BAISHIDENG PUBLISHING GROUP INC

8226 Regency Drive, Pleasanton, CA 94588, USA
Telephone: +1-925-223-8242 Fax: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com http://www.wjgnet.com

Name of Journal: *World Journal of Gastrointestinal Oncology*

ESPS Manuscript NO: 29510

Manuscript Type: Original Article

Observational Study

Explanation of method with more details:

Gastric cancer incidence data is extracted from Iranian annual of national cancer registration report in 2008. Firstly, the Age Standardized Rate (ASR) for gastric cancer (coded according to the 10th revision of the International Classification of Diseases [ICD-10; C16]) is calculated for all provinces of Iran using direct standardization method and the standard population in WHO 2000 for both genders and four age groups (0-14 years, 15-49 years, 50-69 years and over than 70 years old) in order to compare statistics on cancer in Iran with those for the rest of the world. Then expected coverage of cancer cases is calculated for medical university of each province. The process of cancer incidence registration is that diagnosed cancer cases by pathology centers and other diagnostic centers in the country are entered in software that is made by ministry of health. Medical university of each province send the information recorded in this temporary data bank to the ministry of health. Ministry of health after coding the recorded cancers based on 10th revision of international coding of disease and removing duplicates, makes a permanent data bank of cancer cases and sends it back to medical university of each province. So each medical university has an observed number of cancer cases and also has an expected coverage of cancer cases that is considered to be 113 per 100000. By dividing the observed number to the expected number of cancer cases, the percent of expected coverage for each province is calculated. For entering the data to the Bayesian model two vectors y_1 and y_2 were used. Vector $y_1 = [y_{11}, y_{21}, \dots, y_{r1}]'$ for the province that has an expected coverage less than 100% and vector

BAISHIDENG PUBLISHING GROUP INC
8226 Regency Drive, Pleasanton, CA 94588, USA
Telephone: +1-925-223-8242 Fax: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com http://www.wjgnet.com

$y_2 = [y_{12}, y_{22}, \ldots, y_{r2}]'$ for a neighboring province with a more than 100% expected coverage. Subscript r shows the number of covariate patterns that is made by age group and sex group combinations. A Poisson distribution was considered for count data y_1 and y_2. An informative beta prior distribution was assumed for θ as the probability of registering a data in misclassified group; so $\theta \sim beta(a, b)$. For selecting prior value for the parameters of beta distribution, the calculated expected coverage for the medical university which has a lower than 100% expected coverage was used as b and a was calculated with subtracting b from 100. Thus $a/(a+b)$ which is the expectation of beta distribution converges to the misclassified rate. Since θ is unknown, a latent variable approach was employed to correct the misclassification effect. The latent variable U with binomial distribution i.e. $U_i \mid \theta, y_1, y_2 \sim \text{Binomial}(y_{i2}, P_i)$ that $P_i = \frac{\lambda_{i1}\theta}{\lambda_{i1}\theta + \lambda_{i2}}$, was considered as the number of events from the first group that are incorrectly registered in the misclassified group. Finally by using a Gibbs sampling algorithm, the posterior distribution appears in the following form; $\theta \mid U_i, y_1, y_2 \sim \text{Beta}(\sum_i U_i + a, \sum_i y_{i1} + b$. After estimating the misclassification rate between each two neighboring provinces, the rates of gastric cancer incidence for each province were re-estimated. Analyses were carried out using R software version 3.2.0.

Explanation about data:

The reports of cancer incidence is published annually by national cancer registration. But it has some delay. So the report of year 2008 was the last available version to use.