

## RESEARCH ARTICLE

# Investigating the Incidence of Prostate Cancer in Iran 2005 -2008 using Bayesian Spatial Ecological Regression Models

Ahmad Haddad-Khoshkar<sup>1</sup>, Tohid Jafari-Koshki<sup>2</sup>, Behzad Mahaki<sup>1\*</sup>

## Abstract

**Background:** Prostate cancer is the most commonly diagnosed form of cancer and the sixth leading cause of cancer-related deaths among men in the entire world. Reported standardized incidence rates are 12.6, 61.7, 11.9 and 27.9 in Iran, developed countries, developing countries and the entire world, respectively. The present study investigated the relative risk of PC in Iran at the province level and also explored the impact of some factors by the use of Bayesian models. **Materials and Methods:** Our study population was all men with PC in Iran from 2005 to 2008. Considered risk factors were smoking, fruit and vegetable intake, physical activity, obesity and human development index. We used empirical and full Bayesian models to study the relative risk in Iran at province level to estimate the risk of PC more accurately. **Results:** In Iran from 2005 to 2008 the total number of known PC cases was 10,361 with most cases found in Fars and Tehran and the least in Ilam. In all models just human development index was found to be significantly related to PC risk. **Conclusions:** In the unadjusted model, Fars, Semnam, Isfahan and Tehran provinces have the highest and Sistan-and-Baluchestan has the least risk of PC. In general, central provinces have high risk. After adjusting for covariates, Fars and Zanjan provinces have the highest relative risk and Kerman, Northern Khorasan, Kohgiluyeh Boyer Ahmad, Ghazvin and Kermanshah have the lowest relative risk. According to the results, the incidence of PC in provinces with higher human development index is higher.

**Keywords:** Disease mapping - cancer -human development index - bayesian models - regression models

*Asian Pac J Cancer Prev*, 16 (14), 5917-5921

## Introduction

Prostate cancer (PC) is existence of cancer cells in the prostate that can cause a blockage in the urinary tract. It may metastasize to other parts of the body. The prostate is a gland in the male reproductive system that is located below the bladder and in front of the rectum. The size of this gland is about walnut. This gland surrounds part of the urethra and produce part of semen (Afar et al., 2006).

PC is the most commonly diagnosed form of cancer. PC is the sixth leading cause of cancer-related deaths among men in all the world (Jin et al., 2011). It is the most frequently diagnosed malignancy among males in developed countries. In 2002 PC became the fifth most commonly diagnosed cancer in the world. It is estimated that there have been about 233,000 new cases of PC diagnosed in the US in 2014 (Bashir et al., 2014). After non-melanoma skin cancer, PC is the most common cancer among men. This cancer is the second leading cause of cancer death in the US (McGrowder et al., 2012).

The 5-year survival was about 100% for patients with localized PC but in patients in whom distant metastases have occurred, the 5-year survival drops to 31%. PC can metastasize to distant organs such as the liver, lungs and

brain, like most other solid malignancies, but unusually high propensity of metastasizing of this cancer occurred in to the bone. In one autopsy study, about 80% of the men who died from PC had bone metastases (Jin et al., 2011).

Its standardized incidence rates are 12.59, 61.7, 11.9 and 27.9 in Iran, developed countries, developing countries and the entire world, respectively. The peak age for PC is 75-80 (Akbari et al., 2008).

The median age of PC patients is 67 years old. There are 2.5 million PC survivors in the world. It was estimated 241740 men were diagnosed with PC in 2012. Fortunately, most PCs are diagnosed at a non-metastatic stage because of prostate-specific antigen screening, so the survival for PC patients is high. It is estimated that after adjusting for age, sex and race, 97.8% of PC patients survive after 10 years, and 91.4% survive after 15 years (Cheung et al., 2014).

In many studies smoking, fruit and vegetable intake, physical activity, over weightiness and socioeconomic situation was the influencing and measurable factors which affect the incidence of PC (Snowdon et al., 1984; Hiatt 1994; Thune and Lund 1994; Nilsen et al., 2000; Pelucchi et al., 2004; Mahaki, 2011; Cheung et al., 2014; Tewari et al., 2014).

A lot of studies have been conducted about PC and its risk factors (Snowdon et al., 1984; Hiatt, 1994; Thune and Lund 1994; Nilsen et al., 2000; Pelucchi et al., 2004). But most of them were at individual level and those in ecological setting, have not accounted for risk factors (Mahaki, 2011).

Regarding the importance of early diagnosis of PC and ageing process of the Iranian population and increase in the development of the PC and its considerable geographical variation in different parts of the country, it is important to recognize the regions with high incidence and the role of its most significant risk factors to improve the prevention process.

In many studies smoking, fruit and vegetable intake, physical activity, over weightiness and socioeconomic situation was the influencing and measurable factors which affect the incidence of PC (Snowdon et al., 1984; Hiatt, 1994; Thune and Lund, 1994; Nilsen et al., 2000; Pelucchi et al., 2004; Mahaki, 2011; Cheung et al., 2014; Tewari et al., 2014).

Common models for disease mapping is the SMR(SIR) model. But this model has the following disadvantages:

- i) These models are based on estimates of the proportion. Large changes can cause by slight changes in the expected values of the SMR;
- ii) When the expected values are close to zero, these indicators become very large;
- iii) The variance of the index is inversely proportional to the expected value. So in cities with high population, levels of the index will be small and in cities with low population, levels of the index will be small;
- iv) When the value of this index is equal to zero, the difference between the amounts of expected incidence cannot be revealed;
- v) This index doesn't show the basic structure in the data. This index is estimation of relative risk of saturated model (Mahaki et al., 2013).

To avoid these problems we used Bayesian spatial ecological regression model to map the risk of PC in Iran.

The present study investigates the incidence of PC in Iran at province level and also explores the impact of some covariates including smoking, over-weight or obesity, physical activity, fruit and vegetable intakes, and human development index in ecological setting by use of Bayesian models.

## Materials and Methods

All patients were suffering from PC in Iran during 2005 to 2008 were included in this study. These data were extracted from the annually published reports by Non-communicable Disease Management Center of Iranian Ministry of Health and Medical Education.

Data on four risk factors including smoking, fruit and vegetable intake, physical activity level and proportion of population with body mass index (BMI)>25, all measured as average estimates over each province, were extracted from annual reports of Non-communicable Disease Risk Factors Surveillance System of the Ministry of Health. Also human development index (HDI) was extracted from the annual reports of the Central Bank of Iran.

In this study, over weightiness or obesity variable is the percentage of the cases having BMI more than 25

in every province. The fruit and vegetable variable was the total of daily consumed fruit and vegetable amount in every province. Physical activity variable calculated by use of combined index called Metabolic Equivalent (MET). The Human Development Index was geometric mean of achieving success in every province in three main dimensions of human development, including long and healthy life, having access to knowledge and normal life standards. It ranges from zero to one (Bray et al., 2012).

### Spatial Bayesian models consist of two models

**Empirical Bayes:** Empirical Bayes itself consist of two models:

- a) Gamma-Poisson Regression Model:

This model defines the relation between incidence, and risk factors in every province as follows;

$$\log \Theta_i = \alpha + \sum_{h=1}^H \beta_h x_{ih}$$

- b) Lognormal Model

The most important problem of Gamma-Poisson is disregarding spatial correlation. In this model variation of observed counts become more than the expected variability of Poisson model. In this situation for solving this problem, one parameter for accounting over dispersion is added to the model. The rate of overdispersion depends on the score of heterogeneity found among relative risks. In this model it is supposed that  $\Theta_i$  has Gamma distribution. So, the achieved posterior distribution will be Gamma.

Lognormal model is more flexible for relative risk formulation and is defined as follows.

$$\log \Theta_i = \alpha + v_i + \sum_{h=1}^H \beta_h x_{ih}$$

$$v_i \sim N(0, \tau_v^2)$$

(Asmarian NS 2013; Clayton D 1987)

### Full Bayesian model:

**BYM Model:** This model, suggested by Besag, York, and Mollie, considers two sources of changes for justifying the heterogeneity, one for structural and one for non-structural heterogeneity. This model is defined as follows:

$$\log \Theta_i = \alpha + u_i + v_i + \sum_{h=1}^H \beta_h x_{ih}$$

$$[u_i | u_j, i \neq j, \tau_u^2] \sim N(\bar{u}_i, \tau_u^2)$$

$$\bar{u}_i = \frac{1}{\sum_j \omega_{ij}} \sum_j u_j \omega_{ij}, \tau_u^2 = \frac{\tau_u^2}{\sum_j \omega_{ij}}$$

$$\omega_{ij} = 1 \text{ If } i \text{ and } j \text{ are adjacent}$$

$$\omega_{ij} = 0 \text{ If } i \text{ and } j \text{ are not adjacent}$$

Where  $u_i$  and  $v_i$  represent structural and non-structural heterogeneity respectively. They are also supposed to have conditional auto regressive normal distribution and normal distribution n. Regarding conditional auto regressive, the rate of incidence in every region is allowed to depend on the incidence rates in neighboring regions (Lawson et al., 1999; 2003).

These models were fitted using OpenBUGS version 3.2.1. Convergence was checked by using Brooks-Gelman-

Robin plots. Significance tests for parameters were done by the use of Bayesian credible intervals (CrI) which are equivalent to p-value of frequent is approach.

## Results

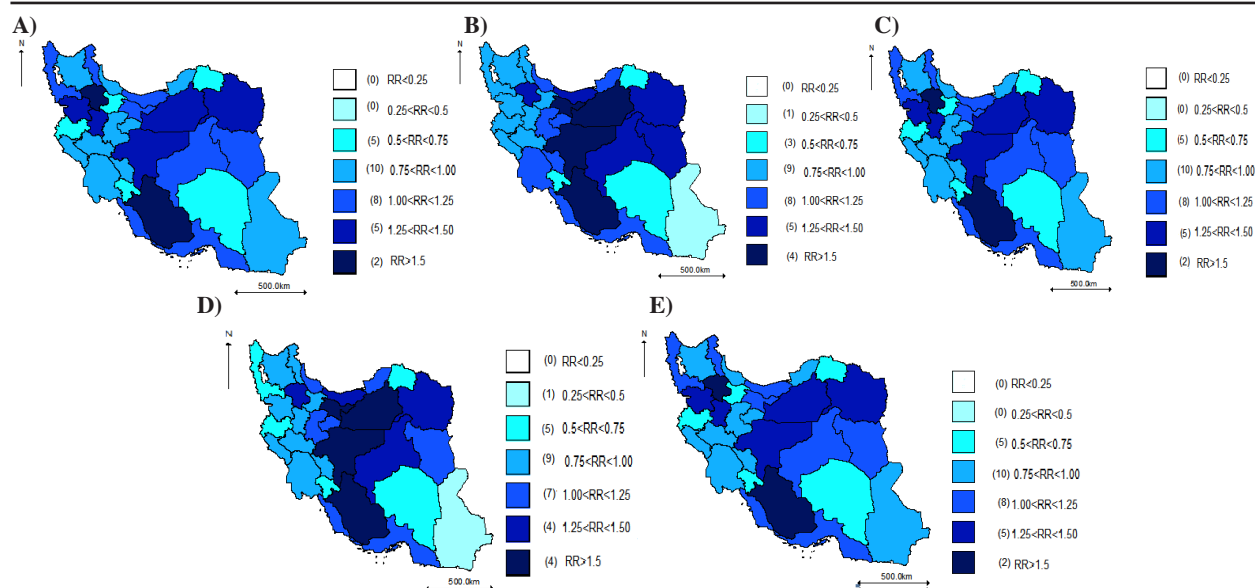
The total number of known PC cases has been 10361

in Iran from 2005 to 2008. The most cases were found in Fars (1995 cases) and Tehran (1880 cases) and the least was in Ilam (60 cases).

According to Table 3, in all of three models (from, just HDI has been significant with positive coefficient; that means there is a direct relationship between HDI and PC incidence.

**Table 1. Overall Prostate Cancer, Populations at Risk and the Expected Number of Prostate Cancer by Provinces in Iran**

Provinces	Overall number of prostate cancer	Populations at risk	Expected number of prostate cancer by provinces
Ardbil	210	617342	302.775
Western_Azrbjan	235	1458581	421.426
Eastern_Azrbjan	276	1839341	401.984
Bousher	282	468062	323.985
ChahrmhlBakhtiary	76	431644	109.19
Isfahan	690	2335399	475.641
Fars	1995	2204852	1132.739
Ghazvin	107	583869	152.477
Gilan	249	1198519	259.613
Golestan	362	804162	448.347
Hamadan	154	856837	202.32
Hormozgan	297	725196	356.952
Ilam	60	27566	101.336
Kermn	131	1298546	281.024
Kermanshah	177	922060	308.578
Southern Khorasan	266	314278	287.649
Razavi Khorasan	630	2783676	586.439
Northern Khorasan	73	490707	132.123
Khozestan	353	2090048	476.495
Kohgiluyeh Boyer Ahmad	83	313316	203.536
Kurdistan	179	729718	232.521
Lorestan	115	873692	194.73
Markazi	177	682367	211.854
Mazndaran	422	1466870	399.433
Ghom	150	536904	187.767
Semnan	286	302433	208.592
Sistan-Baluchestan	83	1221240	297.318
Tehran	1880	6901217	1340.088
Yazd	206	517137	180.642
Zanjan	157	480698	143.45



**Figure 1. Maps of Incidence Risk of PC A) Gamma-Poisson without adjusting risk factor adjustment B) Lognormal without adjusting risk factor adjustment C) Lognormal with adjusting risk factor adjustment D) BYM without adjusting risk factor adjustment E) BYM with adjusting risk factor adjustment**

**Table 2. The Relative Risk of Provinces for Prostate Cancer in Iran**

Provinces	Without considering risk factors		With considering risk factors		
	lognormal	BYM	Gamma-Ppisso	lognormal	BYM
Ardbil	0.883	0.882	1.04	1.04	1.054
Western_Azrbjan	0.714	0.713	1.032	1.031	1.05
Eastern_Azrbjan	0.873	0.872	0.914	0.914	0.924
Bousher	1.1	1.99	1.186	1.185	1.17
ChahrmhlBakhtiary	0.894	0.89	0.981	0.981	0.986
Isfahan	1.83	1.824	1.286	1.287	1.29
Fars	2.228	2.221	1.672	1.672	1.661
Ghazvin	0.896	0.897	0.731	0.731	0.732
Gilan	1.211	1.208	0.987	0.986	0.994
Golestan	1.024	1.022	0.932	0.932	0.932
Hamadan	0.965	0.964	1.365	1.363	1.368
Hormozgan	1.054	1.051	1.119	1.12	1.109
Ilam	0.773	0.768	0.918	0.918	0.907
Kermn	0.607	0.606	0.637	0.636	0.632
Kermanshah	0.738	0.733	0.668	0.667	0.672
Southern Khorasan	1.168	1.165	1.079	1.079	1.06
RazaviKhorasan	1.359	1.354	1.28	1.281	1.267
Northern Khorasan	0.721	0.722	0.729	0.728	0.723
Khuzestan	0.941	0.938	0.809	0.809	0.812
Kohgiluyeh Boyer Ahmad	0.544	0.544	0.713	0.713	0.718
Kurdistan	0.977	0.975	1.474	1.474	1.495
Lorestan	0.762	0.761	0.953	0.952	0.949
Markazi	1.058	1.057	0.947	0.947	0.949
Mazndaran	1.335	1.331	1.201	1.2	1.213
Ghom	1.014	1.013	0.78	0.779	0.783
Semnan	1.717	1.712	1.358	1.358	1.354
Sistan-Baluchestan	0.383	0.38	0.848	0.846	0.84
Tehran	1.775	1.771	1.03	1.03	1.034
Yazd	1.428	1.428	1.013	1.013	1.015
Zanjan	1.367	1.362	1.648	1.648	1.661

**Table 3. Estimation of Coefficients of Relative Risks Effects for Ecological Regression Models**

Risk factors	Poisson-Gamma		Lognormal		BYM	
	coefficient	95%CrI	coefficient	95%CrI	coefficient	95%CrI
Smoking	-0.001	(-0.005,0.003)	-0.001	(-0.005,0.003)	-0.001	(-0.006,0.003)
over weightiness	0	(-0.023,0.023)	-0.004	(-0.023,0.024)	-0.001	(-0.026,0.023)
physical activity	-0.004	(-0.021,0.013)	0	(-0.021,0.013)	-0.004	(-0.021,0.013)
fruit &vegetable intakes	-0.074	(-0.293,0.150)	-0.004	(-0.297,0.153)	-0.069	(-0.294,0.163)
HDI	5.981	(3.467,8.533)	5.971	(3.396,8.562)	6.007	(3.396,8.601)
DIC for models with covariates	276.2		276.2		270.1	
DIC for models without covariates	-		276.4		213.1	

Table 3 also shows the goodness of fit criteria of the models. Here, BYM model without considering risk factors has the best fit with DIC=213.1. Other models do not differ much in the goodness of fit.

Figure 1.A shows the province specific relative risks of PC without adjusting for risk factors in Gamma-Poisson model. Relative risk is the highest for Fars (RR=1.672) and Zanjan (RR=1.648), Kerman, Northern Khorasan, Ghazvin, Kohgiluyeh Boyer Ahmad and Kermanshah have the least relative risk of PC incidence.

RR of unadjusted lognormal model is plotted in Figure 1.B. According to this map, central provinces (Tehran, Semnan, Isfahan and Fars) have the highest adjusted relative risks. Among these provinces, Fars with RR= 2.228 has the highest risk. Sistan-Baluchestan with RR= 0.3826 has the least risk.

Zanjan (RR=1.648) and Fars (RR=1.672) have the

highest adjusted RR of PC in lognormal as shown in Figure 1.C. Kerman, Northern Khorasan, Kohgiluyeh Boyer Ahmad, Zanjan and Kermanshah have the least adjusted risk. Kerman with adjusted relative risk=0.636 has the least risk.

The result of BYM model with no adjustment for risk factors is presented in Figure 1.D where Fars, Isfahan, Tehran, Semnan and Yazd provinces have the highest relative risk. From these provinces, Fars with relative risk = 2.221 has the highest risk and Sistan-Baluchestan with relative risk = 0.3798 has the least risk.

Figure 1.E shows the corresponding BYM model after adjustment for risk factors. According to this map Fars and Zanjan provinces have the highest adjusted relative risks and Kerman, Kermanshah, Kohgiluyeh Boyer Ahmad, Northern Khorasan and Ghazvin have the least risk. From these provinces, Fars with relative risk = 1.661 has the



highest adjusted risk and Kerman with relative risk = 0.632 has the least risk.

## Discussion

According to the results, the incidence of PC in provinces with higher human development index is higher. This is in accordance with the results of Nilsen et al study (Nilsen et al., 2000). This may be due to lifestyle, higher levels of pollution, more contact with carcinogens, or higher detection rates in provinces with higher HDI, .

The comparison of maps A, C and E shows that these maps are exactly the same and maps B and D are very similar. This shows that in this study BYM and lognormal model produce very similar output. This is in accordance with the result of Clayton and Kaldor (1987) and Mahaki et al. (2013) studies.

Allen et al. (2004) performed the study on 18115 men in Hiroshima and Nagasaki, Japan, in order to examine the association between dietary intake and PC risk. They used Poisson regression to calculate incidence rates and found that fish intake was significantly associated with an increased risk of PC in men who consumed fish more than four times per week. No other food items such as soya products were significantly associated with PC risk. They found that dietary factors may not be strong determinants of PC in Japanese men. This is in agreement to results of our study.

Map D is similar to the corresponding map in Mahaki et al. (2011) study for PC. But the differences may be because of differences in cut points or because that study was for 2007, but our study is for 2005-2008.

Ecological bias may confound the relation between PC incidence and risk factors. May be for this reason, other risk factors were not significant.

Rodriguez et al. (2007) performed study on 69,991 men in the Cancer Prevention Study II Nutrition Cohort. They examined body mass index (BMI) and weight change in relation to incident PC by disease stage and grade at diagnosis and participants provided information on height and weight in 1982, and again at enrollment in 1992. They found that association between BMI in 1992 and risk of PC differed by stage and grade at diagnosis. BMI was inversely associated with risk of non-metastatic low-grade PC, but BMI was positively associated with risk of no metastatic high-grade PC. But in our study no significant association was seen between obesity and PC incidence. This may be because we didn't consider stage of PC and the study was not at individual level.

Our study had some limitations. Data was at provinces level and it would be better to study PC distribution on smaller regions such as counties. Also studies can be conducted for a longer period of time on more recent years.

## References

Afar D, Henshall S, Hiller J, et al (2006). Methods of prognosis of prostate cancer. (Google Patents).  
Akbari M, Abachizadeh K, Khayamzadeh M, et al (2008). Iran cancer report. cancer research center shahid beheshti university of medical sciences Tehran, Qom: Darolfekr.

Allen NE, Sauvaget C, Roddam AW, et al (2004). A prospective study of diet and prostate cancer in Japanese men. *Cancer Causes Control*, **15**, 911-20.  
Asmarian NS, Kavousi A, Salehi M, et al (2013). Comparison of point poisson kriging and empirical bayesian methods in mapping of gastrointestinal cancer incidence rate in Iran. *J Health Syst Res*, **9**, 277-85.  
Bashir MN, Ahmad MR, Malik A (2014). Risk factors of prostate cancer: a case-control study in Faisalabad, Pakistan. *Asian Pac J Cancer* [Epub ahead of print].  
Bray F, Jemal A, Grey N, et al (2012). Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. *Lancet Oncol*, **13**, 790-801.  
Clayton D, Kaldor J (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics J*, **43**, 671-81.  
Cheung MR, Kang J, Ouyang D, et al (2014). Association between urinary cadmium and all cause, all cancer and prostate cancer specific mortalities for men: an analysis of National Health and Nutrition Examination Survey (NHANES III) Data. *Asian Pac J Cancer Prev*, **15**, 483.  
Jin JK, Dayyani F, Gallick GE (2011). Steps in prostate cancer progression that lead to bone metastasis. *Int J Cancer*, **128**, 2545-61.  
Lawson A, Biggeri A, Bohning D, et al (1999). Disease mapping and risk assessment for public health (John Wiley & Sons).  
Lawson AB, Browne WJ, Rodeiro CLV (2003). Disease mapping with WinBUGS and MLwiN (11: John Wiley & Sons).  
Mahaki B, Mehrabi Y, Kavousi A, et al (2011). Multivariate disease mapping of seven prevalent cancers in Iran using a shared component model. *Asian Pac J Cancer Prev*, **12**, 2353-8.  
Mahaki B, Mehrabi Y, Kavousi A, et al (2013). Applying gamma-poisson, lognormal, and BYM models in comparing relative risk of suicide among provinces of Ilam, Iran. *J Health Syst Res*, **9**, 86-95.  
McGrowder DA, Jackson LA, Crawford TV (2012). prostate cancer and metabolic syndrome: is there a link. *Asian Pac J Cancer Prev*, **13**, 1-13.  
Nilsen TIL, Johnsen, R, Vatten LJ (2000). Socio-economic and lifestyle factors associated with the risk of prostate cancer. *British Journal of Cancer*, **82**, 1358.  
Pelucchi C, Talamini R, Galeone C, et al (2004). Fibre intake and prostate cancer risk. *Int J Cancer*, **109**, 278-80.  
Rodriguez C, Freedland SJ, Deka A, et al (2007). Body mass index, weight change, and risk of prostate cancer in the cancer prevention study II nutrition cohort. *Cancer Epidemiol Biomarkers Prev*, **16**, 63-69.  
Snowdon DA, PHILLIPS RL, Choi W (1984). Diet, obesity, and risk of fatal prostate cancer. *Am J Epidemiol*, **120**, 244-50.  
Tewari R, Chhabra M, Natu SM, et al. (2014). Significant association of metabolic indices, lipid profile, and androgen levels with prostate cancer. *Asian Pac J Cancer Prev*, **15**, 9841-46.  
Thune I, Lund E (1994). Physical activity and the risk of prostate and testicular cancer: a cohort study of 53,000 Norwegian men. *Cancer Causes Control*, **5**, 549-56.