Determination of effective factors on geographic distribution of the incidence of colorectal cancer in Tehran using geographically weighted Poisson regression model

Kamyar Mansori[1,2], Masoud Solaymani-Dodaran[3], Alireza Mosavi-Jarrahi[4], Ali Ganbary Motlagh[5], Masoud Salehi[6], Alireza Delavari[7], Ali Hosseini[8], Mohsen Asadi-Lari*[2,9]

Abstract

Background: This study aimed to determine effective factors on geographic distribution of the Incidence of Colorectal Cancer (CRC) in Tehran, Iran using Geographically Weighted Poisson Regression Model.

Methods: This ecological study was carried out at neighborhood level of Tehran in 2017-2018. Data for CRC incidence was extracted from the population-based cancer registry data of Iran. The socioeconomic variables, risk factors and health costs were extracted from the Urban HEART Study in Tehran. Geographically weighted Poisson regression model was used for determination of the association between these variables with CRC incidence. GWR 4, Stata 14 and ArcGIS 10.3 software systems were used for statistical analysis.

Results: The total number of incident CRC cases were 2815 in Tehran from 2008 to 2011, of whom, 2491 cases were successfully geocoded to the neighborhood. The median IRR for local variables were : unemployed people over 15 year old (median IRR: 1.17), women aged 17 years or older with university education (median IRR: 1.17), women head of household (median IRR: 1.06), people without insurance coverage (median IRR: 1.10), households without daily consumption of milk (median IRR: 0.85), smoking households (median IRR: 1.07), household's health expenditure (median IRR: 1.39), disease diagnosis costs (median IRR: 1.03), medicines costs of households (median IRR: 1.05), cost of the hospital (median IRR: 1.09), cost of medical visits (median IRR: 1.27).

Conclusion: The spatial variability was observed for most socioeconomic variables, risk factors and health costs that had effects on CRC incidence in Tehran. Spatial variability is necessary when interpreting the results and utterly helpful for implementation of prevention programs.

Keywords: Spatial epidemiology, Colorectal cancer, Socioeconomic, Risk factors, Health expenditures

Introduction

Colorectal cancer (CRC) is the third most common can-cer and the fourth leading cause of cancer-related deaths

Corresponding author: Dr Mohsen Asadi-Lari, mohsen.asadi@yahoo.com

1. Social Determinants of Health Research Center, Semnan University of Medical Sciences, Semnan, Iran
2. Department of Epidemiology, School of Public Health, Iran University of Medical Sciences, Tehran, Iran
3. Minimally Invasive Surgery Research Center, Rasoul Akram Hospital, Iran University of Medical Science, Tehran, Iran
4. Department of Health and Community Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
5. Department of Radiotherapy, Shahid Baheshti University of Medical Sciences, Tehran Iran
6. Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran
7. Digestive Disease Research Institute, Tehran University of Medical Sciences, Tehran, Iran
8. Department of Geography and Urban Planning, University of Tehran, Tehran, Iran
9. Oncopathology Research Centre, Iran University of Medical Sciences, Tehran, Iran

---

↑*What is "already known" in this topic:*
Spatial analysis at small geographical levels can provide useful information about the areas at risk. There have been few studies conducted to carefully consider the impact of various socioeconomic factors, health costs, and risk factors on CRC at the level of small geographical units, such as the neighborhoods of a city.

→*What this article adds:*
The spatial variability was observed for most socioeconomic variables, risk factors and health costs that had effects on CRC incidence in Tehran. Spatial variability is necessary when interpreting the results and utterly helpful for implementation of prevention programs.

in the world. According to a report by the International Agency for Research on Cancer (IARC) in 2012, nearly 1.4 million new cases of CRC were found whereas 700,000 deaths worldwide occurred due to this cancer (1). According to IARC, CRC in Iranian males is the fourth common cancer with an age-standardized incidence rate (ASIR) of 8.7 and an age-standardized mortality rate (ASMR) of 6.3 per 100,000 people. CRC is also the third common cancer among Iranian females with ASIR and ASMR of, 6.4 and 4.6 per 100,000 individuals; respectively (2). In general, CRC has a different geographic distribution, and its incidence varies in different parts of the world, where the highest incidence rates belong to North America, Australia, New Zealand, Western Europe, and Japan, indicating that 63% of the cases occur in developed countries (3, 4). Comparing to developed countries, CRC is lower in Iran; however, its rate in the younger generation of the country is on the rise, which can dramatically increase the burden of disease in the future. On the other hand, there is an uneven geographical distribution of CRC in Iran, where large cities such as Tehran have always been among the cities with the highest incidences of CRC (5-7).

Spatial analysis at small geographical levels can provide useful information about the areas at risk. The spatial analysis of cancer size at geographical levels, such as postal codes and census blocks in developed countries, has been well documented (8-14). This issue has been overlooked in developing countries like Iran where there have been few studies conducted to carefully consider the impact of various socioeconomic factors, health costs, and risk factors on CRC at the level of small geographical units, such as the neighborhoods of a city. These few studies have often been limited to describing the spatial pattern of cancers at the provincial level (15-19).

On the other hand, the incidence of many cancers, especially CRC, varies depending on the geographical situation of the area and has its own spatial pattern (20-24). Usually, Poisson regression is employed as the model for studying the correlation between variables at the level of a geographical unit with an outcome that is count. However, these variables usually have spatial autocorrelation, while the outcomes like cancer rarely occur in some geographical levels which results in problems like small area estimation; therefore, the use of the Poisson regression model, in these cases, leads to over-dispersion. Therefore, the predicted result would not be consistent with the reality of the observations or the diffused dispersion in the model cannot describe the observed dispersion of the variable in reality (25). Hence, in recent years, a new, yet efficient and effective way of studying various spatial relationships has been developed which is called the Geographically Weighted Regression (GWR) (26). Given the above description and considering that no study yet conducted on the simultaneous impacts of various socioeconomic factors, health costs and risk factors on the incidence of CRC in the level of neighborhood in Iran with the use of the Geographically Weighted Poisson Regression (GWPR) model, the present study was designed and conducted to determine the effective factors on geographic distribution of the incidence of colorectal cancer (CRC) in Tehran, Iran by geographically weighted Poisson regression model.

Methods
Study area: The present research is an ecological study that was designed and implemented in the Tehran city (capital of Iran) at 2017-2018. The Tehran metropolitan area 638 square kilometer is situated on the southern slopes of the Alborz Mountains at a latitude of 35°45′N and a longitude of 51° 25′E. This city consists of 22 municipal districts and also the geographical unit of the study was 374 neighborhoods in Tehran city.

Required data for the study
Demographic information: In the present study, demographic data of people aged 50 and older, who were considered as the at-risk population for CRC were extracted from national censuses carried out in 2006 and 2011 and broken down by different parts of Tehran. Then, the population of people aged 50 and older for each neighborhood was calculated as follows:

$$\frac{population\ of\ aged\ 50\ and\ older\ in\ 2006\ census\ +\ population\ of\ aged\ 50\ and\ older\ in\ 2011\ census}{2}$$

Equation 1

Information of CRC incidence: In this study, the data on the incidence of CRC was extracted from the population-based cancer registry data of the Iranian Ministry of Health and Medical Education from 2008 to 2011. Then, according to the postal address of the patients, CRC cases were categorized in 374 neighborhoods of Tehran, and the number of cases of CRC in each neighborhood was determined.

Socio-economic information, risk factors and health costs: In this research, we extracted socio-economic variables, risk factors and health costs from the Equity Assessment Study in Tehran. The details of this study are fully described in the study by Asadi et al. (27). In the present study, the following variables were extracted from the Equity Assessment Study in Tehran by the division of neighborhoods:

1. Socio-economic variables including the percentages of unemployed people over 15 years old, women aged 17 years or older with a university education, women head of household, households without a car, households living in rental houses, households with income below the poverty line and people without insurance coverage.

2. Risk factors including the percentages of households that do not consume fruits daily, households that do not consume milk daily, overweight people aged 15 and older, and smoking households.

3. Health costs variables including the percentages of the cost of household health, cost of the diagnosis of the disease, the cost of household medicine, the cost of the hospital and the cost of medical visits.

Statistical analysis
Geographically Weighted Regression: Since spatial data

is presented in two forms of spatial correlation and non-stationary spatial, it is not possible to investigate such data through ordinary regression methods. Because these traditional methods can only estimate a modest amount of general parameters and are not able to show spatial correlations between variables and their spatial variation in the geographical area, they may lead to biased results. Therefore, a new and effective way has been developed to explore various spatial relationships which is known as Geographically Weighted Regression (GWR) (26). The GWR model, when estimating regression coefficients, permits the spatial dependence of reference location data and unlike global spatial regression, provides regression coefficients locally and in each spatial coordinate, coefficients are not constant in space but vary from one position to another (28, 29).

The general form of the GWR model is shown below. As mentioned above, unlike ordinary regression, the coefficients obtained in GWR in each spatial coordinate (u_i, v_i) have their own special amount. In this model, β_k, x_{ki}, y_i and ε_i are the regression coefficient, the kth independent variable , the dependent variable and the remainder of the model, respectively; they are in the spatial coordinates of i $(u_i, v_{i)}$ (26).

$$y_i(u_i, v_i) = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ki} + \varepsilon_i \quad Equation\ 2$$

Before using the GWR4 model, to identify the most important factors affecting the outcome, the Univariate Poisson Regression Model runs and variables with P-value<0.2 are considered as potential confounders or cofactors to be included in all subsequent analyses. Then all of these variables are investigated by the variance inflation factor (VIF) index for multi-collinearity before entrance to the GWPR model. The VIF = 1/ (1-R2), and generally, if its value is greater than 4, it indicates the strong multi-collinearity (32).

Calibration of GWR4: In ordinary regression models, estimation of coefficients is performed in such a way that the sum of the differences between observed and predicted values for ys would be minimal, but in the GWR model, there is also a weight parameter; so that, in estimation of regression coefficients, in each spatial position less weighs is given to the predicted variables with greater distances. Therefore, the estimated coefficients will be as follows:
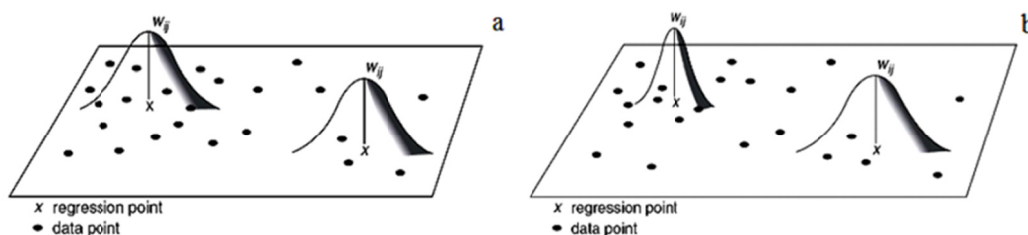
$$\hat{\beta} = (X^TWX)^{-1}X^TWy \quad \text{Equation 3}$$

In the general spatial regression model, the weight matrix elements (w_{ij}s) are constant so that only one calibrate or grading is used to estimate the coefficients; but in the GWR model, the W-matrix changes with the change of spatial position of the observations, and thus for different situations, the calibrate model will also be different, as follows :

$$\hat{\beta}(i) = (X^TW(i)X)^{-1}X^TW(i)y \quad \text{Equation 4}$$

In the GWR model, considering the points of regression on the vector and the approach of spatial proximity, a weighted function is considered for each observation. This can be achieved by determining the bandwidth so that a distance controller threshold from neighboring observations does not significantly affect the local estimation of the parameters. The weighted function of the model is calculated from the two parameters of the distance between the regression points and the data points and the bandwidth. Usually, to calibrate the GWR, a continuous weighted function is used as follows:

$$W_{ij} = exp\left(-\frac{d_{ij}^2}{b^2}\right) \quad \text{Equation 5}$$

In the above formula, W_{ij} is the weight assigned to the observation at position of j when estimating the coefficient of regression point of i . The d_{ij} is the distance between positions i and j and b is the bandwidth of the weighted function core.

Kernel Weighting Function: Using the two methods of Fixed Kernel Spatial and the Adaptive Spatial Kernel, we can calculate the weighted function from the bandwidth. A problem that could occur in the application of the GWR model with the fixed kernel model is that some of the regression points in the data may have a dispersed spatial pattern. The localized regression model with the fixed kernel may be calibrated with a small number of data; therefore, the highest parametric estimate and maximum standard error for the outputs are taken into account, and the surface results are expressed with minimal data smoothing (Fig. 1).

To solve such problems in the GWR model, size variables can be considered with respect to the data density in an adaptive manner, so that such a model is able to allocate more bandwidth for scattered data and less bandwidth

Fig. 1. (a) GWR with Fixed Kernel model, (b) GWR with Adaptive Kernel model

for cumulative frequency data. Therefore, as shown in Figure 1 in Adaptive Kernel, the length of the band can be larger or smaller, depending on the size of the variables in the data density.

If the data are separated, the bandwidth is larger and, if the density is higher, the bandwidth is smaller. Therefore, the Adaptive Kernel model can provide more accurate and realistic estimates and results than the Fixed Kernel model.

In the Adaptive Kernel Model, when the number of bandwidths is high in space, bandwidth is less used. The weight of the jth data point at the ith point of the regression is calculated by the following equation:

$$w_{ij} = \left[1 - \left(\frac{d_{ij}}{h_i}\right)^2\right]^2 \quad when \; d_{ij} \leq h_i 0 \; ;$$

$when \; d_{ij} > h_i$ Equation 6

At the regression point ($d_{ij} = 0$), when the distance is equal to the length of the band ($d_{ij} = h$) the data point weighs zero.

The GWR results are relatively sensitive to the weighted function selection, and the weighted function selection is also sensitive to the selection of the weighted function of the bandwidth, so that the optimal determination of the bandwidth in the GWR model seems necessary. Therefore, bandwidth in the GWR model creates the data-smoothing function. Therefore, according to the above explanation, in this study the Adaptive Kernel Method was used to develop the model and AICc (Akaike Information Criterion) was used to select optimal bandwidth (29, 31). In this study, due to the fact that the dependent variable is the incident cases of CRC in Tehran's neighborhoods and it is count variable , the GWPR model

should be employed. Of course, the above explanation refers to GWR model which can be generalized to a variety of weighted regression models. In the present study, GWR4, Stata 14 and ArcGIS 10.3 softwares were used to analyze the data.

Results

The total number of incident CRC cases were 2815 in the Tehran city from 2008 to 2011, and of them, 2491 cases (88.4%) were successfully geocoded to the neighborhood. Of these, 56.6% were male (1593 cases) and 43.4% (1222 cases) were women and the mean and standard deviation of age for incidence of colorectal cancer were 62.15 and 10.49 years, respectively. The number of CRC cases ranged from 0 to 57 across neighborhoods in Tehran, which prevailed more in northern and central areas of the city (Fig. 2).

Table 1 shows the mean and standard deviations of the studied variables (socioeconomic variables, health costs, and risk factors). To identify the most important factors affecting CRC incidence, the Univariate Poisson Regression Model was performed. All variables had a p<0.2. Then all of these variables were investigated by the variance inflation factor (VIF) index for multi-collinearity before entrance to the GWPR model. The results of the multi-collinearity test showed that the VIF index for all variables is below 4, so there was no multi-collinearity (Table 1).

To select the final variables for the GWPR model, all the variables were first entered simultaneously into the GWPR model, which was named as Global Model 1, then in the next step only variables that had a significant association (p<0.05) with the CRC incidence in Global Model 1 were entered; this model was also called Global Model 2. In the global model 1 with the presence of all variables, the value of AICc was 2268.16 and in the global model 2 with the presence of only significant variables, AICc was
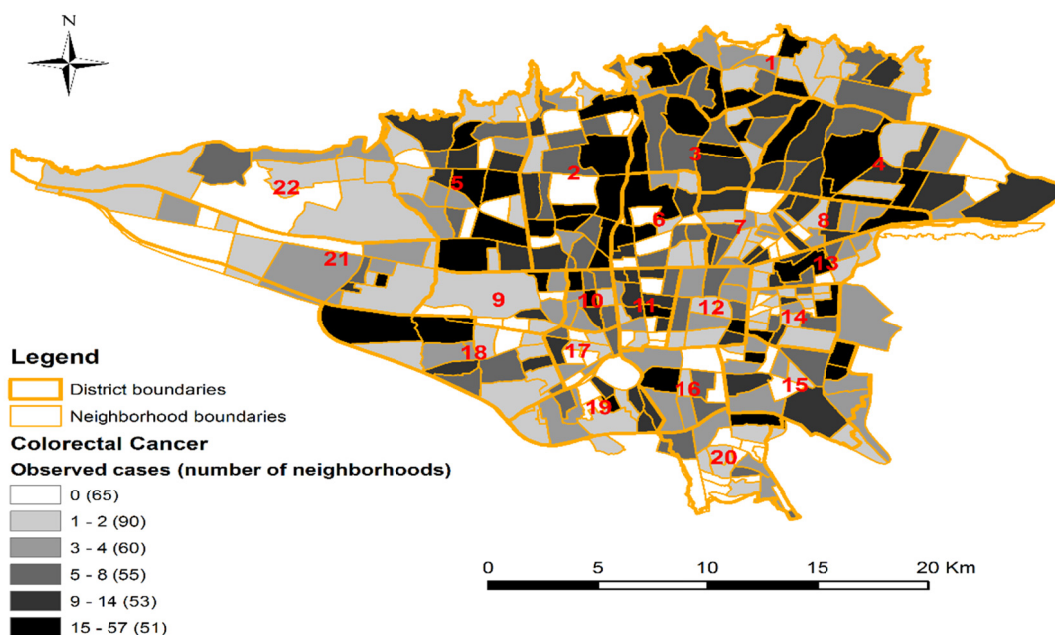

Fig. 2. The number of observed cases of colorectal cancer in the neighborhoods of Tehran (2008-2011)

Table 1. Mean, standard deviation (S.D) and multi-collinearity test for variables of socioeconomic, risk factors and health costs

Variables	Mean (S.D)	VIF	Tolerance
People over 15 years old unemployed (%)	9.09 (1.94)	1.93	0.518
Women aged 17 years or older with university education (%)	30.49 (13.17)	2.46	0.406
Women head of household (%)	11.13 (3.53)	1.91	0.522
Households without car (%)	28.96 (7.64)	2.09	0.478
Households living in rental houses (%)	39.34 (15.29)	1.89	0.528
Households with income below the poverty line (%)	19.28 (14.96)	1.60	0.632
People without Insurance coverage (%)	28.44 (12.62)	1.32	0.754
Households without fruit consumption daily (%)	13.05 (7.36)	1.28	0.777
Households without milk consumption daily (%)	54.94 (13)	1.31	0.761
Overweight people aged 15 and older (%)	32.32 (5.13)	1.03	0.963
Smoking households (%)	24.11 (6.20)	1.16	0.861
Cost of household health (%)	12.68 (6.94)	1.80	0.544
Cost of the diagnosis of the disease (%)	17.91 (9.70)	1.44	0.691
Cost of household medicine (%)	45.86 (14.03)	1.98	0.504
Cost of the hospital (%)	9.67 (9.85)	1.40	0.714
Cost of medical visits (%)	18.17 (8.35)	1.34	0.741

2269.17; the difference was less than 2, so both models could be selected as optimal models. But we chose the global model 1 as the superior model because it involves all the variables under study and also it has a smaller AICc (in generally, AICc is used to select the superior model. The model with a smaller AICc and difference of more than 2 units, is a superior model. But if the difference between AICc of two model is less than and equal to 2, two models are not superior to each other).

Then, for all independent variables geographic variability test was performed to determine which variables have spatial variability, because they should be introduced to the GWRP model in the final analysis. The results of this test showed that the percentages of households without a car, households living in rental houses, households with income below the poverty line, overweight people aged 15 and older, and households who do not consume fruits on a daily basis, do not have significant spatial variability. After performing geographic variability test and determining the position of independent variables (the variables without spatial variability were placed in the Global box and variables with spatial variability in the Local box in GWPR model), the final analysis was performed. Table 2 shows the Global Model, which is the same as the Traditional Poisson model. In this model, it is assumed that all independent variables are spatially homogeneous. As can

be seen, in this model, the percentages of women head of household (IRR: 1.14), households without car (IRR: 0.91), households living in rental houses (IRR: 0.90), households without daily fruit consumption (IRR: 0.92), households without daily milk consumption (IRR: 0.90), overweight people aged 15 and older (IRR: 0.90), smoking households (IRR: 1.08), household health expenditures (IRR: 1.22), the cost of household medicines (IRR: 1.08), hospital admission costs (IRR: 1.10), and the cost of medical visits (IRR: 1.14) had statistically significant association with incidence CRC.

Table 3 also shows the final results of the GWPR model for global and local variables. This table contains the main results of the model. As can be seen, despite the absence of a significant spatial variability among global variables, the percentages of households without car (IRR: 0.89), households living in rental houses (IRR: 0.82), households with income below the poverty line (IRR: 1.10), overweight people aged 15 and older (IRR: 0.87) had statistically significant association with incidence CRC. Among these significant variables, households with income below the poverty line had a direct association with CRC incidence, so that every 10% increase in this variable resulted in a 10% increase in IRR of CRC, but other significant variables had an inverse association with CRC incidence. It also shows Mean, Median, Min, Max and IQR for local

Table 2. The association between socioeconomic variables, risk factors and health costs with incidence CRC with global model

Variables	β- coefficient	*IRR	p
Intercept	-6.561	0.001	<0.001
People over 15 years old unemployed	0.029	1.03	0.260
Women aged 17 years or older with university education	0.049	1.05	0.135
Women head of household	0.128	1.14	<0.001
Households without car	-0.090	0.91	0.033
Households living in rental houses	-0.103	0.90	0.001
Households with income below the poverty line	-0.026	0.97	0.330
People without Insurance coverage	-0.025	0.97	0.219
Households without fruit consumption daily	-0.076	0.92	0.002
Households without milk consumption daily	-0.103	0.90	<0.001
Overweight people aged 15 and older	-0.103	0.90	<0.001
Smoking households	0.081	1.08	0.0016
Cost of household health	0.196	1.22	<0.001
Cost of the diagnosis of the disease	0.007	1.007	0.781
Cost of household medicine	0.079	1.08	0.008
Cost of the hospital	0.096	1.10	<0.001
Cost of medical visits	0.134	1.14	<0.001

*IRR: Incidence Risk Ratio

Table 3. The association between socioeconomic variables, risk factors and health costs with incidence CRC with GWPR model

Variables Global	β- coefficient		IRR		p	
Households without car	0.89		-0.115		0.042	
Households living in rental houses	0.82		-0.193		<0.001	
Households with income below the poverty line	1.10		0.096		0.003	
Households without fruit consumption daily	0.95		-0.049		0.155	
Overweight people aged 15 and older	0.87		-0.135		<0.001	
Variables Global	Mean	SD	Median	IQR	Min	Max
Intercept	0.0013	1.20	0.0012	1.35	0.0009	0.0019
People over 15 year old unemployed (A*)	1.13	1.31	1.17	1.39	0.59	2.02
Women aged 17 years or older with university education (B*)	1.15	1.16	1.17	1.18	0.83	1.59
Women head of household (C*)	1.17	1.24	1.06	1.37	0.90	1.59
People without Insurance coverage (D*)	1.06	1.28	1.10	1.35	0.51	1.63
Households without milk consumption daily (E*)	0.84	1.19	0.85	1.32	0.62	1.17
Smoking households (F*)	1.12	1.18	1.07	1.37	0.90	1.58
Cost of household health (G*)	1.32	1.18	1.39	1.11	0.70	1.62
Cost of the diagnosis of the disease (H*)	1.04	1.27	1.03	1.28	0.54	1.50
Cost of household medicine (I*)	1.04	1.34	1.05	1.45	0.46	1.60
Cost of the hospital (J*)	1.10	1.25	1.09	1.26	0.69	1.80
Cost of medical visits (L*)	1.09	1.34	1.27	1.50	0.72	1.96

*The capital letter is related to the maps of these variables in Figure 2.
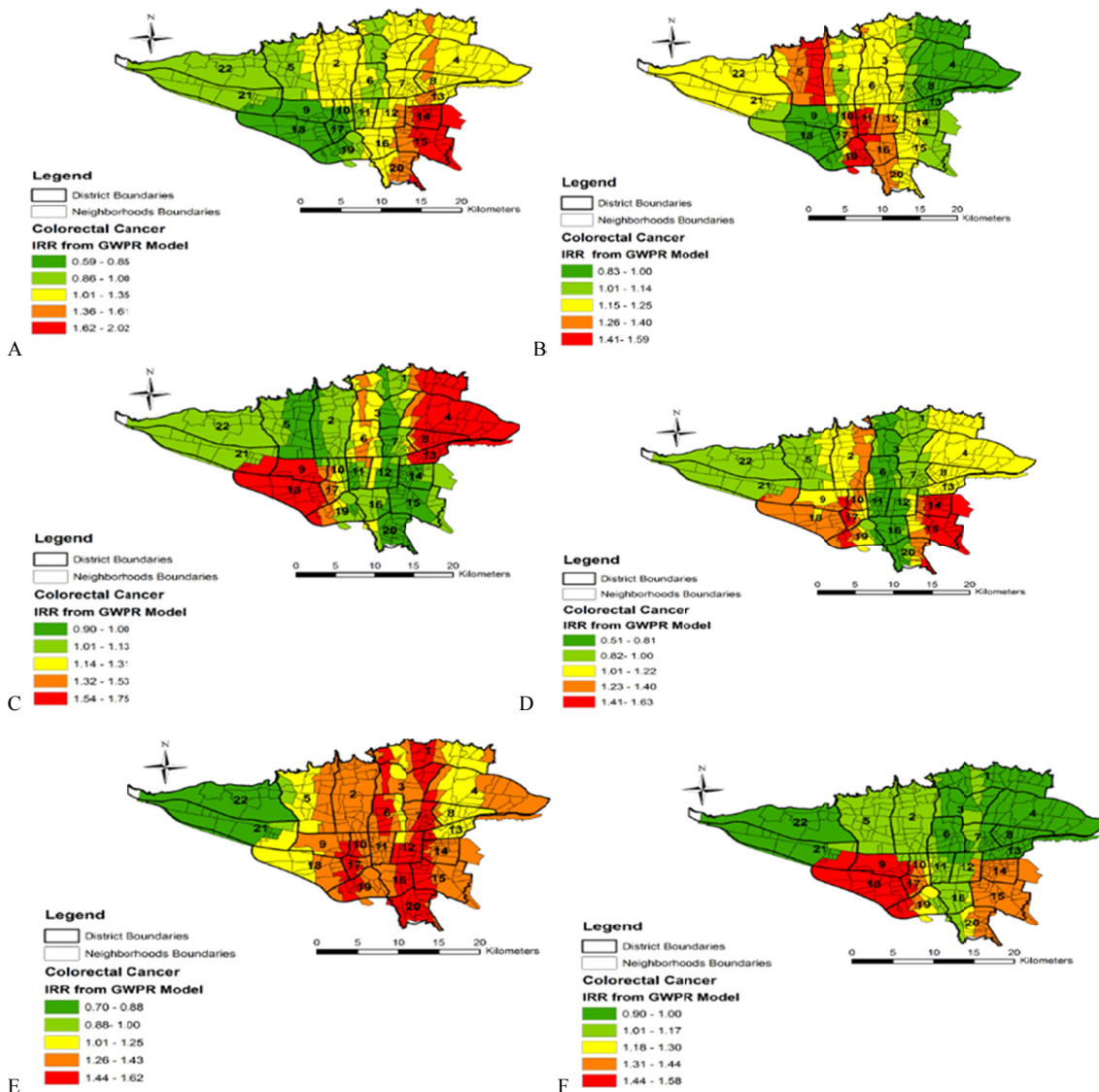

Fig. 3. Spatial Variation of IRR for local variables from GWPR model

variables with significant spatial variability with CRC incidence. Figure 3 also shows the spatial variation of IRR for local variables in the level neighborhoods of Tehran city from GWPR model.

Table 4 shows the performance of both GWPR and Global Models. The GWPR model has a significantly smaller AICc and also a larger percent deviance explained (It is equivalent R2 in regression that shows the explana-
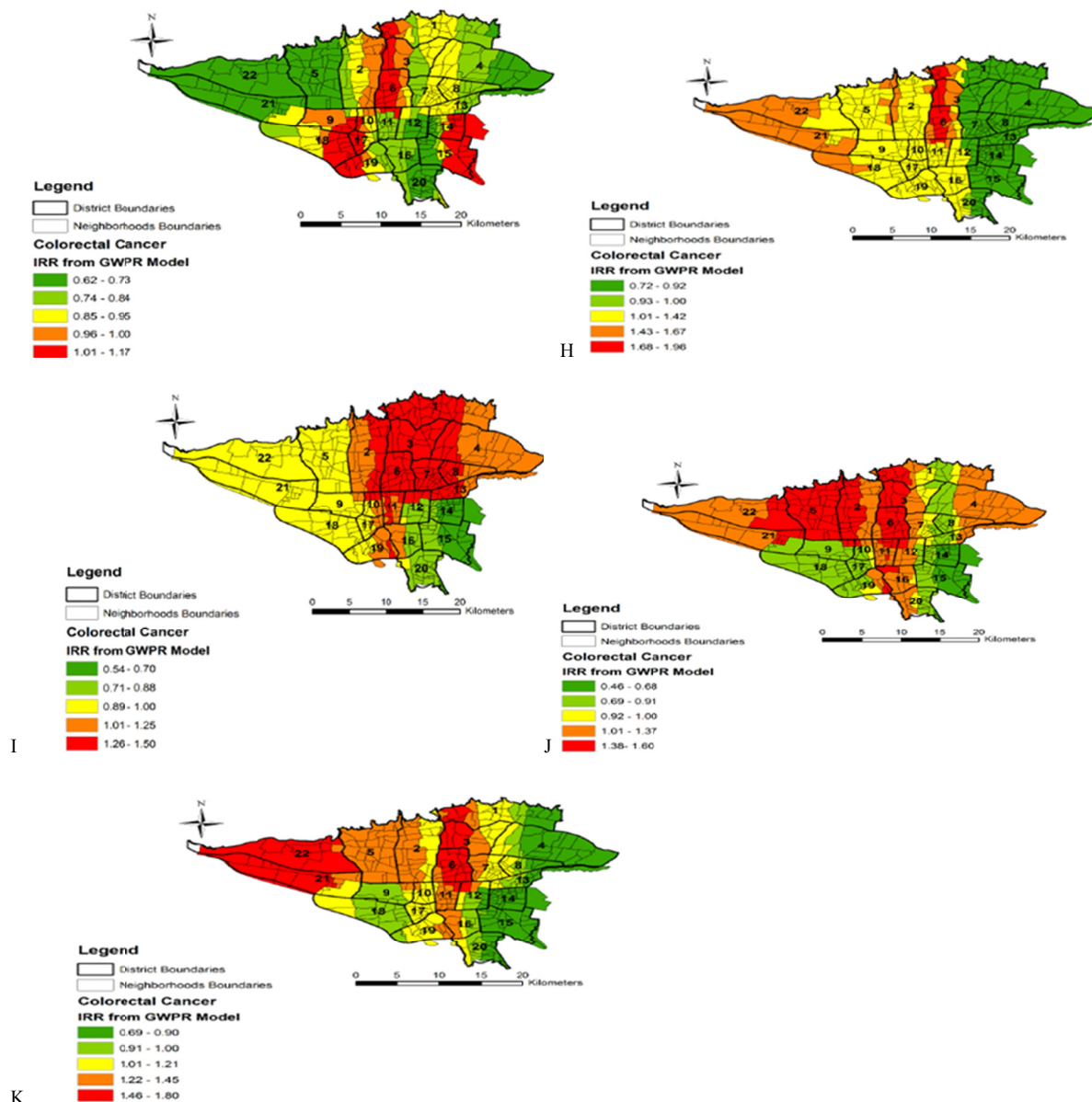
Fig. 3. Continued

tion of variation of dependent variable by independent variables); therefore, in compared to Global Model, this is best model. In this study, the best bandwidth size is 76 with a minimum AICc of 1744.511. Figure 4 shows spatial variation of R-square for CRC incidence from GWPR model. As can be seen, the highest local R-square is related to neighborhoods that located in districts 10, 11, 12, 16, 17, 19, 20, 21, and 22.

Discussion

This study showed that most incidence cases of CRC occurred among the residents of neighborhoods in the northern and central districts of Tehran. This may be due to the fact that the residents of these districts in Tehran are mostly wealthy and highly educated; therefore, they pay more attention to their health status and undergo more on screening and diagnostic programs; on the other hand, residents of these districts have more access to screening, diagnostic and therapeutic services, so more cases of cancer are diagnosed among them (34). The same is true for the direct association of women with 17 year of age or older who have university education (Median IRR: 1.17) with incident CRC. Because various studies have shown that having high educational levels is an important predic-

Table 4. Comparison of the performance of the Global Model and GWPR Model in the estimation of the effect of independent variables on the incidence of CRC

Model	AICc	Percent deviance explained
Global model	2268.16	0.1236
GWPR model	1744.51	0.4251
Difference	523.65	0.3015

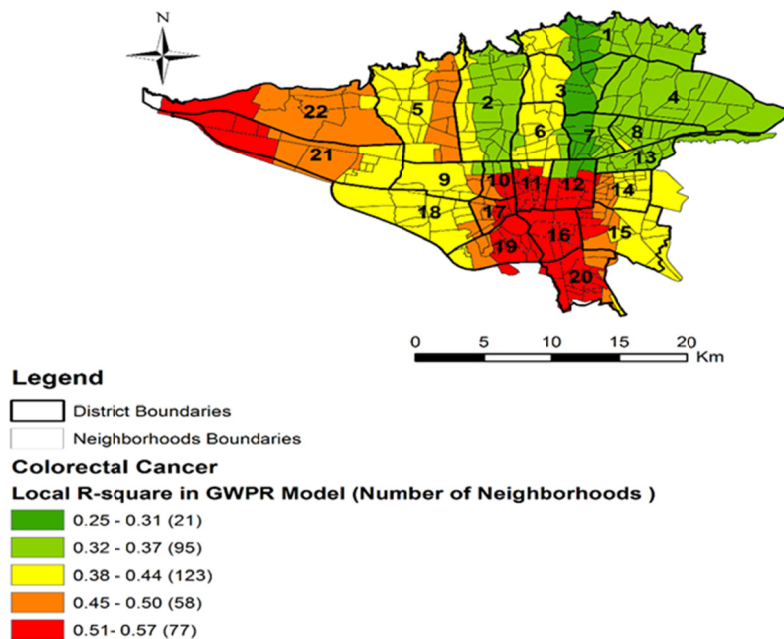The best bandwidth size : 76 with Minimum AICc : 1744.51

Fig. 4. Spatial variation of R-square for incidence of CRC from GWPR mode

tor for participating in colonoscopy screening programs. Increase in the level of education has a direct association with the frequency of conducting colonoscopy and increasing the frequency of participation in screening programs is also associated with a higher incidence and diagnosis of CRC among the participants (35-37).

There was a direct association between the variables of household health expenditures (Median IRR: 1.39), the cost of the diagnosis of the disease (Median IRR: 1.03), the cost of household medicine (Median IRR: 1.05), hospital admission costs (Median IRR: 1.09), and the cost of medical visits (Median IRR: 1.27) with the incidence of CRC, which can be attributed to the fact that these people had utilized health services more, and had spent more as well. Of course, it should be noted that in these patients, the cancer is diagnosed in the initial phase and has a better prognosis (38, 39).

In the present study, the variables of households with lower income (below the poverty line) (IRR: 1.10), the people over 15 year old unemployed (Median IRR: 1.17), women-headed households (Median IRR: 1.06), and people without insurance coverage (Median IRR: 1.10), as the main components of socio-economic status, had a direct statistical associations with CRC incidence. These results were not consistent with some previous studies in this field. For example, a study conducted in Tehran showed that there is no significant association between socioeconomic variables such as employment status, insurance coverage, and income level with the participation of people in colonoscopy screening programs (40). Another study performed by Karen. L. Barclay et al., with the aim of examining the effect of socio-demographic characteristics on the CRC stage in Australia, did not show any significant association between employment status and insurance coverage with CRC stage (41). However, in another research carried out by Susanne Singer in Germany to study the impact of socio-economic status on the cancer stage revealed that unemployment (OR: 1.7, CI: 1.01-2.8), disability pension (OR: 1.8, CI: 1.02 -3.2) and low income (OR: 2.6, CI: 1.1 -6.1) had a significant statistical association with an advanced disease (42).

To better explain the results, two points should be taken into account: first, these variables are directly related to the degree of attention of individuals to their health and participation in screening and diagnostic programs.; so that, those who are less well-off in terms of these variables, are less likely to seek diagnostic and screening programs, and cancers diagnosed are often in advanced stages and do not have good prognoses. Secondly, these variables are closely related to the lifestyle and diet of individuals, and people with inadequate status of these factors usually do not have a healthy lifestyle and diet, resulting in a higher incidence of gastrointestinal cancers, especially CRC. Therefore, these variables should be considered in the design and implementation of screening and diagnostic programs, and mere provision of service without consideration of these factors may not have an appropriate outcome.

In this study, the variable of smoking households (Median IRR: 1.07) had a direct association with CRC incidence, which is consistent with the results of studies conducted in this regard because various meta-analytic studies have indicated that cigarette smoking is associated with an increase in the incidence and mortality of CRC (43, 44). This is also a logical consequence of the fact that cigarettes are a risk factor not only for cancers, but also for many other illnesses. In this study, overweight people with 15 years of age and older (IRR: 0.87) were inversely associated with the incidence of CRC which was not consistent with the studies carried out in this area (45, 46). Obviously, there is no consensus on overweight as a risk factor for CRC. For example, in a meta-analytic study by

Shuangiie Wu, the results showed that obesity (HR: 1.9; CI 95%: 1.06-1.15) was significantly associated with an increase in the risk of CRC; whereas, overweight (HR: 0.92; CI 95%: 0.86 -1) did not show any significant association with it (47). For the variable of households without daily milk consumption, there was an inverse association with CRC which was not consistent with the studies conducted in this area (48, 49). For instance, in a prospective study performed in Europe to study the association between dairy consumption and CRC, after 11 years of follow up, a significant inverse association was observed between milk consumption and the risk of CRC (HR per 200 g / day 0.93, 95% CI: 0.89-0.98) (48). In fact, milk exerts its protective effect through calcium because milk is a rich source of calcium. In this study, there is an inverse relationship between overweight in people aged 15 years and older households without daily milk consumption. Perhaps the most important reason for this is the ecological nature of our study, which is associated with ecological fallacy. The studied unit here is neighborhood, not the individual, therefore, the relationship seen at the level of the neighborhood between the exposure and the outcome, may not be observed at the level of individual (50, 51). Another point that should be mentioned about households without daily milk consumption (Median IRR: 0.85) is that this is a median number, while its IRR range varies from 0.62 to 1.17. So in some neighborhoods located in the central regions, the south-east and west IRR has been reported to 1.17, which is in line with other studies (48, 49).

A number of strengths and weaknesses were there in our study. As was pointed out, the main drawback of this study is its ecological fallacy, which makes it impossible to be certain about the results, only presenting clues and hypotheses. The other limitation was unclear addresses of 15% of the cases with CRC which made it impossible to allocate them into the neighborhoods to be entered into the analysis. Also, In this study, a population of 50 years and over in the neighborhoods of Tehran has been considered as a population at risk for incidence of CRC, while some of the covariates including the overweight people aged 15 and older and unemployed people over 15 years old that were extracted from the Equity Assessment study of Tehran, did not consistent with the age of the population at risk for incidence of CRC, that this could be another limitations of the study. The last limitation of this study, which should be mentioned, is the edge effect phenomenon. This means that the findings for neighborhoods near administrative borders should be interpreted with caution, because it is possible that different indicators outside the understudy neighborhood affect the characteristics of the border residents of neighborhoods. In terms of the strengths, we can say that it is the first ecological research in Iran which simultaneously has been examined the association between various socioeconomic factors, health costs and risk factors with the incidence of CRC in neighborhood level with the GWPR approach in Iran. It also compared the results with the Global (ordinary Poisson) model. If this spatial analysis was conducted based on the geographical unit of the district of Tehran, it could

not show the actual at-risk population, because some low incidence districts may have neighborhoods with a high incidence of cancer and vice versa.

Conclusion

This study showed that there is spatial variability for most socioeconomic variables, risk factors and costs of health that had an effect on the incidence CRC in Tehran city. Considering this spatial variability is necessary when interpreting the results and implementation of prevention programs.

Conflict of Interests

The authors have no conflicts of interest associated with the material presented in this paper.

References

1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1. 0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. 2013; Lyon, France: International Agency for Research on Cancer globocan iarc fr/Default aspx 2014.
2. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010;127:2893-917.
3. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. . Global cancer statistics CA. CA Cancer J Clin. 2011;61:69-90.
4. Etemad K, Gooya M. National report on registered cancer cases in 2009 Tehran, Iran: Cancer Office, Centre for Disease Control, Deputy for Health, Ministry of Health and Medical Education 2009.
5. 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;240:143-7.
6. Khazaei S, Rezaeian S, Khazaei S, Mansori K, Sanjari Moghaddam A, Ayubi E. Effects of human development index and its components on colorectal cancer incidence and mortality: a global ecological study. Asian Pacific journal of cancer prevention: Asian Pac J Cancer Prev. 2015;17:253-6.
7. Shadmani FK, Ayubi E, Khazaei S, Sani M, Hanis SM, Khazaei S, et al. Geographic distribution of the incidence of colorectal cancer in Iran: a population-based study. Epidemiol Health. 2017;39.
8. Bambhroliya AB, Burau KD, Sexton K. Spatial analysis of county-level breast cancer mortality in Texas. J Environ Public Health. 2012;2012:959343.
9. Dai D. Black residential segregation, disparities in spatial access to health care facilities, and late-stage breast cancer diagnosis in metropolitan Detroit. Health Place. 2010;16:1038-52.
10. Henry KA, Niu X, Boscoe FP. Geographic disparities in colorectal cancer survival . Int J Health Geogr.2009;8:48.
11. Wan N, Zhan FB, Lu Y, Tiefenbacher JP. Access to healthcare and

disparities in colorectal cancer survival in Texas. Health Place.2012;18:321-9.
12. Khazaei S, Rezaeian S, Ayubi E, Gholamaliee B, Pishkuhi MA, Mansori K, et al. Global prostate cancer incidence and mortality rates according to the human development index. Asian Pac J Cancer Prev. 2016;17(8):3791-4.
13. Khodadost M, Yavari P, Babaei M, Mosavi-Jarrahi A, Sarvi F, Mansori K, et al. Estimating the completeness of gastric cancer registration in Ardabil/Iran by a capture-recapture method using population-based cancer registry data. Asian Pac J Cancer Prev. 2015;16(5):1981-6.
14. Khazaei S, Mansori K, Soheylizad M, Gholamaliee B, Shadmani FK, Khazaei Z, et al. Epidemiology of lung cancer in Iran: sex difference and geographical distribution. MEJC. 2017 Sep 16;8(4):223-8.
15. Chamanparaaa P, Moghimbeigi A, Faradmal J, Poorolajal J. Exploring the spatial patterns of three prevalent cancer latent risk factors in Iran; Using a shared component model. IJER. 2015;2:68-77.
16. Mahaki B, Mehrabi Y, Kavousi A, Akbari ME, Waldhoer T, Schmid VJ, et al. Multivariate disease mapping of seven prevalent cancers in Iran using a shared component model. Asian Pac J Cancer Prev. 2011;12:2353-8.
17. Goli S, Oroei M, Jalalpour M, Faramarzi H, Askarian M. The Spatial Distribution of Cancer Incidence in Fars Province: A GIS-Based Analysis of Cancer Registry Data. Int J Prev Med. 2013;4:1122-30.
18. 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;15:1557-61.
19. Mansori K, Solaymani-Dodaran M, Mosavi-Jarrahi A, Motlagh AG, Salehi M, Delavari A, et al. Spatial Inequalities in the Incidence of Colorectal Cancer and Associated Factors in the Neighborhoods of Tehran, Iran: Bayesian Spatial Models.J Prev Med Public Health. 2018 Jan;51(1):33.
20. Moradi Y, Jafari M, C ichian S, Khateri S, Akbarian A, Moazzami B, et al.. Trends in ovarian cancer incidence in Iran. Iran J Cancer Prev. 2016;9(6).
21. Jafari M, Moradi Y, Khodadost M, Sekhavati E, Anabad HA, Mansori K, et al. Trend of the esophageal cancer incidence in Iran. Int J Travel Med Glob Health. 2015;3(2):127-31.
22. 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;15:10335-7.
23. Mansori K, Mosavi-Jarrahi A, Motlagh AG, Solaymani-Dodaran M, Salehi M, Delavari A, et al. Exploring Spatial Patterns of Colorectal Cancer in Tehran City, Iran. Asian Pac J Cancer Prev. 2018;19(4):1099.
24. DeChello LM, Sheehan TJ .Spatial analysis of colorectal cancer incidence and proportion of late-stage in Massachusetts residents: 1995–1998. Int J Health Geogr. 2007;6:20.
25. Griffith DA, Haining R. Beyond mule kicks: The Poisson distribution in geographical analysis. Geographical Analysis. 2006 Apr;38(2):123-39.
26. Gao J, Li S .Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using geographically weighted regression. Appl Geogr. 2011;31:292-302.
27. Asadi-Lari M, Vaez-Mahdavi MR, Faghihzadeh S, Cherghian B, Esteghamati A, Farshad AA, et al. Response-oriented measuring inequalities in Tehran: second round of UrbanHealth Equity Assessment and Response Tool (Urban HEART-2), concepts and framework. Med J Islam Repub Iran .2013;27:236.
28. Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical analysis. 1996;28:281-98.
29. Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons; 2003.
30. Saefuddin A, Saepudin D, Kusumaningrum D .Geographically Weighted Poisson Regression (GWPR) for Analyzing The Malnutrition Data in Java-Indonesia. 2013.
31. Tu J, Xia ZG. Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. Sci Total Environ. 2008;407:358-78.
32. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York: Springer; 2013 Feb 11.

33. O'brien RM A. Caution regarding rules of thumb for variance inflation factors Quality & Quantity. 2007;41:673-90.
34. Mandelblatt J, Andrews H, Kao R, Wallace R, Kerner J. The late-stage diagnosis of colorectal cancer: demographic and socioeconomic factors. Am J Public Health. 1996;86:1794-7.
35. Kilickap S, Arslan C, Rama D, Yalcin S. Screening colonoscopy participation in Turkish colorectal cancer patients and their first degree relatives. Asian Pac J Cancer Prev. 2012;13:2829-32.
36. Simmons RG, Lee YC, Stroup AM, Edwards SL, Rogers A, Johnson C, et al. Examining the challenges of family recruitment to behavioral intervention trials: factors associated with participation and enrollment in a multi-state colonoscopy intervention trial. Trials. 2013;14:116.
37. Jacobs LA. Health beliefs of first-degree relatives of individuals with colorectal cancer and participation in health maintenance visits: a population-based survey. Cancer Nurs. 2002;25:251-65.
38. Pignone M, Saha S, Hoerger T, Mandelblatt J. Cost-effectiveness analyses of colorectal cancer screeninga systematic review for the US Preventive Services Task Force. Ann Intern Med. 2002;137:96-104.
39. Maciosek MV, Solberg LI, Coffield AB, Edwards NM, Goodman MJ. Colorectal cancer screening: health impact and cost effectiveness. Am J Prev Med. 2006;31:80-9.
40. Chouhdari A, Yavari P, Pourhoseingholi MA, Sohrabi MR. Association Between Socioeconomic Status and Participation in Colonoscopy Screening Program in First Degree Relatives of Colorectal Cancer Patients. Iran J Cancer Prev. 2016;9.
41. Barclay KL, Goh PJ, Jackson TJ. Socio-economic disadvantage and demographics as factors in stage of colorectal cancer presentation and survival. ANZ J Surg. 2015;85:135-9.
42. Singer S, Roick J, Briest S, Stark S, Gockel I, Boehm A, et al. Impact of socio-economic position on cancer stage at presentation: Findings from a large hospital-based study in Germany. Int J Cancer. 2016;139:1696-702.
43. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. Jama. 2008;300:2765-78.
44. Liang PS, Chen TY, Giovannucci E. Cigarette smoking and colorectal cancer incidence and mortality: Systematic review and meta-analysis. Int J Cancer. 2009;124:2406-15.
45. Giacosa A, Franceschi S, La CV, Favero A, Andreatta R. Energy intake, overweight, physical exercise and colorectal cancer risk. Eur J Cancer Prev. 1999;8:S61.
46. Campbell PT, Jacobs ET, Ulrich CM, Figueiredo JC, Poynter JN, McLaughlin JR, et al. Case–control study of overweight, obesity, and colorectal cancer risk, overall and by tumor microsatellite instability status. J Natl Cancer Inst. 2010;102:391-400.
47. Wu S, Liu J, Wang X, Li M, Gan Y, Tang Y. Association of obesity and overweight with overall survival in colorectal cancer patients: a meta-analysis of 29 studies. Cancer Causes Control. 2014;25:1489-502.
48. Murphy N, Norat T, Ferrari P, Jenab M, Bueno-de-Mesquita B, Skeie G, et al. Consumption of dairy products and colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). PLoS One. 2013;8:e72715.
49. Keum N, Aune D, Greenwood DC, Ju W, Giovannucci EL. Calcium intake and colorectal cancer risk: Dose–response meta-analysis of prospective observational studies. Int J Cancer. 2014;135:1940-8.
50. Krieger N. The real ecological fallacy: epidemiology and global climate change. J Epidemiol Community Health. 2014:jech-2014-205027.
51. Sedgwick P .Understanding the ecological fallacy. BMJ. 2015;351.