Original Article
Middle East Journal of Cancer; January 2018; 9(1): 49-55

Clustering of Breast Cancer Cases among Women from Kurdistan Province, Iran: A Population-based Cross-sectional Study

Seyed Mehdi Hosseini*[+], Masoud Parvin[+], Payam Shokri***, Milad Fadaie****, Bahman Ghaytasi*****, Manoochehr Khondabi*, Meysam Olfatifar***,*[♦], Ebrahim Chavoshi******

*Student Research Committee, Department of Epidemiology, Hamadan University of Medical Sciences, Hamadan, Iran
**Student Research Committee, Department of Dentistry, Hamadan University of Medical Sciences, Hamadan, Iran
***Social Determinants of Health Research Center, Kurdistan University of Medical Sciences, Sanandaj, Iran
****Department of Biotechnology, Hamadan University of Medical Sciences, Hamadan, Iran
*****Department of Public Health and Disease Prevention and Control Center, Health Deputy, Kurdistan University of Medical Sciences, Sanandaj, Iran
******Faculty of Agriculture, Bu Ali Sina University, Hamadan, Iran
[+]Seyed Mehdi Hoseini and Masoud Parvin has equally contributed to this study

Abstract

Background: Spatial analysis is one of the required tools of epidemiology and public health sciences. This study intends to detect significant clusters of breast cancer cases in Kurdistan Province, Iran.

Methods: We obtained data that pertained to breast cancer cases during 2005-2014 from the Health Deputy at Kurdistan University of Medical Sciences. After application of spatial scan statistics to detect the purely spatial (aggregation of cases in particular locations of space) and space-time (diseases clusters in space that depend on the time period) clusters, we calculated the population attribution risk (%) values to better distinguish the detected clusters.

Results: We observed that the second secondary purely spatial cluster (P=0.0051) had the highest population attribution risk (%) of 3.8 and the primary space-time unadjusted cluster (P=0.0019) had the lowest population attribution risk (%) of 0.67 of all the detected clusters. Before we applied the adjustment, both the space-time and purely spatial clusters had similar locations. However, after adjustment for age, the space-time clusters location shifted and population attribution risk (%) values changed (between 0.02 and 0.4).

Conclusion: Population attribution risk (%) value differences and clusters' temporal and spatial variations before and after adjustments can represent disease interventions impact. Additional studies should be conducted to strengthen the registering and reporting system to determine other influencing factors.

[♦]Corresponding Author:
Meysam Olfatifar, MSc
Social Determinants of Health Research Center, Kurdistan University of Medical Sciences, Sanandaj, Iran
Tel: (+) 98 933 312 2006.
Email: ol.meysam92@gmail.com

Keywords: Breast neoplasm, Kurdistan Province, Spatial analysis

Received: January 3, 2017; Accepted: May 24, 2017

Introduction

Chronic non-communicable diseases, especially cancer, along with the control of communicable diseases have been become a worldwide public health concern.[1] According to WHO estimates, cancer will increase to approximately 70% in the next two decades.[2] On the other hand, breast cancer has contributed to 521,000 cases out of 8.2 million cancer-attributed deaths in 2012.[2] In Iran, the breast cancer incidence and mortality age-standardized rate (ASR) is estimated to range from 24.1 to 33.9 and less than 10.1 per 100,000 people respectively in 2012.[3]

Health levels vary between different geographical regions. It is necessary to specify the regions' differences and determine areas that have an accumulation of health problems for epidemiological research and public health decisions.[4] Modern technologies, such as geographic information systems (GIS) and scan statistics, enable not only disease mapping but also spatial analysis such as spatial clustering and cluster detection in epidemiological research.[5-8]

Knowledge about disease status is essential in each society to implement cancer prevention and control programs.[9] Spatial methods can adequately cope with this matter and access characteristics of spatiotemporal clusters to give important information that pertains to disease distribution, transmission patterns, disease risk factors, and prevention program evaluation.[10, 11] Various studies have assessed the spatial and spatiotemporal changes of cancer.[12-14] However, a study to detect significant clusters of breast cancer has not been performed in Kurdistan Province, Iran. Thus, we conducted this study to detect purely spatial and spatiotemporal significant clusters of breast cancer in this province.

Materials and Methods

We obtained breast cancer data from the Health Deputy at Kurdistan University of Medical Sciences from 2005 to 2014. We removed any

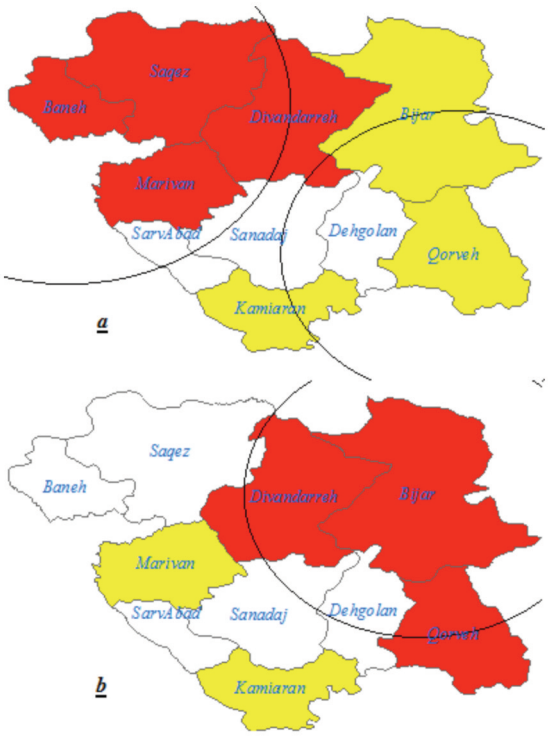

Figure 1. A) Purely spatial cluster before and after adjustment for age and space-time before adjustment for age. B) Space-time cluster after adjustment for age (red: Primary clusters; yellow: Secondary clusters; white: No clusters).

duplicate data and excluded male patients from this study. Spatial scan statistics have been implemented with SaTScan software version 9.4 software was used to detect purely spatial and spatiotemporal clusters. In this method, the software imposes a search window into the study area map which, in turn, is placed on the coordinate center obtained from the GIS for each block. The likelihood ratio (LR) is calculated according to the following formula based on Poisson distribution (used in this study):

$$\left(\frac{c}{EI(c)}\right)^{c}\left(\frac{C-c}{C-E(c)}\right)^{C-c}$$

Where C is the total number of cases, c is the observed number of cases within the window, and E[c] is the covariate adjusted expected number of cases within the window under the null-hypothesis. Note that since the analysis is conditioned on the total number of cases observed, C-E[c] is the expected number of cases outside the window. The software has compared the calculated LR derived from the observed data with its amount calculated based on Monte Carlo simulation, and tested the null hypothesis.

Table 1. Distribution of breast cancer cases in Kurdistan Province by city.

City	Cases	Age (years)	
	(N)	Mean	SD
Saqez	117	46.68	13
Bijar	38	46.84	10.9
Divandarreh	29	47	13.88
Baneh	51	49.47	13.61
Marivan	101	46.73	12.91
Qorveh	86	50.38	11.82
Sanandaj	538	50.05	13.3
Kamiaran	39	46.94	13.03
Total	999	48.99	13.09

As noted, we used discrete Poisson distribution to detect spatial and spatiotemporal clusters. In each case, the analysis was run once by taking into consideration 50% of the population at risk. We repeated the analysis by taking into consideration 25% of the population at risk to detect smaller clusters. Also, the analysis was repeated for

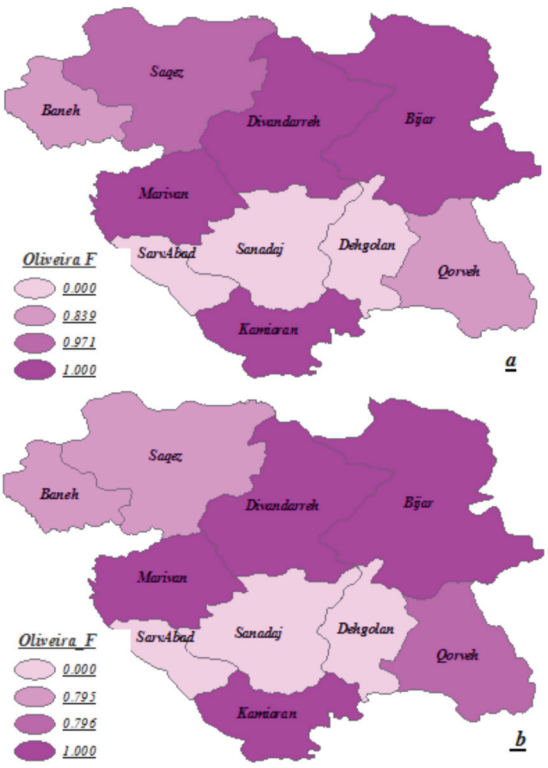

Figure 2. Color map of Oliveir F measure. A) Before adjustment for age. B) After adjustment for age.

Table 2. Significant purely spatial clusters with discrete Poisson distribution clusters for higher rates of breast cancer in Kurdistan Province with a maximum search window size of ≤50% and ≤25% of the risk population, adjusted and unadjusted for age.

Cluster ID	Center coordinates	Radius (km)	Cases (N)	Expa	RRb	LLRc	PAR%
Adjusted							
Primary cluster							
1	35.996N-45.848E	101.65	298	165.88	2.14	53.532	1.37
Secondary clusters							
1	35.229N-47.826E	81.11	124	62.42	2.13	25.598	1.48
2	34.955N-46.914E	0	39	13.65	2.93	15.93	3.69
Unadjusted							
Primary cluster							
1	35.996N-45.848E	101.65	298	162.02	2.2	57.311	1.45
Secondary clusters							
1	35.229N-47.826E	81.11	124	61.8	2.15	26.26	1.51
2	34.955N-46.914E	0	39	13.37	3	16.45	3.8

a: Expected number of cases in clusters; b: Relative risk of clusters; c: Log likelihood ratio; PAR: Population attribution risk

unadjusted and adjusted statuses to estimate the effect of age on the detected clusters.

Similar to other cluster detecting methods, the scan statistic can detect more than one cluster in a given study area. It has difficulty in prioritizing the clusters, especially if the distinction is between the statistical and public health significance. Detected clusters are ranked by their LR values and can neglect regions that have moderately high relative risk (RR) but a large population size. On the other hand, the statistical test is not performed for RR. Significant clusters may or may not have a high RR. Unlike RR and despite its importance from the epidemiologic viewpoint, the proper solution is to use population attributable risk (PAR%) because it takes into consideration both the strength of the relationship and the size of the population at risk. In this study, we used the following equation to calculate PAR%:

$$PAR\% = \frac{P_E(\mathrm{RR}_E - 1)}{P_E(\mathrm{RR}_E - 1) + 1} * 100$$

Where P_E is the proportion of the total number of the infected population and RR is the proportion of disease related to exposure (obtained from SaTScan software results). On the other hand, another limitation of scan statistics is that the exact boundary of the detected cluster will remain uncertain. Hence, some locations within the detected clusters may not have an elevated risk. This is due to the limited sample size in these regions. This situation will occur when the detected cluster is not circular and a circular scan window has been used. However, even if the circular scan window has not been used, this situation may occur and an increase in sample size to obtain a better estimate is not always possible. Therefore, to deal with this situation, the Oliveira's F measure implemented in SaTScan software version 9.4 has been used. This measure is only available for purely spatial analysis with Poisson distribution. This measure will be calculated for each unit (city) based on input data and is a value between 0 and 1. When the unit is closer to 1, the data are more likely to be part of the true cluster. Hence, in this study, we have used the results of the Oliveira's F measure to create a color coded map.

Results

In this study, after excluding the duplicate and male patient cases, we included 999 (0.95) out of 1050 cases. The city of Sanandaj had the highest number of cases at 538(0.53). Qorveh had the highest mean age of 50.38 years, followed by Sanandaj (50.05 years; Table 1). We observed no significant differences in terms of age among the cities (ANOVA; F=2; *P*=0.0529).

Spatial analysis

Purely spatial clusters

In general, we detected one primary and two

Table 3. Significant space-time clusters with discrete Poisson distribution clusters for higher rates of breast cancer in Kurdistan Province with a maximum search window size of ≤50% and ≤25% of the risk population, adjusted and unadjusted for age.

Cluster ID	Center coordinates	Radius (km)	Cases (N)	Expa	RRb	LLRc	PAR%
Adjusted							
Primary cluster							
1	35.944N-47.646E	81.11	75	28.78	2.74	26.737	1.23
Secondary clusters							
1	35.572N-46.354E	0	64	34.53	1.91	10.476	0.71
2	34.955N-46.914E	0	21	6.95	3.07	9.279	2.13
Unadjusted							
Primary cluster							
1	35.996N-45.848E	101.65	169	92.61	1.99	28.579	0.67
Secondary clusters							
1	35.2293N-47.826E	81.11	58	23.83	2.52	18.026	0.93
2	34.955N-46.914E	0	21	6.89	3.09	9.389	2.15

a: Expected number of cases in clusters; b: Relative risk of clusters; c: Log likelihood ratio; PAR: Population attribution risk

secondary clusters in the study areas. The primary cluster with the center in Saqez and 101.56 km radius comprised the cities of Saqez, Baneh, Divandarreh, and Marivan (Table 1, Figure 1a). We noted two secondary clusters in Qorveh and Bijar and Kamiaran. Table 1 and Figure 1a show the characteristics of these clusters. As shown in Table 1, the second of the secondary clusters had the highest PAR% (3.8%) among all of the purely spatial and space-time clusters. Clusters detected with regards to 50% and 25% at risk population were in the same geographical areas. On the other hand, without adjustments for age, the locations of the clusters did not change. Instead, the number of observed and expected cases, LR, and cluster PAR% changed such that the PAR% values changed between 0.1 to 0.2 compared to the corresponding adjusted clusters (Table 1, Figure 1a). Oliveira-F values for Bijar, Divandarreh, Marivan, and Kamiaran before and after the adjustment had values near to 1 (Figure 2).

Space-time clusters

In this case, we detected a primary cluster and two secondary clusters in the study areas. The primary cluster consisted of Bijar, Divandarreh, and Qorveh cities from 2010 to 2014 (Table 3, Figure 1b). Each of the two secondary clusters consisted of Marivan and Kamiaran during 2009 to 2013 and 2006 to 2010 (Table 3, Figure 1b). Before and after adjustment the second secondary clusters have had the highest PAR% at 2.15 and 2.13, respectively. The detected clusters for 50% and 25% at risk population were in the same geographical areas. Without taking into consideration adjustments for age, we observed that the cluster locations were similar to the purely spatial clusters. The PAR% values changed from 0.02 to 0.4 along with the changes in cluster characteristics. However, the time period of the first secondary cluster shifted from 2009-2013 to 2010-2014 (Table 3, Figure 1a).

Discussion

The results of current study revealed the presence of some significant clusters in Kurdistan Province. We have observed that a secondary purely spatial cluster located in Kamiaran had the highest PAR% (3.8%) among all of the detected clusters (Tables 2, 3). A cluster with a high PAR% may reflect the relative similarity of affecting factors on the occurrence of breast cancer in cluster areas and their differences with other areas. Thus, it may be possible to prevent a certain percentage of breast cancers according to PAR% values by simulating the regions in terms of these affecting factors. We have observed that the primary unadjusted space-time cluster composed of Saqez, Baneh, Divandarreh, and Marivan had the lowest PAR% (0.67) among all of the detected clusters (Table 3). Hence, it could be argued that the occurrence of cancer could be prevented by

simulating the cluster areas to each other with regards to PAR% values.

Changes to the in space-time cluster location after adjustment for age can represent the necessity of considering the other influencing factors. Furthermore, the minimal changes in detected cluster PAR% before and after considering adjustments for age can reflect that a small percentage of breast cancer changes in Kurdistan Province have been explained by the age variable. This statement may be verified by the non-significant age differences (ANOVA). However, observed minimal changes in PAR% values and shifting the cluster locations can be attributed to possible changes in the age at disease diagnosis or other influencing factors that considered in this study. Another key result of the current study was that the primary space-time cluster PAR% values after adjustment for age were greater than PAR% values before adjustment. This could be interpreted that after successfully addressing adjustments for the effect of age, we were left with a core of cases in which other affecting factors played major roles. The Olivier-F measure values for significant purely spatial suggested the true high possibility of the cluster units. We compared the current study with other studies that combined the two approaches of scan statistics and PAR%. Yuhui et al.[13] and Yiannakoulias[14] reported on the performance and benefits of this combination. On the other hand, we compared the current study with the studies that have only applied the scan statistic, among which Kavousi et al.[15] conducted a study to detect the high rates of clusters of stomach cancer. Space-time scan statistics detected some significant clusters in the Northern, Northwestern, and Central provinces of the country. For comparison with other abroad studies can also cite to Nordsborg et al.[16] that in early analysis a cluster was detected in northern margin of Copenhagen, however can cite to other conducted studies.[17, 18]

Limitations of the current study included the lack of access to other influencing factors to better explain the detected clusters. Because in various studies[19, 20] have shown that adjustments for influencing factors or covariates, led to changes in the clusters' temporal and spatial characteristics, which supported the current study findings. Another limitation was probably the incorrect registration of breast cancer cases for Sarvabad and Dehgolan, so that during the study period, these cities did not register any cases. A probable explanation can be the recent separation of these areas into two separate cities in Kurdistan Province. Another limitation of this study was the lack of access to breast cancer grade information from 2010 to 2014 to better understand the epidemiology of breast cancer. Hence, additional epidemiological studies would be needed to better understand the epidemiology of breast cancer in Kurdistan Province.

Strengths of this study included the use of Olivier-F to measure PAR% to better distinguish the clusters and their locations.

Finally, according to accuracy in reporting the results and various implications of PAR%, the current study could be a framework to design more appropriate studies and evaluate applied public health interventions in the detected clusters areas with high PAR%. From a practical standpoint, PAR% is more reliable than RR.

Conclusion

The current study detected some high rate significant clusters. The PAR% value differences and cluster temporal/spatial variation can be a marker for disease intervention. Additional epidemiological studies are necessary to strengthen the registering and reporting system, as well as achieve other influencing factors.

Acknowledgements

This work was supported by the Social Determinants of Health Research Center (SDHRC) at Kurdistan University of Medical Sciences.

Funding

The study was funded by the Vice-chancellor for Research and Technology, Kurdistan University of Medical Sciences

(IR.MUK.REC.1395/300).

Conflict of interest

No conflict of interest is declared.

References

1. Taheri NS, Nosrat SB, Aarabi M, Tabiei MN, Kashani E, Rajaei S, et al. Epidemiological pattern of breast cancer in Iranian women: is there an ethnic disparity? *Asian Pac J Cancer Prev*. 2012;13(9):4517-20.
2. World Health Organization. Cancer. [Cited at: December 2016]. Available from: http://www.who.int/mediacentre/factsheets/fs297/en/.
3. International Agency for Research on Cancer. Cancer Fact Sheets 2016 [Cited on: February 6, 2016]. Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
4. Fukuda Y, Umezaki M, Nakamura K, Takano T. Variations in societal characteristics of spatial disease clusters: examples of colon, lung and breast cancer in Japan. *Int J Health Geogr*. 2005;4:16.
5. Green C, Hoppa RD, Young TK, Blanchard J. Geographic analysis of diabetes prevalence in an urban area. *Soc Sci Med.* 2003;57(3):551-60.
6. Jemal A, Kulldorff M, Devesa SS, Hayes RB, Fraumeni JF. A geographic analysis of prostate cancer mortality in the United States, 1970–89. *Int J Cancer*. 2002;101(2):168-74.
7. Kulldorff M, Feuer EJ, Miller BA, Freedma LS. Breast cancer clusters in the northeast United States: a geographic analysis. *Am J Epidemiol.* 1997;146(2):161-70.
8. Rosenberg MS, Sokal RR, Oden NL, DiGiovanni D. Spatial autocorrelation of cancer in Western Europe. *Eur J Epidemiol.* 1999;15(1):15-22.
9. Zahedi A, Rafiemanesh H, Enayatrad M, Ghoncheh M, Salehiniya H. Incidence, trends and epidemiology of cancers in north west of Iran. *Asian Pac J Cancer Prev*. 2015;16(6):7189-93.
10. Randremanana RV, Sabatier P, Rakotomanana F, Randriamanantena A, Richard V. Spatial clustering of pulmonary tuberculosis and impact of the care factors in Antananarivo City. *Trop Med Int Health*. 2009;14(4):429-37.
11. Roza DLd, Caccia-Bava MdCG, Martinez EZ. Spatio-temporal patterns of tuberculosis incidence in Ribeirão Preto, State of São Paulo, southeast Brazil, and their relationship with social vulnerability: a Bayesian analysis. *Rev Soc Bras Med Trop.* 2012;45(5):607-15.
12 Jafari-Koshki T, Schmid VJ, Mahaki B. Trends of breast cancer incidence in Iran during 2004-2008: A Bayesian space-time model. *Asian Pac J Cancer Prev*. 2014;15(4):1557-61.
13. Zhu Y, Xu Q, Lin H, Yue D, Song L, Wang C, et al. Spatiotemporal analysis of infant measles using population attributable risk in Shandong Province, 1999-2008. *PloS One.* 2013;8(11):e79334.
14. Yiannakoulias N. Using population attributable risk to understand geographic disease clusters. *Health Place.* 2009;15(4):1142-8.
15. Kavousi A, Bashiri Y, Mehrabi Y, Etemad K, Teymourpour A. Identifying high-risk clusters of gastric cancer incidence in Iran, 2004-2009. *Asian Pac J Cancer Prev.* 2014;15(15):10335-7.
16. Nordsborg RB, Meliker JR, Ersbøll AK, Jacquez GM, Poulsen AH, Raaschou-Nielsen O. Space-time clusters of breast cancer using residential histories: a Danish case-control study. *BMC Cancer.* 2014;14(1):255.
17. Vieira V, Webster T, Weinberg J, Aschengrau A, Ozonoff D. Spatial analysis of lung, colorectal, and breast cancer on Cape Cod: an application of generalized additive models to case-control data. *Environ Health.* 2005;4(1):1.
18 Vieira VM, Webster TF, Weinberg JM, Aschengrau A. Spatial-temporal analysis of breast cancer in upper Cape Cod, Massachusetts. *Int J Health Geogr.* 2008;7:46.
19. Bambhroliya AB, Burau KD, Sexton K. Spatial analysis of county-level breast cancer mortality in Texas. *J Environ Public Health.* 2012;2012:959343.
20. Sheehan TJ, DeChello LM. A space-time analysis of the proportion of late stage breast cancer in Massachusetts, 1988 to 1997. *Int J Health Geogr*. 2005;4:15.