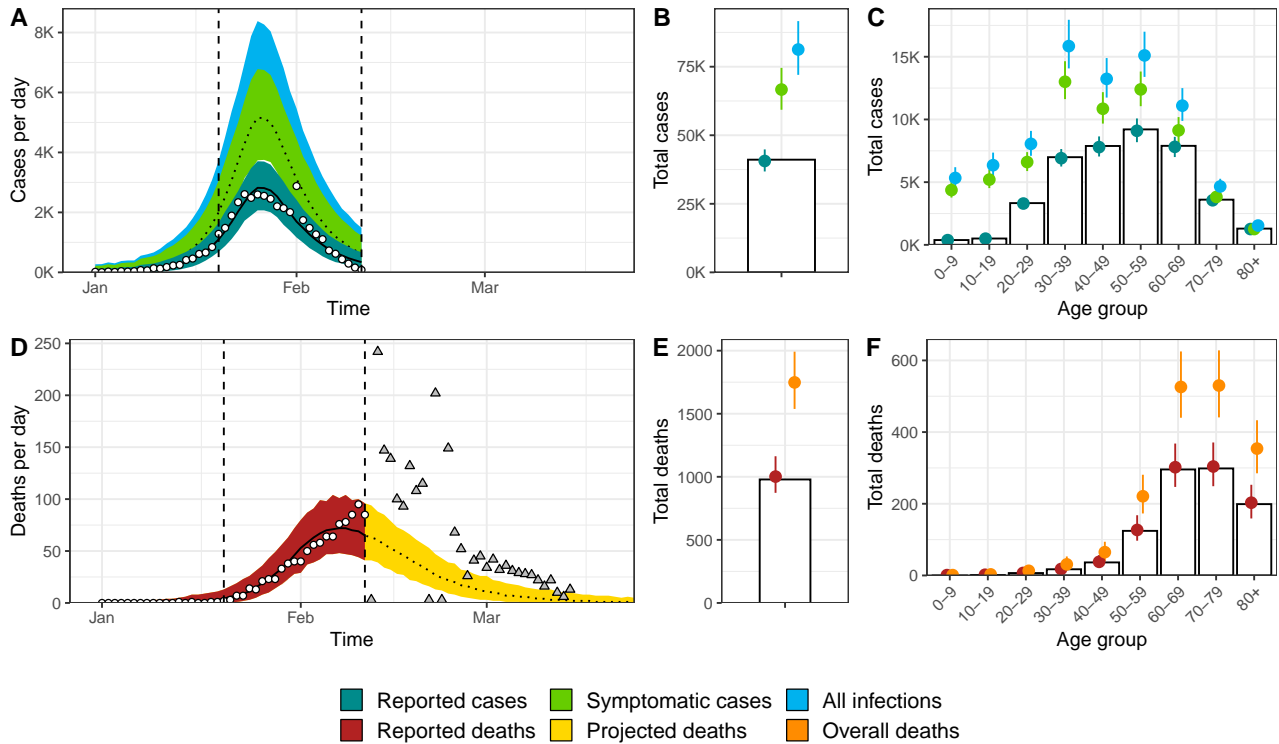


Figure S11: Model fit for Hubei, China **with a time between disease onset and death of 15 days on average instead of 20.2 days**. (A) Incident cases of SARS-CoV-2 infection by date of disease onset, (B) total cases, (C) age distribution of cases, (D) incidence of deaths, (E) total deaths and (F) age distribution of deaths. White circles and bars represent data. Lines and shaded areas or points and ranges show the posterior median and 95% credible intervals for six types of model output: reported cases, symptomatic cases, overall cases (i.e. symptomatic and asymptomatic cases), reported deaths until 11 February 2020, projected deaths after 11 February 2020 and overall deaths.

6 Stan code

Also available as model13.stan on https://github.com/jriou/covid_adjusted_cfr/.

```
functions {
real switch_eta(real t, real t1, real eta, real nu, real xi) {
return(eta+(1-eta)/(1+exp(xi*(t-t1-nu))));
}
real[] SEIR(real t,
real[] y,
real[] theta,
real[] x_r,
int[] x_i
) {
int K = x_i[1];

real tswitch = x_r[1];
real dydt[(4*K)]; //SEIAR, then C and D
real nI; // total infectious
real ntot;

real beta; // transmission rate
real eta; // reduction in transmission rate after quarantine
real xi; // slope of quarantine implementation
real nu; // shift of quarantine implementation
real tau; // incubation period
real mu; // infectious period
real psi; // probability of symptoms
real p_tswitch;
real contact[K*K]; //contact matrix
real n_by_age[K];
real f_inf[K];

real init[K*4];
real age_dist[K];
real pi; // number of cases at t0

// estimated parameters
beta = theta[1];
eta = theta[2];
xi = theta[3];
nu = theta[4];
pi = theta[5];
psi = theta[6];

// Initial conditions
for(k in 1:K){
age_dist[k] = x_r[3+K*K + k];
init[k] = age_dist[k] * (1-pi);
init[K+k] = age_dist[k] * pi;
init[2*K+k] = 0.0;
init[3*K+k] = 0.0;
}

// fixed parameters
tau = 1.0/x_r[2];
mu = 1.0/x_r[3];
contact = x_r[4:(3+K*K)];

//Total number of infectious people
p_tswitch = switch_eta(t,tswitch,eta,nu,xi);
```

```

//Force of infection by age classes:
//beta * p_tswitch * sum((number of infected people by age) / (total number of people by age) *
//(number of contact by age))
for(k in 1:K){
f_inf[k] = beta * p_tswitch * sum(to_vector(y[(2*K+1):(3*K)]) ./ to_vector(age_dist) .*
to_vector(contact[(K*(k-1)+1):(k*K)])); //
}
//print("Force of infection",f_inf);

for (k in 1:K) {
// S
dydt[k] = - f_inf[k] * (y[k]+init[k]);
// E
dydt[K+k] = f_inf[k] * (y[k]+init[k]) - tau * (y[K+k]+init[K+k]);
// I
dydt[2*K+k] = psi * tau * (y[K+k]+init[K+k]) - mu * (y[2*K+k]+init[2*K+k]);
// C
dydt[3*K+k] = psi * tau * (y[K+k]+init[K+k]);
}
return(dydt);
}
}

data {
int K; //number of age classes
vector[K] age_dist;
int pop_t; //total population
real tswitch;

//Data to fit
int D; //number of days with reported incidence
int incidence_cases[D]; //overall incidence for W weeks
int incidence_deaths[D]; //overall incidence for W weeks
int agedistr_cases[K]; //number of cases at tmax for the K age classes
int agedistr_deaths[K]; //mortality at tmax for the K age classes
//Parameters in priors
real p_beta;
real p_eta[2];
real p_pi[2];
real p_epsilon[2];
real p_rho[2];
real p_phi;
real p_xi;
real p_nu;
real p_psi[2];
// fixed parameters
real p_incubation;
real p_infectious;
int G;
real p_gamma[G];
//Simulation
real t0; //starting time
int t_data; //time of first data
int S;
real ts[S]; // time bins
real contact[K*K];

int inference;
int doprint;
}

```

```

transformed data {
real x_r[3+K*K+K]; //4 parameters + K*K contact matrix parameters + K age_dist parameters
int x_i[1] = {K};
real init[K*4] = rep_array(0.0, K * 4); // initial values
x_r[1] = tswitch;
x_r[2] = p_incubation;
x_r[3] = p_infectious;
x_r[4:(3+K*K)] = contact;
for(k in 1:K) {
x_r[3 + K*K + k] = age_dist[k];
}
}

parameters{
real<lower=0,upper=1> beta; // base transmission rate
real<lower=0,upper=1> eta; // reduction in transmission rate after quarantine measures
vector<lower=0,upper=1> [K] epsilon; // age-dependent mortality probability
vector<lower=0,upper=1> [K-1] raw_rho; // age-dependent reporting probability
real<lower=0, upper=1> pi; // number of cases at t0
real<lower=0> phi[2]; // variance parameters
real<lower=0,upper=1> xi_raw; // slope of quarantine implementation
real<lower=0> nu; // shift of quarantine implementation
real<lower=0,upper=1> psi; // proportion of symptomatics
}

transformed parameters {
// transformed parameters
vector[K] rho;
real xi = xi_raw+0.5;
// change of format for integrate_ode_rk45
real theta[6]; // vector of parameters
real y[S,K*4]; // raw ODE output
vector[K] comp_S[S];
vector[K] comp_E[S];
vector[K] comp_I[S];
vector[K] comp_D[S];
vector[K] comp_diffD[S];
vector[K] comp_C[S+G];
vector[K] comp_diffC[S+G];
vector[K] comp_diffM[S+G];
vector[K] comp_M[S+G];
// outcomes
vector[D] output_incidence_cases; // overall case incidence by day
vector[D] output_incidence_deaths; // overall mortality incidence by day
simplex[K] output_agedistr_cases; // final age distribution of cases
simplex[K] output_agedistr_deaths; // final age distribution of deaths

// transformed parameters
for(i in 1:(K-1)){
rho[i]=raw_rho[i];
}
rho[K] = 1.0;
// change of format for integrate_ode_rk45
theta[1:6] = {beta,eta,xi,nu,pi,psi};
// run ODE solver
y = integrate_ode_bdf(
SEIR, // ODE function
init, // initial states
t0, // t0
ts, // evaluation dates (ts)
theta, // parameters

```

```

x_r, // real data
x_i, // integer data
1.0E-10, 1.0E-10, 1.0E3); // tolerances and maximum steps
// extract and format ODE results (1.0E-9 correction to avoid negative values due to unprecise estimates
// of zeros as tolerance is 1.0E-10)
for(i in 1:S) {
comp_S[i] = (to_vector(y[i,1:K]) + to_vector(age_dist) * (1-pi) + 1.0E-9) * pop_t;
comp_E[i] = (to_vector(y[i,(K+1):(2*K)]) + to_vector(age_dist) * pi + 1.0E-9) * pop_t;
comp_I[i] = (to_vector(y[i,(2*K+1):(3*K)]) + 1.0E-9) * pop_t;

comp_C[i] = (to_vector(y[i,(3*K+1):(4*K)]) + 1.0E-9) * pop_t;
// lagged difference of cumulative incidence of symptomatics
comp_diffC[i] = i==1 ? comp_C[i,] : 1.0E-9*pop_t + comp_C[i,] - comp_C[i-1,];
}
//Incidence and cumulative incidence after S
for(g in 1:G){
comp_C[S+g] = comp_C[S];
comp_diffC[S+g] = rep_vector(1.0E-9,K);
}
//Mortality
//set diffM and M to 0
for(i in 1:(G+S)){
comp_diffM[i] = rep_vector(1.0E-9,K);
}
//compute mortality
for(i in 1:S) {
for(g in 1:G) {
comp_diffM[i+g] += comp_diffC[i] .* epsilon * p_gamma[g] ;
}
}
// cumulative sum
for(i in 1:(S+G)) {
for(k in 1:K) {
comp_M[i,k] = sum(comp_diffM[1:i,k]);
}
}
//compute D and diffD
for(i in 1:S){
comp_D[i] = (1.0-psi)/psi * comp_C[i];
comp_diffD[i] = (1.0-psi)/psi * comp_diffC[i];
}
// compute outcomes (again, 1.0E-9 correction to avoid negative values in the lagged differences)
for(i in t_data:S){
output_incidence_cases[i-t_data+1] = sum(comp_diffC[i].*rho);
output_incidence_deaths[i-t_data+1] = sum(comp_diffM[i]);
}
output_agedistr_cases = (comp_C[S,].*rho) ./ sum(comp_C[S,].*rho);
output_agedistr_deaths = (comp_M[S,]) ./ sum(comp_M[S,]);
}
model {
// priors
beta ~ beta(p_beta,p_beta);
eta ~ beta(p_eta[1],p_eta[2]);
for(k in 1:K) epsilon[k] ~ beta(p_epsilon[1],p_epsilon[2]);
for(k in 1:(K-1)) raw_rho[k] ~ beta(p_rho[1],p_rho[2]);
pi ~ beta(p_pi[1],p_pi[2]);
phi ~ exponential(p_phi);
xi_raw ~ beta(p_xi,p_xi);
nu ~ exponential(p_nu);
psi ~ beta(p_psi[1],p_psi[2]);
// debug

```

```

if(doprint==1) {
print("beta: ",beta);
print("eta: ",beta);
print("epsilon: ",epsilon);
print("rho: ",rho);
print("pi: ",pi);
print("y[5,]: ",y[5,]);
print("comp_C[5,]: ",comp_C[5,]);
print("comp_diffC[5,]: ",comp_diffC[5,]);
print("comp_M[5,]: ",comp_M[5,]);
print("comp_diffM[5,]: ",comp_diffM[5,]);
}
// likelihood
if (inference==1) {
for(i in 1:D) {
target += neg_binomial_2_lpmf( incidence_cases[i] | output_incidence_cases[i],
output_incidence_cases[i]/phi[1]);
target += neg_binomial_2_lpmf( incidence_deaths[i] | output_incidence_deaths[i],
output_incidence_deaths[i]/phi[2]);
}
target += multinomial_lpmf(agedistr_cases | output_agedistr_cases);
target += multinomial_lpmf(agedistr_deaths | output_agedistr_deaths);
}
}
}

```

References

- [1] Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. medRxiv. 2020 mar;p. 2020.03.03.20028423.
- [2] Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*;
- [3] Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung Sm, et al. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *Journal of Clinical Medicine*. 2020;9(2):538.
- [4] Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance*. 2020;25(5).
- [5] World Health Organization-China Joint Mission on Coronavirus Disease 2019 Group. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19); 2020. Accessed: 2020-03-22. <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>.
- [6] Team TNCPERE. Vital Surveillances: The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China, 2020. *China CDC Weekly*. 2020;2(8):113–122. Available from: <http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51>.
- [7] Zhang J, Klepac P, Read JM, Rosello A, Wang X, Lai S, et al. patterns of human social contact and contact with animals in Shanghai, china. *Scientific reports*. 2019;9(1):1–11.
- [8] Institute for Health Metrics and Evaluation. Global Health Data Exchange; 2020. Accessed: 2020-03-22. http://ghdx.healthdata.org/ihme_data.
- [9] Civile PdCdMDdP. GitHub repository on COVID-19;. Accessed: 2020-03-22. <https://github.com/pcm-dpc>.
- [10] Istituto Superiore di Sanita. COVID-19 integrated surveillance: the main national data;. Accessed: 2020-03-22. <https://www.epicentro.iss.it/coronavirus/sars-cov-2-sorveglianza-dati>.
- [11] info W. Age-distribution in China;. Available from: <https://www.worldometers.info/demographics/china-demographics/>.

- [12] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*. 2008;5(3).
- [13] UNCTAD. Handbook of Statistics;. Accessed: 2020-03-22. <https://unctadstat.unctad.org/EN/Index.html>.
- [14] Zhang, Juanjuan, Klepac, Petra, Read, M J, et al. Patterns of human social contact and contact with animals in Shanghai and China. *Scientific Reports*. 2019 dec;9(1):1–11.
- [15] Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung Sm, et al. Epidemiological characteristics of novel coronavirus infection: A statistical analysis of publicly available case data. *medRxiv*. 2020;.
- [16] Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung Sm, et al. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *Journal of Clinical Medicine*. 2020;9(2):538.
- [17] Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*. 2020;25(10). Available from: <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.10.2000180>.
- [18] Japanese Ministry of Health. Situation report of 10 March 2020; 2020. Accessed: 2020-03-22. https://www.mhlw.go.jp/stf/newpage_10094.html.
- [19] Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. *Journal of statistical software*. 2017;76(1).