Sedghian *et al. BMC Research Notes* (2025) 18:72
https://doi.org/10.1186/s13104-025-07139-w

BMC Research Notes

DATA NOTE

Nationwide spatial epidemiological dataset of over 100,000 influenza-like illness notifications in Iran by county (2015–2019)

Atieh Sedghian[1,2*], Shahab MohammadEbrahimi[1,2], Benn Sartorius[5,4,3] and Behzad Kiani[3]

Abstract

Objectives This data note documents influenza-like illness (ILI) notifications in Iran by county from 2015 to 2019 as a pre-COVID-19 dataset, providing individual and spatial data for further comprehensive spatiotemporal analysis. Due to the high contagion rate of ILI and global health impact, precise geographic mapping serves as a critical tool for public health officials and researchers to monitor, mitigate, and predict epidemics. By utilizing advanced spatial-temporal epidemiological analysis to study disease occurrence patterns, this geodatabase can enable a better understanding and more effective management of ILIs in the future.

Data description This is the most comprehensive dataset of all individual ILI notifications in Iran between 2015 and 2019 by date of notification and county (398 counties). The database includes two data files, a help file, and a data usage agreement: Data File 1 is an Excel (.xlsx) file detailing demographic and clinical information from 109,919 ILI notifications nationwide, covering county and date of notification, patient demographics, admission details, sample types, differential diagnosis, medical history, mortality details, test results, and symptoms. Data File 2 contains spatiotemporal information in polygon shapefiles (.shp), mapping ILI notification locations by county with data on case counts for each year, total population, gender distribution, and geographic coordinates.

Keywords Influenza-like illness, Space-time, Spatiotemporal, Public health, Iran, Big data

Objective

Influenza-like illness (ILI) is a highly contagious acute respiratory syndrome characterized by symptoms such as fever (≥ 38 °C) and cough within the past ten days [1, 2]. The global burden of ILI is significant, with frequent outbreaks leading to severe health complications, particularly among vulnerable populations like the elderly and those with chronic conditions [3]. During pandemics, the rapid transmission and antigenic variability of ILI pathogens exacerbate the situation, increasing morbidity and mortality rates [4]. Despite its widespread impact, the mechanisms driving the geographical spread and seasonality of ILI remain poorly understood, necessitating

*Correspondence:
Atieh Sedghian
Atieh.sedghian1@gmail.com
[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]UQ Centre for Clinical Research (UQCCR), Faculty of Health, Medicine, and Behavioural Sciences, University of Queensland, Brisbane, Australia
[4]Department of Health Metric Sciences, Faculty of Medicine, University of Washington, Seattle, USA
[5]Centre for Tropical Medicine and Global Health, Faculty of Medicine, University of Oxford, Oxford, UK

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Individual data of ILI cases	MS Excel file (*.xlsx)	Harvard Dataverse https://doi.org/10.7910/DVN/OGK09I [9]
Data file 2	Spatiotemporal data of ILI cases	Polygonal features shapefile format (*.shp)	Harvard Dataverse https://doi.org/10.7910/DVN/OGK09I [9]
Data file 3	Help file	MS Excel file (*.xlsx)	Harvard Dataverse https://doi.org/10.7910/DVN/OGK09I [9]
Data file 4	Data usage agreement	Portable Document Format (*.pdf)	Harvard Dataverse https://doi.org/10.7910/DVN/OGK09I [9]

more precise surveillance and better data to inform public health strategies [5].

Effective management of ILI requires comprehensive datasets that include spatiotemporal information to identify high-risk areas and guide targeted interventions. Such datasets are crucial for precision public health approaches, enabling health officials and researchers to allocate resources efficiently and implement control measures where they are most needed [6]. This study focuses on compiling a detailed geodatabase of ILI notifications in Iran at the county level from 2015 to 2019. The dataset is designed to support spatiotemporal analysis by providing a robust foundation of spatially indexed data with the capability to add additional socio-economic or environmental variables in the future.

While this dataset does not integrate socio-economic or environmental data, it offers a critical resource for initial spatiotemporal analysis of ILI patterns and contributes to the development of spatial tools that enhance our understanding of ILI dynamics. These insights can lay the groundwork for future studies exploring the impact of socio-economic and environmental factors on ILI distribution. The ultimate goal is to equip decision-makers with the tools and knowledge needed to predict future ILI burdens, optimize resource allocation, and implement targeted interventions, particularly in the most vulnerable areas. This comprehensive approach will significantly advance spatial epidemiology and improve our understanding of ILI transmission dynamics.

Data description

In this study, data for the entire country of Iran, located in West Asia, were collected from three different sources. Data on ILI notifications were obtained from Iran's influenza surveillance system (IISS) for four consecutive years, from March 21, 2015, to March 20, 2019. This dataset included 109,919 individuals who visited health centers and hospitals due to ILI symptoms. Population data were gathered from the most recent census conducted in Iran in 2016, sourced from the Population and Housing Census of Iran. These data include the total population, the total male population, and the total female population for each county separately. We geocoded the locations of the notifications to county. To calculate the incidence and case fatality rate (CFR) of ILI based on age and gender in each county, five age groups were used, namely: 0–5, 6–49, 50–64, 65–74, and over 75 years.

This dataset includes four files, each providing essential information for ILI notifications (Table 1). The first file, Data File 1, is an Excel (*.xlsx) file that contains demographic and clinical information for 109,919 ILI cases diagnosed across the country. Each row in this file provides detailed data, including the gender and age group of the patient, the county (one of 398 counties) where the patient was admitted, and the date of admission. Additional fields include the year and month order of admission, the type of sample received, differential diagnosis, medical history, comorbidities, hospital treatment, patient outcome (whether the patient died or not), date of death (if applicable), test result, and symptoms.

Data File 2 offers spatiotemporal data of ILI notifications and includes polygon shapefiles (*.shp) that indicate the locations of all ILI cases collected at the county level. This file provides data organized by county boundaries and includes the identification code and name of each county. It also details the number of ILI notifications in each county by year, the total number of cases, total population, population by gender, and geographic coordinates (latitude and longitude). This geospatial component is crucial for understanding the distribution and spread of ILI notifications within different regions.

The third file is an Excel help file designed to assist users in navigating and utilizing the data provided in the other two files. This additional documentation ensures that researchers can accurately interpret the data and apply it effectively in their analyses. Data File 4 contains the data usage agreement that researchers must accept in order to gain access to Data File 1.

When utilizing our dataset for further research, it is important to consider region-specific factors that may influence the spatiotemporal patterns of ILI in Iran. The country's diverse climatic zones, ranging from humid coastal regions to arid and semi-arid areas, can affect seasonal disease trends. Additionally, disparities in healthcare access, population density, and socio-economic conditions may impact notification rates and reporting accuracy. Population mobility, particularly in provinces with major metropolitan areas, pilgrimage sites, or border crossings, could also contribute to fluctuations in disease incidence. While our dataset does not directly include mobility data, researchers should account for its potential influence when interpreting regional trends.

The dataset presented in this study could be invaluable for researchers in various fields such as spatial epidemiology, public health, and health services research [7]. By leveraging this comprehensive dataset, researchers can uncover critical insights into the spread and impact of ILI and help guide strategies to reduce future burden [8]. The resource and associated tools that will be developed through use of this data, can subsequently inform health policymakers and urban planners in developing appropriate strategies to mitigate the risk of ILI across communities. Table 1 provides an overview of the dataset, which has been uploaded to the Harvard Dataverse [9].

Limitations

We used the 2016 census data as the reference for population data in our data note, which may affect future analyses. However, it should be noted that this was the most up-to-date information available for our study, and given that the census occurred approximately in the middle of period related to this data, we believe that linked census indicators will fairly be representative for this period. However, the lack of annual population estimates prevented us from accounting for potential demographic shifts over time. It is also worth mentioning that this dataset considers all ILI notifications, which differ from ILI patients since one person could appear in the database more than once. This distinction is crucial for accurately interpreting the data and ensuring that analyses account for potential duplicates.

Abbreviations

ILI Influenza-like illness

Acknowledgements

The authors would like to acknowledge the assistance of the Ministry of Health and Medical Education's Center for Communicable Disease Control (CDC) and all the clinicians involved in reporting infectious cases of influenza; without whose support this research would not have been possible. The authors also acknowledge Mashhad University of Medical Sciences (MUMS) for funding this study.

Author contributions

Conceptualization: B.K, and B.S.; Geocoding and cleaning: A.S, Sh.M.; Writing: A.S.; Review and editing: B.K., B.S.; Supervision: B.K.; Funding acquisition: B.K.; All authors read and approved the final draft.

Funding

The research leading to this dataset received funding from Mashhad University of Medical Sciences under Grant Agreement No 4001855.

Data availability

The data described in this data note can be accessed on the Harvard Dataverse under (https://doi.org/10.7910/DVN/OGK09I).

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Mashhad University of Medical Sciences and approved by the Research Ethics Committees of School of Medicine, Mashhad University of Medical Sciences, Iran (#IRMUMS.MEDICAL.REC.1401.032) for studies involving humans.

Consent for publication

All experiments were performed according to relevant guidelines and Regulations. No identifiable individual data were reported and due to the retrospective nature of the study, no consent is needed from the patients.

Competing interests

The authors declare no competing interests.

Received: 5 August 2024 / Accepted: 4 February 2025
Published online: 17 February 2025

References

1. Shah SC, Rumoro DP, Hallock MM, Trenholme GM, Gibbs GS, Silva JC, et al. Clinical predictors for laboratory-confirmed influenza infections: exploring Case definitions for Influenza-Like Illness. Infect Control Hosp Epidemiol. 2015;36(3):241–8.
2. Spencer JA, Shutt DP, Moser SK, Clegg H, Wearing HJ, Mukundan H, et al. Distinguishing viruses responsible for influenza-like illness. J Theor Biol. 2022;545:111145.
3. Chawla D, Benitez A, Xu H, Whitehill V, Tadesse-Bell S, Shapiro A, et al. Predictors of seeking Care for Influenza-Like Illness in a Novel Digital Study. Open Forum Infect Dis. 2023;10(1):ofac675.
4. Peteranderl C, Herold S, Schmoldt C. Human influenza virus infections. Semin Respir Crit Care Med. 2016;37(4):487–500.
5. Li H, Ge M, Wang C. Spatio-temporal evolution patterns of influenza incidence and its nonlinear spatial correlation with environmental pollutants in China. BMC Public Health. 2023;23(1):1685.
6. Pishgar E, Mohammadi A, Bagheri N, Kiani B. A spatio-temporal geodatabase of mortalities due to respiratory tract diseases in Tehran, Iran between 2008 and 2018: a data note. BMC Res Notes. 2020;13(1):469.
7. Kiani B, Lau C, Bergquist R. From Snow's map of cholera transmission to dynamic catchment boundary delineation: current front lines in spatial analysis. Geospat Health. 2023;18(2).
8. Ebrahimi SM, Kiani B, Rahmatinejad Z, Baral S, Hashtarkhani S, Dehghan-Tezerjani M et al. Geospatial epidemiology of hospitalized patients with a positive influenza assay: a nationwide study in Iran, 2016–2018. PLoS ONE. 2022;17(12 December).
9. Sedghian A. Influenza Like Illness in Iran (2015–2019). Harvard Dataverse. 2024. https://doi.org/10.7910/DVN/OGK09I

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.