

Original Paper

Mining the Characteristics of COVID-19 Patients in China: Analysis of Social Media Posts

Chunmei Huang^{1*}, MD; Xinjie Xu^{2*}, BMed; Yuyang Cai^{3*}, PhD; Qinmin Ge^{2*}, PhD; Guangwang Zeng^{1,4*}, BMed; Xiaopan Li^{5,6}, PhD; Weide Zhang⁷, MD; Chen Ji⁸, PhD; Ling Yang¹, PhD

¹Department of Geriatrics, Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

²Department of Emergency, Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

³School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China

⁴The Health Center of Nansheng Town, Wuzhishan, China

⁵Center for Disease Control and Prevention, Pudong New Area, Shanghai, China

⁶Pudong Institute of Preventive Medicine, Pudong New Area, Fudan University, Shanghai, China

⁷Big Data and Artificial Intelligence Center, Zhongshan Hospital, Fudan University, Shanghai, China

⁸Warwick Clinical Trials Unit, Warwick Medical School, Coventry, United Kingdom

*these authors contributed equally

Corresponding Author:

Ling Yang, PhD

Department of Geriatrics

Xinhua Hospital

Shanghai Jiaotong University School of Medicine

1665 Kongjiang Road

Shanghai

China

Phone: 86 13651608005

Email: yangling01@xinhumed.com.cn

Abstract

Background: In December 2019, pneumonia cases of unknown origin were reported in Wuhan City, Hubei Province, China. Identified as the coronavirus disease (COVID-19), the number of cases grew rapidly by human-to-human transmission in Wuhan. Social media, especially Sina Weibo (a major Chinese microblogging social media site), has become an important platform for the public to obtain information and seek help.

Objective: This study aims to analyze the characteristics of suspected or laboratory-confirmed COVID-19 patients who asked for help on Sina Weibo.

Methods: We conducted data mining on Sina Weibo and extracted the data of 485 patients who presented with clinical symptoms and imaging descriptions of suspected or laboratory-confirmed cases of COVID-19. In total, 9878 posts seeking help on Sina Weibo from February 3 to 20, 2020 were analyzed. We used a descriptive research methodology to describe the distribution and other epidemiological characteristics of patients with suspected or laboratory-confirmed SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) infection. The distance between patients' home and the nearest designated hospital was calculated using the geographic information system ArcGIS.

Results: All patients included in this study who sought help on Sina Weibo lived in Wuhan, with a median age of 63.0 years (IQR 55.0-71.0). Fever (408/485, 84.12%) was the most common symptom. Ground-glass opacity (237/314, 75.48%) was the most common pattern on chest computed tomography; 39.67% (167/421) of families had suspected and/or laboratory-confirmed family members; 36.58% (154/421) of families had 1 or 2 suspected and/or laboratory-confirmed members; and 70.52% (232/329) of patients needed to rely on their relatives for help. The median time from illness onset to real-time reverse transcription-polymerase chain reaction (RT-PCR) testing was 8 days (IQR 5.0-10.0), and the median time from illness onset to online help was 10 days (IQR 6.0-12.0). Of 481 patients, 32.22% (n=155) lived more than 3 kilometers away from the nearest designated hospital.

Conclusions: Our findings show that patients seeking help on Sina Weibo lived in Wuhan and most were elderly. Most patients had fever symptoms, and ground-glass opacities were noted in chest computed tomography. The onset of the disease was characterized by family clustering and most families lived far from the designated hospital. Therefore, we recommend the

following: (1) the most stringent centralized medical observation measures should be taken to avoid transmission in family clusters; and (2) social media can help these patients get early attention during Wuhan's lockdown. These findings can help the government and the health department identify high-risk patients and accelerate emergency responses following public demands for help.

(*J Med Internet Res* 2020;22(5):e19087) doi: [10.2196/19087](https://doi.org/10.2196/19087)

KEYWORDS

SARS-CoV-2; COVID-19; coronavirus disease; social media; Sina Weibo; help

Introduction

Background

In December 2019, pneumonia cases of unknown origin were reported in Wuhan City, Hubei Province, China. The illness was identified and officially named as coronavirus disease 2019 (COVID-19), which is caused by a novel viral strain called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1-3] and resembles severe acute respiratory syndrome coronavirus (SARS-CoV) [4]. Since the outbreak, COVID-19 has spread rapidly. Person-to-person transmission in hospital and family settings had occurred due to close contact [5,6]. On January 23, 2020, Wuhan shut down public transportation and was placed under lockdown, and residents were not allowed to leave the city. As of February 20, 2020, the accumulative number of laboratory-confirmed patients in Wuhan was 45,346. The health care system was further overburdened as patients with mild symptoms sought hospitalization instead of self-isolation, mainly due to the anxiety and panic instigated by the epidemic [7]. After failing to be admitted to a hospital, patients sought help on Sina Weibo, a Chinese microblogging site similar to Twitter that allows people to communicate and share information instantly [8]. Social media has become an important channel for promoting risk communication during the crisis [9,10] and can be used to measure public attention given to public health emergencies [11], such as H7N9 [12-14], Ebola [9,15-19], Zika virus [10,20,21], Middle East respiratory syndrome (MERS-CoV) [22], and Dengue fever [23].

Since the COVID-19 outbreak, social media, especially Sina Weibo, has become an important platform for the public to obtain epidemic-related information quickly and effectively. According to the official outbreak data released by Sina Weibo on February 26, 2020, 51.2 million users cumulatively posted 350 million pieces of epidemic-related content. Online readership of epidemic-related topics reached 754.5 billion. Sina Weibo established a communication channel that allowed the government to effectively listen and respond to public opinion quickly. Here, by collecting data from Sina Weibo from February 3 to 20, 2020, we aim to analyze the characteristics of suspected or laboratory-confirmed patients with the SARS-CoV-2 infection.

Objective

In this study, we describe the characteristics of suspected or laboratory-confirmed patients with the SARS-CoV-2 infection, the distribution of patients throughout Wuhan, and the relationship between helpers (eg, relative, friend, spouse, sibling) and patients. Social media was used to obtain timely access to

public demand so that the government and the health department could identify high-risk patients and take measures to help these patients.

Methods

Overview

Sina Weibo launched a platform to provide online help channels for patients infected with SARS-CoV-2. From February 3 to 20, 2020, we obtained 9878 posts by using the keyword 肺炎患者求助 (COVID-19 pneumonia patients seeking help) from Sina Weibo through its application programming interface (API). Python (Python Software Foundation) was used to implement a rule-based screening and classification method on the PyCharm platform. We used the collected posts as a training set, including related posts and unrelated posts. Based on the post-for-help rules formulated by Sina Weibo, we considered the post text, as well as keywords pertaining to name, age, home address, time of illness, and description of illness as a related post; otherwise, it was deemed an irrelevant post. We excluded 6922 irrelevant posts that only described opinions and feelings about help seeking related to COVID-19 and initially collected 2956 related posts that contained mentions of clinical symptoms and/or imaging descriptions. Then, we manually screened out and excluded posts. We excluded 1679 reposted posts, 556 posts with a significant amount of missing valid clinical data, 195 nonpneumonia patient posts, and 41 patient posts with non-Wuhan home addresses. Finally, we selected 485 patient posts that presented clinical symptoms and imaging descriptions (Figures 1 and 2). The number of patient posts on Sina Weibo has been declining because these patients have actively deleted posts upon hospital admission.

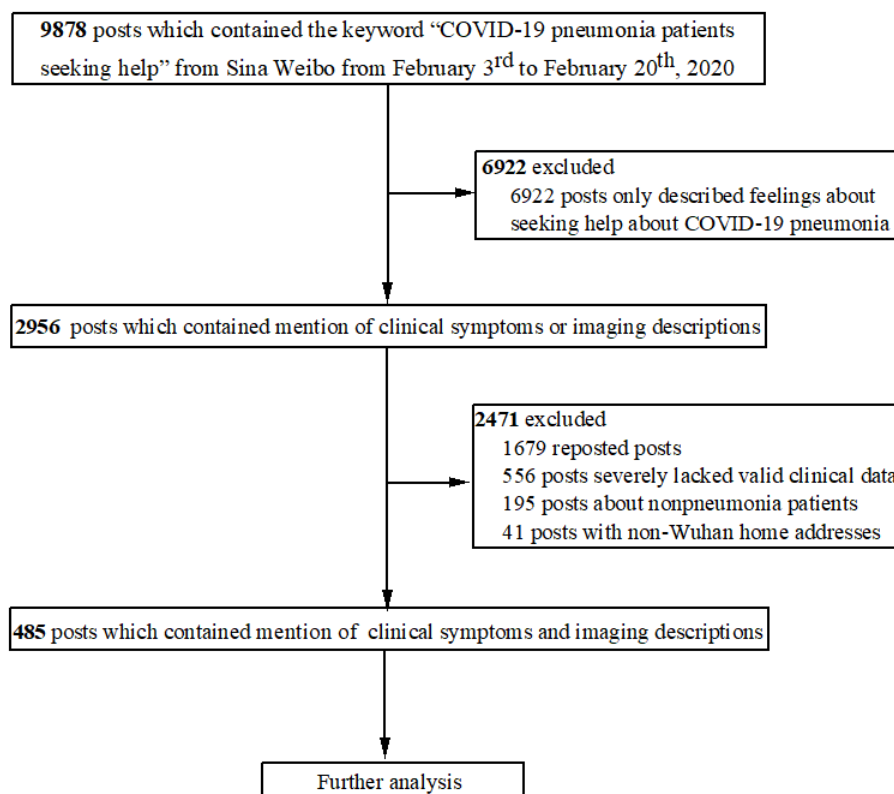
We collected clinical symptoms, chest computed tomography (CT) findings (the chest CT was only summarized for those who provided a clinical report), days from illness onset to online help, days from illness onset to RT-PCR testing, RT-PCR test results, the relationship between helpers and patients, and home address details from Sina Weibo's records. We performed a study on the clinical characteristics of suspected or laboratory-confirmed patients with the SARS-CoV-2 infection seeking help on Sina Weibo. Suspected cases were identified as having fever or respiratory symptoms such as shortness of breath, cough, productive sputum, or chest pain. A laboratory-confirmed case with SARS-CoV-2 infection was defined as a positive result to high throughput sequencing or real-time reverse transcription-polymerase chain reaction (RT-PCR) assay of throat swabs and sputum [2].

We also used a descriptive research methodology to analyze the distribution of patients throughout Wuhan and the relationship between helpers and patients. The distance from patients' home to the nearest designated hospital was calculated using the geographic information system ArcGIS. The data used in the current study is publicly accessible on Sina Weibo and readers can obtain the raw data online [24]. We have effectively

protected the privacy of subjects and strictly adhered to the principle of confidentiality in terms of information collection, storage and transmission, and information use and deletion. The study was approved by the Shanghai Jiaotong University Xinhua Hospital Ethics Committee and was carried out in accordance with the Declaration of Helsinki. We have made an application for exemption from informed consent and obtained approval.

Figure 1. An example of a patient with coronavirus disease (COVID-19) seeking help on Sina Weibo. RT-PCR: reverse transcription-polymerase chain reaction.



Figure 2. Study flow diagram. COVID-19: coronavirus disease.

Statistical Analysis

Continuous variables were expressed as median (IQR) when appropriate. Categorical variables were summarized as counts and percentages in each category. Analysis was conducted using SPSS, version 19.0 (IBM). We used ArcGIS, version 10.2.2, to plot the numbers of patients seeking help on a map.

Results

Demographic and Clinical Characteristics

We selected 485 patients with suspected or laboratory-confirmed SARS-CoV-2 infection with at least clinical symptoms and imaging descriptions from Sina Weibo. The demographic and clinical characteristics were shown in Table 1. The median age was 63.0 years (IQR 55.0 to 71.0), 0.21% (1/470) of patients were below 15 years of age, and 50.10% (243/485) were female. Fever (408/485, 84.12%) was the most common symptom. Other symptoms reported by patients included fatigue (224/485, 46.19%), shortness of breath (261/485, 53.81%), nausea or vomiting (81/485, 16.70%) and diarrhea (61/485, 12.58%). In total, 23.09% (112/485) of patients had at least one underlying disorder (eg, hypertension, chronic obstructive pulmonary disease, etc). All patients underwent chest CT. Of these patients,

35.26% (171/485) reported lung infection on the chest CT but did not provide their clinical reports. In the remaining 64.74% (314/485) of patients, the most common pattern on chest CT was ground-glass opacity (237/314, 75.48%) and bilateral patchy shadowing (191/314, 60.83%). The median time from illness onset to RT-PCR testing was 8.0 days (IQR 5.0-10.0), and the median time from illness onset to online help was 10.0 days (IQR 6.0-12.0). RT-PCR testing was performed in 52.16% (253/485) of patients; 68.38% (173/253) were positive, 1.98% (5/253) were suspected cases, and 10.67% (27/253) were negative.

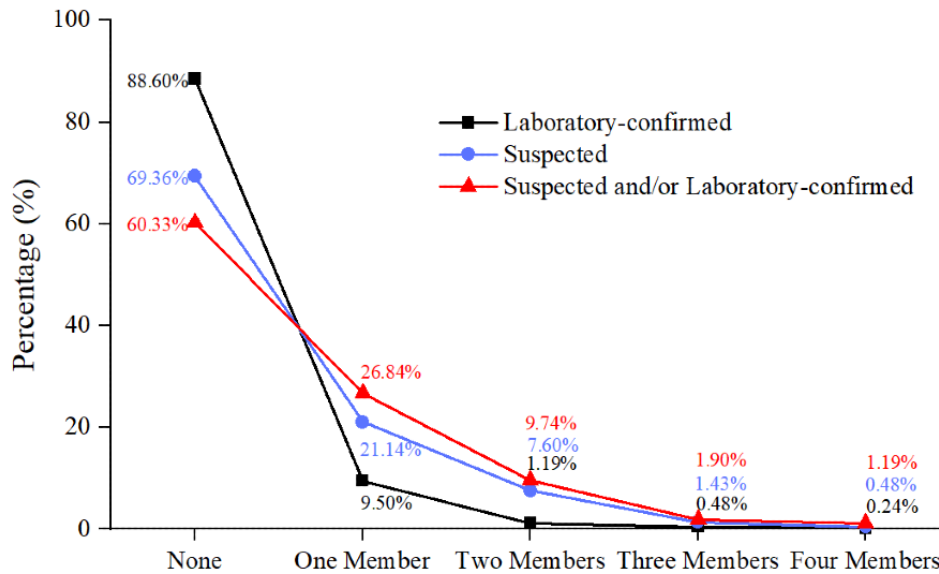
The 485 patients came from 421 families, and 39.67% (167/421) of these families had at least one family member with a laboratory-confirmed and/or suspected diagnosis of SARS-CoV-2; 11.40% (48/421) of families had one laboratory-confirmed family member only. Families with one confirmed case accounted for 9.50% (40/421); two, three, and four confirmed members accounted for 1.19% (5/421), 0.48% (2/421), and 0.24% (1/421), respectively. A suspected diagnosis occurred in 30.64% (129/421) of families; a family with one suspected member accounted for 21.14% (89/421), two suspected members accounted for 7.60% (32/421), three suspected members accounted for 1.43% (6/421), and four suspected members accounted for 0.48% (2/421) (Figure 3).

Table 1. Clinical characteristics of suspected or laboratory-confirmed patients with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection (N=485).

Characteristic	Value
Age (years), median (IQR)	63.0 (55.0-71.0)
Age groups (n=470), n (%)	
0-14 years	1 (0.21)
15-49 years	74 (15.74)
50-64 years	178 (37.87)
≥65 years	217 (46.17)
Sex (female; n=485), n (%)	243 (50.10)
Respiratory symptoms, n (%)	
Fever (temperature ≥37.3°C)	408 (84.12)
Cough	190 (39.18)
Fatigue	224 (46.19)
Shortness of breath	261 (53.81)
Nausea or vomiting	81 (16.70)
Diarrhea	61 (12.58)
Coexisting disorders (n=485), n (%)	
Any	112 (23.09)
Chronic obstructive pulmonary disease	10 (2.06)
Diabetes	43 (8.87)
Hypertension	55 (11.34)
Coronary heart disease	38 (7.84)
Cerebrovascular diseases	12 (2.47)
Cancer ^a	7 (1.44)
Chronic renal diseases	7 (1.44)
Immunodeficiency	2 (0.41)
Hepatitis B infection ^b	3 (0.62)
Radiologic findings: abnormalities on chest CT^c (n=314), n (%)	
Ground-glass opacity	237 (75.48)
Local patchy shadowing	20 (6.37)
Bilateral patchy shadowing	191 (60.83)
Interstitial abnormalities	6 (1.91)
Days from illness onset to online help, median (range)	10 (6-12)
Days from illness onset to RT-PCR ^d testing, median (range)	8 (5-10)
Underwent RT-PCR testing (n=253), n (%)	253 (52.16)
Positive	173 (68.38)
No result	48 (18.97)
Suspect	5 (1.98)
Negative	2 (10.67)

^aCancer referred to any malignancy.^bHepatitis B infection denotes a positive test for hepatitis B surface antigen, with or without elevated alanine or aspartate aminotransferase levels.^cCT: computed tomography.^dRT-PCR: reverse transcription-polymerase chain reaction.

Figure 3. The distribution of family clusters.



The Distribution of Patients Throughout Wuhan and the Distance Between Helpers and Hospitals

All patients were located in Wuhan, but more patients lived in the central districts (Hongshan, Jiang'an, Wuchang, Hanyang, and Qiaokou) compared to outskirt districts (Figure 4). We further analyzed the distance between patients and the nearest

designated hospital. Among these patients, four had missing home address information. We found that 25.57% (123/481) were within 1 kilometer of the nearest designated hospital, 24.74% (119/481) lived within 1-2 kilometers, 17.46% (84/481) lived within 2-3 kilometers, and 32.22% (155/481) lived more than 3 kilometers away (Table 2 and Figure 5).

Figure 4. The distribution of patients throughout Wuhan.

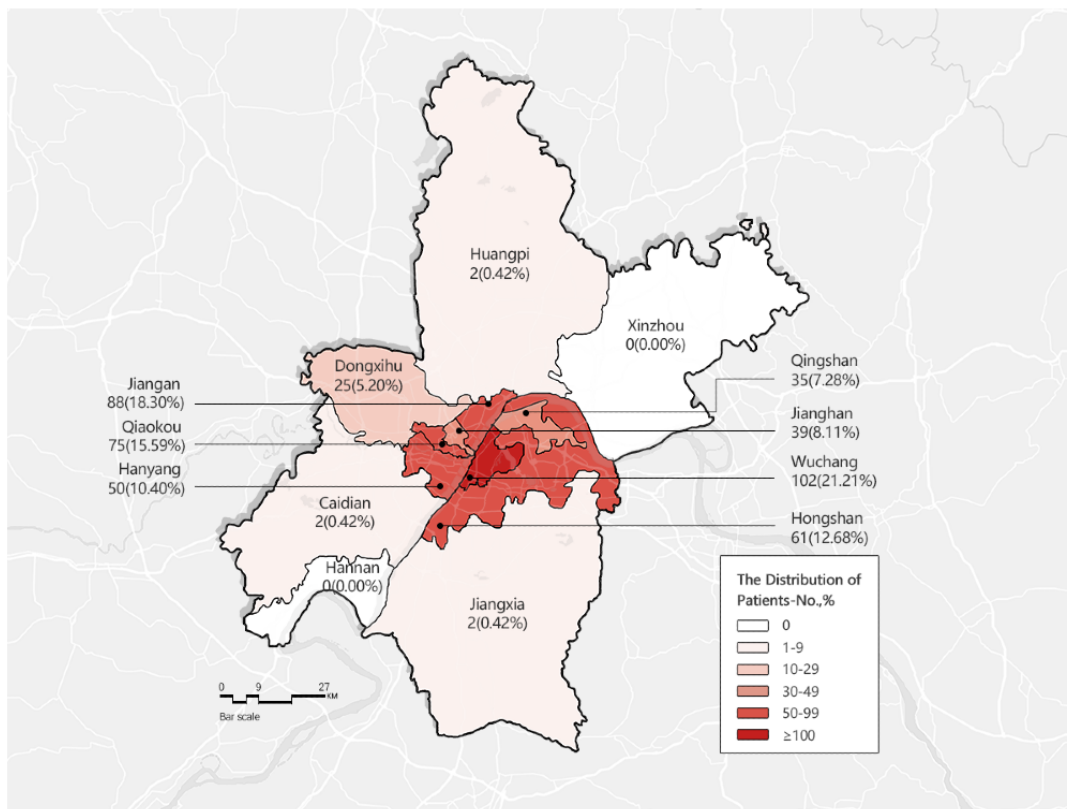
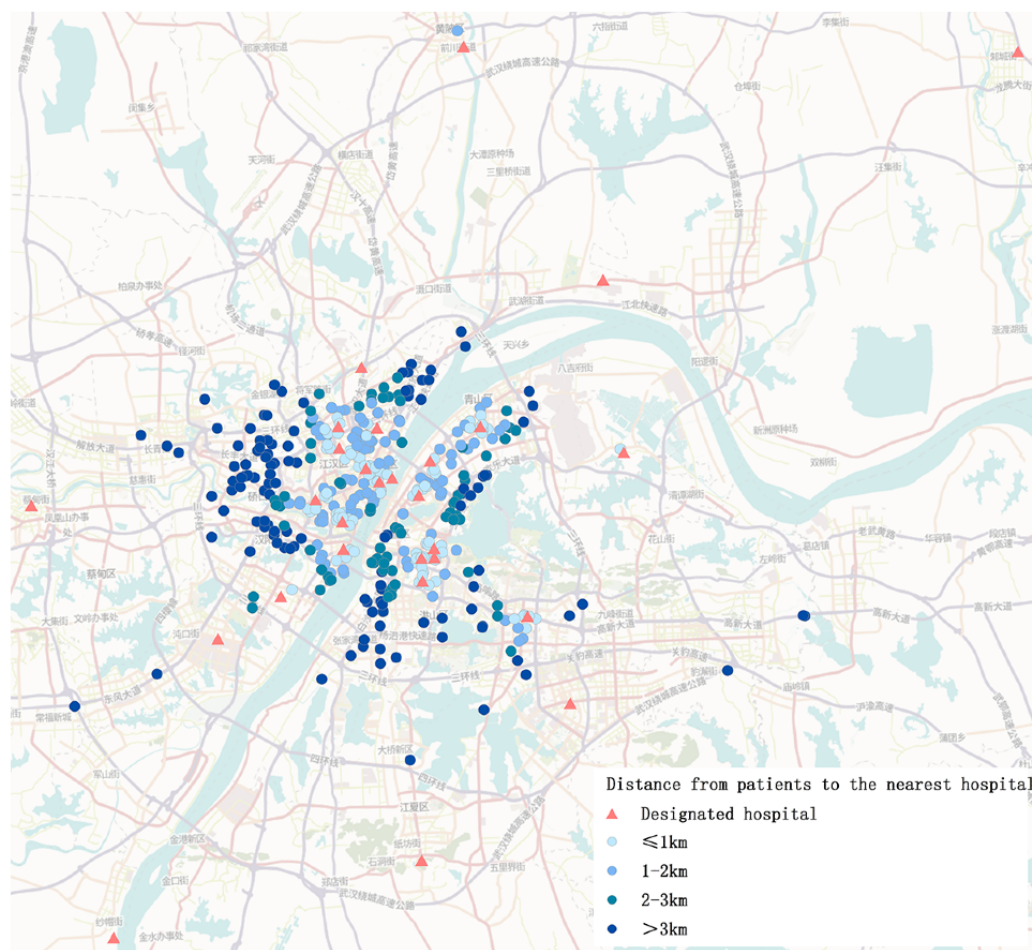


Table 2. The distance between patients and the nearest designated hospital, as well as the relationship between helpers and patients.

Variable	Count, n (%)
Distance (n=481)	
≤1km	123 (25.57)
1-2 km	119 (24.74)
2-3 km	84 (17.46)
≥3km	155 (32.22)
Relationship between helper and patient (n=329)	
Relative	232 (70.52)
Friend	38 (11.55)
Patient themselves	34 (10.33)
Spouse	14 (4.26)
Sibling	11 (3.34)

Figure 5. The distance between patients and their nearest designated hospital.



The Relationship Between Helpers and Patients

We explored the relationship between helpers and patients. During data collection, 156 helpers stated that they were family members of patients, but they did not specify their relationship. The remaining 70.52% (232/329) of helpers were the patients' relatives; 11.55% (38/329) were friends, 4.26% (14/329) were their spouses, 3.34% (11/329) were siblings, and 10.33% (34/329) were the patients themselves (Table 2).

Discussion

Principal Findings

This study has shown that patients seeking help on Sina Weibo lived in Wuhan and most of them were elderly. Our statistical analysis of the age of patients seeking help on Sina Weibo demonstrated that patients on Sina Weibo were older—the proportion of patients who were ≥65 years was as high as

46.17%. Zhong et al [3] reported that only a small proportion (15.1%) of 1099 laboratory-confirmed COVID-19 patients were aged ≥ 65 years. On the other hand, our study has found that the highest incidence was among adults over 50 years of age [25].

Additionally, 23.09% of patients had at least one underlying disorder. Fever was the dominant symptom whereas gastrointestinal symptoms were rare. Ground-glass opacity was the most common pattern on chest CT. Among all laboratory-confirmed COVID-19 patients, the most common pattern on chest CT were ground-glass opacity (56.4%) [3]. Our study has shown that the median time from illness onset to RT-PCR testing was 8 days, and the median time from illness onset to online help was 10 days. A recent study showed that the mean time from onset to hospital admission in 44 patients in Wuhan, with onset before January 1, was 12.5 days; in 189 patients with onset from January 1 to 11, the mean time was 9.1 days [5].

Person-to-person transmission of COVID-19 in hospital and family settings has been increasing [26-29]. Family clustering played an important part in increasing the number of COVID-19 cases [30]. Our study provided further evidence of human-to-human transmission, although 60.33% of families had no clustered onset, indicating that home isolation may be effective for patients. However, 39.67% of families had suspected and/or laboratory-confirmed cases among family members. In addition, 36.58% of families had 1 or 2 suspected and/or laboratory-confirmed family members. This is also in line with the finding that patients, on average, transmit the infection to 2.2 other people [5]. Therefore, home isolation might lead to the risk of COVID-19 outbreaks in family clusters [31]. This means that it is crucial to strictly isolate patients and trace and quarantine contacts as early as possible [32,33]. The most stringent centralized medical observation measures should be taken as soon as possible to avoid outbreaks in family clusters due to home isolation [31], such as a modular hospital to treat patients with mild illness [34].

Our research also found that the number of patients in the Wuchang, Jiang'an, Qiaokou, Hongshan, and Hanyang districts was greater than in other districts. Figure 4 shows a central agglomeration of patients; this may be consistent with the outbreak of the epidemic in the Huanan Seafood Wholesale Market in the Jiangnan district, which was thought to be the initial infection site from an animal source in China [35] or it may be related to the developed economy, convenient transportation, and the population density in the city center. Therefore, close contact with family members and actual population movements from the outbreak source were risk factors for the spread of SARS-CoV-2.

In total, 32.22% (155/481) of patients lived more than 3 kilometers away from their nearest designated hospital. According to Baidu maps, adults can walk 4 kilometers in 1 hour. Considering that the patients in this study were older and their health condition may have slowed them down even more, we estimate that patients could walk 3 kilometers in a 1-hour period. Hence, this indicates that a patient would need to walk more than 1 hour to see a doctor since public transportation was suspended at the time. This may be one of the reasons why

patients wanted to be admitted to a hospital. In addition, on February 5th, the Wuhan municipal health commission designated 28 hospitals for the treatment of laboratory-confirmed patients with the SARS-CoV-2 infection. The empty bed rate of hospitals within the city was only 3.6%. Thus, patients could not be hospitalized for the various reasons above. This also reflected an insufficiency of medical resources during the initial outbreak [36].

We also explored the relationship between helpers and patients. Judging from the content of Sina Weibo posts asking for help, "Mom" and "Dad" were high-frequency words; 70.52% (232/329) of helpers were the patients' relatives, indicating that the publishers of the help information were mostly the children of the elderly. Unfamiliarity with new technology may have hindered elderly people from seeking assistance from the outside world.

With the rapid and effective dissemination of help information, since February 5th, the People's Daily has launched an all-media operation to provide online help channels for patients with the SARS-CoV-2 infection. The government implemented a policy to maximize hospital admissions, which led to a rapid decrease in the number of people seeking help on Sina Weibo on February 6th and remained at low levels since February 8th, indicating that the needs of the public had been met. This also means that it is important to establish new and effective communication mechanisms for the dissemination of important factual information in a timely manner. Through this epidemic, we can see that medical resources are insufficiently allocated. There are substantial regional disparities in health care resource availability and accessibility in China [37]. The rapid increase in the number of patients during the initial outbreak led to a relative shortage of medical resources, which may threaten people with poor self-help capabilities such as the elderly. The government and health departments should pay attention to the elderly population during the outbreak. Social media can be used to understand public demand and aid the government in formulating accurate countermeasures following public demands for help. Although social media can establish effective communication channels, this technology may require a certain threshold, so the government should continue to increase the availability and accessibility of the network to better respond to public health emergencies.

Limitations

Our study has some limitations. First, given that our data was collected from a social media platform, the description of patients' symptoms and laboratory information were likely to be incomplete. Second, the urgent timeline for data extraction and the subjective judgment of the collectors might undermine the data quality to a certain extent. Finally, we learned that most of these patients have been admitted to the hospital with government help and many patients remain in the hospital, so we did not compare the 28-day rate for the composite endpoint.

Conclusions

In summary, our study found that the distance between patients and hospitals and the closure of public transportation further increased the difficulty of hospitalization for the elderly. We

recommend the application of centralized medical observations to avoid the spread of COVID-19 through family clusters. In a public health emergency, making full use of available social media platforms can establish effective, factual communication channels and shorten admission times, helping patients get early

attention during the Wuhan lockdown. These findings can help the government and health departments pay attention to the elderly population during the outbreak and accelerate emergency responses following public demands for help.

Acknowledgments

This study was supported by the Zhejiang University Special Scientific Research Fund for COVID-19 prevention and control (2020XGZX065); the National Natural Science Foundation of China (71432006); the National Social Science Fund of China (grant number 17BSH056); and the Shanghai Jiao Tong University think tank research project (ZKYJ-20200114). The funders had no role in the study design, data collection, data analysis and interpretation, writing of the report, or the decision to submit the paper for publication. The Authors were also supported by the Shanghai Jiaotong University special scientific research fund for "Technology Promotion Project" in 2019 (ZT201903) and the Shanghai science and Technology Fund (18441905200).

Authors' Contributions

CH, XX, YC, QG, and GZ are joint first authors. LY, QG, and YC are joint co-correspondence authors. LY and YC obtained funding. CH, XX, PL, YC, QG, and LY participated in study design; CH, XX, WZ, and GZ collected the data; CH, XX, GZ, WZ, and PL performed data analysis; CH and JC drafted the manuscript; CH, XX, PL, YC, QG, and LY were responsible for study conception. All authors provided critical review of the manuscript and approved the final draft for publication. QG, YC, and LY contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg Microbes Infect* 2018 Sep 12;7(1):154 [FREE Full text] [doi: [10.1038/s41426-018-0155-5](https://doi.org/10.1038/s41426-018-0155-5)] [Medline: [30209269](https://pubmed.ncbi.nlm.nih.gov/30209269/)]
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)] [Medline: [31986264](https://pubmed.ncbi.nlm.nih.gov/31986264/)]
3. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: [10.1056/NEJMoa2002032](https://doi.org/10.1056/NEJMoa2002032)] [Medline: [32109013](https://pubmed.ncbi.nlm.nih.gov/32109013/)]
4. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating Research Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: [10.1056/NEJMoa2001017](https://doi.org/10.1056/NEJMoa2001017)] [Medline: [31978945](https://pubmed.ncbi.nlm.nih.gov/31978945/)]
5. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: [10.1056/NEJMoa2001316](https://doi.org/10.1056/NEJMoa2001316)] [Medline: [31995857](https://pubmed.ncbi.nlm.nih.gov/31995857/)]
6. Huang R, Xia J, Chen Y, Shan C, Wu C. A family cluster of SARS-CoV-2 infection involving 11 patients in Nanjing, China. *The Lancet Infectious Diseases* 2020 May 28;20(5):534-535. [doi: [10.1016/S1473-3099\(20\)30147-X](https://doi.org/10.1016/S1473-3099(20)30147-X)] [Medline: [32119823](https://pubmed.ncbi.nlm.nih.gov/32119823/)]
7. Jakovljevic M, Bjedov S, Jaksic N, Jakovljevic I. COVID-19 Pandemia and Public and Global Mental Health from the Perspective of Global Health Secur. *Psychiatr Danub* 2020;32(1):6-14 [FREE Full text] [doi: [10.24869/psyd.2020.6](https://doi.org/10.24869/psyd.2020.6)] [Medline: [32303023](https://pubmed.ncbi.nlm.nih.gov/32303023/)]
8. Wang Z, Yu G, Tian X, Tang J, Yan X. A Study of Users with Suicidal Ideation on Sina Weibo. *Telemed J E Health* 2018 Sep;24(9):702-709. [doi: [10.1089/tmj.2017.0189](https://doi.org/10.1089/tmj.2017.0189)] [Medline: [29293069](https://pubmed.ncbi.nlm.nih.gov/29293069/)]
9. Househ M. Communicating Ebola through social media and electronic news media outlets: A cross-sectional study. *Health Informatics J* 2016 Sep 26;22(3):470-478. [doi: [10.1177/1460458214568037](https://doi.org/10.1177/1460458214568037)] [Medline: [25656678](https://pubmed.ncbi.nlm.nih.gov/25656678/)]
10. Gui X, Wang Y, Kou Y, Reynolds TL, Chen Y, Mei Q, et al. Understanding the Patterns of Health Information Dissemination on Social Media during the Zika Outbreak. *AMIA Annu Symp Proc* 2017;2017:820-829 [FREE Full text] [Medline: [29854148](https://pubmed.ncbi.nlm.nih.gov/29854148/)]
11. Zhao Y, Cheng S, Yu X, Xu H. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. *J Med Internet Res* 2020 May 04;22(5):e18825 [FREE Full text] [doi: [10.2196/18825](https://doi.org/10.2196/18825)] [Medline: [32314976](https://pubmed.ncbi.nlm.nih.gov/32314976/)]

12. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. *J Med Internet Res* 2014;16(1):e20 [FREE Full text] [doi: [10.2196/jmir.2911](https://doi.org/10.2196/jmir.2911)] [Medline: [24440770](https://pubmed.ncbi.nlm.nih.gov/24440770/)]
13. Chen Y, Zhang Y, Xu Z, Wang X, Lu J, Hu W. Avian Influenza A (H7N9) and related Internet search query data in China. *Sci Rep* 2019 Jul 18;9(1):10434 [FREE Full text] [doi: [10.1038/s41598-019-46898-y](https://doi.org/10.1038/s41598-019-46898-y)] [Medline: [31320681](https://pubmed.ncbi.nlm.nih.gov/31320681/)]
14. Xie T, Yang Z, Yang S, Wu N, Li L. Correlation between reported human infection with avian influenza A H7N9 virus and cyber user awareness: what can we learn from digital epidemiology? *Int J Infect Dis* 2014 May;22:1-3 [FREE Full text] [doi: [10.1016/j.ijid.2013.11.013](https://doi.org/10.1016/j.ijid.2013.11.013)] [Medline: [24576842](https://pubmed.ncbi.nlm.nih.gov/24576842/)]
15. Fung IC, Fu K, Chan C, Chan BSB, Cheung C, Abraham T, et al. Social Media's Initial Reaction to Information and Misinformation on Ebola, August 2014: Facts and Rumors. *Public Health Rep* 2016;131(3):461-473 [FREE Full text] [doi: [10.1177/003335491613100312](https://doi.org/10.1177/003335491613100312)] [Medline: [27252566](https://pubmed.ncbi.nlm.nih.gov/27252566/)]
16. Fung IC, Duke CH, Finch KC, Snook KR, Tseng P, Hernandez AC, et al. Ebola virus disease and social media: A systematic review. *Am J Infect Control* 2016 Dec 01;44(12):1660-1671. [doi: [10.1016/j.ajic.2016.05.011](https://doi.org/10.1016/j.ajic.2016.05.011)] [Medline: [27425009](https://pubmed.ncbi.nlm.nih.gov/27425009/)]
17. Wong R, Harris JK. Geospatial Distribution of Local Health Department Tweets and Online Searches About Ebola During the 2014 Ebola Outbreak. *Disaster Med Public Health Prep* 2018 Jun 24;12(3):287-290. [doi: [10.1017/dmp.2017.69](https://doi.org/10.1017/dmp.2017.69)] [Medline: [28835312](https://pubmed.ncbi.nlm.nih.gov/28835312/)]
18. Feng S, Hossain L, Crawford JW, Bossomaier T. Quantifying Network Dynamics and Information Flow Across Chinese Social Media During the African Ebola Outbreak. *Disaster Med Public Health Prep* 2018 Feb 01;12(1):26-37. [doi: [10.1017/dmp.2017.29](https://doi.org/10.1017/dmp.2017.29)] [Medline: [28760166](https://pubmed.ncbi.nlm.nih.gov/28760166/)]
19. Seltzer E, Jean N, Kramer-Golinkoff E, Asch D, Merchant R. The content of social media's shared images about Ebola: a retrospective study. *Public Health* 2015 Sep;129(9):1273-1277. [doi: [10.1016/j.puhe.2015.07.025](https://doi.org/10.1016/j.puhe.2015.07.025)] [Medline: [26285825](https://pubmed.ncbi.nlm.nih.gov/26285825/)]
20. Seltzer EK, Horst-Martz E, Lu M, Merchant RM. Public sentiment and discourse about Zika virus on Instagram. *Public Health* 2017 Sep;150:170-175. [doi: [10.1016/j.puhe.2017.07.015](https://doi.org/10.1016/j.puhe.2017.07.015)] [Medline: [28806618](https://pubmed.ncbi.nlm.nih.gov/28806618/)]
21. Fu K, Liang H, Saroha N, Tse ZTH, Ip P, Fung IC. How people react to Zika virus outbreaks on Twitter? A computational content analysis. *Am J Infect Control* 2016 Dec 01;44(12):1700-1702. [doi: [10.1016/j.ajic.2016.04.253](https://doi.org/10.1016/j.ajic.2016.04.253)] [Medline: [27566874](https://pubmed.ncbi.nlm.nih.gov/27566874/)]
22. Fung IC, Fu K, Ying Y, Schaible B, Hao Y, Chan C, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. *Infect Dis Poverty* 2013;2(1):31 [FREE Full text] [doi: [10.1186/2049-9957-2-31](https://doi.org/10.1186/2049-9957-2-31)] [Medline: [24359669](https://pubmed.ncbi.nlm.nih.gov/24359669/)]
23. Guo P, Zhang Q, Chen Y, Xiao J, He J, Zhang Y, et al. An ensemble forecast model of dengue in Guangzhou, China using climate and social media surveillance data. *Sci Total Environ* 2019 Jan 10;647:752-762 [FREE Full text] [doi: [10.1016/j.scitotenv.2018.08.044](https://doi.org/10.1016/j.scitotenv.2018.08.044)] [Medline: [30092532](https://pubmed.ncbi.nlm.nih.gov/30092532/)]
24. Xu X. Mining the characteristics of COVID-19 patients from social media in China (Raw data).xlsx (Dataset). Figshare 2020:e [FREE Full text]
25. Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. *Zhonghua Liu Xing Bing Xue Za Zhi* 2020 Feb 17;41(2):145-151. [doi: [10.3760/cma.j.issn.0254-6450.2020.02.003](https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.003)] [Medline: [32064853](https://pubmed.ncbi.nlm.nih.gov/32064853/)]
26. Chan JF, Yuan S, Kok K, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* 2020 Feb 15;395(10223):514-523 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9)] [Medline: [31986261](https://pubmed.ncbi.nlm.nih.gov/31986261/)]
27. Phan LT, Nguyen TV, Luong QC, Nguyen TV, Nguyen HT, Le HQ, et al. Importation and Human-to-Human Transmission of a Novel Coronavirus in Vietnam. *N Engl J Med* 2020 Feb 27;382(9):872-874 [FREE Full text] [doi: [10.1056/NEJMc2001272](https://doi.org/10.1056/NEJMc2001272)] [Medline: [31991079](https://pubmed.ncbi.nlm.nih.gov/31991079/)]
28. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *N Engl J Med* 2020 Mar 05;382(10):970-971. [doi: [10.1056/NEJMc2001468](https://doi.org/10.1056/NEJMc2001468)] [Medline: [32003551](https://pubmed.ncbi.nlm.nih.gov/32003551/)]
29. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020 Feb 29;395(10225):689-697. [doi: [10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)] [Medline: [32014114](https://pubmed.ncbi.nlm.nih.gov/32014114/)]
30. Guan Q, Liu M, Zhuang YJ, Yuan Y, Wang SS, Li J, et al. [Epidemiological investigation of a family clustering of COVID-19]. *Zhonghua Liu Xing Bing Xue Za Zhi* 2020 Mar 09;41:629-633. [doi: [10.3760/cma.j.cn112338-20200223-00152](https://doi.org/10.3760/cma.j.cn112338-20200223-00152)] [Medline: [32149484](https://pubmed.ncbi.nlm.nih.gov/32149484/)]
31. Qiu YY, Wang SQ, Wang XL, Lu WX, Qiao D, Li JB, et al. [Epidemiological analysis on a family cluster of COVID-19]. *Zhonghua Liu Xing Bing Xue Za Zhi* 2020 Mar 05;41(4):506-509. [doi: [10.3760/cma.j.cn112338-20200221-00147](https://doi.org/10.3760/cma.j.cn112338-20200221-00147)] [Medline: [32133831](https://pubmed.ncbi.nlm.nih.gov/32133831/)]
32. Yang HY, Xu J, Li Y, Liang X, Jin YF, Chen SY, et al. [The preliminary analysis on the characteristics of the cluster for the Corona Virus Disease]. *Zhonghua Liu Xing Bing Xue Za Zhi* 2020 Mar 08;41:623-628. [doi: [10.3760/cma.j.cn112338-20200223-00153](https://doi.org/10.3760/cma.j.cn112338-20200223-00153)] [Medline: [32145716](https://pubmed.ncbi.nlm.nih.gov/32145716/)]
33. Tian S, Hu N, Lou J, Chen K, Kang X, Xiang Z, et al. Characteristics of COVID-19 infection in Beijing. *J Infect* 2020 Apr;80(4):401-406. [doi: [10.1016/j.jinf.2020.02.018](https://doi.org/10.1016/j.jinf.2020.02.018)] [Medline: [32112886](https://pubmed.ncbi.nlm.nih.gov/32112886/)]

34. Yusong L, Duanliang P, Jia Y, Dunyan C, Hongbing J, Siyuan Y, et al. Laboratory diagnostics within a modular hospital at the time of Coronavirus disease 2019 (COVID-19) in Wuhan. *Clin Chem Lab Med* 2020 Apr 02:e [FREE Full text] [doi: [10.1515/cclm-2020-0332](https://doi.org/10.1515/cclm-2020-0332)] [Medline: [32242830](https://pubmed.ncbi.nlm.nih.gov/32242830/)]
35. Liu K, Fang Y, Deng Y, Liu W, Wang M, Ma J, et al. Clinical characteristics of novel coronavirus cases in tertiary hospitals in Hubei Province. *Chin Med J (Engl)* 2020 May 05;133(9):1025-1031 [FREE Full text] [doi: [10.1097/CM9.0000000000000744](https://doi.org/10.1097/CM9.0000000000000744)] [Medline: [32044814](https://pubmed.ncbi.nlm.nih.gov/32044814/)]
36. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. *Lancet Glob Health* 2020 Apr;8(4):e480 [FREE Full text] [doi: [10.1016/S2214-109X\(20\)30068-1](https://doi.org/10.1016/S2214-109X(20)30068-1)] [Medline: [32109372](https://pubmed.ncbi.nlm.nih.gov/32109372/)]
37. Yu M, He S, Wu D, Zhu H, Webster C. Examining the Multi-Scalar Unevenness of High-Quality Healthcare Resources Distribution in China. *Int J Environ Res Public Health* 2019 Aug 07;16(16):e [FREE Full text] [doi: [10.3390/ijerph16162813](https://doi.org/10.3390/ijerph16162813)] [Medline: [31394765](https://pubmed.ncbi.nlm.nih.gov/31394765/)]

Abbreviations

API: application programming interface
COVID-19: coronavirus disease
CT: computed tomography
MERS-CoV: Middle East respiratory syndrome
RT-PCR: reverse transcription-polymerase chain reaction
SARS-CoV: severe acute respiratory syndrome coronavirus
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 05.04.20; peer-reviewed by A Benis, J Mueller, D Surian; comments to author 20.04.20; revised version received 03.05.20; accepted 12.05.20; published 17.05.20

Please cite as:

Huang C, Xu X, Cai Y, Ge Q, Zeng G, Li X, Zhang W, Ji C, Yang L
Mining the Characteristics of COVID-19 Patients in China: Analysis of Social Media Posts
J Med Internet Res 2020;22(5):e19087
URL: <http://www.jmir.org/2020/5/e19087/>
doi: [10.2196/19087](https://doi.org/10.2196/19087)
PMID: [32401210](https://pubmed.ncbi.nlm.nih.gov/32401210/)

©Chunmei Huang, Xinjie Xu, Yuyang Cai, Qinmin Ge, Guangwang Zeng, Xiaopan Li, Weide Zhang, Chen Ji, Ling Yang. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 17.05.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.