

The epidemic situation of COVID-19 in Gansu Province, China — a Big Data analysis of the National Health Information Platform

Xuanchen Yan

Health Statistics Information Center of Health Commission of Gansu Province

Jianjian Wang

School of Public Health, Lanzhou University

Jingwen Yao (✉ 18368914063@163.com)

Health Statistics Information Center of Health Commission of Gansu Province

Janne Estill

Institute of Global Health, University of Geneva

Shouyuan Wu

School of Public Health, Lanzhou University

Jie Lu

Health Statistics Information Center of Health Commission of Gansu Province

Baoping Liang

Health Statistics Information Center of Health Commission of Gansu Province

Hongmin Li

Health Statistics Information Center of Health Commission of Gansu Province

Shengxin Tao

Health Statistics Information Center of Health Commission of Gansu Province

Huanli Bai

Health Statistics Information Center of Health Commission of Gansu Province

Hongliang Liu

Health Statistics Information Center of Health Commission of Gansu Province

Yaolong Chen (✉ sinograde@163.com)

School of Public Health, Lanzhou University

Research Article

Keywords: Corona virus disease 2019(COVID-19), Big data, Gansu Province National Health Information Platform

Posted Date: April 16th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-23005/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objective This study is to explore the patient characteristics and transmission chains of COVID-19 in the population of Gansu province, and support decision-making.

Methods We collected data from Gansu National Health Information Platform. We conducted a case-control study, including patients with COVID-19 confirmed between January 23 and February 6, 2020 as cases, and all inpatients, outpatients and emergency patients except those with COVID-19 during the same period as the control group. We compared the gender and age between two groups. We also described the incubation period, consultation time and sources of infection in the cases, and calculated the secondary cases that occurred within Gansu for each imported case.

Results We found no differences in gender ($P=0.107$). COVID-19 cases were more likely to occur among the middle-aged people ($P=0.038$). Twenty-eight (41.8%) of the 67 cases had a history of direct exposure in Wuhan. Twenty-five (52.2%) cases came from ten families, and we found no clear reports of modes of transmission other than family clusters. The largest number of secondary cases linked to a single source was nine.

Conclusion There is no difference in gender between the cases and controls. However, COVID-19 patients were slightly younger than those attending for other reasons. The most common suspected mode of transmission was through family cluster. Gansu and other settings worldwide should continue to strengthen the utilization of big data in epidemic control.

1. Introduction

The first cases of corona virus disease 2019 (COVID-19) appeared in Wuhan, Hubei Province, China, in December 2019, and has since then spread quickly within Wuhan, Hubei, China, and the rest of the world (1,2). The World Health Organization (WHO) officially named the virus that caused the pneumonia epidemic in Wuhan first as "2019 New Coronavirus" (2019-nCoV) on January 12, 2020 (3) and later as "SARS Coronavirus 2" (SARS-CoV-2). The disease was officially named as COVID-19 on February 11, 2020 (4). SARS-CoV-2 is a new type of coronavirus that has not been found in humans before. SARS-CoV-2 is characterized by its rapid spread, high contagiousness, and the high susceptibility of the population (5). The National Health and Health Commission have reported 80,270 cumulative confirmed cases and 2981 deaths in China until 24:00 on March 3, 2020 (6), including 91 confirmed cases and two deaths in Gansu Province (7).

Big data are defined as "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation." (8) In medicine, big data are increasingly used in public health promotion (disease monitoring and population management), healthcare management (quality control and

performance measurement), drug and medical device surveillance, routine clinical practice (risk prediction, diagnosis accuracy, and decision support), and research (9). The combined use of big data resources and new technologies has great potential to solve many existing medical problems and provide better evidence for decision making (10). Big data may be one of the most efficient tools of scientific prevention and control for this public health emergency (11,12). Therefore, we aimed to obtain the epidemic status and characteristics of COVID-19 in Gansu Province, based on the Gansu Province National Health Information Platform, to accumulate the knowledge of this new disease and support epidemic prevention and control in Gansu and elsewhere.

2. Methods

2.1 Data Sources

We obtained data from the Gansu Province National Health Information Platform, and linked these with the data of the COVID-19 epidemic supervision platform, electronic case database, residents' health files, and the entire population information database. In May 2016, the Health Commission of Gansu Province launched the construction of the National Health Information Platform, which includes the following databases: the database of the whole population (basic information collection system for the 26 million permanent residents and 3 million floating population in Gansu Province), the electronic health record database (a total of 28,953,500 electronic health records) and the electronic case database (about 110 million health records daily). After three years of development, Gansu Province has made remarkable progress in health informatization, and accumulated a large number of high-quality health care data together. The construction of the COVID-19 epidemic supervision platform relies on the information database containing the entire population. We wrote dynamic SQL execution scripts, and performed data correlation, quality control, and statistics on Hadoop and Hase calculation engines. We adopted implemented data quality control using multiple verification in a three-in-one manner.

We collected data from the above-mentioned databases on all inpatients, outpatients and emergency patients attending hospitals in Gansu Province within a 14-day period of January 23 to February 6, 2020 (the period corresponding to the estimated maximum incubation period of SARS-CoV-2, starting on the day of the first travel restrictions concerning Wuhan) (13). Patients with a confirmed SARS-CoV-2 infection formed the case group for our analysis, and all other patients the control group. The data on exposure and likely route of SARS-CoV-2 of the patients in the case group were obtained from the official website of the Gansu Provincial Health and Health Committee and media releases from the city and provincial epidemic prevention and joint control working group office. In the National Health Information Platform, all operations are strictly done in accordance with the outline of *Action plan for big data development* (14) and strictly comply with the personal information protection law and ethical standards. Our data was obtained with the consent of the Gansu Provincial Health and Health Commission. To protect the privacy of the individuals, data encryption technology is used in occasion of data use, processing, sharing and interaction. None of the study personnel could see the personal information of individuals.

2.2 Case definition

We defined a confirmed case of COVID-19 according to the "Diagnosis and treatment of COVID-19 (Trial Version 3)" (15). For a case to be confirmed as COVID-19, the detection of either SARS-CoV-2 in a real-time fluorescent PT-PCR of respiratory specimens or blood specimens, or a gene highly homologous with the known gene of SARS-CoV-2 in viral gene sequencing, was required.

2.3 Data extraction and analysis

Data on basic case information such as age, gender, history of exposure, incubation period, and treatment time were collected from the Gansu National Health Information Platform. Continuous variables were expressed as medians, and categorical variables as absolute numbers and percentages. Comparisons between the gender distribution in the case and control groups were performed using the Pearson chi-square test, and the means of age distributions were compared with the t-test. All analyses were conducted using SPSS 25.0 (SPSS Inc., Chicago, Illinois, USA).

In addition, we analyzed the possible source of infection among the confirmed cases in Gansu Province through collecting data on possible routes of exposure and following up the patient. We also explored the possibility of having a "super spreader", defined as an individual who infected more than 10 persons, linked to the cases in Gansu.

3. Results

3.1 The case-control study

We retrieved data from 67 cases of COVID-19 confirmed between January 23 and February 6, 2020. Thirty-six (53.7%) of the patients were women and thirty-one (46.3%) men. The control group included data from 869,406 individuals attending the hospital during the same time period, of whom 549,657 (63.2%) were women and 319,749 (36.8%) men. We found no significant difference in gender ($\chi^2=2.595$, $P=0.107$). The median ages were 40 (IQR 31-53) years for confirmed cases, and 48 (IQR 25-55) years for the control group ($t=-2.123$, $P=0.038$).

Of the 67 cases of COVID-19 confirmed between January 23 and February 6, 2020, forty-one (61.2%) had spent time outside Gansu province, and 28 cases (41.8%) in Wuhan. The time interval from arrival into Gansu Province to the onset of symptoms ranged between 8 days before arriving in Gansu to 12 days after arriving in Gansu, with a mean of 3.5 ± 4.7 days after arrival. The interval from onset of symptoms to the initiation of treatment was 0-18 days (mean 4.4 ± 4.0 days), from onset of symptoms to hospitalization 0-17 days (mean 6.4 ± 3.7 days). In addition, 35 (52.2%) of the confirmed cases come from only ten families, showing obvious family clustering.

3.2 Source of infection

We traced the epidemiological history of the 67 confirmed cases, and found that 35 infections were directly imported from outside the province or from an unknown source. Twenty-three of these patients were properly controlled and caused no further transmissions. Twelve patients continued to spread the infection further, causing therefore 32 onward transmissions. The majority ($n=19$) of these 32 patients were linked to only three sources. This calculation shows that none of the patients from Gansu fulfilled the definition of a “super spreader”. (Figure 1).

4. Discussion

This is the first “big data” study to explore the transmission chains and epidemiological characteristics of COVID-19 in Gansu Province. We found no differences in gender between the patients with COVID-19 and other patients attending hospital in the same period, covering the duration of the estimated maximum incubation period of SARS-CoV-2 starting from 23 January 2020 when the first travel restrictions took place. Most infections in Gansu Province were imported from Wuhan. The most common suspected mode of transmission was through family clusters, and we found no clear reports of other modes of transmission.

4.1 Gender, age and aggregation

More women than men were diagnosed with COVID-19 in Gansu Province, although women formed also the majority of all patients seeking care in hospitals within the province. This finding is inconsistent with the situation at the beginning of the outbreak, men accounting for 56% of the first 425 patients diagnosed in Wuhan (16). A retrospective, single-center study published on February 15, 2020 on 99 cases in Wuhan, also indicated that men were at higher risk than women. The proposed reduced susceptibility of females to viral infections could be attributed to the protection from the X chromosome and sex hormones, which play an important role in innate and adaptive immunity (17,18). However, there is no clear evidence yet for the higher susceptibility of women than men.

Although patients with confirmed COVID-19 tend to be younger than patients seeking care for other reasons, the age range of confirmed cases of COVID-19 in Gansu Province covered almost all age groups. This is in line with previous research findings. People of all ages have been shown to be susceptible to SARS-CoV-2, and are thus at risk of acquiring the infection as long as the conditions necessary for transmission are met (5,19). An analysis of 4021 confirmed cases in China also showed that people of all ages are generally susceptible: 71.5% of the patients were aged 30 to 65 years, and 0.4% were children under the age of 10 years (20). However, the risk of acquiring SARS-CoV-2 may be increased in the elderly and people with chronic underlying diseases such as asthma, diabetes and heart disease (19).

At the early stage of the epidemic, cases of COVID-19 were mainly sporadic. The proportion of clustered epidemics in various locations has continued to increase, which has also changed the development of the epidemic and the sources of exposure. The number of cases linked to clustered epidemics is estimated to account for 50% to 80% of all confirmed cases in several provinces and cities including Beijing, Shanghai, Jiangsu, and Shandong (19). The results of this study, covering a period of 14 days

which is the estimated maximal incubation period, show that although no super spreaders were found, many cases were clustered in families or neighborhoods. This shows that there was ongoing human-to-human transmission also in Gansu Province. These characteristic related to clustering and sources of infection are consistent with reports from Shanxi, Chongqing, and other provinces and cities (21,22). The incubation period of SARS-CoV-2 is generally 4 to 7 days, and a large number of suspected patients and asymptomatic infections become the main source of infection (18,19). Therefore, it is important to track people with asymptomatic infections, and block the occurrence of familial and spacious aggregation. Gansu Province National Health Information Platform is linked with the entire population information database. Such data sources give new opportunities to accurately delineate the group of people with closest contacts and realize early warning to prevent family aggregation of the disease.

4.2 The potential of big data on prevention and control

As a new infectious disease, the spread of COVID-19 was accelerated at the beginning of the epidemic by the delayed in diagnosis, treatment, and epidemic management due to lack of awareness. China's traditional surveillance network mainly relies on reporting and summarizing the situation, which is however too slow to meet the need for rapid response to the epidemic. Big data has a huge potential to help to follow, control and respond to epidemics rapidly. The use of information technology and big data as an effective auxiliary method for epidemiological investigations can not only achieve early detection, early reporting, early isolation, and early treatment of cases, but also quickly map out the current status of the disease, grasp the patients' past medical history, and help to track the sources of infection and control the epidemic. As the big data network allows almost real-time disease monitoring, this comprehensive and rapid surveillance method will make public health surveillance more sensitive, especially to trace the unconscious close contacts and provide the necessary control measures to prevent further infections (23).

Big data in the field of medicine and public health has become one of the most important medical resources, and has played an active role in the prevention and control of COVID-19 in Gansu Province. In particular, it has provided great support for the management of source of infection and the development of epidemiological investigations. In the next step of prevention and control, Gansu Province should continue to strengthen the innovation construction of "big data + epidemiology" (23), prevent the recurrence of imported cases and cluster epidemics, and continuously improve the construction and promotion of big data related platforms, so as to provide a theoretical basis for facing related emergencies to facilitate scientific epidemic prevention and decision-making in the future.

5. Conclusion

In summary, women formed the majority of both COVID-19 cases and all other patients attending hospitals in Gansu during the two-week observation period. The overall age range was broad, confirming that people of all ages are susceptible to the virus. The most common suspected mode of transmission was through family cluster, which is promising since it makes the tracing of close contacts easier. This

study is the first exploratory application of the Gansu Province National Health Information Platform for the prevention and control of COVID-19. We will continue to follow up and investigate the patients with COVID-19 in Gansu province through the platform, and further explore the second generation incidence rate, risk assessment and long-term effect of COVID-19. The results of this study demonstrate the great potential of “big data” in epidemiological research. Gansu and other provinces and settings should continue to strengthen the utilization of “big data” in epidemic prevention and control, to prevent the development of imported cases into clustered epidemics.

Declarations

Acknowledgements: None.

Author Contributions: Xuanchen Yan planned the study and collected the data, Jianjian Wang contributed to writing and editing. Other authors contributed to modifying and reviewing. All authors contributed to the interpretation of the results and approved the final version of the manuscript.

Conflicts of Interest: The authors have no conflicts of interest to declare.

Funding: This study was funded by a research project supported by the Department of Planning and Information of National Health Commission of the People's Republic of China: *Research on the path of national health information construction in underdeveloped areas*.

Ethics Statement: The data of our study was obtained with the consent of the Health Commission of Gansu Province.

References

1. World Health Organization. Novel Coronavirus – China. <https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/>. Accessed 10 March 2020.
2. Wuhan Municipal Health Commission. Report of Wuhan Municipal Health Commission on the current situation of COVID-19 in our city. <http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989>. Accessed 10 March 2020.
3. World Health Organization. Naming the coronavirus disease (COVID-19) and the virus that causes it. [https://www.who.int/zh/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/zh/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it). Accessed 10 March 2020.
4. World Health Organization. WHO Director-General's remarks at the media briefing on 2019-nCoV. <https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020>. Accessed 10 March 2020.
5. National Health Commission of the People's Republic of China. National Health Commission Press Conference. Beijing, 4 February 2020.

6. National Health Commission of the People's Republic of China. Update on the epidemic situation of COVID-19 as of 24:00 on March 3, 2020.
<http://www.nhc.gov.cn/xcs/yqfkdt/202003/7a5f57b3f1b94954b1fc25f81dacc874.shtml>. Accessed 4 March 2020.
7. Health Commission of Gansu Province. No new cases confirmed in Gansu and 91 COVID-19 confirmed. <http://wsjk.gansu.gov.cn/single/10910/84585.html>. Accessed 10 March 2020.
8. Big data. <https://www.gartner.com/it-glossary/big-data/>. Accessed 10 March 2020.
9. The State Council of the People's Republic of China. The People's Republic of China. China issues guideline on artificial intelligence development.
http://english.gov.cn/policies/latest_releases/2017/07/20/content_281475742458322.htm. Accessed 10 March 2020.
10. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375:1216-9. doi:10.1056/NEJMp1606181.
11. Meng RT, Luo Y, Yu CH, Qiu J, Zhou D. Application and Challenges of Healthy Big Data in the Field of Public Health. *Chin Gen Prac*. 2015;18(35):4388-92. doi:10.2478/dim-2018-0014.
12. Mooney SJ, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. *Annu Rev Public Health*. 2018;39:95-112. doi:10.1146/annurev-publhealth-040617-014208.
13. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *Nature*. 2020. doi:10.1038/s41586-020-2012-7.
14. The State Council of the People's Republic of China. Action plan for big data development.
http://www.gov.cn/xinwen/2015-09/05/content_2925284.htm. Accessed 10 March 2020.
15. National Health Commission of the People's Republic of China. Diagnosis and treatment of COVID-19 (Trial Version 3). http://www.gov.cn/zhengce/zhengceku/2020-01/23/content_5471832.htm. Accessed 23 January 2020.
16. Li Q, Guan XH, Wu P, Wang XY, Zhou L, Tong YQ, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus—Infected Pneumonia. *N Engl J Med*. 2020. doi:10.1056/NEJMoa2001316.
17. Chen NS, Zhou M, Dong X, Qu JM, Gong FY, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395(10223):507-13. doi:10.1016/S0140-6736(20)30211-7.
18. Jaillon S, Berthenet K, Garlanda C. Sexual Dimorphism in Innate Immunity. *Clin Rev Allergy Immunol*. 2019;56(3):308-21. doi:10.1007/s12016-017-8648-x.
19. Special Expert Group for Control of the Epidemic of Novel Coronavirus Pneumonia of the Chinese Preventive Medicine Association, The Chinese Preventive Medicine Association. An update on the epidemiological characteristics of novel coronavirus pneumonia (COVID-19). *Chin J Epidemiol*. 2020;41(2):139-44. doi:10.3760/cma.j.issn.0254-6450.2020.02.002.

20. Yang Y, Lu QB, Liu MJ, Wang YX, Zhang AR, Jalali N, et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. medRxiv preprint. 2020.
<https://doi.org/10.1101/2020.02.10.20021675>. Accessed 10 March 2020.
21. Health Commission of Shaanxi Province. 23 new cases of COVID-19 confirmed in Shaanxi, 165 confirmed cases in total. http://sxwjw.shaanxi.gov.cn/art/2020/2/5/art_9_67765.html. Accessed 10 March 2020.
22. Health Commission of Chongqing. Epidemic situation of COVID-19 in Chongqing as at 24 February 15, 2020. <http://wsjkw.cq.gov.cn/yqxxyqtb/20200216/254438.html>. Accessed 10 March 2020.
23. CCTV network. "News 1 + 1"—Today's Epidemic Analysis. 30 January 2020.
<http://tv.cctv.com/2020/01/30/VIDEKbN7yigdsOCj4jnFrQxt200130.shtml?spm=C45404.PLCcHnO0TNnZ.E1jv2XeMQ3Zt.40>. Accessed 10 March 2020.

Figures

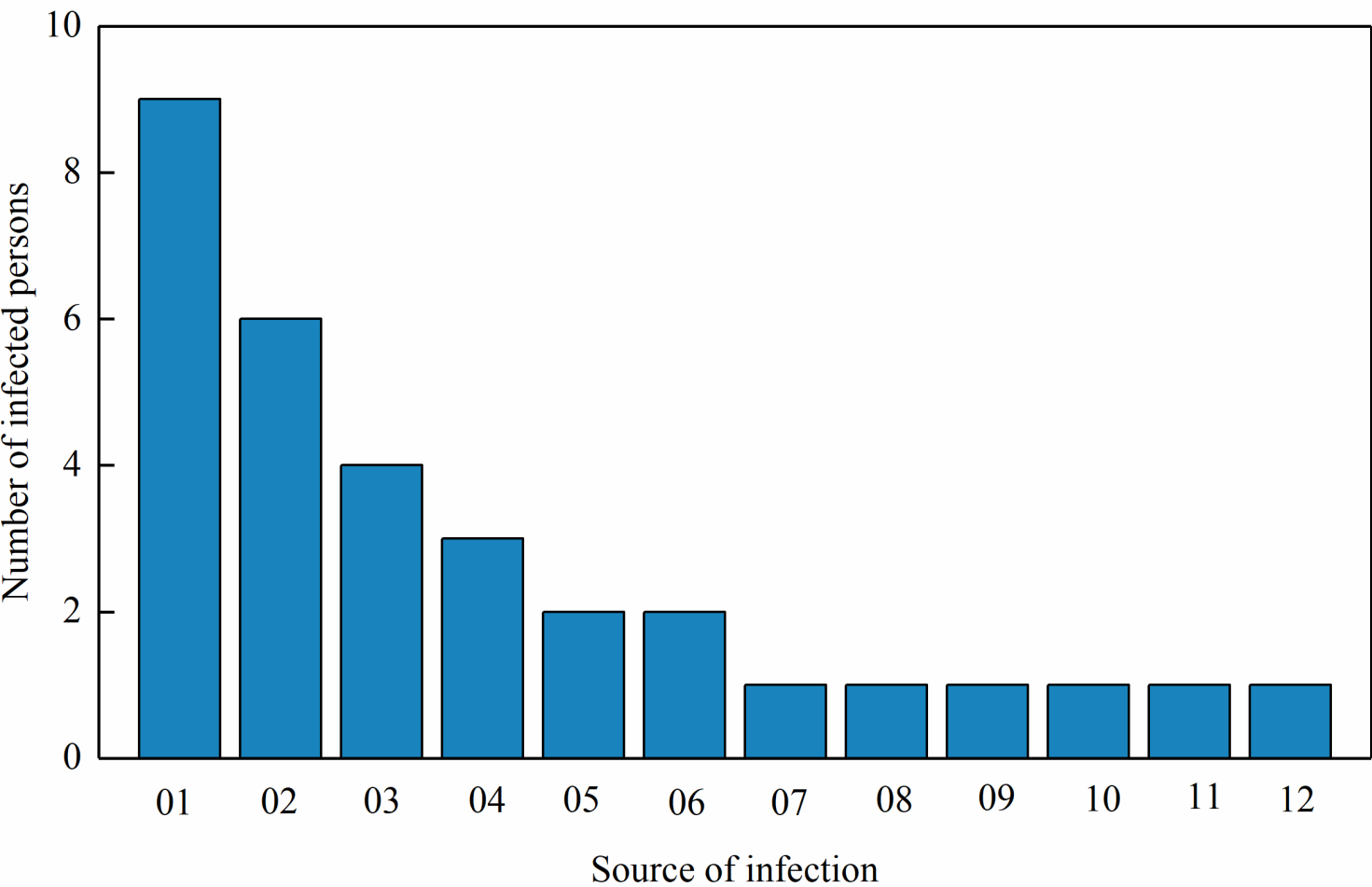


Figure 1

Number of secondary cases linked to the twelve sources who infected at least one other individual