ORIGINAL ARTICLE

Arvand Journal of Health & Medical Sciences
©2016 Abadan University of Medical sciences

Under-estimation and over-estimation in gastric cancer incidence registry, Bayesian analysis for selected provinces

Nastaran Hajizadeh[1], Mohamad Amin Pourhoseingholi[2], Ahmadreza Baghestani[1], Alireza Abadi[1]

[1]Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[2] Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran

✉ **Reprint or Correspondence**: Mohamad Amin Pourhoseingholi; PhD Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran
@✉ Amin_phg@yahoo.com.

ABSTRACT

Incidence is an important index for measuring the burden of cancer in a population. But in medical studies, a difficulty in drawing inference from categorical data is the existence of misclassification error. Misclassification error is the disagreement between the observed and the true value and occurs when new cancer cases diagnosed and registered in neighborhood provinces instead of their hometown due to low facility in their own provinces and difference of quality and quantity of registration system in different provinces. The aim of this study is to use a Bayesian statistical method to assess and correct this misclassification in Gastric cancer incidence, for some selected provinces in Iran. Data extracted from Iranian annual of national cancer registration report in 2008. The Age Standardized Rate (ASR) due to gastric cancer were expressed as rate per/100,000 population for male and female of North, South and Razavi Khorasan provinces. A Bayesian approach was used with Poisson count regression and an informative beta prior distribution assumed for the misclassified parameter. Analyses were carried out using R software version 3.2.0.

The Bayesian analysis showed that, there is 34% misclassification in gastric cancer incidence registry from North and South Khorasans in Razavi Khorasan. After the correction, the rate of ASR decreased for Razavi Khorasan provinces, while, these rate increased for the rest of its neighborhoods. This study indicated that there is a major misclassification among provinces in Iran, in which, from neighborhood provinces and cities, the statistics of gastric cancer incidence misclassified on each other's. In the absence of valid data, Bayesian approach would be a good and flexible alternative to eliminate the effects of Misclassification in incidence registry data for neighboring provinces.

Keywords: Gastric cancer, Misclassification Error, Bayesian Method, Incidence, Cancer Registry.

***Received*: 11 March 2016 *Accepted*: 19 May 2016**

Introduction

Cancer is one of the major causes leading to many disorders, death, and disabilities in the world (1). Among all cancers, gastrointestinal cancers (GI cancers) present an increasing burden for lots of countries (2). One of the most important GI cancers is gastric cancer. This cancer is a disease in which the cells forming the inner lining of the stomach become abnormal

Please cite as: **Hajizadeh N, Pourhoseingholi MA, Baghestani A, Abadi A. Under-estimation and over-estimation in gastric cancer incidence registry, Bayesian analysis for selected provinces. Arvand J Health Med Sci 2016;1(2):90-94.**

and start to divide uncontrollably, forming a mass called a tumor (1). Gastric cancer was 4th most common cancer among men and 5th most common cancer among women diagnosed in 2012 in the world and is first common cancer among Iranian men and the 3th (after breast cancer and colorectal cancer) among Iranian women (4).

There are two major projections to measure the burden of any cancers; mortality and incidence. These statistics are important to monitor the effects of screening programs, earlier diagnosis and other prognostic factors (5). Cancer registration is an important source for measuring the burden of cancer in a population and cancer registry is a systematic collection of any data regarding cancers. But in medical studies, a difficulty in drawing inference from categorical data is the existence of misclassification error. Although among medical indexes, incidence is a familiar projection in the assessment of the burden of diseases (5), the presence of misclassification error makes the registry systems inaccurate and unreliable to use for estimating the burden of disease and other epidemiological criteria, and consequently flaws the planning for cancer prevention (6, 7). Misclassification error is the disagreement between the observed and the true value and occurs when new cancer cases diagnosed and registered in neighborhood provinces instead of their hometown due to low facility in their own provinces and difference of quality and quantity of registration system in different provinces. As the evidence, the expected coverage of cancer incidence in different provinces can be mentioned; that the observed rate of incidence is more than expected in some provinces, on the other hand, it is much less than expected rate in neighboring provinces (8). However it happens while we expect that the rate of cancer incidence be about the same in neighboring provinces that are quite similar in environmental conditions and lifestyle. In the absence of a gold standard, statistical methods help to overcome this problem. There are two approaches to reduce the effects of misclassification error; the first is using a small validation sample (9) and the second is a Bayesian analysis which provides subjective prior information for the subset of the parameters for re-estimate and corrects the statistic (10). In this study, we briefly explained this Bayesian method with an example of misclassified gastric cancer incidence for North, South and Razavi Khorasan provinces.

Materials and Methods

Data for this study extracted from Iranian annual of national cancer registration report in 2008 (8). The Age Standardized Rate (ASR) due to gastric cancer (coded according to the 10th revision of the International Classification of Diseases [ICD-10; C16]) were expressed as rate per/100,000 population for male and female of North, South and Razavi Khorasan. To correct the misclassification effect, a Bayesian approach was used with Poisson count regression. To perform Bayesian inference, we assumed an informative beta prior distribution for the misclassified parameter. Because the misclassified parameter is unknown, a latent variable approach was employed to simplify the full conditional models and estimate the posterior distribution using a Gibbs sampling algorithm (6, 11-17). Expected coverage of each province was used as priors for the parameters of beta distribution. Analyses were carried out using R software version 3.2.0.

Results

All incidence records due to gastric cancer for Khorasan provinces that have registered at Iranian annual of national cancer registration report in 2008 included in this study. The reported percent of expected coverage of cancer incidence for Razavi Khorasan was 155.5%. It means that Razavi Khorasan have covered 55.5% more new cancer cases than its expectation, whereas the North and South Khorasans have just covered respectively 34.8% and 41.4% of their expected coverage; which clearly is an indication of existence of misclassification error. After implementation of the Bayesian method, it was found that there was about 34% misclassification in gastric cancer incidence registry from North and South Khorasans in Razavi Khorasan. After the correction, it is expected to increase in the rate of gastric cancer in north and south Khorasans and decrease in its rate for Razavi Khorasan. So, after this Bayesian adjustment, the rate of ASR decreased for Razavi Khorasan provinces, while, these rate increased for the rest of its neighborhoods (table 1).

Discussion

This study indicated that there is a major misclassification among provinces in Iran, in which, from neighborhood provinces and cities, the statistics of gastric cancer incidence misclassified on each other's. Accurate cancer incidence data are essential to planning, monitoring and evaluating national and regional cancer control programs (18). In Iran, there are provinces with higher or lower incidence of gastric cancers and policy makers employ these data to allocate the facilities and resources according to these incidences statistics. When the cancer incidence data is regionally misclassified, makes underestimation of health risk in some provinces and overestimation for the others. This problem leads to misallocation of resources. So in planning for resource allocation, authorities should notice that, low incidence of gastric cancer in North and South Khorasans, do not mean that they are in a good health situation and gastric cancer incidence is really low in these provinces, but quite the contrary, this may be the effect of misclassification error and it is needed to allocate them more health facilities, equipped health centers, and improve the registration system accuracy, especially in terms of patients permanent residence. Improving the quality of the cancer registry in Iran will require more expert staffing, refining foundations, and powerful hardware and software resources (19). In the absence of valid data, Bayesian approach would be a good and flexible alternative to eliminate the effects of Misclassification in incidence registry data for neighboring provinces (11). By this method, one can correct the incidence or mortality

Table 1. The expected coverage and ASR (per 100,000) for North Khorasan, Razavi Khorasan and South Khorasan provinces, before and after Bayesian correction

	Original data				Bayesian corrected data		
Province	Expected Coverage	Male (ASR)	Female (ASR)	Total (ASR)	Male (ASR)	Female (ASR)	Total (ASR)
North Khorasan	34.80%	12.62	5.37	8.99	28.24	12.02	20.12
Razavi Khorasan	155.50%	24.18	10.24	17.21	20.71	8.77	14.74
South Khorasan	41.40%	9.7	7.62	8.66	19.77	15.53	17.65

underestimation due to misclassified registry for the total of population. Recently this method is widely used to estimate the burden of cancer; Stamey et al employed Bayesian approach in data consisting of the number of deaths due to cancer and non-cancer, among residents of Hiroshima and Nagasaki, Japan, who experienced the atomic bomb disaster (6). Also these models were employed to re-estimate the GI cancer mortality including colorectal cancer and liver cancer (11, 12).

Conclusion

The important assumption in this technique is to select true prior information regarding the misclassified parameter, which could make changes on the results that extended to the rest of data.

Conflict of Interest

Not Declared

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015; 136:359-86.

2. Pourhoseingholi MA, Vahedi M, Baghestani AR. Burden of gastrointestinal cancer in Asia; an overview. Gastroenterol Hepatol Bed Bench 2015; 8:19-27.

3. Pazdur, Richard. Cancer Management: A Multidisciplinary Approach: Medical, Surgical, & Radiation Oncology. 4th ed. Melville, NY: PRR, 2000.

4. Bernard W, Christopher P. World cancer report 2014. International agency for research on cancer. 2014; World Health Organization, Lyon, France.

5. Burnet NG, Jefferies SJ, Benson RJ, Hunt DP, Treasure FP. 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 2005; 92:241-5.

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

7. Arts DG, Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc 2002;9:600-11.

8. Islamic Repulic of Iran. Ministry of Health and Medical Education. Center for Disease Control & Prevention. Noncommunicable Deputy. Cancer Office. Iranian Annual of National Cancer Registration Report, 2009.

9. Lyles RH. A note on estimating crude odds ratios in case–control studies with differentially misclassified exposure. Biometrics 2002; 58:1034-6.

10. McInturff P, Johnson W, Cowling D, Gardner I. Modeling risk when binary outcomes are subject to error. Stat Med 2004; 23: 1095- 1109.

11. Pourhoseingholi MA, Faghihzadeh S, Hajizadeh E, Abadi A, Zali MR. Bayesian estimation of colorectal cancer mortality in the presence of misclassification in Iran. Asian Pac J Cancer Prev 2009;10:691-4.

12. Pourhoseingholi MA, Fazeli Z, Zali MR, Alavian SM. Burden of hepatocellular carcinoma in Iran; Bayesian projection and trend analysis. Asian Pac J Cancer Prev 2010;11:859-62.

13. Pourhoseingholi MA, Abadi A, Faghihzadeh S, et al. Bayesian analysis of esophageal cancer mortality in the presence of misclassification. Ital J Public Health 2010;8:342-7.

14.Paulino C, Soares P, Neuhaus J. Binomial regression with misclassification. Biometrics 2003;59: 670-75.

15. Paulino CD, Silva G, Achcar J. Bayesian analysis of correlated misclassified binary data. Comput Stat Data Anal 2005;49: 1120-31.

16. Liu Y, Johnson WO, Gold EB, Lasley BL. Bayesian analysis of risk factors for an ovulation. Stat Med 2004;23:1901-19.

17. Pourhoseingholi MA. Bayesian Adjustment for Misclassification in Mortality Data. Biom Biostat Int J 2014;1:10.

18. Kamo K, Kaneko S, Satoh K, Yanagihara H, Mizuno S, Sobue T. A mathematical estimation of true cancer incidence using data from population-

based cancer registries. Jpn J Clin Oncol 2007;37: 150-5.

19. Lankarani KB, Khosravizadegan Z, Rezaianzadeh A, Honarvar B, Moghadami M, Faramarzi H, et al. Data coverage of a cancer registry in southern Iran before and after implementation of a population-based reporting system: a 10-year trend study. BMC Health Serv Res 2013;13:169.