

# Age Associated Coronavirus Disease 2019 Incubation Period: Impact on Quarantine Policy

S1. Flow diagram showing selection of confirmed cases for data analysis

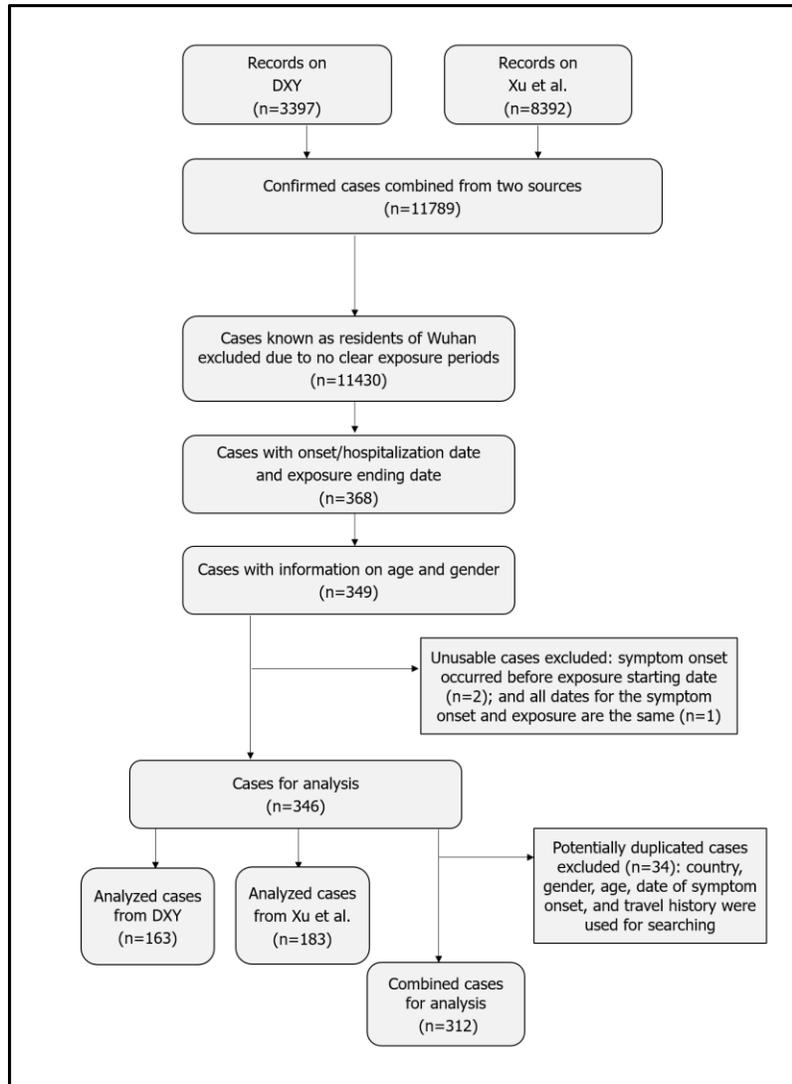


Figure S1. Selection of confirmed cases for data analysis. Two data sources were combined from DXY<sup>5</sup> and Xu et al<sup>6</sup>

## S2. Exposure period and symptom onset for cases with complete information on the exposure period

Patients with starting date of exposure available (n=111)

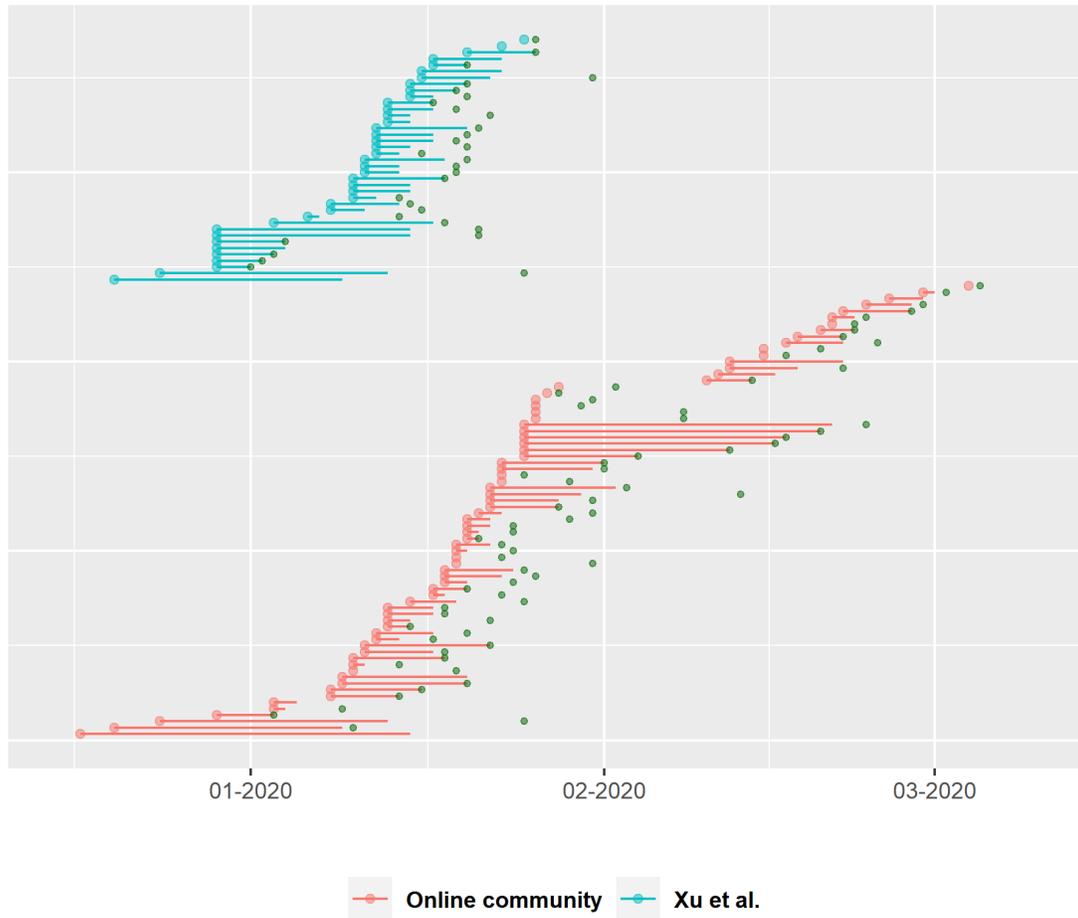


Figure S2. Segments represent the exposure period for online community from DXY (red, n=69) and Xu et al (blue, n=39). Dark green points represent the symptoms onset, when available.

### S3. Modelling interval-censored exposure duration and possibly left-censored symptom onset

Let  $T$  be the incubation period and  $\mathbf{X}$  be a vector of patient covariates. Under the generalized odds-rate class of regression models, the density  $f_T(t|\mathbf{X})$  and survival function  $S_T(t|\mathbf{X})$

conditional on  $\mathbf{X}$  are, respectively,  $f_T(t|\mathbf{x}_i) = \frac{\phi}{\lambda} \left(\frac{t}{\lambda}\right)^{\phi-1} e^{x_i^T \boldsymbol{\beta}} \left\{1 + \rho \left(\frac{t}{\lambda}\right)^\phi e^{x_i^T \boldsymbol{\beta}}\right\}^{-\frac{1+\rho}{\rho}}$  and  $S_T(t|\mathbf{x}_i) = \left\{1 + \rho \left(\frac{t}{\lambda}\right)^\phi e^{x_i^T \boldsymbol{\beta}}\right\}^{-\frac{1}{\rho}}$ ,

where  $\boldsymbol{\beta}$  is a vector of regression parameters for  $\mathbf{X}$ . The model is equivalent to the log-logistic proportional odds model when  $\rho = 1$  and the Weibull proportional hazards model when  $\rho \rightarrow 0$ .

Let  $d$  be days of the exposure duration (the difference in days between the starting date and the ending date). We let  $r$  denote the days to the date of symptom onset and  $h$  denote the days to the date of hospitalization, from the ending date of the exposure period. If a patient's symptom onset date was observed, let  $\delta_s = 1$ ; otherwise,  $\delta_s = 0$ . If a patient had  $d > 0$ , let  $\delta_d = 1$ ; otherwise,  $\delta_d = 0$ . We also let  $A$  denote the periods from the date of infection to the ending date of the exposure period, which is unknown and assumed to occur equally likely during the potential exposure.

The likelihood for each case in Figure 1 is defined as,

- $\delta_s = 1$  and  $\delta_d = 1$  (Case I)

$$L_1 = \int_0^d f_T(a+r|\mathbf{x}) da = \frac{1}{d} \{S_T(r|\mathbf{x}) - S_T(r+d|\mathbf{x})\} \\ \propto S_T(r|\mathbf{x}) - S_T(r+d|\mathbf{x}),$$

- $\delta_s = 0$  and  $\delta_d = 1$  (Case II)

$$L_3 = \int_0^d \{1 - S_T(a+h|\mathbf{X})\} da = 1 - \frac{1}{d} \int_0^d S_T(a+h|\mathbf{x}) da,$$

- $\delta_s = 1$  and  $\delta_d = 0$  (Case III)

$$L_2 = f_T(r|\mathbf{x}), \text{ and}$$

- $\delta_s = 0$  and  $\delta_d = 0$  (Case IV)

$$L_4 = 1 - S_T(h|\mathbf{x}).$$

The log-likelihood for all four types of possible patient cases is then proportional to

$$\log L \propto \sum_{i=1}^n \left[ \delta_{si} \delta_{di} \log \{S_T(r_i|\mathbf{x}_i) - S_T(r_i + d_i|\mathbf{x}_i)\} + \delta_{si} (1 - \delta_{di}) \log \{f_T(r_i|\mathbf{x}_i)\} \right. \\ \left. + (1 - \delta_{si}) \delta_{di} \log \left\{ 1 - \frac{1}{d_i} \int_0^{d_i} S_T(a_i + h_i|\mathbf{x}_i) da \right\} \right. \\ \left. + (1 - \delta_{si}) (1 - \delta_{di}) \log \{1 - S_T(h_i|\mathbf{x}_i)\} \right].$$

#### S4. Estimation of the means of the incubation period

Table S1. Estimated means for the incubation period

Age	Mean	SE	95% CI	
			Lower	Upper
Overall	7.7	0.50	6.71	8.7
≤42	6.7	0.61	5.50	7.9
>42	8.7	0.78	7.2	10.2

#### S5. Parameter estimation and likelihood ratio tests

Table S2. Parameter estimation under the generalized odds-rate class of regression models (full model), Weibull proportional hazards models with gender and age (reduced model I), and Weibull proportional hazards models with age only (reduced model II)

Model*		$\log(\lambda)$	$\log(\phi)$	$\log(\rho)$	$\beta_{male}$	$\beta_{age \leq 42}$
Full model	Estimate	2.300	0.305	-3.766	0.141	0.371
	SE	0.155	0.119	6.569	0.176	0.180
	P-value	<0.001	0.010	0.566	0.423	0.039
Reduced model I	Estimate	2.314	0.290	-	0.140	0.367
	SE	0.121	0.074	-	0.173	0.175
	P-value	<0.001	<0.001	-	0.417	0.036
Reduced model II	Estimate	2.249	0.286	-	-	0.352
	SE	0.091	0.074	-	-	0.174
	P-value	<0.001	<0.001	-	-	0.043

\*The p-values of the likelihood ratio test were 0.999 (full model vs reduced model I) and 0.724 (full model vs reduced model II).

## S6. Descriptive goodness of fit graph for the Weibull regression model

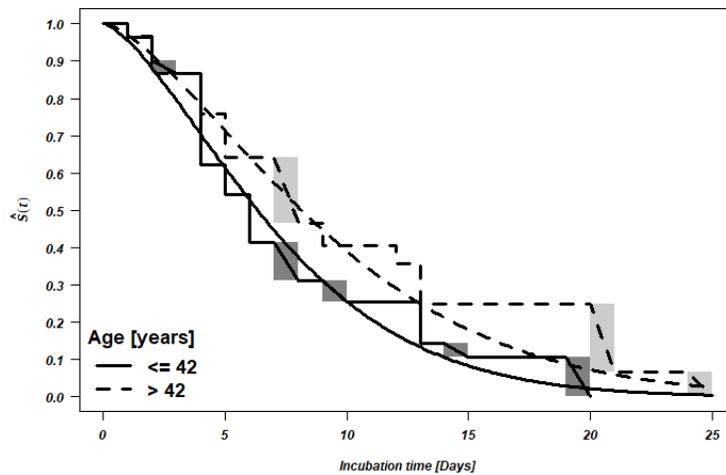


Figure S3. Nonparametric estimation (step functions) and estimation under the Weibull model of the survival functions.

## S7. Sensitivity analysis

Table S3. Estimated difference in the incubation period between two age groups (patients aged > 42 years – patients aged ≤ 42 years) at each percentile by changing the maximum exposure duration to 20 days (scenario I), and by shifting the lower bound for the starting date of exposure to December 15, 2019 (scenario II) and to November 15, 2019 (scenario III)

Scenario	Percentile				
	2.5 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	97.5 <sup>th</sup>
I	0.1 (0, 0.3)	0.7 (0, 1.4)	1.3 (0.0, 2.6)	2.2 (0.1, 4.3)	4.4 (0.1, 8.7)
II	0.1 (0, 0.3)	0.9 (0, 1.7)	1.7 (0.1, 3.2)	2.8 (0.2, 5.4)	5.8 (0.3, 11.2)
III	0.1 (0, 0.3)	0.9 (0, 1.7)	1.7 (0.1, 3.3)	2.8 (0.1, 5.5)	5.9 (0.3, 11.5)

## S8. Subgroup analysis by data sources

Table S4. Summary of two datasets from DXY and Xu et al

Source*	Age	Gender		Total
		Female	Male	
DXY	≤42	40	56	96
	>42	25	42	67
Xu et al.	≤42	42	69	111
	>42	30	42	72

\*Two sources have 34 duplicated cases (22 for age ≤42 and 12 for age >42).

Table S5. Parameter estimation from separate data sets

Source		$\log(\lambda)$	$\log(\phi)$	$\log(\rho)$	$\beta_{male}$	$\beta_{age \leq 42}$
DXY	Estimate	1.773	0.624	-1.226	-0.325	0.630
	SE	0.179	0.173	0.884	0.296	0.306
	P-value	<0.001	<0.001	0.165	0.272	0.039
Xu et al.	Estimate	2.580	0.469	-2.173	0.400	0.370
	SE	0.158	0.123	1.169	0.263	0.265
	P-value	<0.001	<0.001	0.063	0.129	0.162