

DATA NOTE

Open Access



A Geospatial database of gastric cancer patients and associated potential risk factors including lifestyle and air pollution

Fatemeh Hashemi Amin¹, Mahtab Ghaemi², Sayyed Mostafa Mostafavi¹, Ladan Goshayeshi³, Khadijeh Rezaei⁴, Mehdi Vahed⁴ and Behzad Kiani^{5*} 

Abstract

Objectives: Gastric cancer (GC) is a multifactorial disease and the fifth most frequent diagnosed cancer worldwide. It accounts for one third of cancer-related mortalities. Geospatial analysis using geographical information systems (GIS) can provide an efficient solution to identify spatial disparities associated with GC. As such, GIS enables policymakers to control cancer in a better way and identify the regions where interventions are needed. This study aims to publish a comprehensive dataset, which was applied to conduct a spatial analysis of GC patients in the city of Mashhad, Iran.

Data description: We provide a personal geodatabase, a Microsoft Access database that can store, query, and manage both spatial and non-spatial data, which contains four feature classes. "Male_Stomach_Cancer_Patients" and "Female_Stomach_Cancer_Patients" are point feature classes, which show the age and geographical location of 1156 GC cancer patients diagnosed between 2014 and 2017. "Air_Pollution_Mashhad" is another point feature class that reveals the amount of six air pollutants, which was taken from Mashhad Environmental Pollutants Monitoring Center between 2017 and 2018. Finally, "Stomach_Cancer_and_Risk_Factors" is a polygon feature class of neighborhood division of Mashhad, consisting of contributor risk factors including dietary habits, smoking, alcohol use, body mass index and population by age groups for all 165 city neighborhoods.

Keywords: Gastric cancer, Geographical information systems, GIS, Geospatial analysis, Dietary habits, Air pollutants, Environmental factors, Stomach cancer

Objective

Gastric cancer (GC), also known as stomach cancer, is classified as the fifth most frequent diagnosed cancer among both genders and is the third leading cause of cancer mortality [1]. According to GLOBOCAN2018, more than 10.6% of different types of cancer cases in Iran are associated with GC. It also contributed to 16.1% of all cancer-related deaths and accounted for the most common cancer-related mortality [2]. The major GC

risk factors include alcohol drinking, physical inactivity, chronic infections, gender, age, medical history, smoking and unhealthy eating habits [3–6]. Furthermore, the association between environmental risk factors and GC is widely reported in the literature [7–12]. The presence of significant geographical disparities across the world is one of this tumor's epidemiological traits. This implication indicates that environmental exposures can play a key role in the uncertain carcinogenesis of GC [13].

We conducted a spatial analysis of GC incidence at the neighborhoods level in the city of Mashhad, Iran. Dietary habits, smoking, alcohol drinking, BMI, and air pollution were considered in the model. In this study, we aim to offer a comprehensive integrated geodatabase. This

*Correspondence: Kiani.Behzad@gmail.com

⁵ 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



geodatabase is a practical tool for further investigation in future spatial analysis of GC incidence.

Data description

Geospatial approaches and, in particular, GIS describe the spread and etiology of different types of disease. Moreover, they can provide useful strategies for disease control. An important topic of research is the association between geography and cancer incidence, where GIS applications play an important role [14–16]. GIS tools have the potential to accurately measure healthcare resources, track possible regional improvements in disease outcomes and also identify potential differences in cancer care [17, 18].

In this study, a personal geodatabase was created to store all the data files (feature classes). We obtained the data from four different databases. The data of cancer registry of Khorasan-Razavi Province between March 2014 and March 2017 was extracted to obtain individual GC cases. This dataset contains 1156 records of GC cases along with their age, gender and geographical locations in Mashhad, Iran. The address of GC patients was geocoded and this data is made available in “Male_Stomach_Cancer_Patients” and “Female_Stomach_Cancer_Patients” feature classes in the geodatabase (Data file 1). Regarding the patients’ privacy, the point data of patients were randomly moved within a 500 m. The Mashhad Municipal Council provided the neighborhood divisions and their population into different age groups. The age group interval was estimated at five-year intervals for males and females. This data is made available in the “Stomach_Cancer_and_Risk_Factors” feature class stored in the geodatabase (data file 1).

Data on potential risk factors including Body Mass Index (BMI), smoking, alcohol drink, total intake of red meat, processed meat, fruit, vegetable, salt and smoked food were derived from the MASHHAD Cohort study dataset, containing 6388 records [19]. We calculated the percentage of alcoholic and smoker individuals for each neighborhood. Fields regarding intake of vegetables and fruit were measured in terms of grams per day and consumption of red meat and processed meat were calculated in grams per week. The total amount of smoked food consumed per month was recorded in the smoked food field for every neighborhood. “Stomach_Cancer_and_Risk_Factors” reports this data.

Data regarding amount of six air pollutants was prepared by Mashhad Environmental Pollutants Monitoring Center, consisting Ozone (O_3), Particulate matter (PM10, PM2.5), Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), Sulfur Dioxide (SO_2) between March 2017–March 2018. Spatial interpolation method was

employed by ArcGIS 10.6 so as to estimate the amount of air pollutants for each neighborhood where no station was available to calculate the actual amount of particles. These data are available from the “Air_Pollution_Mashhad” feature class stored in the geodatabase (data file 1).

To prepare the geodatabase, the parcel layer of Mashhad was considered as the base layer. Then, the GC patients layer, Persian Cohort layer and heavy metal layer were linked by performing spatial joining. The final layer was a polygon layer with an attribute table containing demographic information, risk factors of Persian Cohort database, characteristics of cancer patients and data related to amount of air pollutants for the neighborhoods of Mashhad. In the Persian Cohort database, the amount of salt consumed by each person was specified by assigning 1 to low salt, 2 to medium salt and 3 to high salt. In the final attribute table, the sum of these numbers in each neighborhood was calculated.

We utilized a Geographical Weighted Regression (GWR) model to explore which risk factors were more related to GC incidence in each neighborhood. GWR can be applied for spatial non-stationary parameter recognition by local parameter estimation [20]. Our dataset can be applied in further research in order to assess the association between other potential risk factors and GC incidence. In addition, the provided dataset is useful for those who attempt to investigate the relation of these available risk factors and other kinds of cancers occurrence. As a result of performing these kinds of analyses, we can mention the impacts of them on implementing more efficient cancer prevention plans and developing some new strategies to reduce the huge burden of cancers. These strategies can be specified for each neighborhood in urban areas. For example, in one neighborhood educating people to improve their dietary habits can be essential but in another neighborhood, reducing air pollution can be the first priority.

Limitations

The data of life style factors aggregated into neighborhood levels were obtained through the PERSIAN Cohort study [19], which is an institutional cohort study. This means that we used a specific sample of the general population, employees of government centers, to estimate life style factors for each neighborhood. This might not be a representative sample of the total population. However, this is the best data available for determining life style factors at neighborhood level in the city of Mashhad, Iran.

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	Stomach_Cancer_Mashhad	MS Access file (.mdb)	Harvard Dataverse (https://doi.org/10.7910/DVN/SOQM4D) [21]

Abbreviations

GC: Gastric cancer; GIS: Geographical information systems.

Acknowledgements

We would like to thank Mashhad University of Medical Sciences for funding this study.

Authors' contributions

FHA and BK drafted the manuscript and provided the population data for sharing. BK was the principal investigator and project leader. LG and MG contributed on identifying GC risk factors. SMM critically revised the text. KR and MV were recorded and cleaned the cancer data. All authors read and approved the final manuscript.

Funding

The study received funding from Mashhad University of Medical Sciences (Fund Number: 980953).

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/SOQM4D>) [21]. Please see Table 1 for details and link to the data.

Declarations**Ethics approval and consent to participate**

This study has been assessed by the research council of Mashhad University of Medical Science (Reference number: IR.MUMS.MEDICAL.REC.1398.785). The study was ethically approved because no identifying data are reported. Regarding the patients' privacy, the point data of patients were randomly moved within a 500 m.

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 Department of Internal Medicine, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. ³ Department of Gastroenterology and Hepatology, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. ⁴ Department of Health, Mashhad University of Medical Sciences, Mashhad, Iran. ⁵ Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.

Received: 3 January 2021 Accepted: 27 February 2021

Published online: 09 March 2021

References

- Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *Lancet Global Health*. 2020;8(2):e191–203.
- Organization WH. *Global Health Observatory*. Geneva: World Health Organization; 2018. p. 2018.
- Alhdiri MA, Samat NA, Mohamed ZM. Mapping Libya's prostate cancer based on the SMR method: A geographical analysis. *Geografia–Malaysian J Soc Space*. 2017;12:9.

- Alhdiri MAS, Samat NA, Mohamed Z. Disease mapping for stomach cancer in libya based on Besag–York–Mollié (BYM) Model. *APJCP*. 2017;18(6):1479.
- Clinton SK, Giovannucci EL, Hursting SD. The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions. *J Nutr*. 2020;150(4):663–71.
- Cancer IAfRo. A review of human carcinogens: personal habits and indoor combustions. *World Health Organization*; 2012.
- Santos-Sánchez V, Córdoba-Doña JA, Viciana F, Escobar-Pujolar A, Pozzi L, Ramis R. Geographical variations in cancer mortality and social inequalities in southern Spain (Andalusia). *Plos one*. 2020;15(5):e0233397.
- Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol Prev Biomark*. 2014;23(5):700–13.
- Rawla P, Barsouk A. Epidemiology of gastric cancer: global trends, risk factors and prevention. *Przeglad Gastroenterol*. 2019;14(1):26.
- Fei X, Lou Z, Christakos G, Ren Z, Liu Q, Lv X. The association between heavy metal soil pollution and stomach cancer: a case study in Hangzhou City. *China Environ Geochem Health*. 2018;40(6):2481–90.
- Yuan W, Yang N, Li X. Advances in understanding how heavy metal pollution triggers gastric cancer. *BioMed Res Int*. 2016;2016:78.
- Yin J, Wu X, Li S, Li C, Guo Z. Impact of environmental factors on gastric cancer: A review of the scientific evidence, human prevention and adaptation. *J Environ Sci*. 2020;89:65–79.
- Aragónés N, Pérez-Gómez B, Pollán M, Ramis R, Vidal E, Lope V, et al. The striking geographical pattern of gastric cancer mortality in Spain: environmental hypotheses revisited. *BMC Cancer*. 2009;9(1):316.
- Goshayeshi L, Pourahmadi A, Ghayour-Mobarhan M, Hashtarkhani S, Karimian S, Dastjerdi RS, et al. Colorectal cancer risk factors in north-eastern Iran: A retrospective cross-sectional study based on geographical information systems, spatial autocorrelation and regression analysis. *Geospat Health*. 2019;14(2):219–28.
- Halimi L, Bagheri N, Hoseini B, Hashtarkhani S, Goshayeshi L, Kiani B. Spatial analysis of colorectal cancer incidence in Hamadan Province, Iran: a retrospective cross-sectional study. *Appl Spatial Anal Policy*. 2020;13(2):293–303.
- 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;20:1.
- Aneja S, Gross CP, Soulos PR, Yu JB. Geographical information systems: applications and limitations in oncology research. *Practice*. 2011;25:12.
- Pickle LW, Szczur M, Lewis DR, Stinchcomb DG. The crossroads of GIS and health information: a workshop on developing a research agenda to improve cancer control. *Int J Health Geogr*. 2006;5(1):51.
- Tohidinezhad F, Khorsand A, Zakavi SR, Rezvani R, Zarei-Ghanavati S, Abrishami M, et al. The burden and predisposing factors of non-communicable diseases in Mashhad University of Medical Sciences personnel: a prospective 15-year organizational cohort study protocol and baseline assessment. *BMC Public Health*. 2020;20(1):1–15.
- Goovaerts P. *Geostatistical analysis of health data: State-of-the-art and perspectives*. *Geoenv VI—Geostatistics for Environmental Applications*. Berlin: Springer; 2008. p. 3–22.
- Kiani, B. Stomach cancer and their related risk factors. *Harvard Dataverse* <https://doi.org/10.7910/DVN/SOQM4D> (2021)

Publisher's Note

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