

Original Paper

Under-estimation and Over-estimation in Gastric Cancer Incidence Registry in Khorasan Provinces, Iran

Nastaran Hajizadeh¹, Mohamad Amin Pourhoseingholi^{2*}, Ahmadrza Baghestani¹, Alireza

Abadi¹, Mohammad Reza Zali²

ABSTRACT

Gastric cancer is a disease which the cells forming the inner lining of the stomach start to divide uncontrollably, forming a mass called a tumor. Patients with gastric cancer from low facility provinces like North and South Khorasans may diagnose and registered in full featured provinces like Razavi-Khorasan; this causes misclassification error. The presence of this error makes the registry systems inaccurate and unreliable for estimating the burden of cancer and policy making. Two approaches are recommended for reducing the effects of misclassification; the first is using a small validation sample and the second is a Bayesian analysis which provides subjective prior information for the subset of the parameters to correct the statistics. Data for this study extracted from Iranian annual of national cancer registration report in 2008. The age standardized rate due to gastric cancer [ICD-10; C16] were expressed as rate per/100,000 population for male and female of North, South and Razavi Khorasans. To correct the misclassification effect, a Bayesian approach was used with Poisson count regression and beta prior. The reported expected coverage of cancer incidence for Razavi-Khorasan was 155.5% and exceeds than what expected, whereas the North and South Khorasans have just observed respectively 34.8% and 41.4% of their expected coverage. The results of the Bayesian analysis indicated that there was about 34% misclassification in gastric cancer incidence registry from North and South Khorasans in Razavi-Khorasan. In planning for resource allocation, authorities should consider that, low incidence of gastric cancer in North and South Khorasans, may be the effect of misclassification and it is needed to allocate them more health facilities and improve their address registration accuracy using national ID, electric bill, etc.

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

Gastric cancer is a disease in which cells forming the inner lining of the stomach become abnormal 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 (2), and is first common cancer among Iranian men and

the 3th (after breast cancer and colorectal cancer) among Iranian women (3). 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 (4), the presence of misclassification error

Author Information

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.

Submitted: 26.01.2016
Accepted: 08.06.2016
Published : 01.08.2016

References

1. Pazdur R. Cancer management: a multidisciplinary approach, medical, surgical & radiation oncology: Cmp; 2004.
2. Stewart B, Wild C. World Cancer Report 2014. International Agency for Research on Cancer. World Health Organization. 2014.
3. Ferly J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. Cancer Incidence and Mortality Worldwide. 2012.

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 (5, 6). 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. 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 (7) 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 (8-10).

Materials & Methods

Data for this study extracted from Iranian annual of national cancer registration report in 2008. 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 (5, 8, 9, 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

References

4. Burnet N, Jefferies S, Benson R, Hunt D, Treasure F. Years of life lost (YLL) from cancer is an important measure of population burden—and should be considered when allocating research funds. *British journal of cancer*. 2005;92(2):241-5.
5. Stamey JD, Young DM, Seaman JW. A Bayesian approach to adjust for diagnostic misclassification between two mortality causes in Poisson regression. *Statistics in medicine*. 2008;27(13):2440-52.
6. Arts DG, De Keizer NF, Scheffer G-J. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *Journal of the American Medical Informatics Association*. 2002;9(6):600-11.
7. Lyles RH. A note on estimating crude odds ratios in case-control studies with differentially misclassified exposure. *Biometrics*. 2002;58(4):1034-6.
8. Whittemore AS, Gong G. Poisson regression with misclassified counts: application to cervical cancer mortality rates. *Applied Statistics*. 1991:81-93.
9. Sposto R, Preston DL, Shimizu Y, Mabuchi K. The effect of diagnostic misclassification on non-cancer and cancer mortality dose response in A-bomb survivors. *Biometrics*. 1992:605-17.
10. McInturff P, Johnson WO, Cowling D, Gardner IA. Modelling risk when binary outcomes are subject to error. *Statistics in medicine*. 2004;23(7):1095-109.
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(4):691-4.

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.

Table1. Number of gastric cancer incidence before and after Bayesian correction for misclassification

| Province | Before Bayesian correction | | | After Bayesian correction | | |
|-----------------|----------------------------|--------|------|---------------------------|--------|------|
| | Male | Female | Both | Male | Female | Both |
| North Khorasan | 42 | 15 | 57 | 94 | 33 | 127 |
| Razavi Khorasan | 550 | 194 | 744 | 471 | 166 | 637 |
| South Khorasan | 22 | 13 | 35 | 45 | 26 | 71 |

Discussion

Accurate cancer incidence data are essential to planning, monitoring and evaluating national and regional cancer control programs (16). 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 some 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 (18). 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).

References

- 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(4):859-62.
- Pourhoseingholi MA, Abadi A, Faghizadeh S, Pourhoseingholi A, Vahedi M, Moghimi-Dehkordi B, et al. Bayesian analysis of esophageal cancer mortality in the presence of misclassification. *Italian Journal of Public Health*. 2012;8(4).
- Daniel Paulino C, Soares P, Neuhaus J. Binomial regression with misclassification. *Biometrics*. 2003;59(3):670-5.
- Paulino CD, Silva G, Achcar JA. Bayesian analysis of correlated misclassified binary data. *Computational statistics & data analysis*. 2005;49(4):1120-31.
- Kamo K-i, Kaneko S, Satoh K, Yanagihara H, Mizuno S, Sobue T. A mathematical estimation of true cancer incidence using data from population-based cancer registries. *Japanese journal of clinical oncology*. 2007;37(2):150-5.
- Liu Y, Johnson WO, Gold EB, Lasley BL. Bayesian analysis of risk factors for anovulation. *Statistics in medicine*. 2004;23(12):1901-19.
- 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 services research*. 2013;13(1):169.