Research Article

Prediction of Colorectal Cancer Incidence Rate in the Counties of Fars Province, Iran: An Application of Small Area Estimation

Reyhane Sefidkar [iD]¹, Amir Kavousi [iD]²,*, Mahmoud Torabi³, Seyyed Vahid Hosseini⁴ and Hojat Alaii¹

¹Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Tehran, Iran
²Workplace Health Promotion Research Center, Department of Epidemiology, Shahid Beheshti University of Medical Sciences, Tehran, Iran
³Department of Community Health Sciences, University of Manitoba, Winnipeg, Canada
⁴Colorectal Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

Corresponding author: Workplace Health Promotion Research Center, Department of Epidemiology, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Email: kavousi_am@yahoo.com

Abstract

Background: Colorectal cancer (CRC) is one of the main causes of mortality and morbidity worldwide. Socio-economic status is one of the most important related factors with CRC.

Objectives: In this study, we used the human development index (HDI) as one of the common measures of socio-economic status to predict the incidence rate of CRC in the counties of Fars Province in Iran.

Methods: In this ecological study, we used the medical records of 108 patients with CRC from Fars province, who referred to Shahid Faghihi Hospital in Shiraz from January 2011 to March 2013. Since sample sizes were not efficient in all the counties, we used the log-normal model within small area estimation framework to have a reliable prediction for the incidence rate in each county. As using related auxiliary variables is necessary in small area models, we considered the HDI of counties as an auxiliary variable.

Results: The findings showed that there was a significant direct relationship between HDI and CRC incidence rate. Furthermore, the highest predicted rates were observed in the northern and eastern parts of the province.

Conclusions: In order to compensate the deficiency of sample size in some of the counties, we used a small area model to predict the CRC incidence rate. The highest incidence rates mostly occurred in the counties with the highest HDI. It is observed that the counties with higher incidence rates are closer to more industrial provinces and the counties with lower incidence rates are closer to less industrial provinces. So, it seems that development disparities strongly affected the incidence rates.

Keywords: Colorectal Cancer, Human Development Index, Incidence Rate, Small Area Estimation

1. Background

Colorectal cancer (CRC) is a major public health problem worldwide. CRC is the third most commonly diagnosed cancer in males, following prostate and lung cancer, which includes 10.9% of new cancer cases, and it is the second most commonly diagnosed cancer in females after breast cancer, including 9.5% of new cancer cases. Totally, it is the second cause of cancer-related deaths (about 9.2% in 2018) (1, 2). The incidence rate of CRC was over 9% of all cancer incidence rates in 2012 with an estimated 1.4 million cases and 700,000 deaths worldwide (3, 4). A substantial global burden is imposed by CRC due to its complications, mortality, side effects of treatment, utilization of healthcare services, and medical costs (5). There are wide geographical variations in CRC incidence and mortality rates and socioeconomic status has a considerable proportion of these variations (4). A common index to measure social

and economic achievement of a country is the human development index (HDI) which combines three important indicators: life expectancy, education, and per capita income (6). Various studies have revealed that the incidence and mortality of CRC have a significant increase in locations (e.g., countries) with medium and high HDI so that in 2012, 55% of CRC cases were related to the highly developed countries. Comparing age-standardized incidence rates, the highest rates have occurred in Australia and New Zealand, and the lowest have occurred in the countries of West Africa (1, 3, 7). Although CRC is more prevalent in developed countries, its incidence rate has an increasing trend in developing regions such as Western Asia (Kuwait) and Eastern Europe (Czech Republic, Slovakia, and Slovenia) (4). According to the literature, the increase in cancer incidence rate has been paralleled by the increase in the HDI in Iran as a developing country (8-11).

2. Objectives

Since CRC is one of the most prevalent cancers and determining the hot spots due to the high incidence rate of cancer helps to reduce the burden of the disease, the aim of this study was to predict the rate of CRC in counties of Fars Province, Iran. Regarding this goal, we used the information of patients from different counties of Fars Province who referred to the Shahid Faghihi Hospital in Shiraz from January 2011 to March 2013. Since there were counties with few or no CRC cases, we used small area estimation techniques to predict the CRC rates. Small area estimation has played a prominent role in survey sampling as it deals with the problem of obtaining reliable estimates for parameters of interest for subpopulations with small or even zero sample sizes (12). Furthermore, as using good auxiliary information is pivotal to improve prediction in small area estimation framework, we considered HDI as a predictor variable in our small area model.

3. Methods

3.1. Data Source

In this ecological study, the incidence of diagnosed patients with CRC in each county of Fars province who were admitted to Shahid Faghihi Hospital in Shiraz, the capital of Fars Province, from January 2011 to March 2013 was regarded as the main outcome. The data was based on the follow-up study which was conducted in the Colorectal Research Center of Shiraz from January 2006. Fars Province had 29 counties in 2011, and out of 108 patients with CRC who referred to Shahid Faghihi Hospital during the study period, the minimum observed sample size was 0 (11 counties) and the maximum sample size was 44 patients who belonged to Shiraz, the capital of Fars Province.

In addition, since using a good auxiliary variable is crucial to improve the model to have a better prediction in small area estimation framework, we used HDI as one of the most important and relevant explanatory variables to predict cancer incidence rate. In this study, we used the HDI which was calculated by Rezaee et al. in 2012 for 27 counties of Fars province. To obtain this index, 22 variables and indexes were used. The value of HDI is between zero and one which expresses the rate of deprivation so that zero shows the most deprivation and 1 indicates the least deprivation (13). The distribution of HDI in counties of Fars province is presented in Figure 1. We also used the population size of counties of Fars provinces based on the 2011 Census population conducted in Iran to adjust its effect in the model.

3.2. Statistical Analysis

In order to estimate the incidence rate of CRC, we used a hierarchical Bayes (HB) approach for our log-normal model which can be expressed in the following steps:

i: $y_i | \theta_i \overset{ind}{\sim} poisson(\theta_i) = 1, \ldots, m,$

ii: $log(\theta_i) = \beta_0 + \beta_1 HDI_i + log(N_i) + v_i$
where $v_i \sim N(0, \sigma_v^2)$,
iii: β and σ_v^2 are mutually independent with $f(\beta) \propto 1$ and $\sigma_v^{-2} \sim G(0.001, 0.001)$,

where m is the number of sampled areas (counties), $\beta = (\beta_0, \beta_1)$ is the vector of regression coefficients, HDI_i is the predictor variable, and N_i is the population size of i-th area, and $v = (v_1, \ldots, v_m)$ is the area-level random effects. We also considered non-informative prior distributions for regression coefficients and variance components in step iii (14). Furthermore, for domains that the predictor variable was not available, we predicted the rate with the direct estimator of the CRC at the provincial level and the counties with zero sample sizes were merged with similar counties. We also reported the credible interval as prediction interval of θ_i (I = 1, ..., m) as a criterion to explain the precision of prediction in each area. We applied the RStudio version 1.0.136 for data analysis.

The Ethic Committee of Shahid Beheshti University of Medical Sciences approved this study (with code number: IR.SBMU.RETECH.REC.1398.395)

4. Results

We studied the incidence rate of CRC in counties of Fars province from January 2011 to March 2013. In sum, 108 patients with CRC were distributed in 29 counties. Among them, 70 patients (64.8%) were men. 54 (50%) were diagnosed with colon cancer, 51 (47.2%) with rectal cancer, and 3 (2.8%) with both colon and rectal cancer. The mean age at the time of admission was 57.39 ± 16.06 with the range of 26 to 87 years. Out of 29 counties, 11 counties (Kharameh, Neyriz, ZarrinDasht, Estahban, Mohr, Gerash, Khorrambid, Farashband, Rostam, Sarvestan, and Pasargad) had no CRC cases. So, in order to get more accurate predictions, we decided to combine these counties (except Kharameh and Gerash) with their neighbors based on their development status. Furthermore, since the HDI was not available for two counties (Gerash and Kharameh), the CRC incidence rates in these two counties were estimated with the CRC rate direct estimator of the province.

Considering HDI as an auxiliary variable, we fitted the log-linear model to predict the rate of CRC in each county. To fit the model, we used the Markov chain Monte Carlo
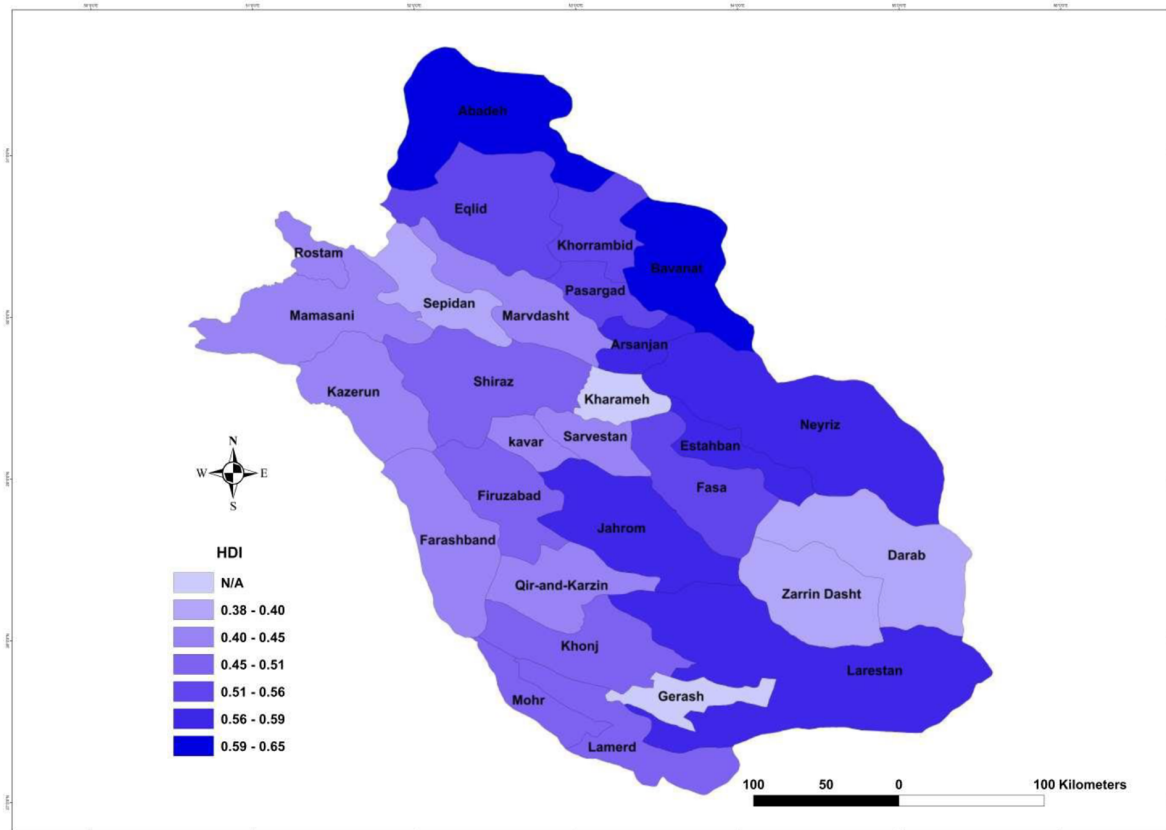
Figure 1. Human development index (HDI) in Fars Province

(MCMC) method. We kept every 5th sample in 80,000 iterations after discarding the first 5,000 iterations as burn-in. Also, in order to check the goodness of fit of the model, we computed the Gelman-Rubin convergence statistic which had acceptable value. Table 1 presents the estimators of the model parameters.

Table 1. Estimates, Standard Deviation, and 95% Credible Interval of the Model Parameters[a]

	Values	95% CI
Intercept	-12.36 ± 0.90	(-14.15, -10.54)
HDI	3.71 ± 1.80	(0.09, 7.24)
σ_v^2	0.04 ± 0.07	(6.15 × 10 - 4, 0.26)

[z]Abbreviation: HDI, human development index.
[a]Values are expressed as mean ± SD.

Regarding the findings in Table 1, the significant variance of random term shows heterogeneity in CRC incidence rate among counties. Also, we observed that HDI has positive and significant effect at 0.05 level so that by 0.1 unit increase in HDI, the rate of CRC incidence will increase by 44% ($e^{0.371} = 1.44$). Based on the results in Table 1, the predicted incidence rates and the related 95% prediction intervals for all the counties are presented in Figure 2 and Table 2, respectively. Based on Figure 2, small-area estimates on the county level ranged from 1.68 to 4.63 per 100,000 population. Khorrambid, Abadeh, Bavanat, and Jahrom are the four counties with the highest CRC incidence rates, respectively and Zarrin Dasht, Darab, Sepidan, and Ghirokarzin are the four counties with the lowest CRC incidence rates, respectively. It is worth mentioning that the counties with the highest rates are the ones with the highest HDI and those with the lowest incidence rates are attributed to the counties with the least HDI.

5. Discussion

In this study, we predicted the CRC incidence rate of counties of Fars province in Iran. As the sample size was

Table 2. Predicted CRC Incidence Rate (Per 100,000 Population) and Related %95 Prediction Interval of Counties of Fars Province

County	Incidence Rate	95% Prediction Intervals
Shiraz	2.68	(2.07, 3.33)
Marvdasht	2.23	(1.43, 3.35)
Kazerun	2.32	(1.49, 3.50)
Fasa	2.67	(1.48, 3.70)
Darab	1.68	(0.81, 2.74)
Jahrom	4.27	(2.72, 6.82)
Larestan	2.78	(1.69, 3.88)
Firuzabad	2.95	(1.95, 4.64)
Mamasani	2.20	(1.27, 3.37)
Abadeh	4.63	(2.13, 7.91)
Lamerd	2.82	(1.79, 5.38)
Eqlid	3.36	(2.01, 5.05)
Kavar	2.19	(1.22, 3.37)
Neyriz	3.60	(1.98, 5.60)
Zarrin Dasht	1.68	(0.81, 2.74)
Ghirokarzin	2.07	(1.14, 3.31)
Estahban	3.60	(1.98, 5.60)
Mohr	2.82	(1.79, 5.38)
Kharameh	2.21	(1.78, 2.64)
Gerash	2.21	(1.78, 2.64)
Sepidan	1.91	(1.03, 3.24)
Khorrambid	4.63	(2.13, 7.91)
Farashband	2.32	(1.49, 3.50)
Rostam	2.20	(1.27, 3.37)
Arsanjan	3.60	(1.98, 5.60)
Khonj	2.34	(1.36, 3.72)
Sarvestan	2.19	(1.22, 3.37)
Pasargad	3.36	(2.01, 5.05)
Bavanat	4.59	(2.45, 7.72)

insufficient in most of the counties, we used a small area model to accurately predict the CRC rate for each county. Furthermore, since using auxiliary variables that have a high correlation with the dependent variable is so necessary in small area models, we used HDI as a predictor variable which has indicated a significant relationship with different types of cancers incidence rates, including CRC, in the literature (8, 9, 11, 15). Our findings showed a significant relationship between CRC incidence rate and HDI, so that, the higher CRC incidence rates occurred in counties with higher HDI. This positive relationship has also been confirmed in various studies. For instance, Arnold et al.'s study (cited in Wong et al. (5)) showed that the

age-standardized rate of incidence in countries with a very high HDI is six times higher than that in countries with low HDI. This positive trend between increasing HDI of counties and rates of CRC incidence and mortality is revealed in other studies too (8, 9, 11, 15). It appears that counties with higher incidence rates of cancer also have higher HDI (8). In several studies, it is observed that the pattern of lifestyle in developed regions is affected by "westernization". Since there is a direct relationship between HDI and smoking, alcohol consumption, using red meat, and obesity, as the risk factors of CRC, it is obvious that increasing the CRC incidence rate is consequence of increasing HDI. In our study, mostly, the counties which were located in northern and
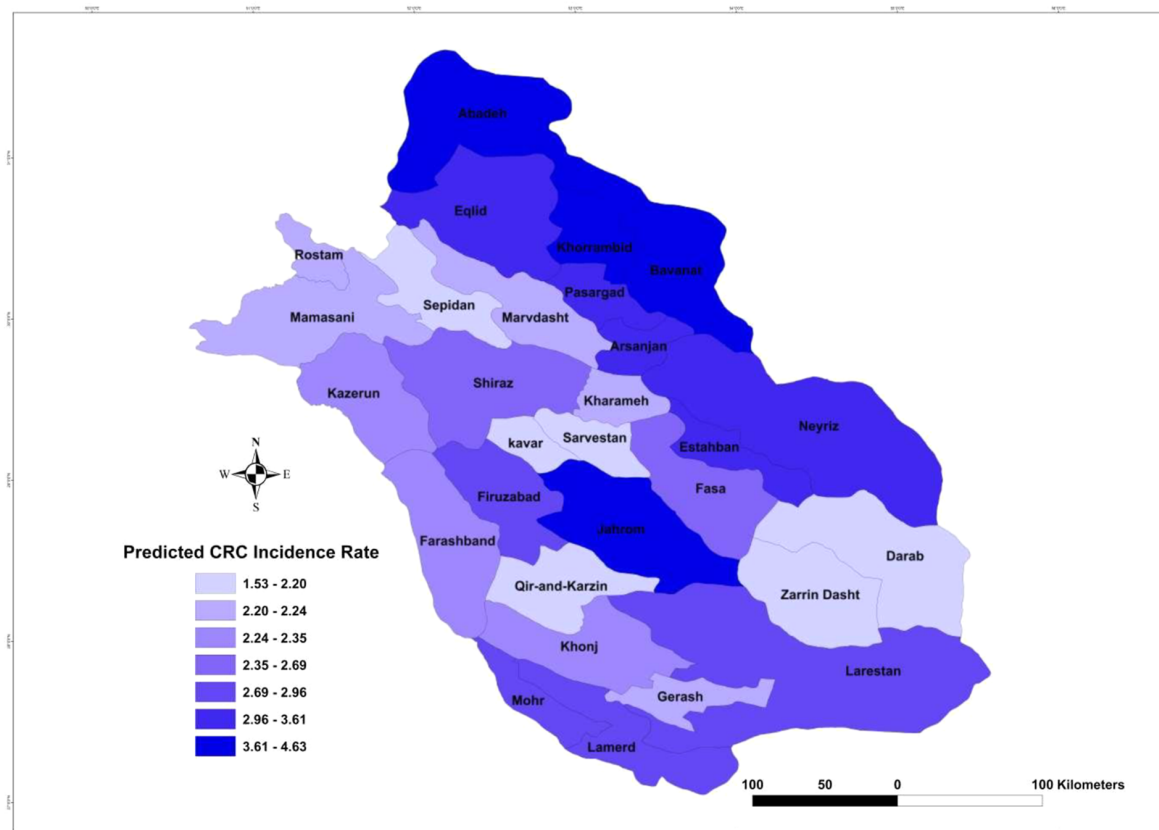
Figure 2. Predicted colorectal cancer (CRC) incidence rate of Fars Province

western parts of the province have a higher risk of CRC. Perhaps, we can relate the development of these counties to their proximity to industrial provinces such as Yazd, Esfahan, and Kerman and since HDI has a positive significant effect on cancer incidence rate, we observed high rates in these counties. In contrast, lower level of HDI was observed in eastern and southern parts of Fars Province due to their proximity to less industrial provinces like Hormozgan, Kohgiluyeh, Boyer-Ahmad, and Bushehr provinces, and as a result, lower incidence rates occurred in these regions (13).

5.1. Conclusions

In summary, the incidence rate of CRC is higher in more developed regions since there is a positive relation between HDI and CRC incidence rate, which reflects the effect of HDI on people's lifestyles. Generally, primary prevention in less developed regions and early detection in developed regions are needed to reduce the incidence rate of CRC.

Footnotes

Authors' Contribution: Reyhane Sefidkar did data analysis and drafting of the manuscript. Amir Kavousi did giving the idea and critical revision of the manuscript. Mahmoud Torabi did data analysis and critical revision of the manuscript. Seyyed Vahid Hosseini did the acquisition of data, Hojat Alaii prepared the maps.

Conflict of Interests: All authors declare that there is no conflict of interest.

Ethical Approval: Code of ethics was IR.SBMU.RETECH.REC.1398.395.

Funding/Support: There is no funding support.

References

1. Monet A, Touzani R, Bouhnik AD, Bendiane MK, Mancini J. Colorectal cancer screening practices among cancer survivors five years after diagnosis. *J Public Health*. 2020. doi: 10.1007/s10389-019-01179-w.

2. Florea A, Sangare L, Lowe K. A multinational assessment of gastric, esophageal, and colorectal cancer burden: A report of disease incidence, prevalence, and fatality. *J Gastrointest Cancer*. 2020;**51**(3):965–71. doi: 10.1007/s12029-019-00328-4. [PubMed: 31784876].

3. Maajani K, Khodadost M, Fattahi A, Shahrestanaki E, Pirouzi A, Khalili F, et al. Survival rate of colorectal cancer in Iran: A systematic review and meta-analysis. *Asian Pac J Cancer Prev*. 2019;**20**(1):13–21. doi: 10.31557/APJCP.2019.20.1.13. [PubMed: 30677864]. [PubMed Central: PMC6485573].

4. Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi RE, Corcione F. Worldwide burden of colorectal cancer: A review. *Updates Surg*. 2016;**68**(1):7–11. doi: 10.1007/s13304-016-0359-y. [PubMed: 27067591].

5. Wong MCS, Huang J, Lok V, Wang J, Fung F, Ding H, et al. Differences in incidence and mortality trends of colorectal cancer worldwide based on sex, age, and anatomic location. *Clin Gastroenterol Hepatol*. 2020. doi: 10.1016/j.cgh.2020.02.026. [PubMed: 32088300].

6. *Human development index*. 2020, [cited 2020 May 5]. Available from: https://en.wikipedia.org/wiki/Human_Development_Index.

7. Navarro M, Nicolas A, Ferrandez A, Lanas A. Colorectal cancer population screening programs worldwide in 2016: An update. *World J Gastroenterol*. 2017;**23**(20):3632–42. doi: 10.3748/wjg.v23.i20.3632. [PubMed: 28611516]. [PubMed Central: PMC5449420].

8. Alizadeh-Navaei R, Hedayatizadeh-Omran A, Janbabaei G. Cancer incidence pattern in Iran provinces and association with human development index. *World J Cancer Res*. 2017;**4**. e913.

9. Enayatrad M, Etemad K, Khodakarim S, Yavari P. Investigation of the relationship between the incidence of colorectal cancer and the human development index in Iran: An ecologic study. *Iran J Epidemiol*. 2017;**13**(3):244–52.

10. Khoshkar AH, Koshki TJ, Mahaki B. Comparison of bayesian spatial ecological regression models for investigating the incidence of breast cancer in Iran, 2005- 2008. *Asian Pac J Cancer Prev*. 2015;**16**(14):5669–73. doi: 10.7314/apjcp.2015.16.14.5669. [PubMed: 26320433].

11. Rohani-Rasaf M, Rohani-Rasaf MR, Asadi-Lari M, Hashemi Nazari SS. Regional differences in cancer incidence trend in Tehran, Iran: A contextual study on the effect of socioeconomic status at regional level. *Int J Cancer Manage*. 2018;**11**(1). doi: 10.5812/ijcm.6641.

12. Torkashvand E, Jafari Jozani M, Torabi M. Pseudo-empirical Bayes estimation of small area means based on James-Stein estimation in linear regression models with functional measurement error. *Can J Stat*. 2015;**43**(2):265–87. doi: 10.1002/cjs.11245.

13. Rezaei MR, ATTAR MA. Rating the level of the interest of the towns of fars province regarding the development indexes by hdi model. *J Reg Plan*. 2012;**1**(4):53–68.

14. Rao JNK, Molina I. *Small area estimation*. John Wiley & Sons, Ltd; 2015. doi: 10.1002/9781118735855.

15. Khazaei S, Rezaeian S, Soheylizad M, Khazaei S, Biderafsh A. Global incidence and mortality rates of stomach cancer and the human development index: An ecological study. *Asian Pac J Cancer Prevent*. 2016;**17**(4):1701–4. doi: 10.7314/apjcp.2016.17.4.1701.