

Bayesian adjustment for misclassification in mortality data

Editorial

In medical studies, a difficulty in drawing inference from categorical data is the existence of misclassification. Misclassification is the disagreement between the observed and the true value. Sick individuals may be diagnosed as healthy or the causes of diseases or death may be misjudged. There is an effect of misclassification on estimation and hypothesis testing, often leading to biased estimates, and can therefore cause one to underestimate health risks.¹ The effect of misclassification was first noted by Bross² and in statistical literature, two approaches are recommended, the first; using a small validation sample³ and the second; Bayesian analysis in which subjective prior information on at least some subset of the parameters is used to re-estimate misclassified statistic.⁴⁻⁶

The difficulty of first approach is that without the presence of additional information beyond the correct data, it is not possible to take into account the effect of misclassification and the difficulty with re-sampling is the necessity of an infallible classifier, which may not exist or may be expensive.⁷ On the other hand, the Bayesian literature on this topic is steadily growing. Most importantly, more complex model could be handled due to development of computational techniques (e.g., Monte Carlo methods).

Among medical indexes, mortality is a familiar projection in the assessment of the burden of diseases. But this aim needs reliable death registry systems which reports death statistics, annually and accurately. Besides, the analysis of death statistics subject to misclassification is a major problem in epidemiological analysis.¹ Although the World Health Organization (WHO) has encouraged member states to introduce systems of death registration involving medical certification of the cause of death, the misclassification or underestimation of mortality data is still happened in official statistics, most of them in developing countries.

Bayesian approach received much attention in the case of misclassification for mortality data. Whittemore and Gong incorporated supplemental data on both true and fallible disease and used this approach to estimate cervical cancer mortality rates in Poisson regression⁴ and Sposto et al.,⁵ developed this likelihood to assess the effect of diagnostic misclassification on non-cancer and cancer mortality dose-response. Stamey et al.,¹ provided a Bayesian approach, which extends the models of Whittemore & Gong⁴ and Sposto et al.,⁵ But their technique dose not assumes that the misclassification parameters are known. Also, the prior information on the misclassification parameters would be used instead of validation data.² They used this Bayesian approach in data consisting of the number of deaths due to cancer and non-cancer among residents of Hiroshima and Nagasaki, Japan. We derived an extension of models proposed by Stamey et al.,¹ to correct and account for misclassification in cancer mortality data.⁸ Suppose there are two sample groups for death classification; $y_1 = [y_{11}, y_{21}, \dots, y_{r1}]$ and $y_2 = [y_{12}, y_{22}, \dots, y_{r2}]$ where r is the covariate pattern, y_1 is the exact cause of death and y_2 is the misclassified group in which the cause of death in the first group was incorrectly labeled, and

Volume 1 Issue 2 - 2014

Mohamad Amin Pourhoseingholi

Gastroenterology and Liver diseases Research Center, Shahid Beheshti University of Medical Sciences, Iran

Correspondence: Mohamad Amin Pourhoseingholi, Gastroenterology and Liver diseases Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran, Tel 982122432515, Fax 982122432517, Email amin_phg@yahoo.com

Received: September 28, 2014 | **Published:** November 11, 2014

$y_1 \sim \text{Poisson}(P_i \mu_{i1})$ and $y_2 \sim \text{Poisson}(P_i \mu_{i2})$ in which μ_i is the observed rate of death mortality for the covariate pattern in Stamey et al.,¹ approach, there is a possibility of two way incorrectly labeled but in our approach, just one group supposed to be misclassified because of the nature of real data). Let θ be the probability that an observation from group 1 is incorrectly labeled in group 2. If the actual rate of death for each group (unknown) is supposed to be as λ_i , the relation between actual rate and observed rate can be written in following form; $\mu_{i1} = \lambda_{i1}(1 - \theta)$ and $\mu_{i2} = \lambda_{i2} + \lambda_{i1}\theta$. The joint distribution of the observable mortality data in this case of misclassification is proportional to;

$$\prod_{i=1}^r [\lambda_{i1}(1 - \theta)^{y_{i1}}][\lambda_{i2} + \lambda_{i1}\theta]^{y_{i2}} \exp\{-P_i[\lambda_{i1}(1 - \theta)] - P_i[\lambda_{i2} + \lambda_{i1}\theta]\}.$$

To perform Bayesian inference, one can assume beta prior distribution for the misclassified parameter, i.e. $\theta \sim \text{beta}(a, b)$. Because θ is an unknown parameter, we employed a latent variable approach according to Paulino et al.,^{9,10} Liu et al.,¹¹ and Stamey et al.,¹ to simplify the full conditional models and estimate the posterior distribution using a Gibbs sampling algorithm.

In this case, we define $U_i | \beta_1, \beta_2, \theta, y_1, y_2 \sim \text{Binomial}(y_{i2}, P_i)$ to be the number of counts from the first group incorrectly labeled as being in the misclassified group. So; $P_i = \frac{\lambda_{i1}\theta}{\lambda_{i1}\theta + \lambda_{i2}}$ and finally the posterior appears in the following form;

$\theta | \beta_1, \beta_2, U_i, y_1, y_2 \sim \text{beta}(\sum_i U_i + a, \sum_i y_{i2} + b)$. This approach was employed to correct the misclassification in cancer mortality data.^{12,13} In the absence of double sampling or valid data, Bayesian approach would be a good alternative to eliminate the effects of misclassification for mortality data, typically for death statistics of developing countries, which data are subject to misclassification or under-reporting. Bayesian technique is flexible and easily handled by computational analysis.

Acknowledgements

None.

Conflict of interests

Authors declare that there is no conflict of interest

Funding

None.

References

1. Stamey JD, Young DM, Seaman JW. A Bayesian approach to adjust for diagnostic misclassification between two mortality causes in Poisson regression. *Statist Med.* 2008;27(13):2440–2452.
2. Bross I. Misclassification in 2×2 tables. *Biometrics.* 1954;10(4):478–486.
3. Lyles RH. A note on estimating crude odds ratios in case–control studies with differentially misclassified exposure. *Biometrics.* 2002;58(4):1034–1036.
4. Whittemore AS, Gong G. Poisson regression with misclassified counts: application to cervical cancer. *J R Stat Soc Ser C Appl Stat.* 1991;40(1):81–93.
5. Sposto R, Preston DL, Shimizu Y, et al. The effect of diagnostic misclassification on non–cancer and cancer mortality dose–response in A–bomb survivors. *Biometrics.* 1992;48(2):605–617.
6. McInturff P, Johnson W, Cowling D, et al. Modeling risk when binary outcomes are subject to error. *Stat Med.* 2004;23(7):1095–1109.
7. Swartz T, Haitovsky Y, Vexler A, et al. Bayesian identifiability and misclassification in multinomial data. *The Canadian Journal of Statistics.* 2004;32(3):1–18.
8. Pourhoseingholi MA, Faghihzadeh S, Hajizadeh E, et al. Bayesian estimation of colorectal cancer mortality in the presence of misclassification in Iran. *Asian Pac J Cancer Prev.* 2009;10(4):691–694.
9. Paulino C, Soares P, Neuhaus J. Binomial regression with misclassification. *Biometrics.* 2003;59(3):670–675.
10. Paulino CD, Silva G, Achcar J. Bayesian analysis of correlated misclassified binary data. *Computational Statistics & Data Analysis.* 2005;49(4):1120–1131.
11. Liu Y, Johnson WO, Gold EB, et al. Bayesian analysis of risk factors for an ovulation. *Statistics in Medicine.* 2004;23(12):1901–1919.
12. Pourhoseingholi MA, Faghihzadeh S, Hajizadeh E, et al. Bayesian Analysis of Gastric Cancer mortality in Iranian Population. *Gastroenterol Hepatol Bed Bench.* 2010;3:15–18.
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;8:342–347.