DATA NOTE

A spatial-epidemiological dataset of subjects infected by SARS-CoV-2 during the first wave of the pandemic in Mashhad, second-most populous city in Iran

Shahab MohammadEbrahimi[1,2] , Alireza Mohammadi[3] , Robert Bergquist[4] , Mahsan Akbarian[5] ,
Mahnaz Arian[6] , Elahe Pishgar[7] and Behzad Kiani[1*]

Abstract

Objective: In March 2020, Iran tackled the first national wave of COVID-19 that was particularly felt in Mashhad, Iran's second-most populous city. Accordingly, we performed a spatio-temporal study in this city to investigate the epidemiological aspects of the disease in an urban area and now wish to release a comprehensive dataset resulting from this study.

Data description: These data include two data files and a help file. Data file 1: "COVID-19_Patients_Data" contains the patient sex and age + time from symptoms onset to hospital admission; hospitalization time; co-morbidities; manifest symptoms; exposure up to 14 days before admission; disease severity; diagnosis (with or without RT-PCR assay); and outcome (recovery vs. death). The data covers 4000 COVID-19 patients diagnosed between 14 Feb 2020 and 11 May 2020 in Khorasan-Razavi Province. Data file 2: "COVID-19_Spatiotemporal_Data" is a digital map of census tract divisions of Mashhad, the capital of the province, and their population by gender along with the number of COVID-19 cases and deaths including the calculated rates per 100,000 persons. This dataset can be a valuable resource for epidemiologists and health policymakers to identify potential risk factors, control and prevent pandemics, and optimally allocate health resources.

Keywords: COVID-19, SARS-CoV-2, Spatial epidemiology, Geographical information systems, Iran

Objective

A novel respiratory infection named coronavirus disease-2019 (COVID-19) originated in November 2019 and produced a major outbreak globally [1]. The pathogen was diagnosed as one of the coronavirus family and is currently known as the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) [2]. Due to the highly contagious nature of this virus, it spread in a short time to almost all countries and eventually being declared a pandemic by the World Health Organization (WHO) on 11 March 2020 [3, 4]. As of 19 July 2021, more than 191 million confirmed cases and approximately 2.15% mortality were identified worldwide. Iran's share at that time was close to 3,523,000 cases with more than 87,000 deaths [5]. Analysing and interpreting the spatio-temporal transmission patterns of the virus are indispensable in order to generate the best-tailored strategies [6]. The strength of Geographic Information Systems (GIS) applications lies in their capability of mapping geographical disease distributions thereby visualizing trends of their spread, which can be utilized for modelling spatial

*Correspondence: Kiani.Behzad@gmail.com
[1] Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
Full list of author information is available at the end of the article

MohammadEbrahimi *et al. BMC Res Notes*   (2021) 14:292

Page 2 of 4

aspects of disease occurrence in relation to the ambient environments [7–9].

The spatial patterns of COVID-19 incidence were investigated by census level with addressing the epidemiological features, during the first wave of the pandemic in Mashhad, Iran [10]. In that study, hotspots and high/low-risk areas were detected by using the Getis-Ord Gi* and Local Moran's *I* statistic [11, 12]. A univariate regression model was developed to quantify the association of COVID-19 mortality with common risk factors [13] including age [14, 15], sex [16, 17], co-morbidities [18, 19], hospitalization length [20, 21] and transfer to an Intensive Care Unit (ICU) [16, 20]. Here, a comprehensive spatial-epidemiological dataset linked to other urban data at the census level is offered for further investigation to identify transmission trends and clustering patterns of the COVID-19 incidence in the densely populated city.

Data description
In the current study, the COVID-19 data were collected in the referral hospitals and health-care centres under the supervision of Mashhad University of Medical Sciences (MUMS). These data were related to all people infected by SARS-CoV-2 in Khorasan-Razavi Province (KRP) and they covered three months with the start coinciding with the beginning of the COVID-19 outbreak in KRP (i.e. from 14 Feb to 11 May 2020). The data included 4000 people referred to the health-care centres and hospitals due to COVID-19 infection with cases either confirmed clinically (n = 2675) or by laboratory tests (n = 1325) using the Reverse Transcription Polymerase Chain Reaction (RT-PCR) assay. Demographic data of all neighbourhoods came from census blocks statistics of 2018–2019 [22].

Addresses of patients in the city of Mashhad with confirmed COVID-19 by RT-PCR test (n = 727) were geocoded manually using the Google MyMaps software (http://www.google.com/mymaps). Mashhad is the capital city of KRP and is the second-most populous city in Iran which has 1,301 census tracts. Five age groups, including 0–14, 15–24, 25–44, 44–64, and > 65 years old,

were used to calculate the age- and sex-adjusted incidence and death rates of COVID-19 in each census tract. In order to avoid the identity of participating COVID-19 cases, the point-density data, expressing patients' physical addresses, were aggregated into each census tract and stratified into ten-day intervals during the study period.

Table 1 shows details of the two data files, a help file and access links. Data file 1 covers the demographic and clinical information of 4000 COVID-19 cases in Excel file format (*.xlsx). Each data row includes patient sex and age + time from symptoms onset to hospital admission; hospitalization time; co-morbidities; manifest symptoms; exposure up to 14 days before admission; disease severity; diagnosis (with or without RT-PCR assay); and outcome (recovery vs. death). Data file 2 covers the spatio-temporal data of those infected by SARS-CoV-2 including polygon shape-files (*.shp) representing the location of all COVID-19 cases aggregated at the census tract level. This data file covers the following information: an identification code of each census; the total population as well as the population by gender in each census; the number of cases and deaths by COVID-19 along with the calculated rates per 100,000 persons separately for each census; and the number of cases and deaths due to COVID-19 based on ten-day intervals. File 3 is a help file for both data files which represents the name and description of each field. Since Mashhad is located at 36° N, 59° E, the projection system of WGS_1984_UTM_Zone_40N was used as the Projected Coordinate System (PCS) for all GIS layers. Due to the need to find better preventive measures and improve hospital care in response to the irreversible psychological and physical effects of COVID-19 [23–25], the data in the current study can be used as a basis for spatial modelling of the disease providing reliable knowledge to other researchers in various fields such as health geography, urban policymaking and healthcare research.

Limitations
Only 33% of all included cases were confirmed by RT-PCR testing, the rest of the cases were clinically approved. Separate data about hypertension and

Table 1 Overview of data sets

Label	Name of data file	File types (Extensions)	Data repository and identifier (DOI or accession number)
Data file 1	COVID-19_Patients_Data	Excel file (*.xlsx)	*Harvard Dataverse* https://doi.org/10.7910/DVN/SSJ7WT [26]
Data file 2	COVID-19_Spatiotemporal_Data	Shape file (*.shp)	*Harvard Dataverse* https://doi.org/10.7910/DVN/SSJ7WT [26]
Data file 3	Help file	Excel file (*.xlsx)	*Harvard Dataverse* https://doi.org/10.7910/DVN/SSJ7WT [26]

MohammadEbrahimi *et al. BMC Res Notes*   (2021) 14:292

Page 3 of 4

recovered cases with remaining disease complications (long COVID) were not collected. Due to the short period covered (3 months), the capabilities of spatial analysis cannot tell us more now. For the future, it is suggested to study the spatial and temporal dynamics of the disease over a longer period in order to provide more operational solutions.

Abbreviations
COVID-19: Corona Virus Disease-2019; SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus-2; GIS: Geographic Information System; KDE: Kernel Density Estimation; LOS: Length of Stay; ICU: Intensive Care Unit; MUMS: Mashhad University of Medical Sciences; KRP: Khorasan-Razavi Province; RT-PCR: Reverse Transcription Polymerase Chain Reaction; KML: Keyhole Markup Language; PCS: Projected Coordinate System.

Authors' contributions
SM and AM contributed to data cleaning and preparing for analyses. S.M and B.K drafted and revised the manuscript. MAK, MAR, and EP contributed to data gathering and geocoding. RB critically reviewed the manuscript. BK is the principal investigator and research leader. All authors read and approved the final manuscript.

Availability of data and materials
The data described in this Data note can be freely and openly accessed on the Harvard Dataverse under (https://doi.org/10.7910/DVN/SSJ7WT) [26]. Please see Table 1 and the reference list for details and links to the data.

Declarations

Ethics approval and consent to participate
The study was approved by the ethical committee of Mashhad University of Medical Sciences with the reference number IR.MUMS.REC.1399.215. As well as, informed consent was not required to be obtained due to the nature of the study.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. [2]Student Research Committee, Mashhad University of Medical Sciences, Mashhad, Iran. [3]Department of Geography and Urban Planning, Faculty of Social Sciences, University of Mohaghegh Ardabili, Ardabil, Iran. [4]Ingerod, Brastad, Sweden (formerly with the UNICEF/UNDP/World Bank/WHO Special Program for Research and Training in Tropical Diseases, World Health Organization), Geneva, Switzerland. [5]Department of Physiology, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. [6]Department of Infectious Diseases, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. [7]Department of Human Geography, Faculty of Earth Science, Shahid Beheshti University, Tehran, Iran.

References
1. Bergquist R, Kiani B, Manda S. First year with COVID-19: Assessment and prospects. Geospatial Health. 2020. https://doi.org/10.4081/gh.2020.953.
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382:727–33.
3. Bergquist R, Stengaard A-S. COVID-19: End of the beginning? Geospat Health. 2020. https://doi.org/10.4081/gh.2020.897.
4. NeJhaddadgar N, Ziapour A, Zakkipour G, Abbas J, Abolfathi M, Shabani M. Effectiveness of telephone-based screening and triage during COVID-19 outbreak in the promoted primary healthcare system: a case study in Ardabil province. Iran J Public Health (Berl). 2020. https://doi.org/10.1007/s10389-020-01407-8.
5. COVID Live Update: 191,231,798 Cases and 4,105,865 Deaths from the Coronavirus Worldometer. https://www.worldometers.info/coronavirus/. Accessed July 19, 2021.
6. Zhang Y, Wang X, Li Y, Ma J. Spatiotemporal analysis of influenza in China, 2005–2018. Sci Rep. 2019;9:1–12.
7. Montazeri M, Hoseini B, Firouraghi N, Kiani F, Raouf-Mobini H, Biabangard A, et al. Spatio-temporal mapping of breast and prostate cancers in South Iran from 2014 to 2017. BMC Cancer. 2020;20:1170.
8. Azimi A, Bagheri N, Mostafavi SM, Furst MA, Hashtarkhani S, Amin FH, et al. Spatial-time analysis of cardiovascular emergency medical requests: enlightening policy and practice. BMC Public Health. 2021;21:7.
9. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci Total Environ. 2020;72:8.
10. Mohammad Ebrahimi S, Mohammadi A, Bergquist R, Dolatkhah F, Olia M, Tavakolian A, et al. Epidemiological characteristics and initial spatiotemporal visualisation of COVID-19 in a major city in the Middle East. BMC Public Health. 2021;21:1373.
11. Ord JK, Getis A. Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal. 1995;27:286–306.
12. Anselin L. Local indicators of spatial association—LISA. Geogr Anal. 1995;27:93–115.
13. Mollalo A, Rivera KM, Vahabi N. Spatial statistical analysis of pre-existing mortalities of 20 diseases with COVID-19 mortalities in the continental United States. Sust Cities Soc. 2021;67:102738.
14. Passamonti F, Cattaneo C, Arcaini L, Bruna R, Cavo M, Merli F, et al. Clinical characteristics and risk factors associated with COVID-19 severity in patients with haematological malignancies in Italy: a retrospective, multicentre, cohort study. Lancet Haematol. 2020. https://doi.org/10.1016/S2352-3026(20)30251-9.
15. Shuja KH, Shahidullah N, Aqeel M, Khan EA, Abbas J. Letter to highlight the effects of isolation on elderly during COVID-19 outbreak. Int J Geriatr Psychiatry. 2020. https://doi.org/10.1002/gps.5423.
16. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: prospective observational cohort study. BMJ. 2020;369:m1985.
17. Suleyman G, Fadel RA, Malette KM, Hammond C, Abdulla H, Entz A, et al. Clinical characteristics and morbidity associated with coronavirus disease 2019 in a series of patients in metropolitan detroit. JAMA Netw Open. 2020;3:e2012270.
18. Pourghasemi HRP. Spatial modelling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between 19 February to 14 June 2020). Int J Infect Dis. 2020;74:547.
19. Bakhshayesh S, Hoseini B, Bergquist R, Nabovati E, Gholoobi A, Mohammad-Ebrahimi S, et al. Cost-utility analysis of home-based cardiac rehabilitation as compared to usual post-discharge care: systematic review and meta-analysis of randomized controlled trials. Expert Rev Cardiovasc Ther. 2020;18:761–76.
20. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA. 2020;91:650.
21. Mohammadebrahimi S, Bayati S, Mardani M, Karim H. Factors Associated with patient length of stay, according to Sina Hospital's admission data—Mashhad. Front Health Inf. 2015;4:1–6.
22. Statistical Center of Iran. Official report of statistical survey of population in Mashhad city archived by the Statistical Center of Iran. 2018–2019.

MohammadEbrahimi *et al. BMC Res Notes*   (2021) 14:292

Page 4 of 4

https://www.amar.org.ir/english/Iran-Statistical-Yearbook/Statistical-Yearbook-2018-2019. Accessed 19 Jul 2021.

23. Abbas J. The impact of coronavirus (SARS-CoV2) epidemic on individuals mental health: the protective measures of Pakistan in managing and sustaining transmissible disease. Psychiatr Danub. 2020;32:472–7.

24. Maqsood A, Abbas J, Rehman G, Mubeen R. The paradigm shift for educational system continuance in the advent of COVID-19 pandemic: Mental health challenges and reflections. Curr Res Behav Sci. 2021;2:100011.

25. Yoosefi Lebni J, Abbas J, Moradi F, Salahshoor MR, Chaboksavar F, Irandoost SF, et al. How the COVID-19 pandemic effected economic, social,

political, and cultural factors: a lesson from Iran. Int J Soc Psychiatry. 2020;207:6402.

26. Kiani B. A dataset of subjects who infected by SARS-CoV-2 in the first wave of the pandemic in the city of Mashhad. 2021. Iran. https://doi.org/10.7910/DVN/SSJ7WT

Publisher's Note