BMC Research Notes

A spatial database of colorectal cancer patients and potential nutritional risk factors in an urban area in the Middle East

Neda Firouraghi[1], Nasser Bagheri[2], Fatemeh Kiani[1], Ladan Goshayeshi[3], Majid Ghayour-Mobarhan[4], Khalil Kimiafar[5], Saeid Eslami[1] and Behzad Kiani[1*]

Abstract

Objectives: Colorectal cancer (CRC) is the third most common cancer across the world that multiple risk factors together contribute to CRC development. There is a limited research report on impact of nutritional risk factors and spatial variation of CRC risk. Geographical information system (GIS) can help researchers and policy makers to link the CRC incidence data with environmental risk factor and further spatial analysis generates new knowledge on spatial variation of CRC risk and explore the potential clusters in the pattern of incidence. This spatial analysis enables policymakers to develop tailored interventions. This study aims to release the datasets, which we have used to conduct a spatial analysis of CRC patients in the city of Mashhad, Iran between 2016 and 2017.

Data description: These data include five data files. The file CRCcases_Mashhad contains the geographical locations of 695 CRC cancer patients diagnosed between March 2016 and March 2017 in the city of Mashhad. The Mashhad_Neighborhoods file is the digital map of neighborhoods division of the city and their population by age groups. Furthermore, these files include contributor risk factors including average of daily red meat consumption, average of daily fiber intake, and average of body mass index for every of 142 neighborhoods of the city.

Keywords: Colorectal cancer, Geographical information systems, Spatial analysis, Red meat, Dietary fiber, Body mass index

Objective

Colorectal cancer (CRC) is the third most frequently diagnosed malignancy and the second most common cause of death from cancer worldwide [1, 2]. CRC incidence varies in the world with the highest incidence rates in Australia, New Zealand, Europe, and North America and the lowest in Africa and South-Central Asia [1, 3]. The incidence rate of CRC was 7–8 per 100,000 for both males and females in Iran from 1996 to 2000 [4]. However, this incidence rate has been increased to 11.8 and

16.5 (per 100,000) for females and males in 2014 [5]. This increasing trend in CRC incidence may related to high rate of urbanization, people's lifestyle and diet change [5, 6].

Both environmental and lifestyle factors contribute to the risk of CRC. Some important such factors include age, high body mass index (BMI), high-fat diet, alcohol consumption, smoking, consumption of red meat, low intake of vegetables and fruit (fiber intake) [2, 7]. Spatial analysis of CRC incidence may provide a new knowledge on the relationships between environmental risk factors and people lifestyle with CRC burden across communities. This will enable policymakers to develop tailored intervention to areas where the CRC risk is greater. Thus, we investigated the spatial variation of CRC incidence in

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

Firouraghi *et al. BMC Res Notes*   (2020) 13:466

Page 2 of 3

Table 1 Overview of data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	CRCcases_Mashhad	Shape file (.shp)	Harvard Dataverse (https://doi.org/10.7910/DVN/RFOCK7) [26]
Data file 2	Mashhad_Neighbourhoods	Shape file (.shp)	Harvard Dataverse (https://doi.org/10.7910/DVN/RFOCK7) [26]
Data file 3	Avg_Daily_Red_Meat_Consumption_Mashhad	Shape file (.shp)	Harvard Dataverse (https://doi.org/10.7910/DVN/RFOCK7) [26]
Data file 4	Avg_Daily_Fiber_Consumption_Mashhad	Shape file (.shp)	Harvard Dataverse (https://doi.org/10.7910/DVN/RFOCK7) [26]
Data file 5	Avg_BMI_Mashhad	Shape file (.shp)	Harvard Dataverse (https://doi.org/10.7910/DVN/RFOCK7) [26]

the city of Mashhad Iran [8]. In that study, we used Local Moran's *I* statistic (an spatial local clustering approach) [9] to identify high-risk and low-risk areas. A linear regression model developed to quantify the relationship of CRC occurrence with common risk factors [10] including age [2, 11], BMI [12–14], daily red meat consumption [15–20] and daily fiber consumption [7, 20–22]. We developed a comprehensive spatial dataset linked to other attribute data and we would like to offer this dataset for further investigation in future spatial analysis of CRC incidence in Mashhad and elsewhere.

Data description

Geographic Information System (GIS) is a powerful tool for visualizing spatial variation and cluster detection in the pattern of CRC incidence to identify unmet areas [23]. GIS can link geo-referenced risk factors and CRC incidence data with other spatial and temporal data to investigate spatial clustering across time and space [24]. Data were extracted from three different databases. Individual CRC cases were obtained from the population-based cancer registry in Khorasan-Razavi Province. There were 695 CRC diagnosed cases in the city of Mashhad between March 2016 and March 2017. This data set contains patients addresses in the Persian language which had to be geocoded manually using the software Google MyMaps (https://www.google.com/mymaps). These geo-coded data were subsequently transformed into a Keyhole Markup Language (KML) file and imported to ArcGIS software version 10.6 (ESRI, Redands, CA, USA) for further spatial analysis. We randomly jittered the latitude and longitude of the patients address into a 100-m buffer to avoid potential identification of CRC cases. The neighborhood divisions and their population separated in age groups were provided from the City Council in Mashhad. The age groups were presented in the categories including, 0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, and over 65. The age data were provided for both gender (male and female separately). Data regarding risk factors like BMI and average of daily consumption of red meat and fibers, were obtained from the MASHHAD cohort study [25], between 2010 and 2020. The original CRC cases data were visualised as point data in Mashhad. We used spatial interpolation technique and calculate the data for each suburb of the city.

Anselin Local Moran's *I* statistic was used to identify the potential clusters in CRC pattern at the neighborhood level based on incidence rate. The CRC incidence rate was calculated by total population and the frequency of cases per 100,000 persons in each neighborhood in Mashhad. This method helps to find high–high (regions as similar clusters with high values) and low–low (regions as similar clusters with low values of CRC incidence), and high–low (HL) and low–high (LH) areas as special outliers with dissimilarity. We used linear regression model to analyse the relationship between CRC incidence and the risk factors of CRC. In this method, we considered CRC frequency as the dependent variable, and the proportion of the population over 50 years of age, average BMI, average consumption of daily red meat, and average of daily fiber intake as independent variables. The coefficient of determination (R^2) was used to establish the performance of regression model [8]. Researchers can link other environmental risk factors such as air pollution and heavy metals to this dataset and investigate their impact on CRC incidence. Table 1 shows the details of each dataset and provides links to access them.

Limitations

The coverage and precision of population-based cancer registry in Iran are not 100% accurate due to insufficient electronic registries, so we may have missed some CRC patients in our study. However, the detection of high-risk and low-risk areas should not be affected by this limitation.

Firouraghi *et al. BMC Res Notes*   (2020) 13:466

Page 3 of 3

Author details
[1] Department of Medical Informatics, School of Medicine, Mashhad University of Medical Science, Mashhad, Iran. [2] Visualization and Decision Analytics (VIDEA) Lab, Centre for Mental Health Research, Research School of Population Health, College of Health and Medicine, The Australian National University, Canberra, Australia. [3] Department of Gastroenterology and Hepatology, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. [4] Metabolic Syndrome Research Center, Mashhad University of Medical Sciences, Mashhad, Iran. [5] Department of Medical Records and Health Information Technology, School of Paramedical Sciences, Mashhad University of Medical Sciences, Mashhad, Iran.

References
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. Macrae FA. Colorectal cancer: epidemiology, risk factors, and protective factors. Uptodate com [ažurirano 9 lipnja 2017; 2016.
3. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Przegląd Gastroenterol. 2019;14(2):89.
4. Ansari R, Mahdavinia M, Sadjadi A, Nouraie M, Kamangar F, Bishehsari F, et al. Incidence and age distribution of colorectal cancer in Iran: results of a population-based cancer registry. Cancer Lett. 2006;240(1):143–7.
5. Roshandel G, Ghanbari-Motlagh A, Partovipour E, Salavati F, Hasanpour-Heidari S, Mohammadi G, et al. Cancer incidence in Iran in 2014: results of the Iranian National Population-based Cancer Registry. Cancer Epidemiol. 2019;61:50–8.
6. Dolatkhah R, Somi MH, Bonyadi MJ, Asvadi Kermani I, Farassati F, Dastgiri S. Colorectal cancer in Iran: molecular epidemiology and screening strategies. J Cancer Epidemiol. 2015. https://doi.org/10.1155/2015/643020.
7. Kunzmann AT, Coleman HG, Huang W-Y, Kitahara CM, Cantwell MM, Berndt SI. Dietary fiber intake and risk of colorectal cancer and incident and recurrent adenoma in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. Am J Clin Nutr. 2015;102(4):881–90.
8. Goshayeshi L, Pourahmadi A, Ghayour-Mobarhan M, Hashtarkhani S, Karimian S, Dastjerdi RS, et al. Colorectal cancer risk factors in northeastern Iran: A retrospective cross-sectional study based on geographical information systems, spatial autocorrelation and regression analysis. Geospat Health. 2019. https://doi.org/10.4081/gh.2019.793.
9. Anselin L. Local indicators of spatial association—LISA. Geogr Anal. 1995;27(2):93–115.
10. Lawson AB, Banerjee S, Haining RP, Ugarte MD. Handbook of spatial epidemiology. Boaca Raton: CRC Press; 2016.
11. Amersi F, Agustin M, Ko CY. Colorectal cancer: epidemiology, risk factors, and health services. Clin Colon Rectal Surg. 2005;18(3):133.
12. Shaukat A, Dostal A, Menk J, Church TR. BMI is a risk factor for colorectal cancer mortality. Dig Dis Sci. 2017;62(9):2511–7.
13. Ning Y, Wang L, Giovannucci E. A quantitative analysis of body mass index and colorectal cancer: findings from 56 observational studies. Obes Rev. 2010;11(1):19–30.
14. Ochs-Balcom HM, Kanth P, Farnham JM, Abdelrahman S, Cannon-Albright LA. Colorectal cancer risk based on extended family history and body mass index. Genet Epidemiol. 2020;44(7):778–84.
15. Aykan NF. Red meat and colorectal cancer. Oncol Rev. 2015;9(1):288.
16. Santarelli RL, Pierre F, Corpet DE. Processed meat and colorectal cancer: a review of epidemiologic and experimental evidence. Nutr Cancer. 2008;60(2):131–44.
17. Klusek J, Nasierowska-Guttmejer A, Kowalik A, Wawrzycka I, Chrapek M, Lewitowicz P, et al. The influence of red meat on colorectal cancer occurrence is dependent on the genetic polymorphisms of s-glutathione transferase genes. Nutrients. 2019;11(7):1682.
18. zur Hausen H. Red meat consumption and cancer: reasons to suspect involvement of bovine infectious factors in colorectal cancer. Int J Cancer. 2012;130(11):2475–83.
19. Lippi G, Mattiuzzi C, Cervellin G. Meat consumption and cancer risk: a critical review of published meta-analyses. Crit Rev Oncol Hematol. 2016;97:1–14.
20. Tuan J, Chen Y-X. Dietary and lifestyle factors associated with colorectal cancer risk and interactions with microbiota: fiber, red or processed meat and alcoholic drinks. Gastrointest Tumors. 2016;3(1):17–24.
21. Dahm CC, Keogh RH, Spencer EA, Greenwood DC, Key TJ, Fentiman IS, et al. Dietary fiber and colorectal cancer risk: a nested case–control study using food diaries. J Natl Cancer Inst. 2010;102(9):614–26.
22. Song M, Wu K, Meyerhardt JA, Ogino S, Wang M, Fuchs CS, et al. Fiber intake and survival after colorectal cancer diagnosis. JAMA Oncol. 2018;4(1):71–9.
23. Sahar L, Foster SL, Sherman RL, Henry KA, Goldberg DW, Stinchcomb DG, et al. GIScience and cancer: state of the art and trends for cancer surveillance and epidemiology. Cancer. 2019;125(15):2544–60.
24. 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 Spat Anal Policy. 2020;13(2):293–303.
25. Ghayour-Mobarhan M, Moohebati M, Esmaily H, Ebrahimi M, Parizadeh SMR, Heidari-Bakavoli AR, et al. Mashhad stroke and heart atherosclerotic disorder (MASHAD) study: design, baseline characteristics and 10-year cardiovascular risk estimation. Int J Public Health. 2015;60(5):561–72.
26. Kiani B. Colorectal cancer cases & related risk factors. Harvard Dataverse. 2020. https://doi.org/10.7910/DVN/RFOCK7.

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