Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study

Qifang Bi, Yongsheng Wu*, Shujiang Mei*, Chenfei Ye*, Xuan Zou, Zhen Zhang, Xiaojian Liu, Lan Wei, Shaun A Truelove, Tong Zhang, Wei Gao, Cong Cheng, Xiujuan Tang, Xiaoliang Wu, Yu Wu, Binbin Sun, Suli Huang, Yu Sun, Juncen Zhang, Ting Ma*, Justin Lessler*, Tiejian Feng**

Summary

Background Rapid spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Wuhan, China, prompted heightened surveillance in Shenzhen, China. The resulting data provide a rare opportunity to measure key metrics of disease course, transmission, and the impact of control measures.

Methods From Jan 14 to Feb 12, 2020, the Shenzhen Center for Disease Control and Prevention identified 391 SARS-CoV-2 cases and 1286 close contacts. We compared cases identified through symptomatic surveillance and contact tracing, and estimated the time from symptom onset to confirmation, isolation, and admission to hospital. We estimated metrics of disease transmission and analysed factors influencing transmission risk.

Findings Cases were older than the general population (mean age 45 years) and balanced between males (n=187) and females (n=204). 356 (91%) of 391 cases had mild or moderate clinical severity at initial assessment. As of Feb 22, 2020, three cases had died and 225 had recovered (median time to recovery 21 days; 95% CI 20–22). Cases were isolated on average 4·6 days (95% CI 4·1–5·0) after developing symptoms; contact tracing reduced this by 1·9 days (95% CI 1·1–2·7). Household contacts and those travelling with a case were at higher risk of infection (odds ratio 6·27 [95% CI 1·49–26·33] for household contacts and 7·06 [1·43–34·91] for those travelling with a case) than other close contacts. The household secondary attack rate was 11·2% (95% CI 9·1–13·8), and children were as likely to be infected as adults (infection rate 7·4% in children <10 years *vs* population average of 6·6%). The observed reproductive number (*R*) was 0·4 (95% CI 0·3–0·5), with a mean serial interval of 6·3 days (95% CI 5·2–7·6).

Interpretation Our data on cases as well as their infected and uninfected close contacts provide key insights into the epidemiology of SARS-CoV-2. This analysis shows that isolation and contact tracing reduce the time during which cases are infectious in the community, thereby reducing the *R*. The overall impact of isolation and contact tracing, however, is uncertain and highly dependent on the number of asymptomatic cases. Moreover, children are at a similar risk of infection to the general population, although less likely to have severe symptoms; hence they should be considered in analyses of transmission and control.

Funding Emergency Response Program of Harbin Institute of Technology, Emergency Response Program of Peng Cheng Laboratory, US Centers for Disease Control and Prevention.

Introduction

Since emerging in Wuhan, China, in December, 2019,[1] the coronavirus disease 2019 (COVID-19) epidemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has progressed rapidly into a pandemic. COVID-19 is characterised by fever, cough, fatigue, shortness of breath, pneumonia, and other respiratory tract symptoms,[2–4] and in many cases progresses to death. As of April 15, 2020, there have been 1914916 confirmed cases and 123010 deaths reported worldwide.[5] Most cases were initially confined to Hubei province in China, but there has since been substantial spread not only elsewhere in China but worldwide. A rapid and robust response by the global scientific community has described many important aspects of SARS-CoV-2 transmission and natural history,[1,2,6–8] but key questions remain.

If well tracked, early introductions of an emerging pathogen provide a unique opportunity to characterise its transmission, natural history, and the effectiveness of screening. Careful monitoring of cases and low probability of infection from the general community enables inferences, important to modelling the course of the outbreak, that are difficult to make during a widely disseminated epidemic. In particular, we can make assumptions about when and where cases were likely to have been infected that are impossible when the pathogen is widespread. Furthermore, during these early phases, uninfected and asymptomatic contacts are often closely tracked, providing important information about transmission and natural history. Combined, these data on early introductions can be used to give insights into the natural history of the disease,[9]

Lancet Infect Dis 2020; 20: 911–19

Published Online April 27, 2020 https://doi.org/10.1016/S1473-3099(20)30287-5

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

See Comment page 876

For the Chinese translation of the abstract see Online for appendix 1

*Contributed equally

Department of Epidemiology (Q Bi MHS, J Lessler PhD), **and Department of International Health** (S A Truelove PhD), **Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; Department of Public Health Information** (Yongsheng Wu MPhil, Z Zhang MPhil, X Liu MPhil, L Wei PhD), **Department of Communicable Diseases Control and Prevention** (S Mei MPhil, W Gao MPhil, C Cheng MPhil, X Tang MPhil, X Wu MPhil), **Department of School Health** (Yu Wu MD), **and Department of Environment and Health** (S Huang MD), **Shenzhen Center for Disease Control and Prevention, Shenzhen, China** (X Zou MPhil, T Feng MPhil); **School of Electronic and Information Engineering** (C Ye PhD, T Ma PhD), **and School of Computer Science** (B Sun), **Harbin Institute of Technology at Shenzhen, Shenzhen, China; and Artificial Intelligence Research Center, Peng Cheng Laboratory, Shenzhen, China** (C Ye, T Zhang PhD, B Sun MPhil, Y Sun MPhil, J Zhang BE, T Ma)

Correspondence to:
Dr Ting Ma, Department of Electronic and Information Engineering, Harbin Institute of Technology Campus, Shenzhen University Town, Xili, Shenzhen, Guangdong 518055, China
tma@hit.edu.cn

or

Dr Justin Lessler, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
justin@jhu.edu

or

Mr Tiejian Feng, Shenzhen Center for Disease Control and Prevention, Nanshan, Shenzhen, Guangdong 518055, China
fengtiej@126.com

Research in context

Evidence before this study

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been accompanied by a substantial increase in research and publications ranging from transmissibility of SARS-CoV-2 to clinical characteristics of the infection. As of March 23, 2020, our search of PubMed using keywords ("COVID-19" OR "SARS-nCoV-2" OR "novel coronavirus") AND ("serial interval" OR "incubation period" OR "attack rate") yielded 14 articles that have estimated either the serial interval or incubation period of coronavirus disease 2019 (COVID-19) and one that investigated the secondary attack rate. However, most of these estimates have come from either Wuhan or publicly available case data; in both instances, incomplete capture of infections and cases might have produced biased estimates.

Added value of this study

This study is, to our knowledge, the first analysis of SARS-CoV-2 transmission and COVID-19 natural history based on a large primary dataset of cases and close contacts, for which the mode of surveillance (ie, symptom-based versus contact-based) was sufficiently documented and RT-PCR testing was nearly universal. We present one of the first estimates of the serial interval, secondary household attack rate, and dispersion (ie, tendency towards super spreading) for SARS-CoV-2 based on active surveillance data. We found that the attack rate does not differ significantly by age, with on average 7% of close contacts becoming infected, around 80% of these contacts showing any symptoms, and 3% of infections manifesting severe disease at initial assessment. We also found that contact-based surveillance in Shenzhen reduced the duration an infected individual transmits in the community by 2 days. These findings are important for understanding the burden of COVID-19 and for strategic planning across the world.

Implications of all the available evidence

These results shed further light on how SARS-CoV-2 is transmitting, how severe it is, and how effective control measures can be in specific contexts. We provide a key piece of evidence supporting intensive contact tracing and highlighting that children might be an important target for interventions aimed at reducing transmission, even if they do not get sick.

	Contact-based surveillance (n=87)	Symptom-based surveillance (n=292)	Unknown or other (n=12)	Total (n=391)	p value
Sex					
Female	63 (72%)	131 (45%)	10 (83%)	204 (52%)	<0·0001
Male	24 (28%)	161 (55%)	2 (17%)	187 (48%)	..
Age					
0–9 years	13 (15%)	6 (2%)	1 (8%)	20 (5%)	<0·0001
10–19 years	5 (6%)	6 (2%)	1 (8%)	12 (3%)	..
20–29 years	11 (13%)	23 (8%)	0 (0%)	34 (9%)	..
30–39 years	15 (17%)	71 (24%)	1 (8%)	87 (22%)	..
40–49 years	9 (10%)	49 (17%)	2 (17%)	60 (15%)	..
50–59 years	10 (12%)	63 (22%)	1 (8%)	74 (19%)	..
60–69 years	20 (23%)	60 (21%)	6 (50%)	86 (22%)	..
≥70 years	4 (5%)	14 (5%)	0 (0%)	18 (5%)	..
Severity					
Mild	18 (21%)	82 (28%)	2 (17%)	102 (26%)	0·03
Moderate	66 (76%)	180 (62%)	8 (67%)	254 (65%)	..
Severe	3 (3%)	30 (10%)	2 (17%)	35 (9%)	..
Symptomatic					
No	17 (20%)	8 (3%)	0 (0%)	25 (6%)	<0·0001
Yes	70 (80%)	284 (97%)	12 (100%)	366 (94%)	..
Fever					
No	25 (29%)	34 (12%)	2 (17%)	61 (16%)	0·0002
Yes	62 (71%)	258 (88%)	10 (83%)	330 (84%)	..

Table 1: **Demographic and clinical characteristics of cases by contact-based versus symptom-based surveillance**

See Online for appendix 2

transmission characteristics,[10] and the unseen burden of infection.[11]

Here, we use data collected by the Shenzhen Center for Disease Control and Prevention (Shenzhen CDC) on 391 cases of COVID-19 and 1286 of their close contacts to characterise key aspects of its epidemiology outside of Hubei province. We characterise differences in demographics and severity between cases identified through symptom-based surveillance and monitoring of close case contacts, and estimate the time to key events, such as confirmation, isolation, and recovery. Using data from contact tracing, we characterise SARS-CoV-2 transmission by estimating key values, such as the household secondary attack rate, serial interval, and observed reproductive number (R).

Methods

Case identification

On Jan 8, 2020, Shenzhen CDC identified the first case of pneumonia with unknown cause and began monitoring travellers from Hubei province for symptoms of COVID-19. Over the next 2 weeks this surveillance programme expanded to include travellers from Hubei regardless of symptoms, patients at local hospitals, and individuals detected by fever screening in neighbourhoods and at local clinics. Suspected cases and close contacts were tested for SARS-CoV-2 by RT-PCR of nasal swabs at 28 qualified local hospitals, ten district-level CDCs, and two third-party testing organisations, with final confirmation done at the Guangdong Provincial Center for Disease Control and Prevention (Guangdong CDC) or Shenzhen CDC (appendix 2 p 11). Close contacts were identified through contact tracing of a confirmed case and were defined as those who lived in the same apartment, shared a meal, travelled, or socially interacted with an index case 2 days before symptom onset. Casual

	Outcome: symptom-based surveillance						Outcome: moderate or severe assessment					
	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	3·23	1·93	5·53	3·06	1·77	5·44	1·30	0·82	2·07	1·37	0·85	2·22
Age												
0–9 years	0·07	0·02	0·23	0·08	0·02	0·25	0·88	0·30	2·77	0·67	0·22	2·23
10–19 years	0·19	0·05	0·77	0·17	0·04	0·76	1·08	0·28	5·27	0·88	0·22	4·40
20–29 years	0·33	0·12	0·89	0·37	0·13	1·02	0·65	0·28	1·56	0·61	0·26	1·49
30–39 years	0·75	0·31	1·77	0·8	0·32	1·93	1·17	0·58	2·37	1·15	0·57	2·33
40–49 years	0·86	0·32	2·33	0·79	0·29	2·19	1·27	0·58	2·83	1·21	0·55	2·72
50–59 years	Ref	..	..	Ref	..	..	Ref	..	..	Ref	..	..
60–69 years	0·48	0·2	1·08	0·44	0·18	1·04	1·50	0·72	3·16	1·43	0·68	3·03
≥70 years	0·56	0·16	2·25	0·45	0·12	1·89	1·41	0·45	5·43	1·28	0·40	4·96
Severity												
Mild	Ref	..	..	Ref	..	..	..	..	..	..	..	..
Moderate	0·6	0·33	1·05	0·51	0·26	0·94	..	..	..	..	..	..
Severe	2·2	0·68	9·84	1·51	0·42	7·23	..	..	..	..	..	..
Fever												
No	Ref	..	..	..	..	..	Ref	..	..	..	..	..
Yes	3·06	1·69	5·49	..	..	..	0·94	0·49	1·75	..	..	..
Symptomatic												
No	Ref	..	..	..	..	..	Ref	..	..	..	..	..
Yes	8·62	3·68	21·89	..	..	..	0·51	0·15	1·39	0·55	0·29	1·01
Surveillance method												
Contact-based	..	..	..	..	..	..	Ref	..	..	..	..	..
Symptom-based	..	..	..	..	..	..	0·67	0·37	1·17	..	..	..

OR=odds ratio.

Table 2: **Association of clinical and demographic factors with mode of detection and severity at initial assessment**

contacts (eg, other clinic patients) and some close contacts (eg, nurses) who wore a mask during exposure were not included in this group.

Symptomatic cases were isolated and treated at designated hospitals regardless of RT-PCR test results. Asymptomatic individuals who tested positive were quarantined at centralised facilities. Close contacts and travellers from Hubei who tested negative were quarantined at home or a central facility, and monitored for 14 days. RT-PCR testing was required for all close contacts at the beginning of isolation, and release was conditional on a negative RT-PCR result. Basic demographics, signs and symptoms, clinical severity, and exposure history were recorded for all confirmed cases.

Here, we analyse confirmed cases identified by the Shenzhen CDC between Jan 14 and Feb 12, 2020, and close contacts of cases confirmed before Feb 9, 2020.

This work was done in support of an ongoing public health response, and hence was determined not to be human subjects research after consultation with the Johns Hopkins Bloomberg School of Public Health institutional review board. Data collection is part of the continuing public health investigation of an emerging outbreak and therefore the individual informed consent was waived. The study was approved by the ethics committees of Shenzhen CDC. Analytical datasets were constructed in an anonymised manner, and all analysis of personally identifiable data took place onsite at the Shenzhen CDC.

Epidemiological and clinical characteristics of cases

We defined symptom-based surveillance to include symptomatic screening at airport and train stations, community fever monitoring, home observation of recent travellers to Hubei, and testing of patients admitted to hospital. Contact-based surveillance is the identification of cases through monitoring and testing of close contacts of confirmed cases, independently of their symptom presentation. By protocol, those in the contact-based group were tested for SARS-CoV-2 infection regardless of symptoms, whereas those in the other categories were tested only if they showed signs or symptoms of disease.

At the first clinical assessment, data were recorded on 21 signs and symptoms (appendix 2 p 1), and disease severity was assessed. Cases with fever, respiratory symptoms, and radiographic evidence of pneumonia were classified as having moderate symptoms. Cases

were classified as having severe symptoms if they had any of the following: breathing rate 30 breaths per min or higher; oxygen saturation level 93% or lower at rest; oxygen concentration level PaO_2/FiO_2 (ratio of arterial oxygen partial pressure to fractional inspired oxygen) 300 mm Hg (1 mm Hg=0·133 kPa) or lower; lung infiltrates higher than 50% within 24–48 h; respiratory failure requiring mechanical ventilation; septic shock; or multiple organ dysfunction or failure. All other symptomatic cases were classified as mild.

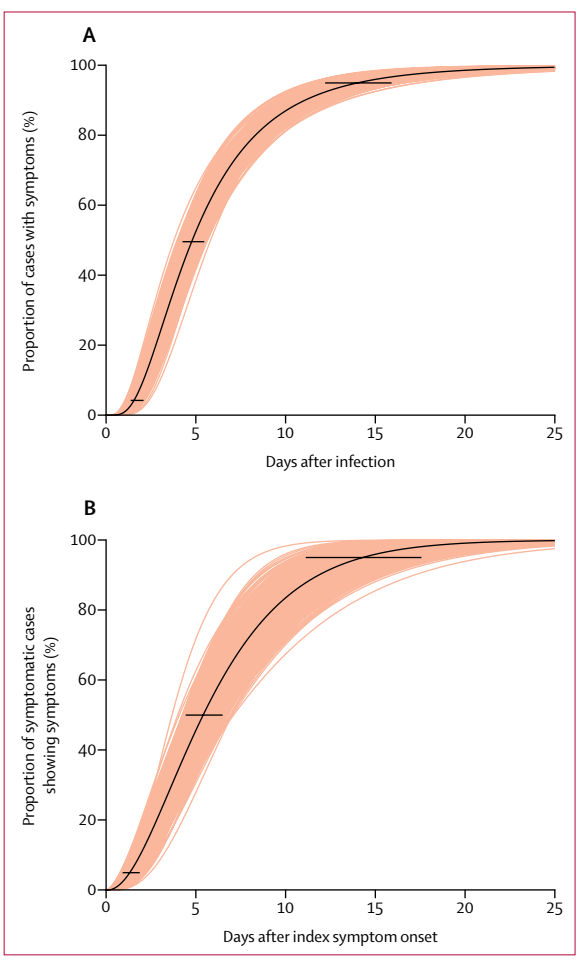

Figure 1: **Incubation period and serial interval of COVID-19**
(A) Proportion of cases who developed symptoms of coronavirus disease 2019 (COVID-19) by days after infection (ie, the cumulative distribution function of the incubation period). (B) Proportion of cases infected by an index case who developed symptoms by a given number of days after the day of symptom onset of the index case (ie, the cumulative distribution function of the serial interval). The maximum-likelihood estimates for the parametric distribution of the cumulative distribution function are shown, along with 1000 parametric bootstrap estimates of the cumulative distribution function. The median incubation period of COVID-19 is estimated to be 4·8 days (95% CI 4·2–5·4). 5% of cases who develop symptoms will do so by 1·6 days (95% CI 1·3–2·0) after infection, and 95% by 14·0 days (12·2–15·9). We estimated that the median serial interval of COVID-19 is 5·4 days (95% CI 4·4–6·5). 5% of infected cases who develop symptoms will do so by 1·3 days (95% CI 0·9–1·9) after symptom onset of the index case, and 95% by 14·3 days (11·1–17·6).

Relationships between demographics, mode of detection, and symptom severity were assessed and characterised with χ^2 tests, and simple and multiple logistic regression.

Timing of key events

Distributions were fit to the timing of key events in each confirmed case's course of infection and treatment. The time from infection to symptom onset (incubation period) was assumed to be log-normally distributed and estimated as previously described.[12–14] We determined the left and right boundaries on the possible exposure and symptom onset times. Cases who recently travelled to Hubei were assumed to have been exposed while there. Cases without a recent travel history but with exposure to a confirmed case were assumed to be exposed from the time of earliest to latest possible contact with that case. Only cases for whom we could identify the earliest and latest possible time of exposure and who had a date of symptom onset were included in the analysis.

Time between symptom onset and recovery was estimated by use of parametric survival methods. Patients who had not recovered were considered to be censored on Feb 22, 2020, or at the time of death. All other delay distributions were estimated by directly fitting parametric distributions to time between symptom onset or arrival in Shenzhen, and confirmation, isolation, or admission to hospital. Confidence intervals were calculated with bootstrapping or standard parametric estimators.[15]

Transmission characteristics

Transmission was characterised by examining the relationship between confirmed cases and their infected and uninfected close contacts. The household secondary attack rate was calculated as the percentage of household contacts (those sharing a room, apartment, or other sleeping arrangement) who were later confirmed to have SARS-CoV-2 infection. The distribution of serial intervals (the time between symptom onset in the confirmed case and their infected contacts) was calculated by fitting parametric distributions to the time of symptom onset in clear case–contact pairs. The mean R and distribution of individual reproductive numbers (ie, the number of secondary infections caused by each case) were calculated from the number of secondary infections observed among close contacts of each index case, with ambiguities resolved through multiple imputation (appendix 2 p 11). The relative odds of transmission among contacts of various types were estimated by use of conditional logistic regression and random-effects models, to account for differing numbers of possible infected individuals in each risk group. When assessing the impact of characteristics of infected individuals, we only included risk sets where a single potential infected individual was clearly identifiable. Confidence intervals were estimated by use of bootstrapping or standard parametric approaches.

Role of the funding source

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. Shenzhen CDC had full access to all the data in the study, and all corresponding authors share the final responsibility for the decision to submit for publication.

Results

Between Jan 14 and Feb 12, 2020, the Shenzhen CDC confirmed 391 cases of SARS-CoV-2 infection (table 1). Of 379 with a known mode of detection, 292 (77%) were detected through symptom-based surveillance. Overall, there were approximately equal numbers of male and female cases (187 *vs* 204). The mean age of the population was 45 years, and 307 (79%) of 391 cases were adults aged 30–69 years. At the time of first clinical assessment, most cases were mild (102 [26%] of 391) or moderate (254 [65%] of 391), and only 35 (9%) were severe. 330 (84%) of 391 cases had fever at the time of initial assessment, while 25 (6%) of 391 had no signs or symptoms. As of Feb 22, 2020, final clinical outcomes were known for 228 of 391 cases in our data, with three who had died (all captured through symptom-based surveillance) and 225 who had recovered.

A larger proportion of cases detected through symptom-based surveillance were male (161 [55%] of 292 *vs* 24 [28%] of 87) and aged 20–69 years (266 [91%] of 292 *vs* 65 [75%] of 87) than were those detected through contact-based surveillance (tables 1, 2). At the time of the first clinical assessment, 25 (29%) of 87 cases in the contact-based surveillance group did not have fever, and 17 (20%) of 87 had no symptoms. By contrast, 258 (88%) of 292 in the symptom-based surveillance group had fever, and only eight reported no symptoms.

In multiple logistic regression, male sex was associated with severe symptoms (odds ratio [OR] 2·5 [95% CI 1·1–6·1]). The probability of severe symptoms increased slightly with age, although only individuals aged 60–69 years had a significantly increased risk compared with the reference category, individuals aged 50–59 years (OR 3·4 [95% 1·4–9·5]).

Based on 183 cases with a well defined period of exposure and symptom onset (appendix 2 p 8), we estimated the median incubation period for COVID-19 to be 4·8 days (95% CI 4·2–5·4; figure 1, appendix 2 p 2),

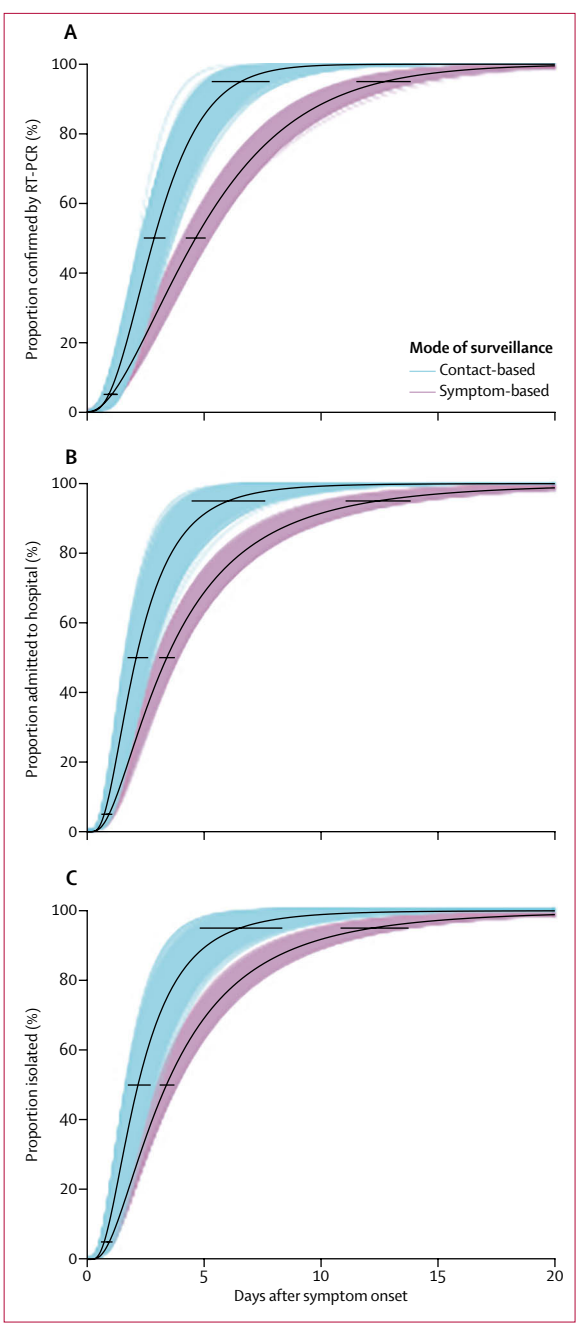

Figure 2: **Time between symptom onset and SARS-CoV-2 confirmation (A), admission to hospital (B), and isolation among cases (C) detected by contact-based and symptom-based surveillance**

The maximum-likelihood estimates for the parametric distribution of the cumulative distribution function are shown, along with 1000 parametric bootstrap estimates of the cumulative distribution function. Panel A shows estimates of the proportion of cases who are confirmed by RT-PCR, according to the number of days after symptom onset. We estimated that 50% of cases detected through symptom-based surveillance were confirmed by RT-PCR within 4·6 days (95% CI 4·2–5·0) after symptom onset, and 95% were confirmed by RT-PCR within 12·7 days (11·5–13·8) after symptom onset. Contact-based surveillance reduced the days from symptom onset to RT-PCR confirmation to 2·9 days (95% CI 2·4–3·4) in 50% of cases and to 6·6 days (5·3–8·0) in 95% of cases. Panel B shows estimates of the proportion of cases who were admitted to hospital, according to the number of days after symptom onset. We estimated that 50% of the cases detected through symptom-based surveillance were admitted to hospital by 3·4 days (95% CI 3·1–3·8) after symptom onset, and 95% by 12·4 days (10·9–13·8). Contact-based surveillance reduced the days from symptom onset to hospital admission to 2·1 days (95% CI 1·7–2·6) in 50% of cases, and 6·0 days (95% CI 4·5–7·5) in 95% of cases. Panel C shows estimates of the proportion of cases isolated, according to number of days after symptom onset. We estimated that 50% of cases detected through symptom-based surveillance were isolated by 3·4 days (95% CI 3·1–3·7) after symptom onset, and 95% by 12·2 days (95% CI 10·8–13·6). Contact-based surveillance reduced the days from symptom onset to isolation to 2·2 days (95% CI 1·7–2·6) in 50% of cases, and to 6·5 days (4·7–8·2) in 95% of cases.

	Number of cases*	Number infected	Attack rate (95% CI)	Univariate regression			Multivariate regression		
				Odds ratio	2·5%	97·5%	Odds ratio	2·5%	97·5%
Sex									
Female	558	58	10·4% (8·1–13·2)	Ref	..	..	..	..	..
Male	486	26	5·3% (3·7–7·7)	0·43	0·21	0·86	..	..	..
Age									
0–9 years	148	11	7·4% (4·2–12·8)	2·33	0·38	14·05	..	..	..
10–19 years	85	6	7·1% (3·3–14·6)	3·50	0·53	23·24	..	..	..
20–29 years	114	7	6·1% (3·0–12·1)	4·91	0·74	32·64	..	..	..
30–39 years	268	16	6·0% (3·7–9·5)	1·84	0·34	9·80	..	..	..
40–49 years	143	7	4·9% (2·4–9·8)	3·46	0·55	21·92	..	..	..
50–59 years	110	10	9·1% (5·0–15·9)	Ref	..	..	..	..	..
60–69 years	130	20	15·4% (10·2–22·6)	5·68	1·01	32·09	..	..	..
≥70 years	72	7	9·7% (4·8–18·7)	4·26	0·64	28·44	..	..	..
Contact type: household									
No	456	4	0·9% (0·3–2·2)	Ref	..	..	Ref	..	..
Yes	686	77	11·2% (9·1–13·8)	15·10	3·69	61·69	6·27	1·49	26·33
Contact type: travel									
No	824	63	7·6% (6·0–9·7)	Ref	..	..	Ref	..	..
Yes	318	18	5·7% (3·6–8·8)	9·13	1·85	45·08	7·06	1·43	34·91
Contact type: meal									
No	435	20	4·6% (3·0–7·0)	Ref	..	..	Ref	..	..
Yes	707	61	8·6% (6·8–10·9)	23·01	2·51	211·2	7·13	0·73	69·32
Contact frequency									
Rare	230	1	0·4% (0·02–2·4)	<0·0001	0	Inf	..	..	..
Moderate	305	9	3·0% (1·6–5·5)	Ref	..	..	..	..	..
Often	555	71	12·8% (10·3–15·8)	8·8	2·58	30·06	..	..	..

SARS-CoV-2=severe acute respiratory syndrome coronavirus 2. Inf=infinity. *15 confirmed close contacts were excluded from this analysis because contact tracing reports for the negative close contacts in the same clusters were missing. Close contacts with missing data on sex, age, contact types, or contact frequency not shown.

Table 3: **Group-specific attack rates and risk factors for SAR-CoV-2 infection among close contacts**

and estimated that 95% of those who develop symptoms will do so within 14·0 days (95% CI 12·2–15·9) of infection. We estimated that about 5·0% of cases who develop symptoms would not show symptoms until 14 days after infection.

Based on 228 cases with known outcomes, we estimated that median time to recovery was 20·8 days (95% CI 20·1–21·5). We estimated that the median time to recovery was 22·4 days (95% CI 20·8–24·1) in individuals aged 50–59 years, and was estimated to be significantly shorter in younger adults (eg, 19·2 days in individuals aged 20–29 years; appendix 2 pp 3, 10). In multiple regression models including sex, age, baseline severity, and method of detection, in addition to age, baseline severity was associated with time to recovery (appendix 2 p 3). Compared to cases with mild symptoms, those with severe symptoms had a 41% (95% CI 24–60) longer time to recovery (appendix 2 p 3). As of Feb 22, 2020, three cases had died. These deaths occurred 35–44 days from symptom onset and 27–33 days from confirmation.

Cases detected through symptom-based surveillance were confirmed on average 5·5 days (95% CI 5·0–5·9) after symptom onset (figure 2, appendix 2 p 2), whereas those detected by contact-based surveillance were confirmed on average 3·2 days (95% CI 2·7–3·8) after symptom onset. 17 (5%) of 342 cases with a known onset date and start date of quarantine were isolated before developing symptoms. Among those isolated after developing symptoms, the symptom-based surveillance group was, on average, isolated 4·6 days (95% CI 4·1–5·0) after symptom onset, whereas the contact-based surveillance group was isolated 2·7 days (2·1–3·3) after symptom onset. Hence, contact-based surveillance was associated with a 2·3-day (95% CI 1·5–3·0) decrease in time to confirmation and a 1·9-day (1·1–2·7) decrease in time to isolation. The mean time between symptom onset and admission to hospital was similar to time between symptom onset and isolation in both the symptom-based and contact-based surveillance groups (figure 2, appendix 2 p 2).

191 (64%) of 298 travellers developed symptoms after arriving in Shenzhen, with a mean time from arrival to symptom onset of 4·9 days (95% CI 4·2–5·5; appendix 2 p 2). Those developing symptoms before arrival or on the day of arrival were confirmed as cases on average 4·5 days (95% CI 3·8–5·1) after arrival, and isolated on average 3·1 days (2·5–3·7) after arrival.

Overall, 1286 close contacts were identified for index cases testing positive for SARS-CoV-2 between Jan 14 and Feb 9, 2020, with 244 (84%) of 292 cases having at least one close contact. 622 (95%) of 653 close contacts with known dates for the period when they were under quarantine were followed up for 12 days or longer. 98 of the close contacts tested positive for SARS-CoV-2 infection by RT-PCR, and one had presumptive infection. Assuming those with a missing test result were uninfected, we found that the secondary attack rate was 11·2% (95% CI 9·1–13·8) among household contacts and 6·6% (5·4–8·1) overall (the secondary attack rate increased to 14·9% [12·1–18·2] among household contacts and 9·7% [7·9–11·8] overall if those with missing results were removed from the denominator). In multiple conditional logistic regression analysis of contact types, household contact (OR 6·3; 95% CI 1·5–26·3) and travelling together (OR 7·1; 1·4–34·9) were significantly associated with infection (table 3). Reporting contact that occurred often was also associated with increased risk of infection compared with moderate-frequency contact (OR 8·8; 95% CI 2·6–30·1; table 3).

Attack rates were similar across all age categories of infected contacts (table 3), although we observed some indication of elevated attack rates in older age groups (figure 3). Notably, the rate of infection in children younger than 10 years (7·4%) was similar to the population average (6·6%). There was no significant association between probability of infection and age of the index case. Surprisingly, in univariate analysis a longer time in the community before isolation was associated with a reduced risk of causing infections (data not shown). However, this association was no longer significant after adjusting for contact frequency and type.

Based on 48 pairs of cases with a clear relationship between the index case and secondary case and time of symptom onset, we estimated that the serial interval is gamma distributed with a mean of 6·3 days (95% CI 5·2–7·6) and an SD of 4·2 days (95% CI 3·1–5·3; figure 1B, appendix 2 p 2). Hence, 95% of secondary cases were expected to develop symptoms within 14·3 days (95% CI 11·1–17·6) of their infector. This estimate includes the effect of isolation on truncating the serial interval. Stratified results show that if the infected individual was isolated less than 3 days after infection the average serial interval was 3·6 days, increasing to 8·1 days if the infected individual was isolated on the third day after symptom onset or later (appendix 2 p 3).

The mean number of secondary cases caused by each index case (ie, the observed reproductive number, *R*), was 0·4 (95% CI 0·3–0·5). The distribution of individual *R* values was highly over-dispersed, with 80% of infections being caused by 8·9% (95% CI 3·5–10·8) of cases (negative binomial dispersion parameter 0·58; 95% CI 0·35–1·18).

We examined the potential impact of surveillance and isolation through truncating the infectious period.

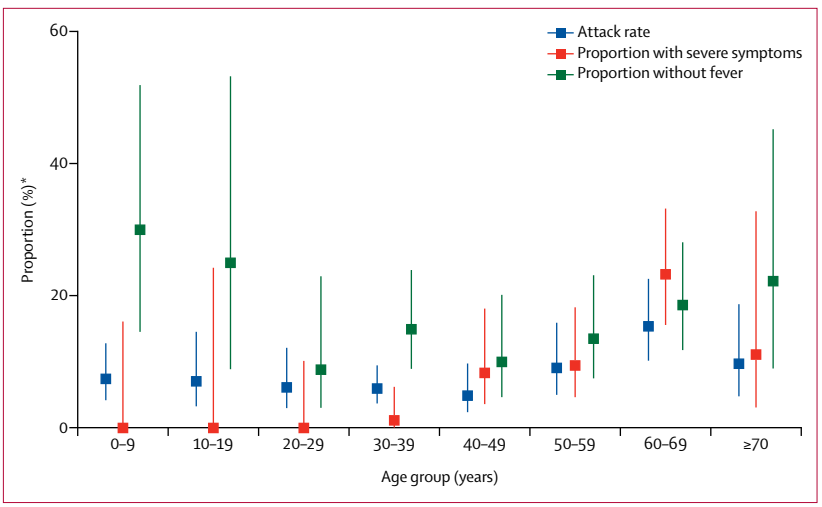

Figure 3: **Attack rate among close contacts, baseline severity, and proportion of cases without fever at initial assessment by age group**

*Proportion of close contacts for attack rate; proportion of all cases for those with severe symptoms or no fever at initial assessment.

Because of the scarce understanding of infectious periods following SARS-CoV-2 infection, we considered a range of possible infectious periods where infectiousness varies over time and follows a gamma distribution (appendix 2 p 11). We defined the mean infectious day (ie, the average number of days after symptom onset that an infected individual is expected to infect a secondary case) as the weighted mean of the infectious period, where each day is weighted by relative infectiousness. We considered periods where the mean infectious day is less than 15 days after symptom onset (roughly the period of SARS and early SARS-CoV-2 reports),[16,17] and assumed that R_0=2·6 and that isolation effectively ends the infectious period. Under these assumptions we found that if the mean infectious day is greater than 5 days, then it might be possible to bring *R* below one in cases detected by symptom-based surveillance, and the same can be accomplished by contact-based surveillance if the mean infectious day is greater than 3 days (appendix 2 p 9). For the impact of passive surveillance alone to achieve our observed *R* of 0·4, we projected that the mean infectious day must be at least 5·5 days (and likely greater) after symptom onset.

Even if transmission is completely eliminated in the group captured by surveillance (eg, if we could get perfect surveillance on the day of symptom onset), assuming R_0=2·6, the cases captured by surveillance must, if not isolated, be expected to cause 61% of onward transmission to achieve local elimination by surveillance and isolation alone (appendix 2 p 9).

Discussion

This analysis of early SARS-CoV-2 cases and their close contacts in Shenzhen, China, provides insight into the natural history, transmission, and control of this disease.

The values estimated provide the evidentiary foundation for predicting the impact of this virus, evaluating control measures, and guiding the global response. Analyses of how cases are detected, and use of data on individuals exposed but not infected, indicate that infection rates in young children are not lower than the population average (even if rates of clinical disease are). We were able to directly estimate important transmission parameters, and show that, at least among observed contacts, transmission rates are low. Estimates of the distribution of time between symptom onset and case isolation by surveillance type reveal that heightened surveillance combined with case isolation could plausibly account for these low rates of transmission. These results paint a positive picture of the impact of heightened surveillance and isolation in Shenzhen. However, uncertainty in the number of asymptomatic cases missed by surveillance and their ability to transmit SARS-CoV-2 must temper any hopes of stopping the COVID-19 pandemic by these measures.

This work further supports the understanding of COVID-19 as a disease with a fairly short incubation period (mean 4–6 days) but a long clinical course,[2,7,18] with patients taking many weeks to die or recover. Notably, however, we estimate a higher proportion of cases taking 14 days or more to develop symptoms (5%) than estimated in the study by Lauer and colleagues (1%).[6]

Focusing on cases detected through contact-based surveillance adds nuance to previous characterisations of COVID-19. Since RT-PCR testing of contacts is near universal, we can assume these cases are more reflective of the average SARS-CoV-2 infection than cases detected through symptomatic surveillance. In the contact-based surveillance group, any tendency for cases to be male or older (beyond the underlying population distribution) disappears. Furthermore, in this group, 20% of cases were asymptomatic at the time of first clinical assessment and nearly 30% did not have a fever. This observation is consistent with a reasonably high rate of asymptomatic carriage, but lower than that suggested by some modelling studies,[19] although RT-PCR has imperfect sensitivity.[20]

In Shenzhen, SARS-CoV-2 transmission most probably occurred between very close contacts, such as individuals sharing a household. However, even in this group fewer than one in six contacts (ie, secondary attack rate 11–15%) were infected, and overall we observed far fewer than one (0·4) onward transmission per primary case. As noted above, low transmission levels might in part be due to the impact of isolation and surveillance, but it is equally likely that unobserved transmission has some role. We also estimated reasonably high rates of overdispersion in the number of cases caused by each infected individual, leaving open the possibility that large COVID-19 clusters can occur even if surveillance and isolation are forcing R below one—events that could potentially overwhelm the surveillance system.

This work has numerous limitations. As in any active outbreak response, the data were collected by multiple teams under protocols that, by necessity, changed as the situation developed. Hence, there might be noise and inconsistency in definitions. Notably, the definition of a confirmed case changed to require symptoms near the end of our analysis period (Feb 7), but sensitivity analyses show that truncating the data at this point does not qualitatively influence results (appendix 2 pp 5–6). It is, likewise, impossible to identify every potential contact an individual has, so contact tracing focuses on those close contacts who are most likely to be infected; hence our R is assuredly lower than the true reproductive number in the population. Asymptomatic travellers will be missed by symptom-based surveillance and, even if they are tested, some asymptomatic contacts might be missed because of the imperfect sensitivity of the RT-PCR test.[20] Recovery time in Shenzhen is likely to be inflated because cases are required to be isolated for 2 weeks and release is conditional on a negative RT-PCR test. Although guidelines for contact tracing and case detection are supposed to be implemented across the country, whether the impact of contact tracing can be generalised to other parts of China depends on a range of factors, including local testing capacity and surveillance resources.

As SARS-CoV-2 continues to spread, it is important that we continue to expand our knowledge about its transmission and natural history. Data from the early phase of local outbreaks, when detailed contact tracing is possible and sources of infection can still be reliably inferred, are particularly powerful for estimating critical values pertinent to describing transmission and the natural history of a disease. This is especially true when information about uninfected contacts and mode of detection is used, as we have done here. The resulting estimates provide important inputs for interpreting surveillance data, evaluating interventions, and setting public health policy.

Contributors

YoW, SM, XZ, ZZ, XL, LW, WG, CC, XT, XW, YuW, SH, and TF collected the data. JL, QB, SAT, and CY did statistical analyses, and drafted the manuscript and figures. TZ, BS, YS, JZ, TM, and CY collected and cleaned data. QB, TM, JL, and TF conceived the study and supervised data collection.

Declaration of interests

We declare no competing interests.

Acknowledgments

TM, CY, TZ, BS, YS, and JZ were funded by the Emergency Response Program of Harbin Institute of Technology (HITERP010) and Emergency Response Program of Peng Cheng Laboratory (PCLERP001). JL, SAT, and QB were funded by a grant from the US Centers for Disease Control and Prevention (NU2GGH002000). We thank A Azman, D Cummings, S Lauer, J Wallinga, and M Mina for advice and input on the manuscript and analyses. We thank all patients, close contacts, and their families involved in the study, as well as the front-line medical staff and public health workers who collected these important data.

References

1 Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020; **382:** 1199–207.

2 Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020; **395:** 507–13.

3 Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 2020; **323:** 1061–69.

4 Yang Y, Lu Q-B, Ming-Jin L, et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. *MedRxiv* 2020; published online Feb 21. DOI:10.1101/2020.02.10.20021675 (preprint).

5 WHO. Coronavirus disease (COVID-19) situation reports. 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (accessed Feb 21, 2020).

6 Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Int Med* 2020; published online March 10; DOI: 10.7326/M20-0504.

7 Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Euro Surveill* 2020; **25:** 2000062.

8 Chan JF-W, Yuan S, Kok K-H, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020; **395:** 514–23.

9 Lessler J, Reich NG, Cummings DAT, Nair HP, Jordan HT, Thompson N. Outbreak of 2009 pandemic influenza A (H1N1) at a New York City school. *N Engl J Med* 2009; **361:** 2628–36.

10 Lipsitch M, Cohen T, Cooper B, et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science* 2003; **300:** 1966–70.

11 Lessler J, Salje H, Van Kerkhove MD, et al. Estimating the severity and subclinical burden of Middle East respiratory syndrome coronavirus infection in the Kingdom of Saudi Arabia. *Am J Epidemiol* 2016; **183:** 657–63.

12 Reich NG, Lessler J, Azman A, Kamvar ZN. Analysis of coarsely observed data [R package coarseDataTools version 0·6–5]. Version 0.6–5. Dec 6, 2019. https://CRAN.R-project.org/package=coarseDataTools (accessed Feb 16, 2020).

13 Reich NG, Lessler J, Cummings DAT, Brookmeyer R. Estimating incubation period distributions with coarse data. *Stat Med* 2009; **28:** 2769–84.

14 Reich NG, Lessler J, Cummings DAT, Brookmeyer R. Estimating absolute and relative case fatality ratios from infectious disease surveillance data. *Biometrics* 2012; **68:** 598–606.

15 Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am Stat* 1983; **37:** 36–48.

16 Peiris JSM, Chu CM, Cheng VCC, et al. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* 2003; **361:** 1767–72.

17 Zou L, Ruan F, Huang M, et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med* 2020; **382:** 1177–79.

18 Pan F, Ye T, Sun P, et al. Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. *Radiology* 2020; published online Feb 13. DOI:10.1148/radiol.2020200370.

19 Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S, Ferguson NM. Report 3: transmissibility of 2019-nCoV. Jan 25, 2020. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-transmissibility-25-01-2020.pdf (accessed Feb 24, 2020).

20 Fang Y, Zhang H, Xie J, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* 2020; published online Feb 19. DOI:10.1148/radiol.2020200432.