RESEARCH COMMUNICATION

Bayesian Estimation of Colorectal Cancer Mortality in the Presence of Misclassification in Iran

Mohamad Amin Pourhoseingholi[1], Soghrat Faghihzadeh[1]*, Ebrahim Hajizadeh[1], Alireza Abadi[2], Mohammad Reza Zali[3]

Abstract

Background: Colorectal cancer is the third most common cause of cancer-related deaths in the world. Addressing the burden of related mortality is therefore important but according to the Iranian mortality registry, about 20% of death statistics were still recorded in misclassified categories. The aim of this study is to re-estimate the CRC mortality rate for Iranian population, using a Bayesian approach in order to revise this misclassification. Methods: National Death Statistics Reported by the Ministry of Health and Medical Education (MOH&ME) from 1995 to 2003 were included in this analysis. The Bayesian approach to correct and account for misclassification effects in Poisson count regression was employed to estimate the mortality rates by age and sex group. Results: According to the Bayesian re-estimate there were between 30 to 40 percent underreported mortality records in death due to colorectal cancer and the rate for related mortality had moderately increased through recent years. Conclusion: Our findings suggest a substantial undercount of colorectal cancer mortality in the Iranian population. Therefore healthcare policy makers who determine research and treatment priorities from death rates as indicators of public health problems should pay notice to this underreporting.

Key Words: Colorectal cancer - mortality data - Baynesian estimation - Iran

Asian Pacific J Cancer Prev, **10**, 691-694

Introduction

With its high incidence and mortality, colorectal cancer (CRC) constitutes a public health burden in most industrialized countries (Sonnenberg et al., 2000). CRC is the third most common cause of cancer-related deaths in the world (Parkin, 2001). The incidence has increased in Iran in recent years (Moghimi et al., 2008) and the disease will become increasingly important in the next decade. Also Iranian data suggest a younger age distribution compared to Western reports (Azadeh et al., 2008; Pourhoseingholi et al, 2008; 2009). So it is predicted that CRC will impose a substantial burden on Iranians.

A familiar projection to address the burden of non communicable diseases is the estimation of mortality rates. With regards to cancer mortality, data are important to monitor the effects of screening program, earlier diagnosis, demographic data and other prognostic factors (Burnet et al., 2005). On the other hand the analysis of death statistic subject to misclassification is a major problem in epidemiological analysis leading to biases estimates, and can therefore cause one to underestimate health risks (Stamey et al., 2008). As like as other developing countries, Iran has incomplete mortality information (Khosravi et al., 2007). According to the Iranian death registry, between 15% to 20% death statistics were recorded in misclassified categories such as septicemia, senility without mention of psychosis symptoms and other ill-defined conditions (Naghavi, 2004).

Two approaches to correct for misclassification are recommended. One uses a small validation sample yielding more accurate parameter estimates (Lyles, 2002) and the alternative method is the Bayesian approach in which subjective prior information on at least some subset of the parameters is applied to estimate a misclassified parameter and then re-estimate death statistics (Whittemore and Gong, 1991; Sposto et al., 1992).

The aim of this study was to re-estimate CRC mortality rates for the Iranian population, using the Bayesian approach.

Patients and Methods

National death statistics reported by the Ministry of Health and Medical Education (MOH&ME) from 1995 to 2003, stratified by age group, sex, and cause of death (coded according to the 9th revision of the International Classification of Diseases [ICD-9]) were included in this analysis. CRC mortality [ICD-9; 153-154] was expressed as the mortality rate for 100,000 people.

The Bayesian approach we considered here was derived from models recommended by Stamey et al to correct and account for misclassification effects in Poisson count regression (Stamey et al., 2008). Stamey's technique

*[1]Dept of Biostatistics, Tarbiat Modares University, [2]Dept of Biostatistics, Shahid Beheshti University M.C, [3]Research Center of Gastroenterology and Liver Diseases, Shahid Beheshti University M.C, Tehran, Iran *For correspondence: Aminphg@gmail.com*

Table 1. Bayesian CRC Mortality and Frequentist Rates/100,000 Adjusted for Sex and Age Groups

Date/Age category		<5 Years		5-14 Years		15-49 Years		≥50 Years		All ages		Total
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	
1995	FR	0.00	0.00	0.01	0.04	0.17	0.07	2.96	2.96	0.46	0.41	0.44
	BR	0.04	0.04	0.02	0.05	0.24	0.09	4.27	4.27	0.68	0.59	0.64
1996	FR	0.17	0.04	0.02	0.05	0.28	0.27	5.05	3.49	0.81	0.59	0.70
	BR	0.25	0.08	0.04	0.08	0.42	0.40	7.29	5.04	1.17	0.86	1.02
1997	FR	0.16	0.12	0.05	0.00	0.46	0.21	6.13	4.18	1.04	0.64	0.85
	BR	0.25	0.16	0.07	0.02	0.67	0.31	8.86	6.08	1.51	0.94	1.23
1998	FR	0.16	0.16	0.06	0.04	0.39	0.39	7.12	5.54	1.13	0.91	1.02
	BR	0.24	0.24	0.08	0.06	0.55	0.57	10.18	7.82	1.62	1.31	1.47
1999	FR	0.24	0.23	0.01	0.04	0.55	0.49	8.89	5.75	1.43	1.00	1.22
	BR	0.36	0.35	0.02	0.06	0.78	0.72	12.71	8.24	2.05	1.45	1.76
2000	FR	0.55	0.22	0.07	0.05	0.69	0.51	10.71	8.15	1.73	1.25	1.49
	BR	0.83	0.32	0.10	0.07	1.01	0.73	15.44	11.91	2.50	1.83	2.17
2001	FR	0.14	0.11	0.10	0.04	0.71	0.50	10.34	8.96	1.86	1.52	1.69
	BR	0.21	0.14	0.14	0.07	1.06	0.75	13.72	12.80	2.54	2.24	2.39
2002	FR	0.00	0.00	0.00	0.00	0.90	0.70	18.6	11.97	2.86	1.95	2.42
	BR	0.13	0.14	0.04	0.05	1.37	1.06	25.0	17.91	3.96	2.96	3.47
2003	FR	0.12	0.18	0.09	0.02	1.03	0.60	19.59	11.97	3.15	1.90	2.54
	BR	0.18	0.31	0.12	0.04	1.48	0.88	28.65	17.29	4.59	2.76	3.70

FR: Frequentist Rate, BR: Bayesian Rate

extended the model recently proposed to overcome the problem of misclassification in cancer data (Whittemore and Gong, 1991; Sposto et al., 1992). Whittemore and Gong (1991) used a likelihood approach to estimate regression parameters when the counts are underreported and Sposto et al (1992) developed this likelihood to allow for misclassification across two groups. Stamey et al (2008) extended these approaches but did not rely on asymptotic results in order to perform inferences and also did not assume that the misclassification parameters are known in a Poisson regression model . We studied Iranian death statistic in a Bayesian Poisson regression using Stamey's approach to re-estimate mortality rate of CRC. All analysis performed by a Macro, developed in S-Plus according to Stamey's approach.

Results

We considered data consisting of all deaths due to CRC from 1995 to 2003, (up to 6804 records). The misclassification probability estimate we proposed here was based on Iranian death registration which introduced up to 20% misclassified records in total deaths. So a beta prior assumed to re-estimate death statistic of CRC from misclassified groups. The rates of CRC mortality classified by sex and age, generated from original database (Frequentist Rate) and their Bayesian corresponding projections (Bayesian Rate) are shown in Table 1. According to the Bayesian re-estimate there were between 30 to 40 percent underreported mortality records in death due to CRC (Figure 1). The rate of CRC mortality moderately increased from 1995 to 2003. Also CRC mortality was higher for older age (Table 1). Figure 2 shows CRC mortality and its Bayesian projection according to gender, indicating that the mortality rate for male was high comparing to female considerably.

Discussion

Response misclassification of counted data for death statistics leads to biases and underestimates. In the new Iranian Death Registration System, data on causes of death are collected from various sources and have been assessed to be about 80% complete (Khosravi et al., 2007). In spite of this new registry system, there is still up to 20% undefined death records that categorized as misclassification. In a study conducted by Khosravi et al, validation data from hospital death used to find the Impact of misclassification on measures of cardiovascular disease mortality (Khosravi et al., 2008). But they didn't employ

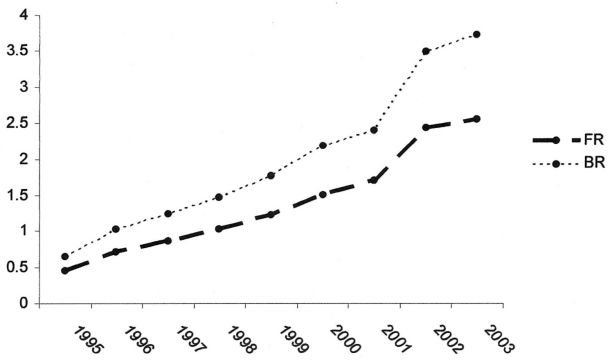

Figure 1. Bayesian CRC Mortality Rates over Time. FR: Frequentist Rate, BR: Bayesian Rate

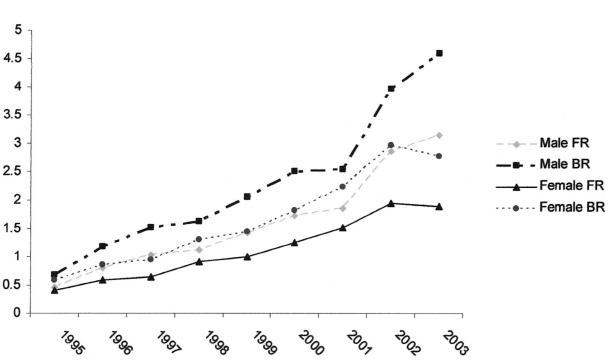

Figure 2. Bayesian CRC Mortality Rates Adjusted for Sex Groups. FR: Frequentist Rate, BR: Bayesian Rate

any Bayesian projections. So our study is the first Bayesian analysis on Iranian mortality data. Recently Bayesian approach received much attention in the case of misclassification. McInturff et al used a Bayesian approach to estimate the parameters of a binomial regression with misclassification (McInturff et al., 2004). Whittemore and Gong used this approach to estimate cervical cancer mortality rates (Whittemore and Gong, 1991) and Sposto et al developed this likelihood to assess the effect of diagnostic misclassification on non-cancer and cancer mortality dose–response in A-bomb survivors (Sposto et al, 1992). Stamey et al (which we considered their producer in here) used Bayesian approach in data consisting of the number of deaths due to cancer and non-cancer among residents of Hiroshima and Nagasaki, Japan, who were present during the atomic bombings in August of 1945 (Stamey et al, 2008).

Our study indicated that although the mortality rate of CRC seems to be low, up to 40% of its death statistics underreported. In addition the linear increase of mortality rate during these recent years and predicting to experience a higher incidence in future indicated this fact that the population may be experiencing an acceleration of the disease burden (Ansari et al, 2006).

Our findings suggested a substantial undercount of CRC mortality in Iranian population. So healthcare policy makers who determine research and treatment priorities on death rates as an indicator of public health priorities should notice to this under-reported data.

References

Ansari R, Mahdavinia M, Sadjadi A, et al (2006). Incidence and age distribution of colorectal cancer in Iran: results of a population-based cancer registry. *Cancer Lett*, **240**,143-7.

Azadeh S, Moghimi-Dehkordi B, Fatemi SR, et al (2008). Colorectal cancer in Iran: an epidemiological study. *Asian Pac J Cancer Prev*, **9**, 123-6.

Burnet NG, Jefferies SJ, Benson RJ, Hunt DP, Treasure FP (2005). Years of life lost (YLL) from cancer is an important measure of population burden – and should be considered when allocating research funds. *Br J Cancer*, **92**, 241-5.

Khosravi A, Taylor R, Naghavi M, Lopez AD (2007). Mortality in the Islamic Republic of Iran, 1964–2004. *Bull WHO*, **85**, 607-14.

Khosravi A, Rao C, Naghavi M, et al (2008). Impact of misclassification on measures of cardiovascular disease mortality in the Islamic Republic of Iran: a cross-sectional study. *Bull WHO*, **86**, 688-96.

Lyles RH (2002). A note on estimating crude odds ratios in case–control studies with differentially misclassified exposure. *Biometrics*, **58**, 1034-6.

McInturff P, Johnson W, Cowling D, Gardner I (2004). Modeling risk when binary outcomes are subject to error. *Statist Med*, **23**, 1095-109.

Moghimi-Dehkordi B, Safaee A, Zali MR (2008). Prognostic factors in 1,138 Iranian colorectal cancer patients. *Int J Colorectal Dis*, **23**, 683-8.

Naghavi M (2004). Death report from 23 provinces in Iran. 1st edition. Tehran: Ministry of Health.

Parkin DM (2001). Global cancer statistics in the year 2000. *Lancet Oncol*, **2**, 533-43.

Pourhoseingholi A, Pourhoseingholi MA, Vahedi M, et al (2008). Relation between demographic factors and type of gastrointestinal cancer using probit and logit regression. *Asian Pac J Cancer Prev*, **9**, 753-5.

Pourhoseingholi MA, Vahedi M, Moghimi-Dehkordi B, et al (2009). Burden of hospitalization for gastrointestinal tract cancer patients - Results from a cross-sectional study in Tehran. *Asian Pac J Cancer Prev*, **10**, 107-10.

Sonnenberg A, Delcò F, Inadomi JM (2000). Cost-effectiveness of colonoscopy in screening for colorectal cancer. *Ann Intern Med*, **133**, 573-84.

Sposto R, Preston DL, Shimizu Y, Mabuchi K (1992). The effect of diagnostic misclassification on non-cancer and cancer mortality dose–response in A-bomb survivors. *Biometrics*, **48**, 605-17.

Stamey JD, Young DM, Seaman Jr JW (2008). A Bayesian approach to adjust for diagnostic misclassification between two mortality causes in Poisson regression. *Statist Med*, **27**, 2440-52.

Whittemore AS, Gong G (1991). Poisson regression with misclassified counts: application to cervical cancer mortality rates. *Applied Statistics*, **40**, 81-93.