Archives of Advances
in Biosciences

Original Article:

Iranian Stomach Cancer Registry Data Analysed through Multivariate Adaptive Geographically Weighted Generalized Poisson Regression Spline

Zahra Nasseriakbar [1], Maryam Khayamzadeh[2], Amir Kavousi[3*]

[1.] Master Student, Paramedical Faculty,Biostatistics Department,Shahid Beheshti University of Medical Science, Tehran,Iran
[2.] Community Medicine Specialist, Academy of Medical Sciences ,Cancer Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[3.] Safety Promotion and Injury Prevention Research Center, Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Cite this article as: Nasseriakbar, Z., Khayamzadeh, M., & Kavousi, A. (2024). Iranian Stomach Cancer Registry Data Analysed through Multivariate Adaptive Geographically Weighted Generalized Poisson Regression Spline. *Archives of Advances in Biosciences*, *15*(1), 1–10. https://doi.org/10.22037/aab.v15i1.45989

https://doi.org/10.22037/aab.v15i1.45989

Article info:
Received: 12 Aug 2024
Accepted: 30 Nov 2024
Published: 01 Dec 2024

*** Corresponding author:**

Amir Kavousi, PhD.

Address: Safety Promotion and Injury Prevention Research Center, Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

E-mail: kavousi @sbmu.com

Abstract

Introduction: In 2022, gastric and breast cancer had a mortality rate of 6.8 cases, ranking fourth despite intervention efforts. A 2020 study by the International Agency for Research on Cancer found global variations in incidence rates. This study examines key risk factors and preventive measures [1-2].

Materials and Methods: This applied ecological study uses the MAGWGPRS (Multivariate Adaptive Geographically Weighted Generalized Poisson Regression Spline) model, integrating MARS and GWGPR, to analyze cancer registry data. The model identifies geographic variations and hotspots in cancer risk. Data sources include pathology reports, death records, biopsy data, and a non-communicable disease risk factor survey. The dataset comprises patient age, location, cancer case counts, and relevant risk factors. Analysis is conducted using R, with ArcGIS for map visualization.

Results: According to the International Agency for Research on Cancer, key risk factors for stomach cancer include obesity, smoking, physical inactivity, poor nutrition, age, and population density. The MAGWGPRS model, a Geographically Weighted Model, identifies regional variations in these factors by weighting observations based on distance using a kernel function and optimizing the model with the GCV criterion. Our analysis highlights vegetable consumption, smoking, low physical activity, and age as the primary determinants of gastric cancer risk.

Conclusion: Our model identifies vegetable consumption, smoking, low physical activity, and aging as significant risk factors for gastric cancer. Further research is needed to refine obesity risk based on BMI criteria. The MAGWGPRS model is a valuable tool for identifying high-risk regions, enabling targeted interventions and prioritizing key risk factors across diverse geographic areas.

Keywords: stomach Neoplasms/pathology, stomach Neoplasms/ diagnosis, Aged, Body Mass Index, obesity/complications,Toxicity, (Diet,Vegetarian) /statistics and numerical data, Analyses Spatial, Geographically weighted regression, Poisson Distribution,Spatial Regression, Iran

1. Introduction

Analyzing and assessing disease rates, especially cancer, is a critical issue in public health and epidemiology. This data can be a valuable tool for planning preventive actions. However, previous studies have overlooked the impact of location on their findings, focusing instead on independent observations [3-4].

In order to make better predictions and estimations, it is essential to consider the impact of nearby regions, making spatial analysis more beneficial. According to Tobler (1969), the first law of geography states that everything is related to everything else, but things that are closer together are more closely related [5]. As a result, various spatial models have been utilized, including spatial autoregressive models (SAR) (Francisca Corpas-Burgos et al. 2021, Leiwen Gao et al. 2020), spatial errors model, and Geographically Weighted Regression (GWR)[6], which have prompted us to enhance our understanding and utilize MARGWGPRS for these purposes [7-11].
MARGWGPRS is an expansion of geographically weighted Regression, incorporating count data (GWPR) and the adoption of MARS. This makes it more precise for forecasting unsampled regions based on nearby sampled regions. It aims to recognize spatial heterogeneity and systematic variation between observation locations. MARGWGPRS has been applied to cancer registry data in Iran and the attachment of non-communicable disease risk factors prevalence survey. It calculates observation vicinity using Euclidean distance and creates a distance matrix. Geographically weighted Regression is responsible for weighing the observations, giving more weight to observations near the regression point, and less weight to those further away [11,12-18].

Calculating the bandwidth is controversial, with smaller bandwidths causing more variation and longer ones leading to ordinary models (Global model) (E.D Priyana and Purhadi 2014). Therefore, various kernel functions have been introduced to weigh the statement with respect to the bandwidth, including "fixed" and "adaptive" categories [19-20].

Adaptive kernels, proportional to studies, are more persuasive, with Gaussian adaptive kernel being selected. MARS adds some assumptions like minimum observation, maximum interaction, and basis function, with the optimum developed using cross-validation criteria for selecting the best models.[15,19,21]

The rest of the essay is structured as follows: model development and methods in Sec2, results Sec3, discussion in Sec 4, and conclusion and remarks in Sec5.

2. Materials and Methods

This study employs the MAGWGPRS (Multivariate Adaptive Geographically Weighted Generalized Poisson Regression Spline) model to analyze trends in gastric cancer occurrence and identify key risk factors. Cancer registry data from 2016 was used due to its relevance and the increasing prevalence of cancer in Iran. The study focuses on spatial regression techniques, specifically Geographically Weighted Poisson Regression, to detect hotspots and understand spatial patterns of gastric cancer. The dataset includes information on patient age, location, and risk factors derived from non-communicable disease surveys. Cases were identified through biopsy, pathology, and death reports. Data was converted into a spatial format, with geographic methods applied to improve accuracy and analytical depth. In 2016, a total of 10.382 cases of gastric cancer were reported in both sexes, making it the fifth most common cancer in Iran.

Data

The data was exported from Iran's 2016 gastric cancer registry, along with covariates from the non-communicable disease risk factors survey. Registry data was accumulated based on population with reports of the biopsy, pathology and cancer death. Six covariates include the percentage of people who consume vegetables more than twice a week, the percentage of smokers in different regions of Iran, the percentage of people struggling with obesity, the percentage of people over fifty years old, population, and the percentage of people with low physical activity. These covariates, chosen based on previous studies, have important environmental and ecological reasons for their inclusion. Prior research (2019) and has highlighted the significant roles of smoking and aging in increasing cancer risk.[22-23] Studies have demonstrated that consuming an adequate amount of vegetable can have a protective effect against gastric cancer.[24]The population is not selected as an individual factor since cancer is not a virus-based disease; rather, we aimed to analyze the ultimate model to confirm whether it is truly significant or not.

It is noteworthy that the study area is ecological and R-Open sources software is used for in-depth analysis. ArcGIS has been utilized for improved map visualization. Due to the failure of ordinary regression assumptions on the current data, GWPR models and their extension are implemented. The subsequent paragraph elaborates on these models and their progress in development.

MAGWGPRS Models

First of all, using ordinary regression is one of the approaches generally considered in our data related to different regions. We have exhausted these methods, and if we continue down this path due to the presence of heteroscedasticity, we will achieve poor results. We are exploring spatial methods, as confirmed by the Breusch-Pagan test introduced by Trevor. The heterogeneity among the data has been validated; therefore, we are moving towards using GWR and its generalized form.[25]

The GWR is a spatial regression that employs a Kernel function to weigh neighboring observations. It considers observations around the regression points, assigning greater weight to those closer to the regression point, while applying less weight to those further away. In this study, the Gaussian Kernel function has been adopted. Weighing the observations according to distance and bandwidth – distance between regression point and the neighboring observation – is a crucial part of this regression.

$$w_{ij} = \begin{cases} exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right) & \text{i=1,2,…,n} \\ 0 & \text{o.w} \end{cases}$$

where

$$d_{ij} = \sqrt{(\mu_i - \mu_j)^2 + (\upsilon_i - \upsilon_j)^2}$$

Bandwidth selection by focusing on the lowest CV has been chosen.

By focusing on data where the response variable is in count form, it becomes clear that using GWPR, which relates to the Poisson variant of the GWR model, which the response variable is countable, in other words dependent variable is following Poisson distribution means that events repeat in time period by specified rate λ_i .[14,26]

$$y_i \sim Poisson\left[exp\left(\sum_k \beta_k(\mu_i, \upsilon_i) x_{ik}\right)\right]$$

x_{ik} is the kth explanatory variable, β_k geographically varying coefficient, (μ_i, υ_i) describing the longitude and latitude of the location i.

requires testing the assumption of equidispersion. This assumption often fails to hold in real data, leading to a lack of confidence in the results. To address this issue, an alternative form of GWPR is necessary. According to Adrianta et al., the generalized form of the model, GWGPR, has been introduced and can effectively help mitigate this problem[27]. As some neighbors exist that require more precise predictions, particularly in areas without samples, the spline method has been applied to the GWGPR, yielding better results and establishing it as MAGWGPRS.

$$y_i \sim \left(\mu(\mu_i, \upsilon_i), \varphi(\mu_i, \upsilon_i)\right) \qquad i = 1,2,\cdots,n$$

$$\mu(\mu_i, \upsilon_i) = exp\left({b_i}^T \gamma(\mu_i, \upsilon_i)\right)$$

Where ${b_i}^T = \left(1, B_1(x_i), B_2(x_i), \cdots, B_M(x_i)\right)$, $\gamma(\mu_i, \upsilon_i) = \left(\gamma_0(\mu_i, \upsilon_i), \gamma_1(\mu_i, \upsilon_i), \cdots, \gamma_M(\mu_i, \upsilon_i)\right)^T$

Are parameters of basis function.

MAGWGPRS model is a combination of the Geographically Weighted Regression model and Multivariate Adaptive Regression Splines. To clarify MARS, it should emphasize its functionality, which involves elaborating a polynomial of basis functions (BF), minimum observations (MO), and maximum interactions (MI). This process determines knots within the domain of the explanatory variables, using forward stepwise selection to add basis truncated functions until reaching the maximum BF, followed by backward stepwise elimination of non-significant ones according to GCV. The model which remains after the process is the best model with the lowest GCV:

$$GCV(M) = \frac{MSE}{\left(1 - \frac{C(\widetilde{M})}{n}\right)^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}_M(x_i)\right)^2}{\left(1 - \frac{C(\widetilde{M})}{n}\right)^2}$$

Which $C(\widetilde{M}) = C(M) + \frac{dM}{2}$, let $C(\widetilde{M})$ be a complex function, $C(M)$ the number of parameters to estimate, d the interaction degree, optimal within $2 \le d \le 4$, and $\hat{f}_M(x_i)$ the estimated response variable value on basis function M and the ith observation.

All GWR generalized methods follow the same process but come with additional assumptions that have been mentioned. Certain criteria must be introduced to evaluate the functionality of the models for which AIC (Sakamoto and Kitagawa 1986), Ishiguro and AICc [28-30] :

$$AIC(h) = D(h) + 2K(h)$$

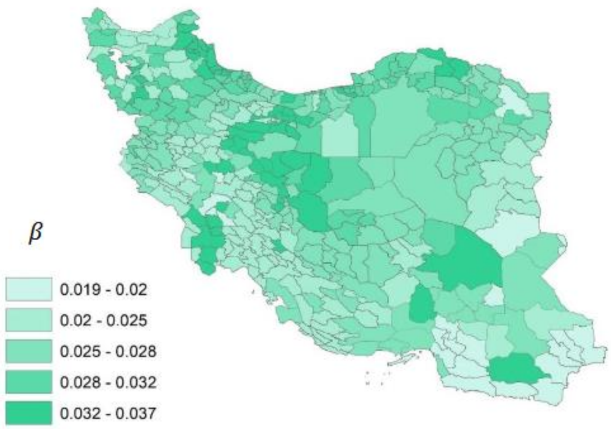
Which D is the deviance of the model, K effective number of parameters, and h is the bandwidth

3. Results

The data has been examined and found to have a $\varphi = 85.24 > 1$, indicating equidispersion. By applying the GWPR model with a bandwidth of 85 and CV=82188, we reached AIC=110.92. All factors were significant, but the population had a fluctuating effect, so further investigation is needed to confirm the existence of heterogeneity. Thus, the Breusch-Pagan test statistic of 35.19 was used to verify it. The spatial distribution of the data, which was found through our previous examination, supports the use of our new method to reach more reliable results. A bandwidth value of 78, with a cross-validation (CV) value of 84269, was selected with careful attention to the selection process for the ultimate model.

Following the Friedman rule model, the best model was determined to have BF=24, MI=3, MO=2, and GCV=16.08. In this model, obesity and population were found not to be significant and were removed from the algorithm. By comparing the AIC of the GWPR model to the ultimate model with an AIC of 109.26, we achieved a more reliable result and enhanced functionality. As can be seen in the maps below, the effects of factors over fifty years old on gastric cancer are more pronounced in the central part of Iran. Vegetable consumption impacts the western region; low physical activity affects the northern area, and smoking is prevalent in the northwestern part of Iran. Additionally, it should be clarified that the effect of vegetable consumption is protective, meaning that higher vegetable intake leads to a lower occurrence of gastric cancer.

Standard error (SE) maps are displayed to verify the accuracy of the parameter estimates, indicating that a lower SE correlates with better accuracy. In areas where the map is darker, SE accuracy is diminished.



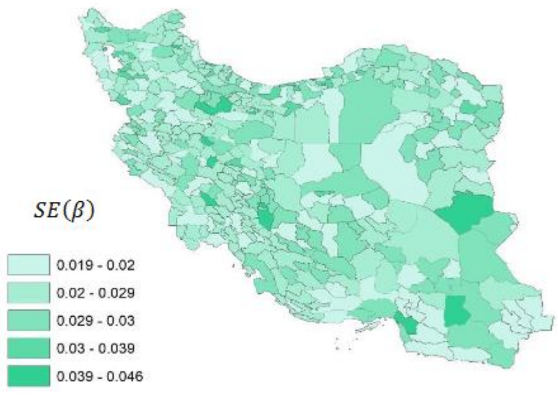


Figure 5-1(a) Age Effect map on Gastric cancer in Iran geographical locations

Figure 5-1 (b) Standard error of Age Effect map on Gastric cancer in Iran geographical locations

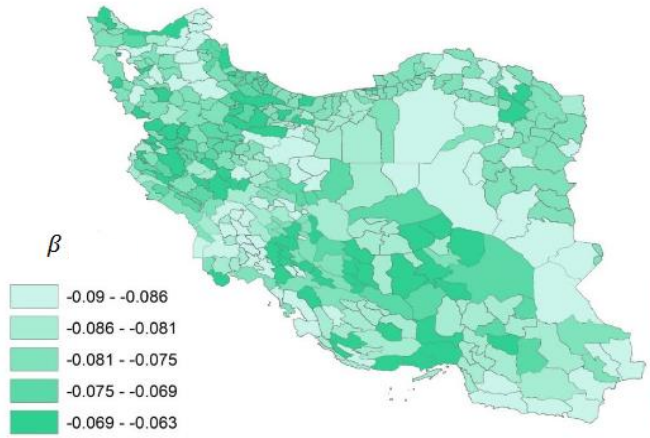


Figure 5-2 (a) Vegetable Effect map on Gastric cancer in Iran geographical locations

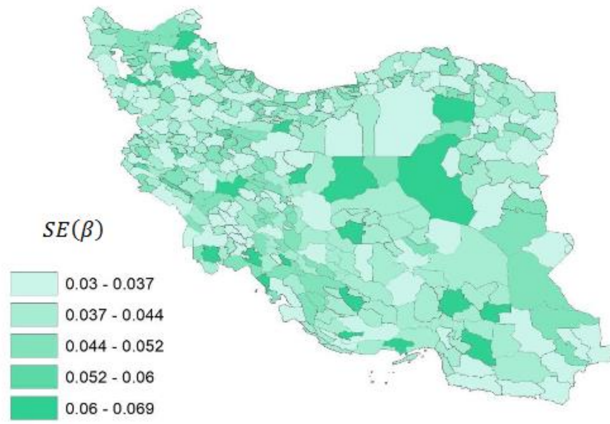


Figure 5-2 (b) Standard error of Vegetable Effect map on Gastric cancer in Iran geographical locations

Individuals in northern Iran with lower weekly activity levels have a significantly higher risk of gastric cancer mortality and morbidity.

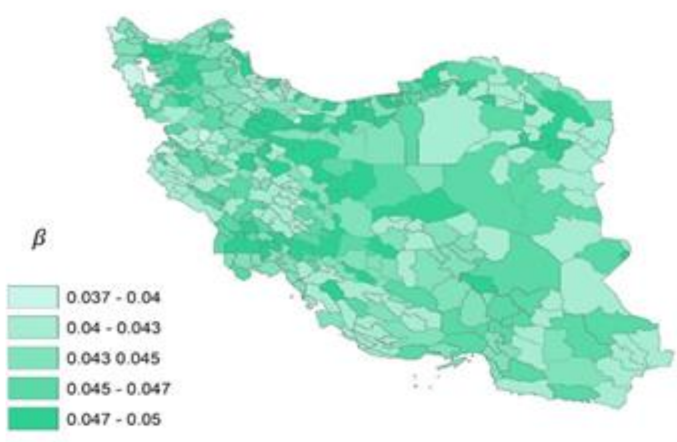


Figure 5-3(a) Low Physical activity map on Gastric cancer in Iran geographical locations

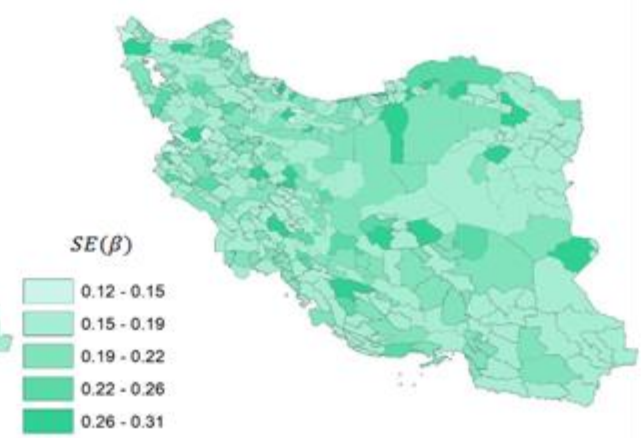
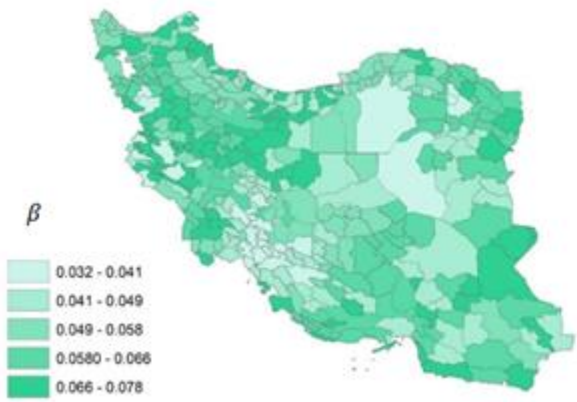


Figure 5-3(b) Standard error of Low Physical activity map on Gastric cancer in Iran geographical locations



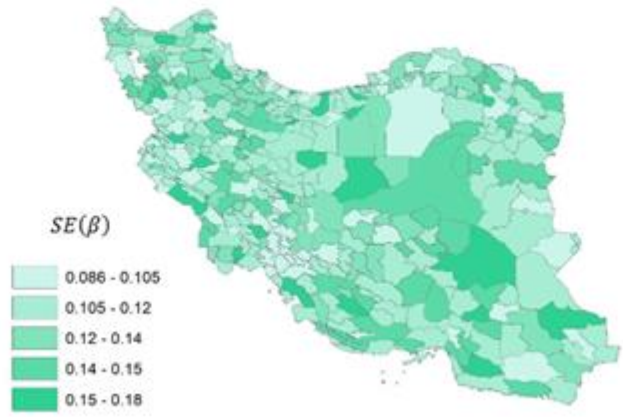


Figure 5-4 (a) Smoking Effect map on Gastric cancer in Iran geographical locations

Figure 5-4 (b) Standard error of Smoking Effect map on Gastric cancer in Iran geographical locations

4. Discussion

According to our model, all variables, except obesity and population, were significant. In general, smoking, obesity, vegetable consumption, low physical activity, and age are the most important factors for cancer risk. The outputs of the ultimate model highlight four key risk factors.

As the model estimations relate to various regions, the coefficients of the optimal model should be shown on a map to indicate which area has a higher occurrence of gastric cancer linked to specific risk factors.

As illustrated in Figure 5-1(a), the covariate “over fifty years old” exhibits a significant association with gastric cancer occurrence, based on non-communicable disease surveys of individuals aged fifty and above, concentrated in the central part of Iran. This region demonstrates a notably higher incidence rate of gastric cancer compared to other areas. Figure 5-2(a) illustrates the impact of the appropriate consumption of vegetables and fruits, which have protective effects against the occurrence of gastric cancer. The darker areas of the map indicate stronger protective effects. Figure 5-3 (a) indicates the impact of low physical activity on gastric cancer. According to the visualization, it is evident that individuals in the northwest of Iran exhibit higher levels of procrastination and a greater incidence of cancer. Figure 5-4 (a) illustrates the impact of smoking on gastric cancer. As noted by Han et al. (2019), its effects have been confirmed and are prevalent in western and southeastern Iran [31].

Previous studies highlighted the significant roles of smoking and aging in elevating cancer risk. Our model corroborated these findings. It is evident that aging is responsible for a greater cancer risk because of increased sensitivity to harm and delayed gastric mucosa recuperation [23,32] Moreover adequate vegetable consumption can have a protective effect against gastric cancer, and recent research indicates that low physical activity leads to higher body fat levels and increases cancer risk [33].

Environmental factors like physical disability, family issues, economic crisis, and environmental insecurity are the main causes of aging in central part of Iran; hence, the effects of that on gastric cancer is obvious[34-36].

Unhealthy lifestyle, family disease history, and dependence on personal vehicles are often cited as reasons for low physical activity in western Iran, which is responsible for more gastric cancer occurrence in that region [37-39].

Factors such as low educational level, low income level, aging, stress, and social distress contribute to reduced utilization of vegetables, the coefficient estimates clearly indicate that the negative sign preceding them accounts for the protective effects of vegetable consumption on gastric cancer.[40-42] smoking has been identified as a significant contributing factor to the prevalence of gastric cancer in the western and southern regions of Iran, potentially attributed to lower levels of education and literacy.[43-48]The aforementioned studies place greater emphasis on the identified risk factors that we have discussed.

Based on various risk factors and environmental and ecological aspects, different occurrences of gastric cancer can be observed. However, there is generally more evidence of gastric cancer occurrence in the northwest of Iran.

5. Conclusion

The study aimed to apply the best MAGWGPRS model with Gaussian adaptive bandwidth to Iran's Cancer registry data 2016. The model has been applied in the generalized form of the GWR by addressing the issues present in previous iterations through the use of additional assumptions observed in the ultimate version. Transparently employing ordinary regression without considering the correlation of neighboring regions yielded poor results. Applying GWPR and its extensions ensures that the correlation between nearby areas and the theory of the impact of adjacent regions on one another are taken into account, leading to more reliable results.

We assumed that vegetable consumption would be a significant factor in the morbidity and mortality of gastric cancer, and according to our model we reached resemble conclusion and emphasized protective effects of that. while in the other studies has been shown Obesity is one of the risk factors on affecting Gastric cancer our model evidence rejected the direct influence of this risk factor, while aging was the most significant factor in the center. Smoking was visualized in the north, and population as an environmental factor was nonsignificant due to the noncommunicable disease it is obvious this explanatory variable individually cannot play real vital role According to our model, The main factors affecting cancer risk are typically smoking, aging, vegetable consumption, low physical activity, according to our model.

Different kernel functions can be used to achieve different results, and we found that the best fit for our data was the Gaussian function. In this study, we aimed to identify the location of gastric cancer occurrences and their trends in that area. However, we recommend that utilizing information about incidence and prevalence can yield more comprehensive results. Analyzing data related to consequence reports and minimizing deviations for further investigation can also be beneficial. The availability of more significant factors, such as the effects of Helicobacter pylori, which were beyond the scope of our study on gastric cancer occurrence, could enhance prediction accuracy.

Ethical Considerations

Compliance with ethical guidelines

This study received ethical approval from the Ethics Committee of Shahid Beheshti University of Medical Sciences (IR.SUMS.RETECH.REC.1402.407). Furthermore, we affirm that no interventions were conducted as part of this research. Additionally, we confirm that all methodologies employed in this study adhere to the relevant guidelines and regulations of the AAB journal.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Author's contributions

The authors equally contributed to preparing this article.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We would like to extend our appreciation and thanks to all participants, staff, and managers who made this study possible.

References

[1] Bray, Freddie, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Isabelle Soerjomataram, and Ahmedin Jemal. 2024. "Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide

for 36 Cancers in 185 Countries." CA: A Cancer Journal for Clinicians 74 (3): 229–63. [DOI: 10.3322/caac.21834]

[2] "IARC – INTERNATIONAL AGENCY for RESEARCH on CANCER." n.d. [LINK]

[3] Xu, Wei, Tianxiao Jiang, Kanger Shen, Dongxu Zhao, Man Zhang, Wenxin Zhu, Yunfei Liu, and Chunfang Xu. 2023. "GADD45B Regulates the Carcinogenesis Process of Chronic Atrophic Gastritis and the Metabolic Pathways of Gastric Cancer." *Frontiers in Endocrinology* 14 (August). [DOI: 10.3389/fendo.2023.1224832]

[4] Li, Guo-Cai, Hong-Wei Zhang, Hong-Gang Tian, Yun Zhao, Qin-Xian Huang, Ze-Yu Xu, Xiao-Hui Xi, and Kai Zhang. 2022. "Clinicopathological Difference between Gastric Cancer in the Lesser Curvature and Gastric Cancer in the Greater Curvature." Medicine 101 (33): e29984. https://doi.org/10.1097/md.0000000000029984. [DOI: 10.1097/md.0000000000029984]

[5] Waters, Nigel. 2017. "Tobler's First Law of Geography." International Encyclopedia of Geography: People, the Earth, Environment and Technology, March, 1–13. [DOI:10.1002/9781118786352.wbieg1011]

[6] Banerjee, Srikanta, and Matt Jones. 2022. "Using Geographically Weighted Linear Regression for County-Level Breast Cancer Modeling in the United States." MedRxiv (Cold Spring Harbor Laboratory), April.

[DOI: 10.1101/2022.03.28.22272969]

[7] Corpas-Burgos, Francisca, and Miguel A Martinez-Beneito. 2021. "An Autoregressive Disease Mapping Model for Spatio-Temporal Forecasting." Mathematics 9 (4): 384–84. https://doi.org/10.3390/math9040384. [DOI:10.3390/math9040384]

[8] Gao, Leiwen. 2021. Bayesian Hierarchical Models for Multivariate Regionally Aggregated Data Using Directed Acyclic Graph Auto-Regressive (DAGAR) Models.

[9] Luo, Lan, Sara McLafferty, and Fahui Wang. 2010. "Analyzing Spatial Aggregation Error in Statistical Models of Late-Stage Cancer Risk: A Monte Carlo Simulation Approach." International Journal of Health Geographics 9 (1): 51. [DOI:10.1186/1476-072x-9-51]

[10] Sun, Hong-Bo, and Bradley P Carlin. 1998. "Spatio-Temporal Models with Errors in Covariates: Mapping Ohio Lung Cancer Mortality." Statistics in Medicine 17 (18): 2025–43.

[11] Riry Sriningsih, Bambang Widjanarko Otok, and None Sutikno. 2023. "Determination of the Best Multivariate Adaptive Geographically Weighted Generalized Poisson Regression Splines Model Employing Generalized Cross-Validation in Dengue Fever Cases," April, 102174–74. https://doi.org/10.1016/j.mex.2023.102174.[DOI: 10.1016/j.mex.2023.102174]

[12] C. Brunsdon. Geographically Weighted Regression-Modelling Spatial Non-Stationarity. Journal of the Royal Statistical Society Series A (Statistics in Society). 1998 Jan 1;47:431–43.[DOI:10.1111/1467-9884.00145]

[13] Nakaya T, Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted Poisson regression for disease association mapping. Statistics in medicine. 2005 Sep 15;24(17):2695-717. [DOI:10.1002/sim.2129] [PMID]

[14] Friedman JH. Multivariate adaptive regression splines. The annals of statistics. 1991 Mar;19(1):1-67.[DOI:10.1214/aos/1176347963]

[15] A Stewart Fotheringham, Chris Brundson, and Martin Charlton. 2002. Geographically Weighted Regression : The Analysis of Spatially Varying Relationships. Chichester, West Sussex, England ; Hoboken, Nj: Wiley. ISBN: 978-0-471-49616-8

[16] Brunsdon, Chris, A. Stewart Fotheringham, and Martin E. Charlton. 1996. "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity." *Geographical Analysis* 28 (4): 281–98. [DOI: 10.1111/j.1538-4632.1996.tb00936.x]

[17] Khaulasari H. Modelling mixed geographically weighted poisson regression for tuberculosis disease in Surabaya. Journal of Physics: Conference Series. 2020 Mar;1490:012004. [LINK]

[18] Davies, Tilman M, and Andrew B Lawson. 2019. "An Evaluation of Likelihood-Based Bandwidth Selectors for Spatial and Spatiotemporal Kernel Estimates." Journal of Statistical Computation and Simulation 89 (7): 1131–52. [DOI:10.1080/00949655.2019.1575066]

[19] W. John Braun, and Valentin Rousson. 2000. "An Autocorrelation Criterion for Bandwidth Selection in Nonparametric Regression∗" 68 (1): 89–101. [DOI:10.1080/00949650008812057]

[20] Wang, Yang, An-Na Wang, Qing Ai, and Hai-Jing Sun. 2017. "An Adaptive Kernel-Based Weighted Extreme Learning Machine Approach for Effective Detection of Parkinson's Disease." Biomedical Signal Processing and Control 38 (September): 400–410. https://doi.org/10.1016/j.bspc.2017.06.015.. [DOI: 10.1016/j.bspc.2017.06.015]

[21] Davies, Tilman M., and Martin L. Hazelton. 2010. "Adaptive Kernel Estimation of Spatial Relative Risk." *Statistics in Medicine* 29 (23): 2423–37. [DOI:10.1002/sim.3995]

[22] Li WY, Han Y, Xu HM, Wang ZN, Xu YY, Song YX, Xu H, Yin SC, Liu XY, Miao ZF. Smoking status and subsequent gastric cancer risk in men compared with women: a meta-analysis of prospective observational studies. BMC cancer. 2019 Dec;19:1-2. [DOI: 10.1186/s12885-019-5601-9] [PMID] [PMCID]

[23] Ferro A, Morais S, Rota M, Pelucchi C, Bertuccio P, Rossella Bonzi, et al. Tobacco smoking and gastric cancer: meta-analyses of published data versus pooled analyses of individual participant data (StoP Project). European Journal of Cancer Prevention. 2018 May 1;27(3):197–204. [DOI:10.1097/cej.0000000000000401]

[24] Wan Q, Li N, Du L, Zhao R, Yi M, Xu Q, Zhou Y. Allium vegetable consumption and health: An umbrella review of meta-analyses of multiple health outcomes. Food science & nutrition. 2019 Aug;7(8):2451-70. [DOI: 10.1002/fsn3.1117] [PMID][PMCID]

[25] Breusch, T. S., and A. R. Pagan. 1979. "A Simple Test for Heteroscedasticity and Random Coefficient Variation." Econometrica 47 (5): 1287–94. https://doi.org/10.2307/1911963.[LINK]

[26] Li, Zhibin, Wei Wang, Pan Liu, John M. Bigham, and David R. Ragland. 2013. "Using Geographically Weighted Poisson Regression for County-Level Crash Modeling in California." *Safety Science* 58 (October): 89–97. [DOI: 10.1016/j.ssci.2013.04.005]

[27] Sischa Wahyuning Tyas, None Gunardi. Geographically weighted generalized poisson regression model with the best kernel function in the case of the number of postpartum maternal mortality in east java. MethodsX (Amsterdam). 2023 Jan 1;10:102002–2. [DOI: 10.1016/j.mex.2023.102002]

[28] Sakamoto, Y, M Ishiguro, and G Kitagawa. 1984. *Akaike Information Criterion Statistics*. ISBN:9027722536

[29] Liu, Xinwang, Sihang Zhou, Yueqing Wang, Miaomiao Li, Yong Dou, En Zhu, and Jianping Yin. 2017. "Optimal Neighborhood Kernel Clustering with Multiple Kernels." *Proceedings of the AAAI Conference on Artificial Intelligence* 31 (1). [DOI:10.1609/aaai.v31i1.10895aaai.v31i1.10895]

[30] Cavanaugh, Joseph E. 1997. "Unifying the Derivations for the Akaike and Corrected Akaike Information Criteria." *Statistics & Probability Letters* 33 (2): 201–8.[DOI:10.1016/s0167-7152(96)00128-9 10.1101/2022.03.28.22272969]

[31] Li WY, Han Y, Xu HM, Wang ZN, Xu YY, Song YX, Xu H, Yin SC, Liu XY, Miao ZF. Smoking status and subsequent gastric cancer risk in men compared with women: a meta-analysis of prospective observational studies. BMC cancer. 2019 Dec;19:1-2 [PMID] [PMCID]

[32] Lee JG, Kim SA, Eun CS, Han DS, Kim YS, Choi BY, Song KS, Kim HJ, Park CH. Impact of age on stage-specific mortality in patients with gastric cancer: A long-term prospective cohort study. PloS one. 2019 Aug 1;14(8):e0220660. [PMID] [PMCID]

[33] Wan Q, Li N, Du L, Zhao R, Yi M, Xu Q, Zhou Y. Allium vegetable consumption and health: An umbrella review of meta-analyses of multiple health outcomes. Food science & nutrition. 2019 Aug;7(8):2451-70. [DOI: 10.1002/fsn3.1117] [PMID][PMCID]

[34] Ayoubi-Mahani S, Eghbali-Babadi M, Farajz-adegan Z, Keshvari M. Facilitators and Barriers to Active Aging in Iran: A Gualitative Study. Iranian Journal of Ageing. 2023 Oct 10;18(3):438-59.[Doi: 10.3389/fpubh.2023.1121761] [PMID][PMCID]

[35] Gietel-Basten S, Marois G, Torabi F, Kabiri K. Reframing policy responses to population aging in Iran. Genus. 2024 May 28;80(1). [DOI: 10.1186/s41118-023-00210-z]

[36] Mahmoudzadeh H, Aghayari Hir T, Hatami D. Study and Analysis of the Elderly Population of the Iran. Geographical Researches. 2022 Jan 10;37(1):111-25.[LINK]

[37] Nejadghaderi SA, Ahmadi N, Rashidi MM, Ghanbari A, Noori M, Abbasi-Kangevari M, Nasserinejad M, Rezaei N, Yoosefi M, Fattahi N, Ghasemi E. Physical activity pattern in Iran: Findings from STEPS 2021. Frontiers in Public Health. 2023 Jan 4;10:1036219.

[DOI:10.3389/fpubh.2022.1036219]

[38] Mohebi F, Mohajer B, Yoosefi M, Sheidaei A, Zokaei H, Damerchilu B, Mehregan A, Shahbal N, Rezaee K, Khezrian M, Nematollahi Dehmoosa A. Physical activity profile of the Iranian population: STEPS survey, 2016. BMC public health. 2019 Dec;19:1-7.

[DOI:10.1186/s12889-019-7592-5] [PMID][PMCID]

[39] Sahebkar M, Miri HH, Noormohammadpour P, Tiyuri A, Pakzad R, Mansournia N, Heidari Z, Mansournia MA, Stamatakis E. Geographical patterning of physical activity prevalence in Iran: Spatial analysis of 4 pooled national health surveys among 119,560 adults. Journal of Physical Activity and Health. 2019 Dec 1;16(12):1071-7.[DOI: 10.1123/jpah.2019-0053] [PMID]

[40] World Cancer Research Fund, American Institute For Cancer Research. Diet, nutrition, physical activity and cancer : a global perspective : a summary of the Third expert report. London: World Cancer Research Fund International; 2018. ISBN(print):978-1-912259-47-2 ISBN(pdf) 978-1-912259-47-2

[41] Sabzghabaee AM, Mirmoghtadaee P, Mohammadi M. Fruit and vegetable consumption among community dwelling elderly in an Iranian population. International journal of preventive medicine. 2010;1(2):98. [PMID][PMCID]

[42] Saghafian F, Malmir H, Saneei P, Keshteli AH, Hosseinzadeh-Attar MJ, Afshar H, Siassi F, Esmaillzadeh A, Adibi P. Consumption of fruit and vegetables in relation with psychological disorders in Iranian adults. European journal of nutrition. 2018 Sep;57:2295-306. [DOI:10.1007/s00394-018-1652-y]

[43] Abbasi-Kangevari M, Masinaei M, Fattahi N, Rahimi Y, Rezaei N, Azadnajafabad S, Ghanbari A, Fakhimi R, Jourahmad Z, Ghamari SH, Malekpour MR. Current inequities in smoking prevalence on district level in Iran: a systematic analysis on the STEPS Survey. Journal of research in health sciences. 2021 Dec 28;22(1):e00540. [DOI: 10.34172/jrhs.2022.75] [PMCID]

[44] Shuja M, Sarrafzadegan N, Roohafza HR, Sadeghi M, Ghafari M, Mohammadian M, Hafshejani AM. Factors associated with cigarette smoking in central parts of Iran. Asian Pacific journal of cancer prevention: APJCP. 2017;18(3):647.

[DOI: 10.22034/APJCP.2017.18.3.647] [PMID]

[45] Chaman R, Khosravi A, Sajedinejad S, Nazemi S, Mohasseli KF, Valizade B, Vahedi H, Hosseinzadeh E, Amiri M. Smoking and its related factors among Iranian high school students. Iranian journal of psychiatry and behavioral sciences. 2015 Dec;9(4).[PMID]

[46] Shahmohamadi E, Yousefi M, Mohammadi E, Ghanbari A, Shaker E, Azadnajafabad S, Abbasi-Kangevari M, Rashidi MM, Rezaei N, Fateh SM, Mehr EF. National and Provincial Prevalence of Cigarette Smoking in Iran; A Systematic Analysis of 12 Years of STEPS Experience. Archives of Iranian Medicine. 2023 Sep;26(9):472. [PMCID]

[47] Aslani M, Nazari M, Jamaati H, Naseri O, Saedmoucheshi S, Singh PK, Gheshlagh RG. Prevalence of Cigarette Smoking and its Related Factors among Students in Iran: A Meta-Analysis. Tanaffos. 2022 Mar;21(3):271. [PMID]

[48] Rad EH, Hajizadeh M, Rezaei S, Reihanian A. Cigarette smoking and its financial burden among Iranian households: evidence from household income and expenditures survey. Journal of Research in Health Sciences. 2020;20(4):e00494. [PMID] [PMCID]