Estimating the Esophagus Cancer Incidence Rate in Ardabil, Iran: A Capture-Recapture Method

Mahmoud Khodadost,[1,2] Parvin Yavari,[3,4,*] Behnam Khodadost,[5] Masoud Babaei,[6,7] Fatemeh Sarvi,[8] Seyed Reza Khatibi,[2] and Saeed Barzegari[9]

[1]Gastroenterology and Liver Diseases Research Center, Baqiyatallah University of Medical Sciences, Tehran, IR Iran
[2]Department of Epidemiology, Faculty of Health, Iran University of Medical Sciences, Tehran, IR Iran
[3]Department of Health and Community Medicine, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran
[4]Cancer Research Center, Shahid Beheshti University of Medical Sciences, Tehran , IR Iran
[5]Department of Epidemiology, Faculty of Health, Shiraz University of Medical Sciences, Shiraz, IR Iran
[6]Health Deputy, Ardabil University of Medical sciences, Ardabil, IR Iran
[7]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
[8]Epidemiology and Biostatistics Department, School of Public Health, Hamadan University of Medical Sciences, Hamadan, IR Iran
[9]Department of Health Information Technology, Amol Faculty of Paramedical Sciences, Mazandaran University of Medical Sciences, Sari, IR Iran

Corresponding author: Parvin Yavari, Department of Health and Community Medicine, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran. Tel: +98-9121887593, Fax: +98-2122439936, E-mail: p.yavari-grc@sbmu.ac.ir

Received 2015 August 31; **Accepted** 2015 September 22.

Abstract

Background: Accurate cancer registry and awareness of cancer incidence rate is essential in order to define strategies for cancer prevention and control programs. Capture-recapture methods have been recommended for reducing bias and increase the accuracy of cancer incidence estimation.

Objectives: This study aimed to estimate the esophagus cancer incidence by capture-recapture method based on Ardabil population-based cancer registry data.

Patients and Methods: Total new cases of esophagus cancer reported by three sources of pathology reports, medical records, and death certificates to Ardabil province cancer registry center in 2006 and 2008 were enrolled in the study. All duplicated cases between three sources were identified and removed using Excel software. Some characteristics such as name, surname, father's name, date of birth and ICD codes related to their cancer type were used for data linkage and finding the common cases among three sources. The incidence rate per 100,000 was estimated based on capture-recapture method using the log-linear models. We used BIC, G^2 and AIC statistics to select the best-fit model.

Results: After removing duplicates, total 471 new cases of esophagus cancer were reported from three sources. The model with linkage between pathology reports, medical record sources and independence with the death certificates source was the best fitted model. The reported incidence rate for the years 2006 and 2008 was 18.77 and 18.51 per 100,000, respectively. In log-linear analysis, the estimated incidence rate for the years 2006 and 2008 was 49.71 and 53.87 per 100,000 populations, respectively.

Conclusions: Based on the obtained results, it can be concluded that none of the sources of pathology reports, death certificates and medical records individually or collectively were fully covered the incidence cases of esophagus cancer and need to apply some changes in data abstracting and case finding.

Keywords: Capture-Recapture, Esophagus Cancer, Completeness, Esophageal Neoplasm

1. Background

Esophagus cancer was the eighth most common cancer in the world in 2008. Also, in the same year, it was the sixth most common cause of death from cancer worldwide and most of these incidences and mortalities occurred in developing countries (1). The south-eastern coast of Caspian Sea is one of the highest incidence areas for esophageal cancer in the world (2). Cancer is the third cause of mortality in Iran. So this disease is an important public health issue in this country. Esophagus cancer has high incidence in north and northwestern regions of Iran (3). Accurate cancer incidence data are essential for planning, monitoring and evaluating national and regional cancer control programs (4).The purpose of population based cancer registries is to estimate the cancer burden in the area covered, to observe trends and regional differences and to provide a data-

base for epidemiological research (5).

As the decision makers in health authorities set their policies according to the results of data registration, they need to know how much the data registration process is reliable. Therefore, completeness of registration is used as one of the measures of quality of a cancer registry (5, 6). Completeness is defined as the proportion of incident cancer cases that is registered (5). Completeness level of cancer registration is one of the main parts of quality control in such registration (7). Since most cancer registries employ more than one data source for case finding, capture-recapture methods may be used to estimate the number of incident cases in the population, and hence to assess completeness of case ascertainment (8).

Capture-recapture is the method widely used in wildlife population censuses (9). Another important application of this method in epidemiology is to estimating the prevalence of a particular disease and also estimating the completeness of ascertainment of disease registers. However, capture-recapture method can principally be applied to any situation where there are two or even more incomplete lists (10). Two assumptions have to be made when using the simple capture-recapture method. Firstly, the sources are independent, and secondly, all individuals within the same source have an equal chance of being included (9, 11). The use of capture-recapture methods is very efficient for reducing the costs of disease registration as well as reducing bias in incidence estimations and in the case of comparing population subgroups (7). Ardabil province is located in Northwestern Iran, an area 70 km inland from the western Caspian coastline. The esophagus cancer is the second prevalent cancer in both males and females. Also, upper digestive tract cancers are the leading cause of 43% death in Ardabil province (2). The pathology-based cancer registry has been established in Ardabil province since 1999. For the first time in Iran, the population based cancer registry was established in Ardabil in 2003. The Ardabil Cancer Registry (ACR) actively collects information of cancer incidence from Pathology-based, hospital-based and death certificates. The main goal of the ACR is to measure cancer incidence and mortality in residents of Ardabil province (12).

2. Objectives

This study aims to estimate the incidence rate of esophagus cancer in Ardabil by capture-recapture method using log linear model.

3. Patients and Methods

This study was conducted in Ardabil Province which is located in the north-west of Iran. All new cases of esophagus cancer reported by three sources: pathology reports, death certificates, and medical records

reported to Ardabil population-based cancer registry in 2006 and 2008 were enrolled in the study. All duplicates in each source were identified and removed using EXCEL software. Some characteristics such as name, surname, father's name, date of birth and ICD codes related to their cancer type were used to identify the common cases among three sources. The incidence rate of esophagus cancer was estimated by the capture-recapture method and log-linear models. To use capture-recapture method, two main assumptions should be considered, sources of information should be independently and all people who are in every data source should have an equal chance to presence in the study (9, 11). However, in most human populations and medical science studies, usually these assumptions are not established and different sources are not independent. Three source capture-recapture and log-linear model were used to estimate the completeness and more accurate incidence rate of esophagus cancer in Ardabil province. With three registers, there are eight possible combinations of these registers in which cases do or do not appear. The general model uses eight parameters: the common parameter (the logarithm of the number expected to be in all lists), three 'main effects' parameters (the log odds ratios against appearing in each list for cases who appear in the others), three 'two-way interactions' or second order effect parameters (the log odds ratios between pairs of lists for cases who appear in the other), and a 'three-way' interaction parameter. For three registers, A with i levels, B with j levels, C with k levels, the natural logarithm (ln or log$_e$) of expected frequency F_{ijk} for cell ijk, $\ln F_{ijk}$, can be denoted as:

(1)

$$\ln F_{ijk} = \theta + \lambda_i^A + \lambda_j^B + \lambda_K^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

where θ is the common parameter, λ^A, λ^B, and λ^C are the main effect parameters, λ^{AB}, λ^{AC} and λ^{BC} are the second order effect (two-way interaction) parameters and λ^{ABC} is the highest order effect (three-way interaction) parameter. The value of this last three-way interaction parameter cannot be tested from the study data and is assumed to be zero (13).To assess how the various log-linear models fit the data (model fitting) and select the best model, we used the log likelihood-ratio test, also known as G² or deviance, Akaike's information criterion (AIC) and Bayesian information criteria (BIC) which they can be expressed as:

(2)

$$G^2 = -2\sum \text{Obs}_j \ln\left[\text{Obs}_j/\text{Exp}_{ji}\right]$$

Where Obs_j is the observed number of individuals in each

cell j, and Exp_{ji} is the expected number of individuals in each cell j under model i.

(3)

$$AIC = G^2 - 2\,[df]$$

Where the first term, G^2, is a measure of how well the model fits the data and the second term, 2[df], is a penalty for the addition of parameters (and hence model complexity).

(4)

$$BIC = G^2 - \left[\ln N_{obs}\right][df]$$

Where N_{obs} is the total number of observed individuals.

The lower the value of G^2, AIC and BIC the better is the fit of the model (13). AIC is the more appropriate criteria which is used by researchers for model selection (14-16). Therefore, we used these criteria for evaluating the fitting quality. Finally the model with lower amount of the AIC was chosen as the best model.

Estimated incidence was calculated based on the estimated new cases of esophagus cancer (by use of selected model in log-linear analysis) at a certain time over the number of the population at risk in Ardabil province at that time. All the incidences are reported based on the incidence per one hundred thousand populations. Also the completeness was calculated over age groups and calendar time, respectively. The informed consent was written by Ardabil cancer registry center and the authors certified that. In all stages of this study the individual's information such as name, surname and other characteristics were kept confidential. The P Values, less than 0.05, were considered significant. We used STATA software, version 12 (StataCorp, Texas, USA) for all computations.

4. Results

After investigating and removing duplicate cases between three sources, a total number of 471 new cases of esophagus cancer were reported to Ardabil population-based cancer registry in 2006 and 2008. The pathology source, hospital records and death certificates were reported 277, 193 and 152 new cases of esophagus cancer, respectively. Of 471 subjects 266 (56.50%) were male. The mean age of participants was 67.60 (± 12.91) years for men and 65.48 (± 13.58) years for women.

Venn diagram shows the common cases between pathology reports, hospital records and death certificate (Figure 1). In three source capture-recapture analysis, the log-linear model, a model in which two sources of pathology and medical records were mutually inter-

dependent and death certificates source were independent, was chosen as the best model with the lowest value of Akaike's Information Criterion and Bayesian Information Criterion (Table 1). The estimated total number of esophageal cancer in 2006 and 2008 was 1308.27 (95%CI: 1053.34 - 1674.80). The completeness of registration for all three sources after removing duplicates was 36% (471 cases) and also for pathology reports, hospital records and death certificates were 21.17% (277 cases), 14.75% (193 cases) and 11.61% (152 cases), respectively. The incidence of esophagus cancer was estimated generally and to the sex subgroups based on the estimated new cases of an esophagus cancer (by use of selected model in log-linear analysis) (Table 2). The estimated incidence rate of esophagus cancer for 2006 and 2008 was 49.71 (95% CI: 37.79 - 69.12) per 100,000 populations and 53.87% (95% CI: 38.97 - 7964.82) per 100,000 populations, respectively. In sex groups, the estimated incidence rate in male for 2006 and 2008 was 68.22 (95% CI: 45.17 - 113.32) per 100,000 populations and 58.81 (95% CI: 38.93 - 99.50) per 100,000 populations, respectively. The female estimated incidence rate for 2006 and 2008 was 33.17 (95% CI: 24.38 - 51.57) per 100,000 populations and 48.15 (95% CI: 30.31 - 89.88) per 100,000 populations, respectively. Also, the estimatedcompleteness of esophagus cancer registration for 2006 and 2008 was 37.76% and 34.36%, respectively.
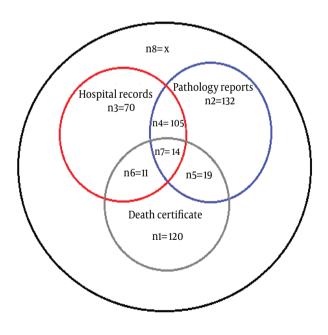

Figure 1. Venn Diagram of the Common Cases of Esophagus Cancer Between Pathology Reports, Hospital Records and Death Certificates

Table 1. Model Selection in Log-Linear Analysis by AIC, BIC and G^{2}Statistics

Model	X[a]	N[b]	95% CI for N	DF[c]	G^2[d]	BIC[d]	AIC[d]
P/C/D [e]	258.71	729.71	(673.84 - 800.96)	4	102.25	145.48	145.80
PC/D	837.27	1308.27	(1053.34 - 1674.80)	5	0.14	49.33	49.60
PD/C	145.92	616.92	(577.59 - 670.75)	5	51.29	108.43	108.70
CD/P	192.26	663.26	(616.92 - 724.31)	5	71.51	126.62	126.89
PC/PD	763.63	1234.63	(865.52 - 1949.07)	6	0.04	51.17	51.50
PC/CD	833.68	1304.68	(970.05 - 1863.68)	6	0.14	51.28	51.60
PD/CD	88	559	(533.18 - 595.53)	6	23.62	70.62	70.95
PC/PD/CD	707.36	1178.36	(734.06 - 2373.06)	7	0.00	53.08	53.46

[a]The estimated number of esophagus cancer that were not recorded in any of three sources.
[b]The estimated total number of esophagus cancer in Ardabil province in 2006 and 2008.
[c]Degree of freedom.
[d]Akaike's Information Criterion/ Bayesian Information Criterion/ Goodness of fit.
[e]D, death certificates source; P, pathology reports Source; C, hospital records; Model P/C/D, A model where all available resources are independent; Model PC/D, A model where sources P and C are dependent and independent of the source D; Model PD/C, A model where sources P and D are dependent and independent of the source C; Model CD/P, A model where sources C and D are dependent and independent of the source P; Model PC/PD, A model where two sources P and C and also two sources P and D are mutually interdependent and two sources C and D are independent; Model PC/CD, A model where two sources P and C and also two sources C and D are mutually interdependent and two sources P and D are independent; Model PD/CD, A model where two sources P and D and also two sources C and D are mutually interdependent and two sources P and C are independent; Model PC/PD/CD, A model where all two-way interaction between resources are exist.

Table 2. Estimated Incidence of Esophagus Cancer by Log-Linear Model Based on Ardabil Population in 2006 and 2008

Subgroups for Each Year	Reported Incidence Rate, per 100,000[a]	Estimated Incidence Rate, per 100,000	95% CI for Estimated Incidence Rate
2006			
Male	21.08	68.22	(45.17 - 113.32)
Female	16.33	33.17	(24.38 -51.57)
Total	18.77	49.71	(37.79 - 69.12)
2008			
Male	19.95	58.81	(38.93 - 99.50)
Female	16.99	48.15	(30.31 - 89.88)
Total	18.51	53.87	(38.97 - 79.64)

[a]The incidence rate reported by pathology reports, hospital records and death certificates after removing duplicates

5. Discussion

In this study, the incidence rate of esophagus cancer was estimated by the capture-recapture method and log-linear models. The mean age of all subjects was 66.68 ± 13.23 years (67.60 ± 12.91 for men and 65.48 ± 13.58 for women). The age distribution does not show the difference between men and women and slightly higher than the average age reported by studies conducted in other parts of Iran that reported average age 58 - 65 years for esophagus cancer (17-19), but reported average age in the study conducted in Tehran metropolitan was 70.40 (70.5 for men and 70.3 for women) that is slightly higher than present study (7). Male to female ratio was 1.29 that is consistent with international and regional reports (2, 7, 20).

In log-linear analysis, the model where sources pathology and hospital records are dependent and independent of the source death certificates was selected. In this study, we select the best model using AIC, BIC and G^2 statistics. The description of this relationship in situations as to what happens in society seems also logical, especially about dependencies between pathology reports and hospital records. The reported incidence rate of esophagus cancer based on population based cancer registry in Ardabil province, after removing duplicate cases between pathology reports, hospital records and death certificates, for 2006 and 2008 years was 18.77 per 100,000 and 18.51 per 100,000 populations, respectively. Also the registered incidence rate in men and women, after removing the duplicates between sources, in 2006 was 21.08 and 16.33 per 100,000 populations, and in 2008 was 19.95 and 16.99 per 100,000 populations, respectively. Babaei et al. reported esophagus cancer incidence as 19.5 per 100,000 populations during 2004 to 2006 years (21). This result is consistent with the reg-

istered incidence rate in this study. The estimated incidence rate of esophagus cancer in log-linear analysis, with consideration of the number of cases not registered in any sources, for 2006 and 2008 was 49.71 (95% CI: 37.79 - 69.12) per 100,000 populations and 53.87(95% CI: 38.97 - 79.64) per 100,000 populations, respectively. In a study conducted in Tehran, the estimated incidence of esophagus cancer was reported 10.5 per 100,000 populations, respectively (7). Also the estimated incidence rate reported by Globocan for man and woman in Iran was 9 and 8 per 100,000 populations, respectively (22). But this estimation is for the country in general, while in Iran, most northern and northwestern areas are at high risk for esophagus cancer. Also the central and western provinces are at medium risk and the southern regions are at a low risk (23). As Ardabil is located in northwestern of Iran and has the highest incidence of esophagus cancer in Iran (23), the higher incidence in this study seems reasonable. The estimated completeness in log-linear analysis for 2006 and 2008 was 37.76% and 34.36%, respectively. The completeness of cancer registries in our study is much lower than other countries that reported 96 % to 99.6% for all types of cancers in overall (8, 24, 25) and also for gastrointestinal cancers in Canada was reported 95.83% (8). The completeness of esophagus cancer registry in Tehran was reported 27.3% (7) and is consistent with results of our study. Thus the results of our study confirmed that the quality of cancer registry in Iran is inappropriate and needs more attention to amplify its quality. Some strategies such as use of national code of patients instead of name and surname, universal online registry using electronic health records, training cancer registry staffs and also training of death registry staffs about the immediate or underlying cause of death must be used to improve the quality of the cancer registry system.

Acknowledgments

Footnotes

Authors' Contribution: Mahmoud Khodadost designed the study, analyzed the data and wrote the paper. Parvin Yavari contributed to study design and analysis. Masoud Babaei and Saeed Barzegari contributed to the data collection and literature review. Behnam Khodadoust contributed to the data cleaning, data linkage and writing-up process. All authors read and approved the final manuscript.

References

1. Kiadaliri AA. Gender and social disparities in esophagus cancer incidence in Iran, 2003-2009: a time trend province-level study. *Asian Pac J Cancer Prev.* 2014;**15**(2):623–7. [PubMed: 24568468]

2. Amani F, Ahari SS, Akhghari L. Epidemiology of esophageal cancer in ardabil province during 2003-2011. *Asian Pac J Cancer Prev.* 2013;**14**(7):4177–80. [PubMed: 23991972]

3. Asmarian NS, Ruzitalab A, Amir K, Masoud S, Mahaki B. Area-to-Area Poisson Kriging analysis of mapping of county- level esophageal cancer incidence rates in Iran. *Asian Pac J Cancer Prev.* 2013;**14**(1):11–3. [PubMed: 23534706]

4. 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**(2):150–5. doi: 10.1093/jjco/hyl143. [PubMed: 17272318]

5. Schmidtmann I. Estimating completeness in cancer registries--comparing capture-recapture methods in a simulation study. *Biom J.* 2008;**50**(6):1077–92. doi: 10.1002/bimj.200810483. [PubMed: 19067337]

6. Bhurgri Y, Bhurgri A, Hasan SH. Comparability and Quality Control in Cancer Registration; Karachi (data monitoring 1995-2001). *J Pak Med Assoc.* 2002;**52**(7):301–7. [PubMed: 12481661]

7. Mosavi-Jarrahi A, Ahmadi-Jouibari T, Najafi F, Mehrabi Y, Aghaei A. Estimation of esophageal cancer incidence in Tehran by loglinear method using population-based cancer registry data. *Asian Pac J Cancer Prev.* 2013;**14**(9):5367–70. [PubMed: 24175827]

8. Robles SC, Marrett LD, Clarke EA, Risch HA. An application of capture-recapture methods to the estimation of completeness of cancer registration. *J Clin Epidemiol.* 1988;**41**(5):495–501. [PubMed: 3367181]

9. Suwanrungruang K, Sriplung H, Attasara P, Temiyasathit S, Buasom R, Waisri N, et al. Quality of case ascertainment in cancer registries: a proposal for a virtual three-source capture-recapture technique. *Asian Pac J Cancer Prev.* 2011;**12**(1):173–8. [PubMed: 21517253]

10. Poorolajal J, Haghdoost AA, Mahmoodi M, Majdzadeh R, Nasseri-Moghaddam S, Fotouhi A. Capture-recapture method for assessing publication bias. *J Res Med Sci.* 2010;**15**(2):107–15. [PubMed: 21526067]

11. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. *Eur J Cancer.* 2009;**45**(5):756–64. doi: 10.1016/j.ejca.2008.11.033. [PubMed: 19128954]

12. Babaei M, Pourfarzi F, Yazdanbod A, Chiniforush MM, Derakhshan MH, Mousavi SM, et al. Gastric cancer in Ardabil, Iran--a review and update on cancer registry data. *Asian Pac J Cancer Prev.* 2010;**11**(3):595–9. [PubMed: 21039022]

13. Van Hest NAH. *Capture-recapture methods in surveillance of tuberculosis and other infectious diseases.* Rotterdam: University Medical Center Rotterdam; 2007.

14. Motevalian A, Holakoei naeini K, Mahmoodi M, Majdzadeh R, Akbari M. Estimating deaths due to traffic accidents in Kerman using capture_recapture method. *J health fac health res.* 2007;**5**(2):61–72.

15. Hook EB, Regal RR. Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation. *Am J Epidemiol.* 1997;**145**(12):1138–44. [PubMed: 9199544]

16. Hook EB, Regal RR. Accuracy of alternative approaches to capture-recapture estimates of disease frequency: internal validity analy-

sis of data from five sources. *Am J Epidemiol.* 2000;**152**(8):771–9. [PubMed: 11052556]

17. Samadi F, Babaei M, Yazdanbod A, Fallah M, Nouraie M, Nasrollahzadeh D, et al. Survival rate of gastric and esophageal cancers in Ardabil province, North-West of Iran. *Arch Iran Med.* 2007;**10**(1):32–7. [PubMed: 17198451]

18. Akbari MR, Malekzadeh R, Nasrollahzadeh D, Amanian D, Sun P, Islami F, et al. Familial risks of esophageal cancer among the Turkmen population of the Caspian littoral of Iran. *Int J Cancer.* 2006;**119**(5):1047–51. doi: 10.1002/ijc.21906. [PubMed: 16570268]

19. Mohagheghi MA, Mosavi-Jarrahi A, Malekzadeh R, Parkin M. Cancer incidence in Tehran metropolis: the first report from the Tehran Population-based Cancer Registry, 1998-2001. *Arch Iran Med.* 2009;**12**(1):15–23. [PubMed: 19111024]

20. Ke L. Mortality and incidence trends from esophagus cancer in selected geographic areas of China circa 1970-90. *Int J Cancer.* 2002;**102**(3):271–4. doi: 10.1002/ijc.10706. [PubMed: 12397650]

21. Babaei M, Jaafarzadeh H, Sadjadi AR, Samadi F, Yazdanbod A, Fal-

lah M, et al. Cancer incidence and mortality in Ardabil: Report of an ongoing population-based cancer registry in Iran, 2004-2006. *Iran J Public Health.* 2009;**38**(4):35–45.

22. Globocan. *Estimated Cancer Incidence, Mortality and Prevention Worldwide in 2012. 2012.* Available from: http://globocan.iarc.fr/Pages/fact_sheets_population.aspx.

23. Kolahdoozan S, Sadjadi A, Radmard AR, Khademi H. Five common cancers in Iran. *Arch Iran Med.* 2010;**13**(2):143–6. [PubMed: 20187669]

24. Crocetti E, Miccinesi G, Paci E, Zappa M. An application of the two-source capture-recapture method to estimate the completeness of the Tuscany Cancer Registry, Italy. *Eur J Cancer Prev.* 2001;**10**(5):417–23. [PubMed: 11711756]

25. Gajalakshmi V, Swaminathan R, Shanta V. An Independent Survey to Assess Completeness of Registration: Population Based Cancer Registry, Chennai, India. *Asian Pac J Cancer Prev.* 2001;**2**(3):179–83. [PubMed: 12718628]