Bayesian adjustment of gastric cancer mortality rate in the presence of misclassification

Nastaran Hajizadeh, Mohamad Amin Pourhoseingholi, Ahmad Reza Baghestani, Alireza Abadi, Mohammad Reza Zali

Bayesian adjustment of gastric cancer mortality rate in the presence of misclassification

Aim
To correct for misclassification error in registering causes of death in Iran death registry using Bayesian method.

Methods
National death statistic from 2006 to 2010 for gastric cancer which reported annually by the Ministry of Health and Medical Education included in this study. To correct the rate of gastric cancer mortality with reassigning the deaths due to gastric cancer that registered as cancer without detail, a Bayesian method was implemented with Poisson count regression and beta prior for misclassified parameter, assuming 20% misclassification in registering causes of death in Iran.
RESULTS
Registered mortality due to gastric cancer from 2006 to 2010 was considered in this study. According to the Bayesian re-estimate, about 3%-7% of deaths due to gastric cancer have registered as cancer without mentioning details. It makes an undercount of gastric cancer mortality in Iranian population. The number and age standardized rate of gastric cancer death is estimated to be 5805 (10.17 per 100000 populations), 5862 (10.51 per 100000 populations), 5731 (10.23 per 100000 populations), 5946 (10.44 per 100000 populations), and 6002 (10.35 per 100000 populations), respectively for years 2006 to 2010.

CONCLUSION
There is an undercount in gastric cancer mortality in Iranian registered data that researchers and authorities should notice that in sequential estimations and policy making.

Key words: Misclassification; Bayesian method; Cause of death; Gastric cancer; Iran © The Author(s) 2017. Published by Baishideng Publishing Group Inc. All rights reserved.

Core tip: In some mortality cases, causes of deaths are registered as causes that cannot or should not be considered as the underlying causes of death like cancer without mentioning the type. These cases are not included in the estimations of cause specific mortality rates and leads to under-estimate health risks and burden of disease. The aim of this study is to correct the misclassification of gastric cancer deaths in cancer without label group using a Bayesian method.

INTRODUCTION
Cancer is one of the major health problems in the world and is the third cause of death (after cardiovascular disease and injuries) in Iran[1]. Gastric cancer is a disease in which the cells of the inner lining of the stomach start to divide abnormally and uncontrollably, that forming a mass called tumor[2]. Gastric cancer is the seventh cause of all deaths in Iran and is the first cause of cancer death in Iranian men and the second cause of cancer death (after breast cancer) in Iranian women[3]. The mortality of gastric cancer is high because this cancer does not show symptoms in early stages and diagnosed when the cancer is in its final stages[4]. Burden of disease is used to evaluate the health status of a country and determining priority of risk factors in order to setup cancer control programs. Cancer registry data are important to estimate the burden of disease, monitoring the screening programs effects, early diagnostics and other prognostic factors, and can be used to guide policy makers to appropriate cancer prevention programs. Among medical indices, mortality is a familiar projection to assess the burden of diseases. But achieving this aim requires a reliable death registry systems that reports death statistics accurately and completely[5-7]. In Iran, among four vital events (births, marriages, divorces and mortality) which were registered by the National Organization for Civil Registration (NOCR), mortality was the worst in quality. There was some progress in registering deaths but some problems like delayed registration and inaccurate recording of causes of death remained until 2002, that Ministry of health and medical education Deputy of Research and Technology, started up a system to record the causes of deaths. This system did not allow to delayed deaths registry, but the causes of death were susceptible to information bias due to misclassification[8]. Most high-income and many middle-income countries have a complete vital registration system in which the majority of deaths get a death certificate completed by a physician[9]. But still, a number of causes of death in the process of completing death certificates and the coding of underlying cause of death based on standardized international rules, remains challenging[10-13]. In some cases, especially in developing countries, the cause of death is recorded with error[14,15]. For example if a death due to gastric cancer being labeled as a death due to any other cause, the misclassification error in outcome is occurs. Misclassification error makes the registered data inaccurate and often leads to major problems like biased estimates of burden and health risks in epidemiological analysis[16,17].

According to the Iranian death registry, about 15% to 20% of death statistics are recorded in misclassified categories such as cardiopulmonary arrest, old age without dementia, septicemia, unknown, cancer without mention of details, and other ill-defined conditions. Murray and Lopez in 1996, for the first time, introduced the term “garbage coding” for assigning deaths to causes that are not useful for public health analysis of cause-of-death data[18-21].

In developing countries like Iran that registration is not completely accurate, statistical methods can be very helpful to overcome this problem. Two statistical approaches are recommended to deal with misclassification; first is using a small valid sample and extending the results to the population[22] and the second is Bayesian analysis which is a flexible method that makes the possibility of combining the prior information regarding the subset of the parameters with the observed data to achieve a posterior distribution which will be the basis of inferences to correct the statistics. Bayesian models also can easily accommodate unobserved variables such as an individual’s true information in the presence of Misclassification error[23]. The aim of this study is to use
The initial value for the parameter of beta distribution for the misclassified parameter, mortality from gastric cancer as mortality due to cancer considered to be the probability of incorrectly registering a cause of death instead of a specific cancer like gastric cancer with no label is most likely to be registered as cancer without label because of gastric cancer in Iran’s cancer registry system.

Bayesian method to estimate the rate of misclassification that occurs by registering cancer (with no label) as the cause of death instead of deaths that have occurred because of gastric cancer in Iran’s cancer registry system.

MATERIALS AND METHODS

Mortality rates due to gastric cancer and also cancer without label from 2006 to 2010 are extracted from Iranian annual of death statistics which reported annually by Iran’s Ministry of Health and Medical Education, in two sex groups (male and female) and four age groups (under 15 years - 15 to 49 years - 50 to 69 years - 70 years and more).

To reassign deaths from garbage codes to valid causes, the approach can be divided into three steps: The first is identifying garbage codes. The second is identifying the target causes where the deaths assigned to a garbage code should in principle be reassigned to; for example if a death cause is registered as cancer and the type of cancer is not mentioned, we face with a garbage code that should be reassigned to a specific cancer. The third step is choosing the fraction of deaths that are assigned to the garbage code that should be reallocated to the target cause. In this study we consider cancer without label as garbage code because cancer with no label is most likely to be registered as cause of death instead of a specific cancer like gastric cancer. The data were entered to the Bayesian model by two vectors $y_1 = [y_{11}, y_{12}, \ldots, y_{1r}]$ for gastric cancer and $y_2 = [y_{21}, y_{22}, \ldots, y_{2r}]$ for cancer without label. Both $y_1$ and $y_2$ are count data and follow the Poisson distribution. The subscript $r$ shows the number of covariate patterns that is made by age and sex group combinations. $\theta$ is considered to be the probability of incorrectly register a mortality from gastric cancer as mortality due to cancer without label group. To perform Bayesian inference, an informative beta prior distribution was assumed for the misclassified parameter, i.e., $\theta \sim \text{Beta}(a, b)$. The initial value for the parameter of beta distribution are taken to be $a = 20$ and $b = 80$, based on Iranian annual cancer registration reports. Since $\nu$ (misclassified parameter) is an unknown parameter, a latent variable approach was employed to simplify the full conditional models; considering $U_i | \theta_1, \theta_2 \sim \text{Binomial}(y_{i2}, P)$ as the number of counts from the first group that are incorrectly labeled as being in the misclassified group that $P = (\lambda_1\theta)/(\lambda_1\theta + \lambda_2)$, finally the posterior distribution appears in the following form; $\theta \sim \text{Beta}(\Sigma U + a, \Sigma y + b)$. The misclassified parameter is estimated using a Gibbs sampling algorithm and averaging of the outcome. Analyses were done using R software version 3.2.0.

RESULTS

Mortality data consisting of all deaths due to gastric cancer from 2006 to 2010 were considered in this study. Age standardized rate (ASR) of gastric cancer mortality was 9.69 per 100000 populations in 2006, 10.2 per 100000 populations in 2007, 9.93 per 100000 populations in 2008, 9.76 per 100000 populations in 2009 and 9.67 per 100000 populations in 2010 respectively. According to the Bayesian estimation, in year 2006, there was between 3% to 7% misclassification in registering cause of death as cancer without mentioning details while the underlying cause of death has been gastric cancer. The estimated percent of misclassification based on implemented Bayesian method for year 2006 to 2010 is shown in Table 1. This percent were subtracted from deaths that had registered as cancer without mentioning details and added to the number of deaths due to gastric cancer. The age standardized rate per 100000 populations for gastric cancer was estimated to be 10.17 in 2006, 10.51 in 2007, 10.23 in 2008 10.44 in 2009 and 10.35 in 2010, after Bayesian correction respectively. The age standardizes rate of gastric cancer before and after Bayesian correction for 2006 to 2010 is visualized in Figure 1. The number of gastric cancer death before and after Bayesian correction of misclassification for years 2006 to 2010 is shown in Table 1 and its trend is shown in Figure 2.

DISCUSSION

Iran’s death registry is subject to misclassification in reporting the underlying cause of death. About 3%-7% of deaths due to gastric cancer are registered as cancer without mentioning the type of cancer. After correcting misclassification error in death registry data, the number of deaths due to gastric cancer and its age standardized rate were increased. Gastric cancer crude mortality count in Iran had an increasing trend from year 2006 to 2010 except for 2008 that might be because of incompleteness of data; but the age standardized rate of gastric cancer was decreasing from year 2007 onward (except for 2008). About two-thirds of gastric cancer occurs in developing countries and its rates are generally about twice as high in men as in women. The age standardized rate (ASR) of gastric cancer incidence and
mortality per 1000000 populations based on GLOBOCAN report 2012 is shown in Table 2. The rates show that the ASR of gastric cancer incidence (15.8 per 100000) and also the ASR of gastric cancer mortality (11.7 per 100000) is highest in Asia compared to other continents; It is moderate in Europe and South America and lowest in Northern America and most parts of Africa[3,28].

The age standardized rates of incidence and mortality per 1000000 populations in different regions of Asia based on GLOBOCAN report 2012 are shown in Table 3. The incidence and mortality rates are also higher in Eastern Asia in comparison with other Asian regions. This region includes China, Japan and South Korea, that are three countries with the highest gastric cancer incidence and mortality rates[29]. Gastric cancer is the most frequently diagnosed form of cancer in Iran[30], with incidence rate 15.3 per 100000 and mortality rate 12.9 per 100000 populations based on GLOBOCAN report 2012[31]. A steady decline has been observed in gastric cancer incidence and mortality rates in the most of countries in Northern America and Europe since the middle of the 20th century[31,32]. In recent years similar decreasing trends have been noted in areas with high rates of gastric cancer history, including some countries in Asia (Japan, China, and South Korea), Latin America (Colombia and Ecuador), and Europe (Ukraine)[33]. This reduction maybe due to improved sanitation and antibiotics and consequently reduction in chronic H. pylori infection[34]. Although the age-adjusted rates have been decreased, it is estimated to have a substantial rise in the crude rates between the years 2000 to 2020 because of the increasing the size and age of the world population, especially in developing countries[14,36].

Gastric cancer is a major health problem in the world, especially in Asia. So it is needed to make appropriate policy making for allocation of resources for gastric cancer control and prevention. To achieve this aim an accurate registry system is needed, while there are some misclassifications in registering causes of death especially in developing countries[14,35]. Misclassification of causes of death has been a concern in cancer trends analysis and researches on cancer epidemiology for decades[34]. Misclassification error leads to under-estimation of cause specific mortality rates and consequently under-estimation in burden of disease and influences the policy makings in burden of disease and influences the policy makings especially in Asia. So it is needed to make appropriate policy making for allocation of resources for gastric cancer control and prevention. To achieve this aim an accurate registry system is needed, while there are some misclassifications in registering causes of death especially in developing countries[14,35]. Misclassification of causes of death has been a concern in cancer trends analysis and researches on cancer epidemiology for decades[34]. Misclassification error leads to under-estimation of cause specific mortality rates and consequently under-estimation in burden of disease and influences the policy makings especially in developing countries[14,35]. Misclassification of causes of death has been a concern in cancer trends analysis and researches on cancer epidemiology for decades[34]. Misclassification error leads to under-estimation of cause specific mortality rates and consequently under-estimation in burden of disease and influences the policy makings especially in developing countries[14,35]. Misclassification of causes of death has been a concern in cancer trends analysis and researches on cancer epidemiology for decades[34]. Misclassification error leads to under-estimation of cause specific mortality rates and consequently under-estimation in burden of disease and influences the policy makings especially in developing countries[14,35]. Misclassification of causes of death has been a concern in cancer trends analysis and researches on cancer epidemiology for decades[34]. Misclassification error leads to under-estimation of cause specific mortality rates and consequently under-estimation in burden of disease and influences the policy makings especially in developing countries[14,35]. Misclassification of causes of death has been a concern in cancer trends analysis and researches on cancer epidemiology for decades[34]. Misclassification error leads to under-estimation of cause specific mortality rates and consequently under-estimation in burden of disease and influences the policy makings especially in developing countries[14,35]. Misclassification of causes of death has been a concern in cancer trends analysis and researches on cancer epidemiology for decades[34]. Misclassification error leads to under-estimation of cause specific mortality rates and consequently under-estimation in burden of disease and influences the policy makings especially in developing countries[14,35]. Misclassification of causes of death has been a concern in cancer trends analysis and researches on cancer epidemiology for decades[34]. Misclassification error leads to under-estimation of cause specific mortality rates and consequently under-estimation in burden of disease and influences the policy makings especially in developing countries[14,35]. Misclassification of causes of death has been a concern in cancer trends analysis and researches on cancer epidemiology for decades[34].
By using the Bayesian approach, it is not needed to valid the data for estimating the rate of misclassification. Data validation is very costly and time consuming. By using the Bayesian method, it is not needed to valid the data for estimating rapid and cost-effective. The hotspot of this study is using the Bayesian method for registering causes of death. Misclassification is lack of agreement between the observed value and the true value in categorical data. Bayesian method is one of the statistical approaches that assign a distribution or a probability to events or parameters based on previous experience or an expert's idea and revise those probabilities and distributions after obtaining experimental data with applying Bayes' theorem.

In conclusion there is an undercount of gastric cancer mortality in Iranian registration system. Because of misclassification error in registering causes of death. Although it seems that the misclassification rate has been reduced, it still exists as a major problem. So, policy makers who use mortality data to determine priorities for disease control and prevention, should notice to this underreported data and registration of causes of deaths should be done more accurately. Increase in data accuracy, requires more expert staffing, refining foundations, and powerful hardware and software resources. In the absence of valid data, Bayesian approach is a good and flexible alternative to reduce the effects of Misclassification in registered cancer mortality data.

**REFERENCES**

1. **Pourhoseingholi MA**, Fazeli Z, Ashouri S, Bavand-Pour FS. Mortality trends of gastrointestinal cancers in Iranian population. *Gastroenterol Hepatol Bed Bench* 2013; 6: S52-S57 [PMID: 24834288]
2. **Pazdur R**, Coia LR, Hoskins WJ. Cancer management: a multi-disciplinary approach: Medical, surgical and radiation oncology. Cmmp. 2004
3. **Ferlay J**, Soerjomataram I, Ervik M, Dikshit R, Mathers C, Rebelo M, Forman D, Bray, F. GLOBCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: URL: http://publications.iarc.fr/Databases/Iarc-Cancerbases/Globocan-2012-Estimated-Cancer-Incidence-Mortality-And-Prevalence-Worldwide-In-2012-V1-0-2012
4. **Pourhoseingholi MA**, Moghim-Dekhkordi B, Safae A, Hajizadeh E, Soltipour A, Zali MR. Prognostic factors in gastric cancer using log-normal censored regression model. *Indian J Med Res* 2009; 129: 262-267 [PMID: 19491418]
5. **Stamey JD**, Young DM, Seaman JW. A Bayesian approach to adjust for diagnostic misclassification between two mortality causes in Poisson regression. *Stat Med* 2008; 27: 2440-2452 [PMID: 17979218 DOI: 10.1002/sim.3134]
6. **Pourhoseingholi MA**, Vahedi M, Moghimi-Dekhkordi B, Pourhoseingholi A, Ghafarnejad F, Masenat E, Safae A, Mansoori BK, Zali MR. Burden of hospitalization for gastrointestinal tract cancer patients - Results from a cross-sectional study in Tehran. *Asian Pac J Cancer Prev* 2009; 10: 107-110 [PMID: 19469635]
7. **Sharifian A**, Pourhoseingholi MA, Baghestani A, Hajizadeh N, Gholizadeh S. Burden of gastrointestinal cancers and problem of the incomplete information; how to make up the data? *Gastroenterol Hepatol Bed Bench* 2016; 9: 12-17 [PMID: 26744699]
8. **Khosravi A**, Taylor R, Naghavi M, Lopez AD. Mortality in the Islamic Republic of Iran, 1964-2004. *Bull World Health Organ* 2007; 85: 607-614 [PMID: 17768519 DOI: 10.2471/BLT.06.038802]
9. **Mathers CD**, Fat DM, Inoue M, Rao C, Lopez AD. Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bull World Health Organ* 2005; 83: 171-177 [PMID: 15798840]
10. **Perey C**, Stanek E, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. *Am J Public Health* 1981; 71: 242-250 [PMID: 7468855 DOI: 10.2105/AJPH.71.3.242]
11. **Gobbato F**, Vecchiet F, Barbierato D, Melato M, Marconi R. Inaccuracy of death certificate diagnoses in malignancy: an analysis of 1,405 autopsied cases. *Hum Pathol* 1982; 13: 1036-1038 [PMID: 7152507 DOI: 10.1016/S0046-8177(82)80096-8]
12. **Selikoff II**, Seidman H. Use of death certificates in epidemiological
studies, including occupational hazards: variations in discordance of different asbestos-associated diseases on best evidence ascertainment. Am J Ind Med 1992; 22: 481-492 [PMID: 1442783]

13 Naghavi M, Maleka S, Foreman K, O’Brien J, Poumalae F, Lozano R. Algorithms for enhancing public health utility of national causes-of-death data. Popul Health Metr 2010; 8: 9 [PMID: 20459720 DOI: 10.1186/1478-7954-8-9]

14 Edwards JK, Cole SR, Chu H, Olshan AF, Richardson DB. Accounting for outcome misclassification in estimates of the effect of occupational asbestos exposure on lung cancer death. Am J Epidemiol 2014; 179: 641-647 [PMID: 24352593 DOI: 10.1093/aje/kvt509]

15 Yavari P, Sadlohefazi B, Mohangleh MA, Madani H, Mosazadeh A, Naqiyjoo A, Meneb Y, Pourheingholi MA. An epidemiological analysis of cancer data in an Iranian hospital during the last three decades. Asian Pac J Cancer Prev 2008; 9: 145-150 [PMID: 18439094]

16 Burnet NG, Jefferies SJ, Benson R, Hunt DP, Treasure FF. 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: 241-245 [PMID: 15655548 DOI: 10.1038/sj.bjc.6602221]

17 Aris DG, De Keizer NF, Scheffer GI. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc 2002; 9: 600-611 [PMID: 12386611 DOI: 10.1197/jamia.M1087]

18 Lopez AD, Murray CJ, editors. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Harvard: Harvard School of Public Health, 1996

19 Johansson LA, Pavillon G, Anderson R, Glenn D, Griffiths C, Hoyert D, Jackson G, Notzon FS, Rooney C, Rosenberg HM, Walker S, Weber S. Counting the dead and what they died of. WJGO 2015; 65: 1034-1036 [PMID: 12499516 DOI: 10.1111/j.0006-341X.2002.10.2471]

20 Mclnturf P, Johnson WO, Cowling D, Gardner IA. Modelling risk when binary outcomes are subject to error. Stat Med 2004; 23: 1095-1109 [PMID: 15705880 DOI: 10.1002/sim.1656]

21 Khorasani A, Aghashamani S, Kazemi E, Pour Malek F, Shariati M. Mortality Profile in Iran (29 Provinces) over the Years 2006 to 2010. Int J Cancer 2015; 125: 145-150 [PMID: 24382179 DOI: 10.1002/ijc.24290]

22 Parkin DM. The global health burden of infection-associated cancers in the year 2002. Int J Cancer 2006; 118: 3030-3044 [PMID: 16404738 DOI: 10.1002/ijc.21731]

23 Pisani P, Parkin DM, Bray F, Ferlay J. Estimates of the worldwide mortality from 25 cancers in 1990. Int J Cancer 1999; 83: 18-29 [PMID: 10449602]

24 Pourheingholi MA, Vahedi M, Baghestani AR. Burden of gastrointestinal cancer in Asia: an overview. Gastroenterol Hepatol Bed Bench 2015; 8: 19-27 [PMID: 25584172]

25 Cameron HM, McGooan E. A prospective study of 1152 hospital autopsies: II. Analysis of inaccuracies in clinical diagnoses and their significance. J Pathol 1981; 133: 285-300 [PMID: 7241268 DOI: 10.1002/path.1711330403]

26 Khorasani A, Rao C, Naghavi M, Taylor R, Jafari N, Lopez AD. Impact of misclassification on measures of cardiovascular disease mortality in the Islamic Republic of Iran: a cross-sectional study. Bull World Health Organ 2008; 86: 688-696 [PMID: 18797644 DOI: 10.2471/BLT.07.046532]

27 Whittemore AS, Gong G. Poisson regression with misclassified counts: application to cervical cancer. J R Stat Soc Ser C Appl Stat 1991; 40: 81-93 [PMID: 12157994]

28 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; 48: 605-617 [PMID: 1637983 DOI: 10.2307/25232315]

29 Stamey JD, Bratcher TL, Young DM. Parameter subset selection and multiple comparisons of Poisson rate parameters with misclassification. Comput Stat Data An 2004; 45: 467-479 [DOI: 10.1016/j.csda.2003.08.008]

30 Pourheingholi MA, Abadi A, Faghizadeh S, Pourheingholi A, Vahedi M, Moghim-Dekhori B, Safaei A, Zali MR. Bayesian analysis of esophageal cancer mortality in the presence of misclassification. Italian Journal of Public Health 2012; 8 [DOI: 10.2427/5680-7486-1-PB]

31 Pourheingholi MA. Bayesian adjustment for misclassified death registration data. Transl Gastrointest Cancer 2014; 3: 144-148 [DOI: 10.3978/j.issn.2224-4778.2014.08.08]

32 Pourheingholi MA, Faghizadeh S, Haji-Zadeh E, Abadi A. Bayesian analysis of gastrointestinal cancer mortality in Iran. Gastroenterol Hepatol Bed Bench 2010; 3: 15-18

33 Lankarani KB, Khosravizadegan Z, Rezaianzadeh A, Honavar B, Moghadami M, Faramarzi H, Mahmoodi M, Farahmand M, Masoompour SM, Nazemzadeh B. 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 [PMID: 23647828 DOI: 10.1186/1472-6963-13-169]

P- Reviewer: Ayoghi K, Deans C, Shen LZ  S- Editor: Kong JX  L- Editor: A  E- Editor: Lu YJ
