

Supplementary appendix to:
Adjusted age-specific case fatality ratio during the COVID-19 epidemic in Hubei,
China, January and February 2020

March 9, 2020

1 Introduction

We model the dynamics of the COVID-19 epidemic in Hubei by date of disease onset from 1 January 2020 to 11 February 2020. 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/.

2 Data

Table 1: Summary of data sources.

| Symbol | Comment | Value | Unit | Source |
|----------|---|-----------------------|----------------|----------|
| A | Confirmed cases of COVID-19 in Hubei from 1 January to 11 February 2020 | - | - | - |
| B | Age distribution of all confirmed cases in China up to 11 February 2020 (85% of these are in Hubei, we assume the age distribution is the same in overall China) | - | - | - |
| C | Deaths among patients with confirmed COVID-19 infection by day of report in China from 1 January to 11 February 2020 (95% of these are in Hubei, so the numbers were scaled down by 5%) | - | - | - |
| D | Age distribution of the deaths among patients with confirmed COVID-19 infection that occurred in China up to 11 February 2020 (we assume the age distribution is the same in overall China) | - | - | - |
| E | Age distribution of the population in China | - | - | - |
| F | 9×9 matrix with the potentially infectious contacts between each age class in Shanghai | - | Daily contacts | - |
| G | Distribution of time from disease onset to death | lognormal(20.2, 11.6) | Days | - |
| $1/\tau$ | Incubation period | 5.95 | Days | Bi et al |
| ψ | Proportion of symptomatics | Beta(71, 18) | - | Bi et al |
| $1/\mu$ | Time from disease onset to isolation | 2.4 | Days | Bi et al |

3 Model

3.1 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 Hubei province (59,020,000), so that the sum of all compartments is 1. Susceptible individuals may become infected through contact with infected individuals with symptoms (assuming that both incubating and 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. We use the following logistic function:

$$f(t, \eta, \nu, \xi) = \eta + \frac{1 - \eta}{1 + \exp(\xi * (t - t_1 - \nu))} \quad (1)$$

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 23 January 2020) until the control measures are 50% effective.

The force of infection λ_k can then be expressed as:

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

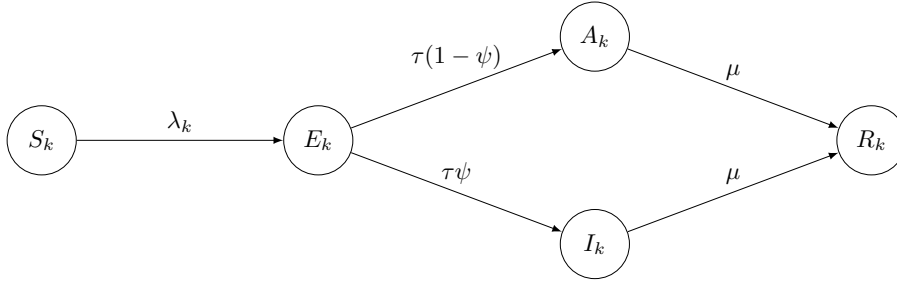


Figure 1: Schematic description of the COVID-19 transmission model.

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: $\{\beta, \tau, \psi, \mu\}$

3.2 Initial conditions and simulation settings

The simulation start ($t = 0$) is set on 30 December 2019, just before the closure of the market in Wuhan, after which all infections can be assumed to come from human-to-human transmission. At this date, most individuals are assumed to be susceptible and are distributed across age groups according to the population age distribution in China \mathbb{E} . The number of people in the exposed compartment at $t = 0$ is controlled by a parameter π , and also distributed according to \mathbb{E} . Other compartments are set to 0.

3.3 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:

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

and the cumulative incidence of new asymptomatic infections::

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

Outputs from the ODE model include:

- the total number of new infections per day in each age group:

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

- the total number of new reported infections per day, summed over age groups, according to an age-specific reporting rate ρ_k :

$$E_t = \sum_k^9 \rho_k D_{k,t} \quad (6)$$

- the age distribution of all reported cases between simulation start and 11 February 2020 ($t = 43$):

$$F_k = \frac{\rho_k C_k^I(43)}{\sum_k^9 \rho_k C_k^I(43)} \quad (7)$$

- the total number of infections with symptoms:

$$G = \sum_k^9 C_k^I(43) \quad (8)$$

- the total number of infections with or without symptoms:

$$H = \sum_k^9 C_k^I(43) + C_k^A(43) \quad (9)$$

To allow identifiability of the parameters, the reporting rate of the highest age group ρ_9 is fixed to 1. This assumes that all individuals aged 80 and more that are infected with symptoms seek medical care, are identified and reported to the authorities.

3.4 Delayed mortality

Mortality is considered outside of the system of ODEs, using an age-specific mortality ϵ_k . In accordance to data, we only consider the mortality among cases with a date of disease onset up to 11 February 2020. We consider all infected individuals with symptoms by date of disease onset, and include the discrete distribution of time from disease onset to death \mathbb{G} . The following outputs are computed at this point:

- the daily number of deaths in age group k for each day between simulation start and 100 days after 11 February 2020 ($t = 143$), to account for the delay between disease onset and deaths:

$$M_{k,t} = \epsilon_k \sum_d^{60} D_{k,t-d} \mathbb{G}_d \quad (10)$$

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

$$N_t = \sum_k^9 M_{k,t} \quad (11)$$

- the age distribution of all deaths occurring up to $t = 43$:

$$O_k = \frac{\sum_t^{43} M_{k,t}}{\sum_k^9 \sum_t^{43} M_{k,t}} \quad (12)$$

- the total number of deaths occurring up to 11 February 2020:

$$P = \sum_k^9 \sum_t^{43} M_{k,t} \quad (13)$$

- the total number of deaths occurring among cases with disease onset up to 11 February 2020 ($t = 43$), which can occur up to 100 days after 11 February 2020 ($t = 143$):

$$Q = \sum_k^9 \sum_t^{143} M_{k,t} \quad (14)$$

3.5 Inference

The objective of this step is to estimate the following parameters: $\theta = \{\beta, \eta, \xi, \nu, \psi, \epsilon_k, \rho_k, \pi, \phi_1, \phi_2\}$. We set weakly-informative priors for these parameters.

Table 2: Summary of data sources.

| Symbol | Comment | Prior |
|--------------|--|-----------------------------------|
| β | Probability of transmission upon contact | Exp(1) Should be beta(1,1) |
| η | Reduction in transmission after control measures | Beta(1, 1) |
| ξ | Slope of logistic function for control measures implementation | $0.5 + \text{Beta}(1, 1)$ |
| ν | Shift of logistic function for control measures implementation | Exp(1, 1) |
| ψ | Proportion of symptomatics | Beta(71, 18) |
| ϵ_k | Proportion of deaths among symptomatic cases (by age group) | Beta(1, 1) |
| ρ_k | Reporting rate of symptomatic cases (by age group) | Beta(1, 1) |
| π | Initial proportion of exposed (at $t = 0$) | Beta(1, 999) |
| ϕ_1 | Overdispersion for cases | Exp(1/100) |
| ϕ_2 | Overdispersion for deaths | Exp(1/100) |

We simultaneously fitted our model to four data sets:

1. the number of confirmed cases of CIVD-19 by day of disease onset from January 1 to February 11:

$$\Pr(\theta|\mathbb{A}) = \text{Neg-Bin}(\mathbb{A}|E, \phi_1) \quad (15)$$

2. the age distribution of all confirmed cases up to 11 February 2020:

$$\Pr(\theta|\mathbb{B}) = \text{Multinomial}(\mathbb{B}|F) \quad (16)$$

3. the number of deaths among patients with confirmed COVID-19 infection by day of report from January 1 to February 11:

$$\Pr(\theta|\mathbb{C}) = \text{Neg-Bin}(\mathbb{C}|N, \phi_2) \quad (17)$$

4. the age distribution of the deaths among patients with confirmed COVID-19 infection that occurred up to 11 February 2020 in China:

$$\Pr(\theta|\mathbb{D}) = \text{Multinomial}(\mathbb{D}|O) \quad (18)$$

3.6 Final outputs

| Output | Formula |
|--|--------------|
| <i>Case fatality ratio among symptomatic infections</i> | |
| Crude | $P / \sum E$ |
| Adjusted for delayed mortality | $Q / \sum E$ |
| Adjusted for unidentified symptomatic cases | P / G |
| Adjusted for both | Q / G |
| <i>Case fatality ratio among all symptomatic and asymptomatic infections</i> | |
| Adjusted for both | Q / H |