Supplementary information for: Modeling the Epidemic Dynamics and Control of COVID-19 Outbreak in China

Shilei Zhao[1,2,3], Hua Chen[1,2,3,4,*]

[1] CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

[2] China National Center for Bioinformation, Beijing 100101, China

[3] School of Future Technology, University of Chinese Academy of Sciences, Beijing 100049, China

[4] CAS Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China

[*]Correspondence: chenh@big.ac.cn

Supplementary data

Table S1: Number of confirmed infections from January 20th, 2020 to February 21th, 2020

Date	Wuhan	Hubei	China	Beijing	Shanghai	Guangzhou	Shenzhen
1-20	258	270	291	5	2	-	9
1-21	363	375	440	10	9	2	14
1-22	425	444	571	22	16	5	15
1-23	495	549	830	26	20	7	15
1-24	572	729	1287	36	33	14	20
1-25	618	1052	1975	51	40	14	27
1-26	698	1423	2744	68	53	39	36
1-27	1590	2714	4515	80	66	51	49
1-28	1905	3554	5974	91	80	63	63
1-29	2261	4586	7711	111	101	79	86
1-30	2639	5806	9692	132	128	106	110
1-31	3215	7153	11791	156	153	137	170
2-1	4109	9074	14380	183	177	175	196
2-2	5142	11177	17236	191	193	189	226
2-3	6384	13522	20471	228	208	216	269
2-4	8351	16678	24363	253	233	237	289
2-5	10117	19665	28060	274	254	255	314
2-6	11618	22112	31211	297	269	284	334
2-7	13603	24953	34598	315	281	298	351
2-8	14982	27100	37251	326	292	304	364
2-9	16902	29631	40235	337	295	313	368

2-10	18454	31728	42708	342	302	317	375
2-11	19558	33366	44730	352	306	323	386
2-12	32994	48206	59882	366	313	327	391
2-13	35991	51986	63932	372	318	328	400
2-14	37914	54406	66576	375	326	335	406
2-15	39462	56249	68584	380	328	338	414
2-16	41152	58182	70635	381	331	339	415
2-17	42752	59989	72528	387	333	339	416
2-18	44412	61682	74279	393	333	339	416
2-19	45027	62457	75101	395	333	339	416
2-20	45346	63088	75993	396	334	339	416
2-21	45660	63454	76392	399	334	343	417

Data is collected from National Health Commission of China and the local Health Commissions.

Supplementary Figures

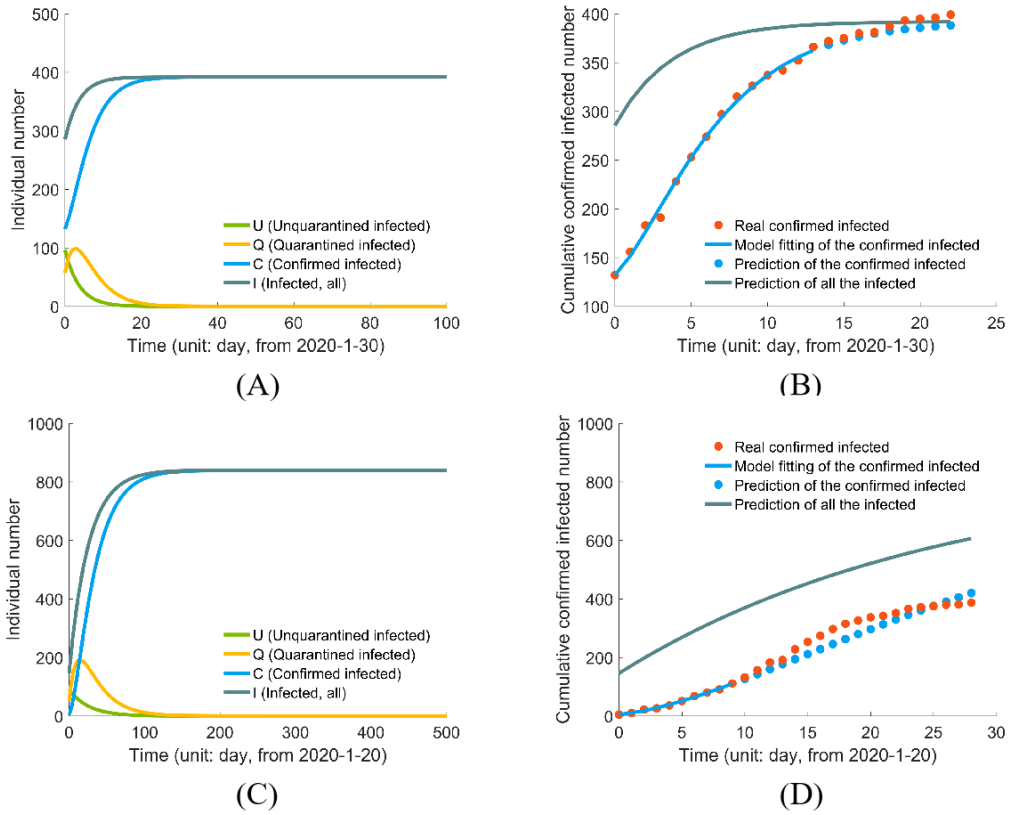

Figure S1. Inferring epidemic dynamics in Beijing: (A) prediction using stage Ⅱ data; (B)model-fitting and testing with stage Ⅱ data. The first 15 data points (from 1-30) are used to optimize the parameters, and the remaining data points are used to

test the model; (C) prediction using stage Ⅰ data; (D) model-fitting and testing with stage Ⅰ data. The first 10 data points (from 1-20) are used to optimize the parameters, the remaining data points are used to test the model.

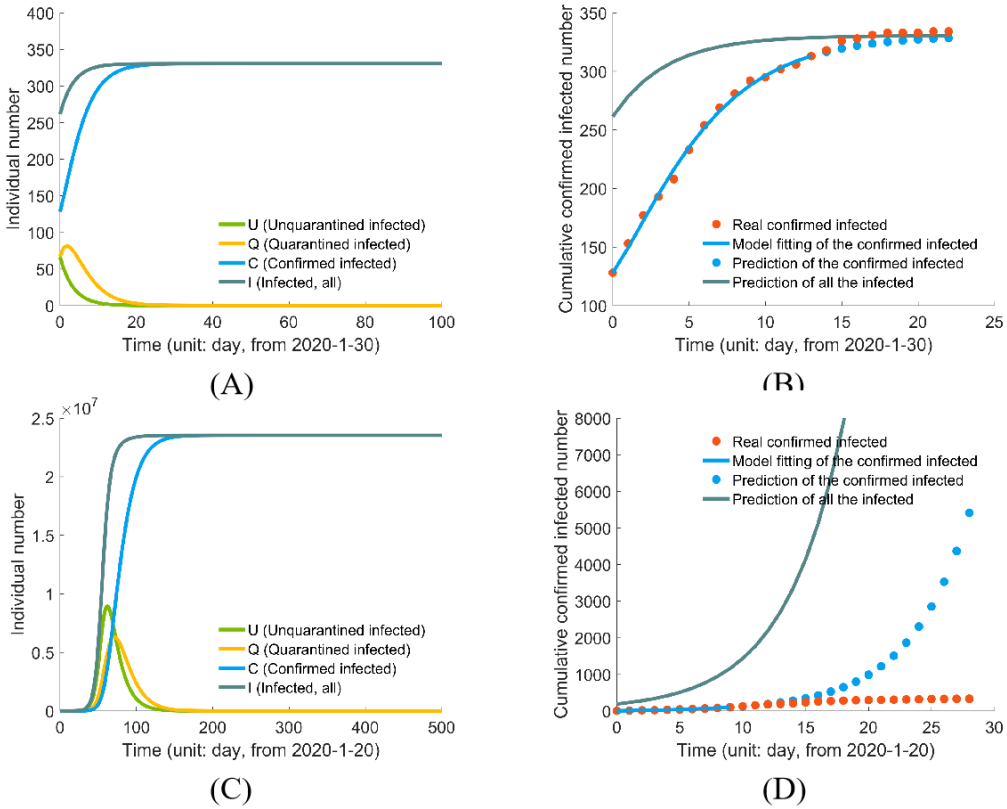

Figure S2. Inferring epidemic dynamics in Shanghai: (A) prediction using stage Ⅱ data; (B) model-fitting and testing with stage Ⅱ data. The first 15 data points (from 1-30) are used to optimize the parameters, and the remaining data points are used to test the model; (C) prediction using stage Ⅰ data; (D) model-fitting and testing with stage Ⅰ data. The first 10 data points (from 1-20) are used to optimize the parameters, the remaining data points are used to test the model.

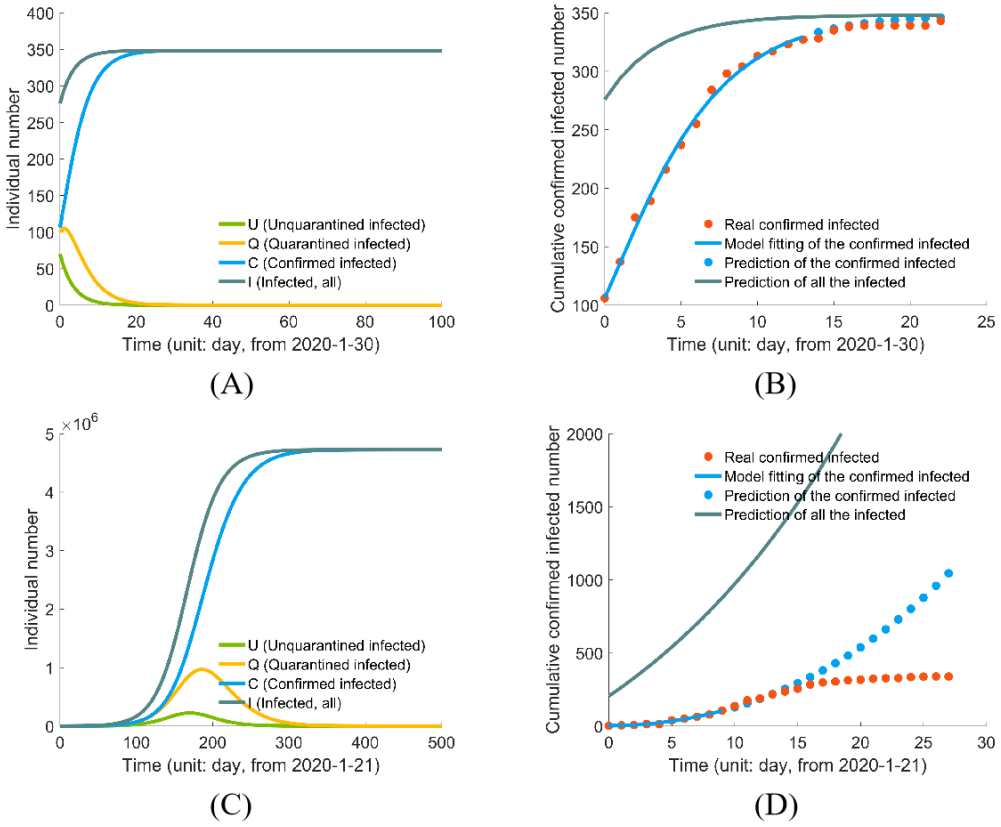

Figure S3. Inferring epidemic dynamics in Guanzhou: (A) prediction using stage Ⅱ data; (B)model-fitting and testing with stage Ⅱ data. The first 15 data points (from 1-30) are used to optimize the parameters, and the remaining data points are used to test the model; (C) prediction using stage Ⅰ data; (D) model-fitting and testing with stage Ⅰ data. The first 10 data points (from 1-20) are used to optimize the parameters, the remaining data points are used to test the model.

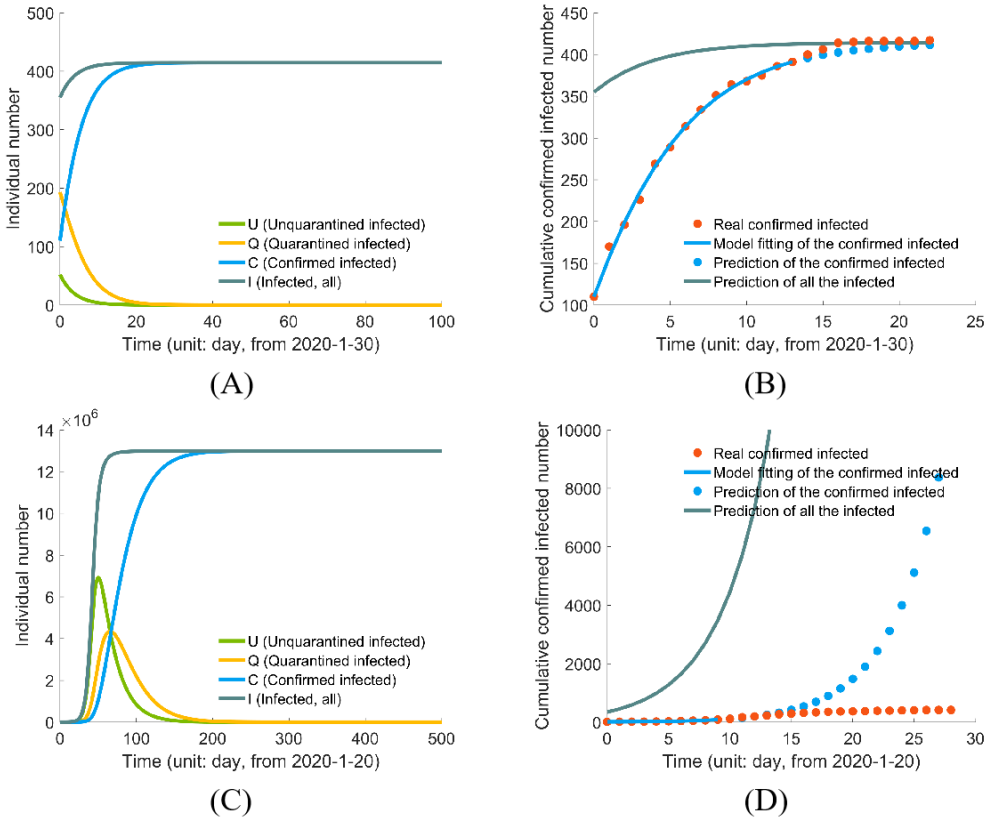

Figure S4. Inferring epidemic dynamics in Shenzhen: (A) prediction using stage Ⅱ data; (B) model-fitting and testing with stage Ⅱ data. The first 15 data points (from 1-30) are used to optimize the parameters, and the remaining data points are used to test the model; (C) prediction using stage Ⅰ data; (D) model-fitting and testing with stage Ⅰ data. The first 10 data points (from 1-20) are used to optimize the parameters, the remaining data points are used to test the model.

Comparison of loss functions

We compared three different loss functions (see Eqns S1, S2, S3 below). The first two are weighted by the confirmed infection numbers, and the third one is the un-weighted loss function we used in the paper. The results with the three loss functions are shown in Figure S5 for China (excluding Hubei) and in Figure S6 for Hubei (excluding Wuhan). The estimated values of the final infected in China (excluding Hubei) are 13638.9891, 13621.8964, and 13322.1858 respectively. The estimations seem robust. The estimations have a larger deviation in Figure R2 with the values 19814.9048, 19523.3124, 17678.7119 respectively. The weighted loss function may better integrate information across the whole epidemic. On the other hand, the observed confirmed

numbers contain more uncertainty in the early time of the epidemic for many reasons. Thus in practice, we tried different types of loss functions and chose the one according to their performance on the test data (the data points of the last several days, see Figs S5-S6 below).

The three loss functions are:

$$err(\gamma_1, \beta, U_0 Q_0) = \| (C - \hat{C}) / C \|_2 \tag{S1}$$

$$err(\gamma_1, \beta, U_0 Q_0) = \| (C - \hat{C}) / C \|_1 \tag{S2}$$

$$err(\gamma_1, \beta, U_0 Q_0) = \| (C - \hat{C}) \|_2 \tag{S3}$$

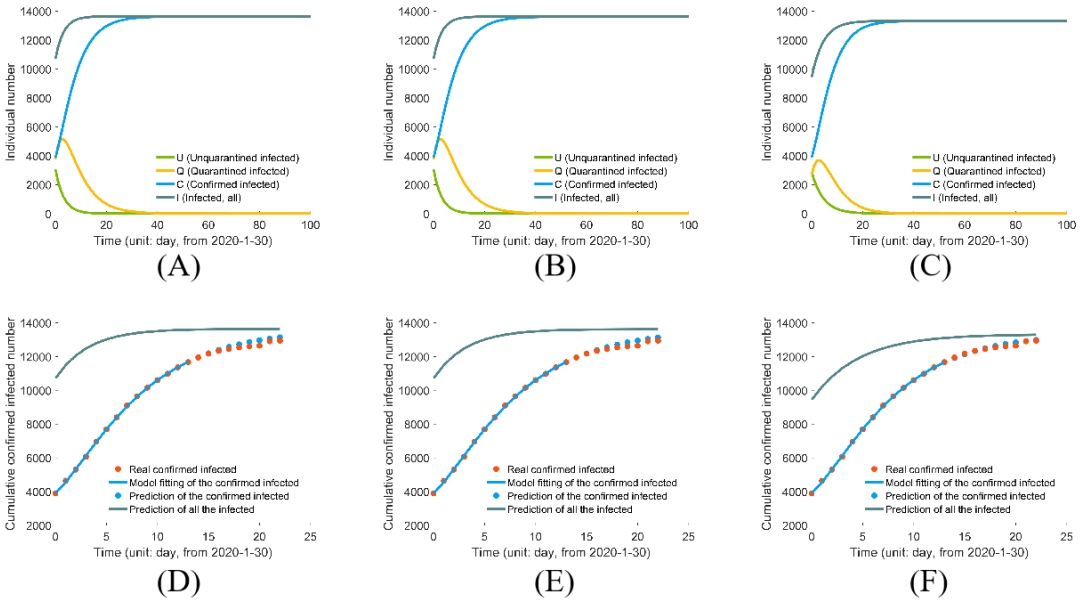

Figure S5: Comparing the estimation of the epidemic in China (excluding Hubei) using different loss functions: (A) prediction using loss function (S1); (B) prediction using loss function (S2); (C) prediction using loss function (S3); (D) model-fitting and testing with loss function (S1); (E) model-fitting and testing with loss function (S2); (F) model-fitting and testing with loss function (S3).

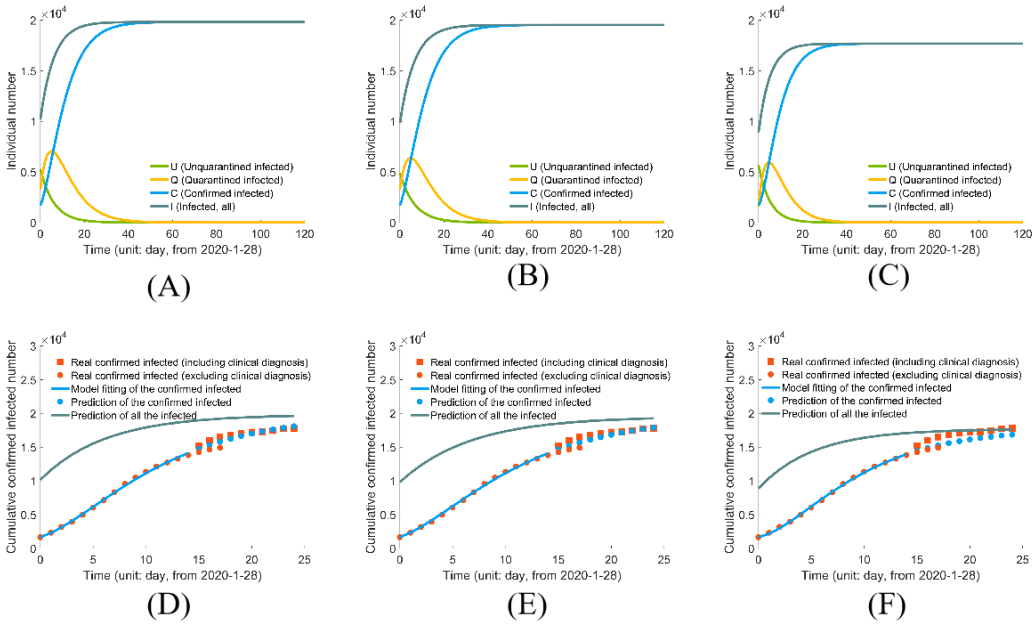

Figure S6: Comparing the estimation of the epidemic in Hubei (excluding Wuhan) using different loss functions: (A) prediction using loss function (S1); (B) prediction using loss function (S2); (C) prediction using loss function (S3); (D) model-fitting and testing with loss function (S1); (E) model-fitting and testing with loss function (S2); (F) model-fitting and testing with loss function (S3).