THE LANCET Infectious Diseases

Supplementary webappendix

This webappendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

This online publication has been corrected. The corrected version first appeared at thelancet.com on May 5, 2020.

Supplement to: Bi Q, Wu Y, Mei S, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect Dis* 2020; published online April 27. https://doi.org/10.1016/S1473-3099(20)30287-5.

SUPPLEMENTARY MATERIAL

Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts: a retrospective cohort study

Table S1: Additional clinical characteristics by mode of case detection

Symptom	Value	Contact-based (N=87)	Symptom-based (N=292)	Unknown/ other (N=12)	Total (N=391)	P-value
chill	no	85 (97.7%)	272 (93.2%)	12 (100.0%)	369 (94.4%)	0.27
	yes	2 (2.3%)	20 (6.8%)	0 (0.0%)	22 (5.6%)	
shortness of breath	no	86 (98.9%)	281 (96.2%)	12 (100.0%)	379 (96.9%)	0.42
	yes	1 (1.1%)	11 (3.8%)	0 (0.0%)	12 (3.1%)	
difficulty breathing	no	87 (100.0%)	285 (97.6%)	11 (91.7%)	383 (98.0%)	0.35
	yes	0 (0.0%)	7 (2.4%)	1 (8.3%)	8 (2.0%)	
chest tightness	no	85 (97.7%)	290 (99.3%)	12 (100.0%)	387 (99.0%)	0.17
	yes	2 (2.3%)	2 (0.7%)	0 (0.0%)	4 (1.0%)	
chest pain	no	85 (97.7%)	281 (96.2%)	12 (100.0%)	378 (96.7%)	0.77
	yes	2 (2.3%)	11 (3.8%)	0 (0.0%)	13 (3.3%)	
conjunctivitis	no	87 (100.0%)	290 (99.3%)	12 (100.0%)	389 (99.5%)	1.00
	yes	0 (0.0%)	2 (0.7%)	0 (0.0%)	2 (0.5%)	
nausea	no	87 (100.0%)	292 (100.0%)	12 (100.0%)	391 (100.0%)	1.00
	yes	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
vomit	no	87 (100.0%)	289 (99.0%)	12 (100.0%)	388 (99.2%)	0.64
	yes	0 (0.0%)	3 (1.0%)	0 (0.0%)	3 (0.8%)	
diarrhea	no	86 (98.9%)	290 (99.3%)	12 (100.0%)	388 (99.2%)	0.57
	yes	1 (1.1%)	2 (0.7%)	0 (0.0%)	3 (0.8%)	
stomachache	no	83 (95.4%)	275 (94.2%)	12 (100.0%)	370 (94.6%)	0.78
	yes	4 (4.6%)	17 (5.8%)	0 (0.0%)	21 (5.4%)	
cough	no	67 (77.0%)	165 (56.5%)	6 (50.0%)	238 (60.9%)	<0.01
	yes	20 (23.0%)	127 (43.5%)	6 (50.0%)	153 (39.1%)	
runny nose	no	86 (98.9%)	273 (93.5%)	12 (100.0%)	371 (94.9%)	0.14
	yes	1 (1.1%)	19 (6.5%)	0 (0.0%)	20 (5.1%)	
sore throat	no	83 (95.4%)	273 (93.5%)	9 (75.0%)	365 (93.4%)	0.60
	yes	4 (4.6%)	19 (6.5%)	3 (25.0%)	26 (6.6%)	
headache	no	77 (88.5%)	241 (82.5%)	10 (83.3%)	328 (83.9%)	0.25
	yes	10 (11.5%)	51 (17.5%)	2 (16.7%)	63 (16.1%)	
fatigue	no	80 (92.0%)	247 (84.6%)	11 (91.7%)	338 (86.4%)	0.17
	yes	7 (8.0%)	45 (15.4%)	1 (8.3%)	53 (13.6%)	
muscle soreness	no	81 (93.1%)	223 (76.4%)	11 (91.7%)	315 (80.6%)	0.01
	yes	6 (6.9%)	69 (23.6%)	1 (8.3%)	76 (19.4%)	
joint soreness	no	83 (95.4%)	247 (84.6%)	12 (100.0%)	342 (87.5%)	0.01
	yes	4 (4.6%)	45 (15.4%)	0 (0.0%)	49 (12.5%)	

Table S2: Distributional fits to key COVID-19 distributions. 95%CIs of the distributional fits were shown in the brackets.

Time lag	Distribution	Parameter 1	Parameter 2	Mean	5%	50%	95%
Incubation period	lognormal	1.57 (1.44,1.69)	0.65 (0.56,0.73)	5.95 (4.94,7.11)	1.64 (1.33,2.04)	4.80 (4.22,5.44)	14.04 (12.19,15.90)
Serial interval	gamma	2.29 (1.77,3.34)	0.36 (0.26,0.57)	6.29 (5.17,7.56)	1.32 (0.92,1.87)	5.41 (4.43,6.49)	14.3 (11.12,17.57)
Onset to PCR confirmation; among contact-based	gamma	3.2 (2.43,4.76)	1.00 (0.71,1.57)	3.18 (2.65,3.76)	0.92 (0.7,1.26)	2.86 (2.4,3.37)	6.56 (5.25,8.01)
Onset to PCR confirmation; among symptom-based	gamma	2.12 (1.87,2.45)	0.39 (0.33,0.46)	5.46 (4.99,5.92)	1.04 (0.88,1.26)	4.63 (4.23,5.03)	12.71 (11.51,13.82)
Onset to hospitalization; among contact-based	lognormal	0.74 (0.55,0.95)	0.64 (0.55,0.71)	2.57 (2.06,3.16)	0.73 (0.61,0.93)	2.09 (1.73,2.58)	6.03 (4.49,7.53)
Onset to hospitalization; among symptom-based	lognormal	1.23 (1.12,1.33)	0.79 (0.74,0.83)	4.64 (4.13,5.1)	0.93 (0.82,1.08)	3.41 (3.06,3.78)	12.42 (10.89,13.77)
Onset to isolation; among contact-based	lognormal	0.77 (0.53,0.97)	0.67 (0.56,0.75)	2.71 (2.08,3.31)	0.72 (0.58,0.96)	2.17 (1.71,2.64)	6.52 (4.69,8.24)
Onset to isolation; among symptom-based	lognormal	1.22 (1.12,1.31)	0.78 (0.73,0.83)	4.58 (4.13,5.02)	0.94 (0.82,1.08)	3.38 (3.07,3.69)	12.19 (10.79,13.62)
Arrival to symptom onset; among onset after arrival	lognormal	1.22 (1.1,1.34)	0.85 (0.79,0.91)	4.87 (4.24,5.49)	0.83 (0.71,0.97)	3.38 (2.99,3.81)	13.79 (11.75,15.86)
Arrival to confirmation; among onset on or before arrival	weibull	1.28 (1.04,1.59)	4.8 (4.05,5.58)	4.5 (3.81,5.07)	0.47 (0.25,0.83)	3.61 (2.92,4.35)	11.32 (9.56,13.06)
Arrival to isolation; among onset on or before arrival	gamma	0.39 (0.3,0.54)	0.13 (0.1,0.18)	3.05 (2.49,3.67)	0 (0,0.02)	1.09 (0.67,1.66)	12.75 (10.57,14.84)

Table S3: Comparison of age distribution of cases with Shenzhen 2010 census (source http://www.stats.gov.cn/english/)

Age category	N	Proportion	Contact-based	Symptom-based
0-9 years	736,978	7%	13 (14.9%)	6 (2.1%)
10-19 years	1,058,098	10%	5 (5.7%)	6 (2.1%)
20-29 years	3,783,127	37%	11 (12.6%)	23 (7.9%)
30-39 years	2,528,979	24%	15 (17.2%)	71 (24.3%)
40-49 years	1,478,974	14%	9 (10.3%)	49 (16.8%)
50-59 years	466,403	5%	10 (11.5%)	63 (21.6%)
60-69 years	192,595	2%	20 (23.0%)	60 (20.5%)
70+ years	113,227	1%	4 (4.6%)	14 (4.8%)
Total	10,358,381	100%	87 (100.0%)	292 (100.0%)

Table S4: Comparison of observed serial intervals by time from symptom onset to isolation.

Time to isolation	Mean serial interval (95% CI)
0-2 days	3.6 (3.0, 4.2)
3-5 days	8.1 (5.3, 11.0)
6 or more days	8.0 (6.2, 9.7)

Table S5. Time to recovery from symptom onset in days.

Variable	Value	Time to recovery	2.5%	97.5%
sex	female	20.3	19.4	21.3
	male	21.2	20.2	22.3
age	0-9	17.5	15.3	20.0
	10-19	19.1	15.8	22.9
	20-29	19.2	17.5	21.0
	30-39	19.2	18.0	20.5
	40-49	21.6	20.0	23.4
	50-59	22.4	20.8	24.1
	60-69	22.9	21.2	24.7
	70+	22.5	19.1	26.3
severity	mild	20.1	19.0	21.3
	moderate	20.3	19.5	21.1
	severe	28.3	25.3	31.6
mode of detection	contact-based	19.3	17.9	20.9
	symptom-based	21.2	20.4	22.0
Total	Total	20.8	20.1	21.5

Table S6. Sensitivity analysis of age specific attack rates and risk for SAR-CoV-2 infection among household close contacts.

				Univariate regression		
Age group of contacts	**N**	**Infected**	**Attack rate % (95% CI)**	**Odds ratio**	**2.5%**	**97.5%**
0-9	109	10	9.17 (5.06,16.07)	2.26	0.20	26.13
10-19	57	6	10.53 (4.91,21.12)	5.48	0.44	67.79
20-29	48	7	14.58 (7.25,27.17)	7.28	0.50	105.61
30-39	156	14	8.97 (5.42,14.50)	1.93	0.18	21.12
40-49	69	7	10.14 (5.00,19.49)	8.76	0.69	110.63
50-59	59	10	16.95 (9.48,28.46)	Ref	..	..
60-69	86	18	20.93 (13.67,30.68)	2.53	0.23	27.59
70+	44	5	11.36 (4.95,23.98)	6.44	0.45	92.93

Table S7. Sensitivity analyses showing demographic and clinical characteristics of cases by contact-based vs. symptom-based surveillance after truncating data after Feb 7th 2020 when the definition of a confirmed case changed to require symptom.

Variable	Value	Contact-based surveillance (N=74)	Symptom-based surveillance (N=270)	Unknown/other (N=10)	Total (N=354)	P-value
sex	F	54 (73.0%)	124 (45.9%)	9 (90.0%)	187 (52.8%)	0.0001
	M	20 (27.0%)	146 (54.1%)	1 (10.0%)	167 (47.2%)	
age	0-9	13 (17.6%)	6 (2.2%)	1 (10.0%)	20 (5.6%)	<0.0001
	10-19	5 (6.8%)	5 (1.9%)	0 (0.0%)	10 (2.8%)	
	20-29	7 (9.5%)	22 (8.1%)	0 (0.0%)	29 (8.2%)	
	30-39	12 (16.2%)	66 (24.4%)	1 (10.0%)	79 (22.3%)	
	40-49	6 (8.1%)	45 (16.7%)	2 (20.0%)	53 (15.0%)	
	50-59	10 (13.5%)	57 (21.1%)	1 (10.0%)	68 (19.2%)	
	60-69	18 (24.3%)	58 (21.5%)	5 (50.0%)	81 (22.9%)	
	70+	3 (4.1%)	11 (4.1%)	0 (0.0%)	14 (4.0%)	
severity	mild	16 (21.6%)	76 (28.1%)	2 (20.0%)	94 (26.6%)	0.083
	moderate	55 (74.3%)	166 (61.5%)	6 (60.0%)	227 (64.1%)	
	severe	3 (4.1%)	28 (10.4%)	2 (20.0%)	33 (9.3%)	
symptomatic	no	15 (20.3%)	8 (3.0%)	0 (0.0%)	23 (6.5%)	<0.0001
	yes	59 (79.7%)	262 (97.0%)	10 (100.0%)	331 (93.5%)	
fever	no	21 (28.4%)	31 (11.5%)	2 (20.0%)	54 (15.3%)	0.0006
	yes	53 (71.6%)	239 (88.5%)	8 (80.0%)	300 (84.7%)	

Table S8. Sensitivity analyses showing the association of clinical and demographic factors with mode of detection and severity at initial assessment after truncating data after Feb 7th 2020 when the definition of a confirmed case changed to require symptom.

		Outcome: symptom-based surveillance						**Outcome: moderate/severe symptom**					
		Univariate regression			**Multivariate regression**			**Univariate regression**			**Multivariate regression**		
		OR	**2.5%**	**97.5%**	**OR**	**2.5%**	**97.5%**	**OR**	**2.5%**	**97.5%**	**OR**	**2.5%**	**97.5%**
sex	female	ref	..	..	ref	..	..	ref	..	..	ref	..	..
	male	3.18	1.83	5.71	3.18	1.75	6.01	1.38	0.85	2.24	1.37	0.85	2.22
age	0-9	0.08	0.02	0.25	0.08	0.02	0.27	0.86	0.29	2.74	0.67	0.22	2.23
	10-19	0.18	0.04	0.73	0.14	0.03	0.63	0.92	0.23	4.63	0.88	0.22	4.40
	20-29	0.55	0.19	1.69	0.6	0.2	1.89	0.49	0.2	1.21	0.61	0.26	1.49
	30-39	0.96	0.38	2.4	0.99	0.38	2.52	1.15	0.55	2.4	1.15	0.57	2.33
	40-49	1.32	0.45	4.12	1.16	0.39	3.73	1.16	0.51	2.68	1.21	0.55	2.72
	50-59	ref	..	..	ref	..	..	ref	..	..	Ref	..	..
	60-69	0.57	0.23	1.31	0.51	0.2	1.23	1.61	0.74	3.54	1.43	0.68	3.03
	70+	0.64	0.16	3.2	0.52	0.13	2.72	1.45	0.4	6.93	1.28	0.40	4.96
severity	mild	ref	..	..	ref	..	..	..	..	..	..	..	..
	moderate	0.64	0.33	1.16	0.53	0.26	1.03	..	..	..	..	..	..
	severe	1.96	0.6	8.9	1.37	0.37	6.74	..	..	..	..	..	..
fever	no	ref	..	..	..	..	..	ref	..	..	..	..	..
	yes	3.05	1.62	5.71	..	..	..	0.9	0.44	1.73	..	..	..
symptomatic	no	ref	..	..	..	..	..	ref	..	..	..	..	..

	yes	8.33	3.45	21.54	..	..	..	0.39	0.09	1.18	..	..	..
Surveillance method	Contact-based	..	..	..	..	..	..	ref	..	..	..	..	..
	symptom-based	..	..	..	..	..	..	0.7	0.37	1.28	0.55	0.29	1.01

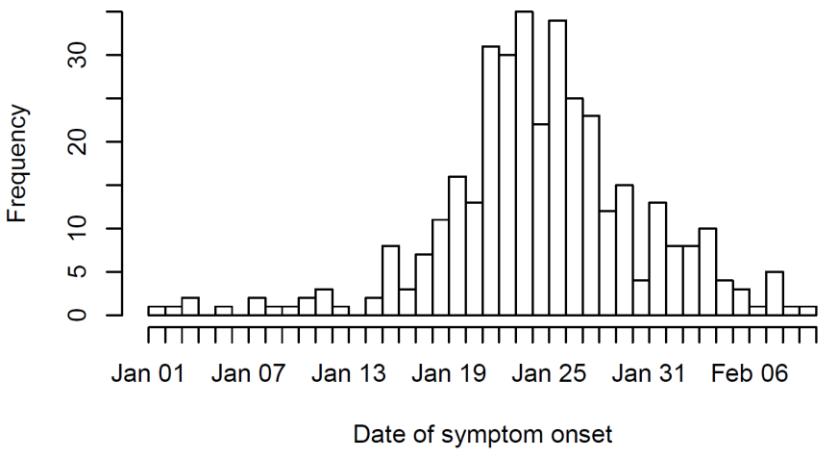

Figure S1. Epidemic curve of cases in Shenzhen identified before Feb 12, 2020.

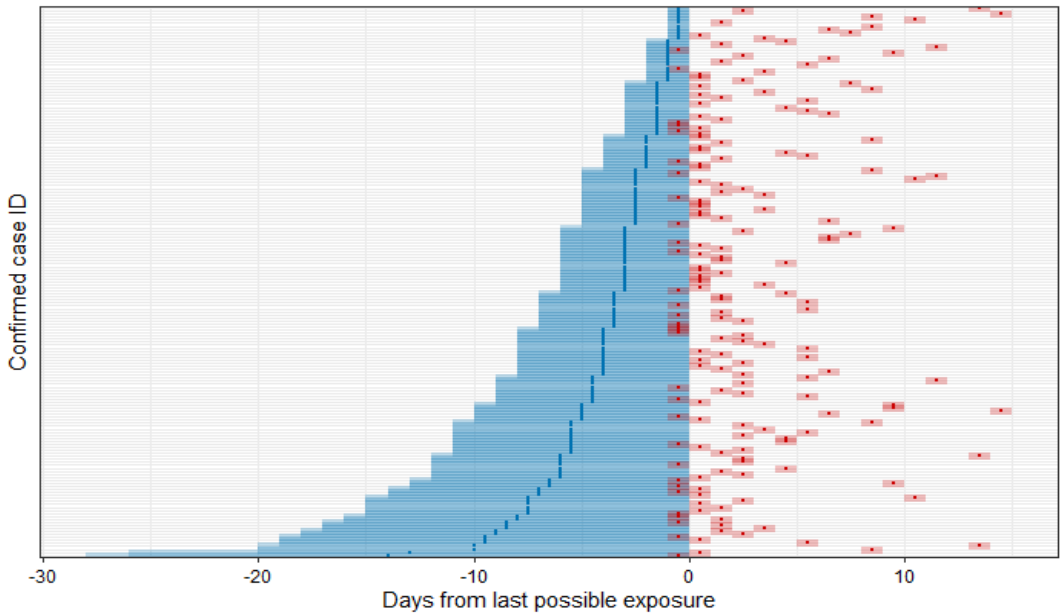

Figure S2: The exposure and symptom onset windows 339 confirmed cases from Shenzhen, China. Shaded regions represent the full possible interval of exposure (blue) and of symptom onset (red); points represent the midpoint of these intervals. The exposure and symptom onset windows are aligned relative to the right-bound of the exposure window.

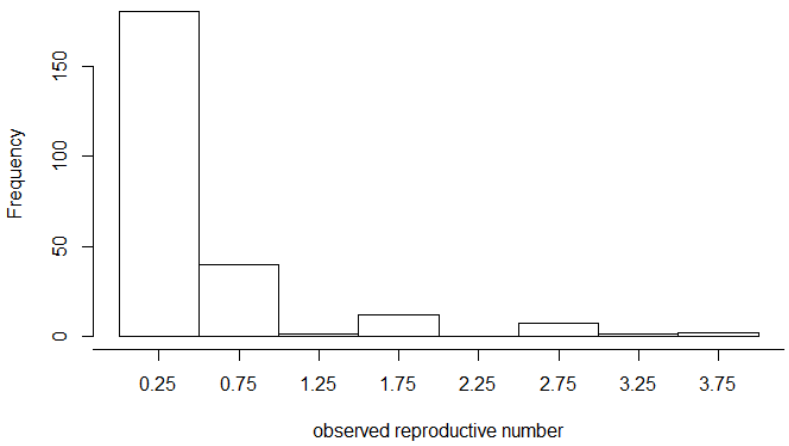

Figure S3. Distribution of the observed individual reproductive numbers in Shenzhen.

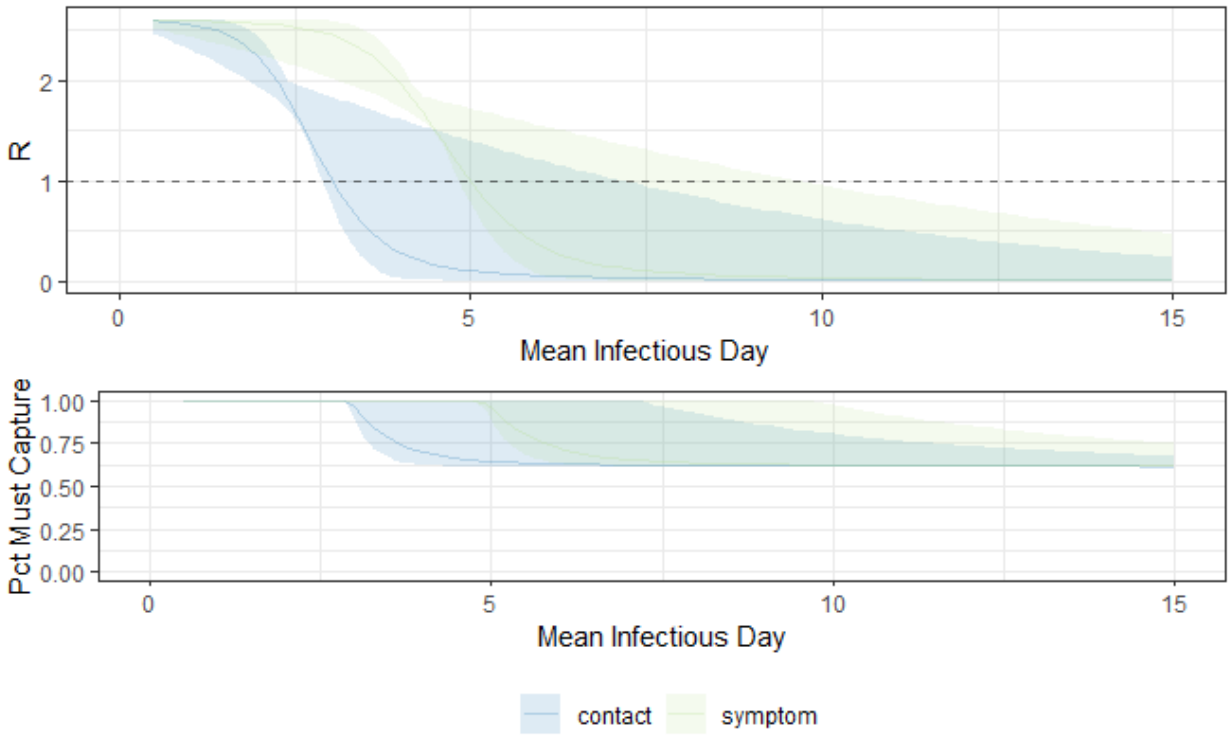

Figure S4: Effective *R* among those captured by surveillance (top) and proportion needed to be captured by surveillance to drive *R* less than one (bottom) by the weighted mean day of the infectious period. Weighting is by relative infectiousness, which is assumed to follow a gamma distribution. The shaded area covers all gamma distributions with a mean of that day and a rate parameter in the range of 0.1-10. See Appendix Page 10 for detailed methods.

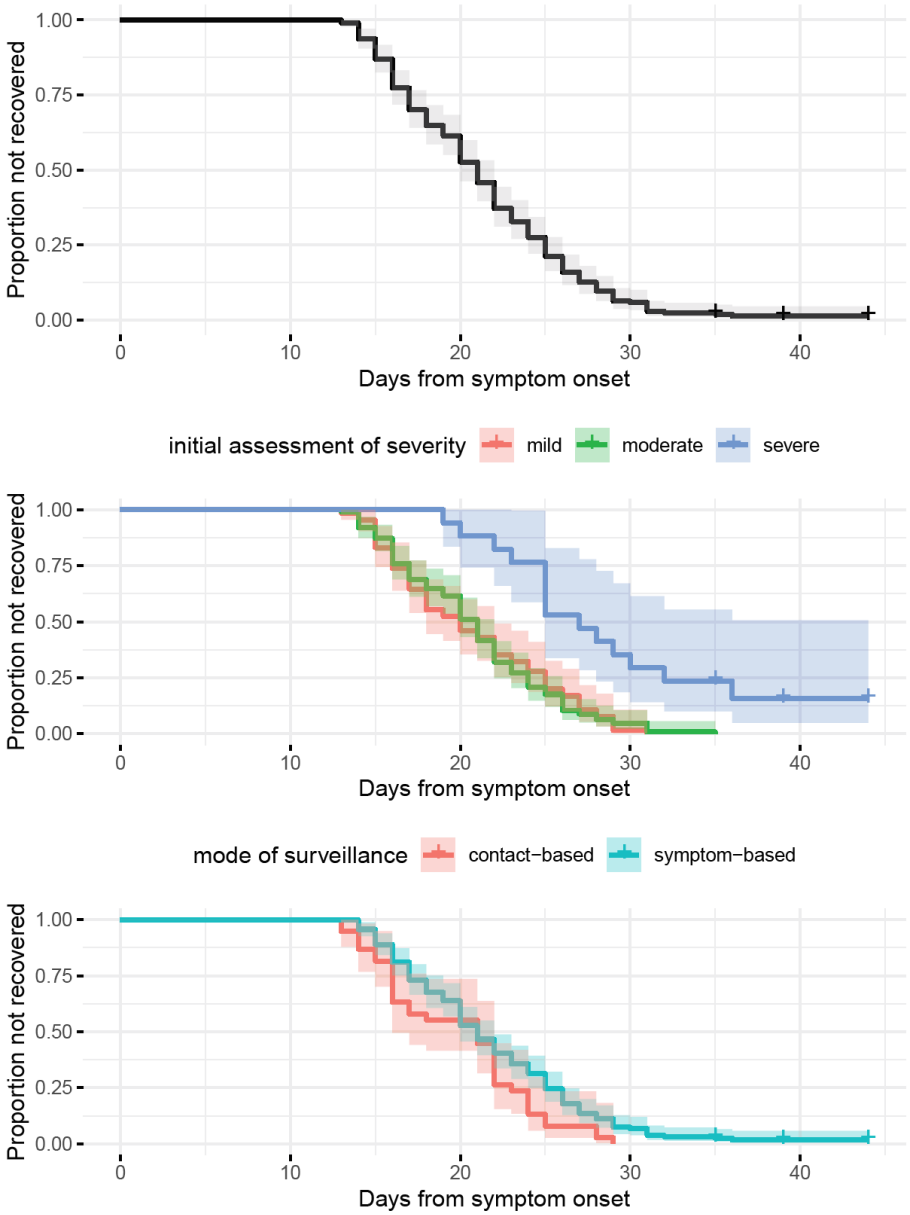

Figure S5: Time from symptom onset to recovery for all cases (top), by clinical severity at initial assessment (middle), and by mode of surveillance (bottom). Time from symptom onset to death was marked by "+" for the three patients who died.

Text S1: Data extraction and confirmation details

By categorizing COVID-19 as a notifiable disease Class B, Chinese Law on the Prevention and Treatment of Infectious Diseases required all cases to be immediately reported to China's Infectious Disease Information System. Each case was recorded into the system by local epidemiologists and public health professionals who did the field investigation and collected possible exposure related information. All data on COVID-19 case reported in Shenzhen were extracted from the Infectious Disease Information System by the end of February 12, 2020. Then personal information including demographics, symptoms, clinical outcome and severity and so on, were stripped to construct an anonymous dataset. All cases were included without sampling and no eligibility criteria were needed. Aggregated data may be available upon request.

All laboratory confirmation of SARS-CoV-2 were done by Guangdong Center for Disease Prevention and Control (CDC) before Jan 30, 2020, and then only need to be done by Shenzhen CDC, when it obtained the qualification of laboratory-confirmation of 2019-nCoV from the authority. The RT-PCR assay was conducted in the BSL-2 laboratory of Shenzhen CDC, using the protocol established by the World Health Organization and China CDC.

Text S2: Observed reproductive number calculation

We calculated the mean observed reproductive number as the ratio of the number of infectees (N) to the number of potential infectors (M) across all risk sets, so $\bar{R}_{obs} = N/M$. To calculate overdispersion, we took a multiple imputation approach, where in each imputation any infectee who could have been infected by multiple individuals was randomly assigned one as their actual infector. Independent imputations were conducted with each bootstrap draw.

Text S3: Supplemental calculation

Let R_0 be the basic reproductive number, ρ be the percent of transmission due to cases potentially reachable by the surveillance system, and γ be the relative effective infectious period of those captured by surveillance. Then:

$$R = \rho\gamma R_0 + (1-\rho)R_0$$

When R is below one, sustained outbreaks are impossible. Hence, for a known R_0 and γ such that $\gamma R_0 < 1$, we can calculate the proportion of transmission that must be from people who can be captured by surveillance as:

$$\rho = \frac{1 - R_0}{R_0(\gamma - 1)}$$

Assuming an R_0 of 2.6 and that surveillance reduces R_0 by a factor of 0.18, we find control is possible if 75% of people can be captured by surveillance.

Text S4: Analysis of impact of surveillance

We calculated the potential impact of symptom and contact based surveillance by calculating the impact of a truncated infectious period on R0. As this calculation is heavily dependent on the distribution of infectiousness over time, and that distribution is largely unknown, we explored large number of infectious period distributions, considering those where the mean day of infectiousness (i.e., the average day after an infectors symptom onset on which a secondary case they caused would be infected) and the overall distribution of infectiousness follows a gamma distribution with a rate parameter ranging between 0.1 and 10. We then calculated the expected resulting R if individuals were removed from the population based on symptom or contact based surveillance (Fig S4), thereby truncating their infectious period.