RESEARCH ARTICLE

Estimating the Completeness of Gastric Cancer Registration in Ardabil/Iran by a Capture-Recapture Method using Population-Based Cancer Registry Data

Mahmoud Khodadost[1], Parvin Yavari[2,3*], Masoud Babaei[4,5], Alireza Mosavi-Jarrahi[2], Fatemeh Sarvi[2,4], Kamyar Mansori[8], Behnam Khodadost[9]

Abstract

    <u>Background</u>: Knowledge of cancer incidences is essential for cancer prevention and control programs. Capture-recapture methods have been recommended for reducing bias and increasing the accuracy of cancer incidence estimations. This study aimed to estimate the completeness of gastric cancer registration by the capture-recapture method based on Ardabil population-based cancer registry data. <u>Materials and Methods</u>: All new cases of gastric cancer reported by three sources, pathology reports, death certificates and medical records that reported to Ardabil population-based cancer registry in 2006 and 2008 were enrolled in the study. The duplicate cases based on the similarity of first name, surname and fathers names were identified between sources. The estimated number of gastric cancers was calculated by the log-linear method using Stata 12 software. <u>Results</u>: A total of 857 new cases of gastric cancer were reported from three sources. After removing duplicates, the reported incidence rates for the years 2006 and 2008 were 35.3 and 32.5 per 100,000 population, respectively. The estimated completeness calculated by log-linear method for these years was 36.7 and 36.0, respectively. <u>Conclusions</u>: These results indicate that none of the sources of pathology reports, death certificates and medical records individually or collectively fully cover the incident cases of gastric cancer. We can obtain more accurate estimates of incidence rates using the capture-recapture method.

Keywords: Capture-recapture - incidence estimation - log-linear - gastric cancer - cancer registration - completeness

Introduction

    Accurate cancer incidence data are essential like to planning, monitoring and evaluating national and regional cancer control programs (Kamo et al., 2007). 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 (Schmidtmann, 2008). Decision makers in health authorities need to know how reliable the data is on which they base their policies. Therefore, completeness of registration is used as one of the measures of quality of a cancer registry (Parkin et al., 1994; Schmidtmann, 2008).

    Completeness is defined as the proportion of incident cancer cases that is registered (Schmidtmann, 2008). Completeness level of cancer registration is one of the main parts of quality control in such registration (Mosavi-Jarrahi et al., 2013). In the literature, several methods are described to evaluate completeness, which

are divided in two categories: qualitative methods and quantitative methods. The qualitative methods used were Mortality/Incidence ratios and the proportion of microscopically verified cases and, among quantitative methods, the ones applied were the capture recapture, the death certificates and M:I ratios method and the flow method (Castro, 2011). 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 (Robles et al., 1988). Capture-recapture is the method widely used in wildlife population censuses (Suwanrungruang et al., 2011). Another important application for this method is in epidemiology for estimating prevalence of a particular disease and 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 (Poorolajal et al., 2010). Two assumptions have to be made when using

*[1]Department of Epidemiology, Faculty of Health, [2]Dept. of Health and Community Medicine, [7]Paramedical Department, School of Medicine, Shahid Beheshti University of Medical Sciences, [3]Iranian Epidemiological Association, [8]Department of Epidemiology, Faculty of Public Health, Iran University of Medical Sciences, Tehran, [5]Health Deputy, Ardabil University of Medical Sciences, Ardabil, [9]Faculty of Health, Shiraz University of Medical Sciences, Shiraz, Iran, [4]German Cancer Research Center (DKFZ), Heidelberg, Germany *p.yavari-grc.sbmu.ac.ir*

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 (Parkin and Bray, 2009; Suwanrungruang et al., 2011). 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 (Mosavi-Jarrahi et al., 2013; Ghojazadeh et al., 2013).

The pathology-based cancer registry has been established in Ardabil province since 1999. For the first time in Iran, the population based cancer registry 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 (Babaei et al., 2010). Ardabil, a northwestern province, has the highest incidence of gastric cancer in Iran with an ASR of 49.1 and 25.4 in men and women, respectively. In Iran, most northern and northwestern areas are at high risk for gastric cancer, whereas the central and western provinces are at medium risk and the southern regions are at a low risk (Radmard, 2010). This study aims to estimate the completeness of gastric cancer in Ardabil by three source capture-recapture method using log linear model.

Materials and Methods

This study was conducted in Ardabil Province which is located in north-west of Iran. All new cases of gastric cancer reported by three sources; pathology reports, death certificates and medical records that 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 gastric 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 (Parkin and Bray, 2009; Suwanrungruang, Sriplung et al., 2011). 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 was used to estimate the completeness and more accurate incidence rate of gastric 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 leve ls, the natural logarithm (ln or loge) of expected frequency Fijk for cell ijk, ln Fijk, can be denoted as:

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

where θ is the common parameter, λA, λB and λC are the main effect parameters, λAB, λAC and λBC are the second order effect (two-way interaction) parameters and λABC 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 (Van Hest, 2007). 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 G^2 or deviance, akaike's information criterion (AIC) and bayesian information criteria (BIC) which they can be expressed as:

$$G^2 = -2 \sum Obs_j \ln[Obs_j / Exp_{ji}]$$

where Obsj is the observed number of individuals in each cell j and Expji is the expected number of individuals in each cell j under model i.

AIC = G^2–2df

where the first term, G2, 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).

BIC = G^2 – [lnN$_{obs}$][df]

Where Nobs is the total number of observed individuals. The lower the value of G2, AIC and BIC the better is the fit of the model (Van Hest, 2007). AIC is the more appropriate criteria which is used by researchers for model selection (Hook and Regal, 1997; Hook and Regal, 2000; Motevalian et al., 2007). Therefore, we used this criteria for evaluating the goodness of fit. Finally the model with lower amount of the AIC was chosen as the best model.

Estimated completeness was calculated based on the estimated new cases of gastric 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 thous and populations. Also the completeness was calculated by age groups and calendar time respectively. In all stages of this study the individual's information such as name, surname and other characteristics were kept confidential. We used STATA software, version 12 (StataCorp, Texas, USA) for all computations.

Results

After investigating and remove duplicate cases between three sources, total 857 new cases of gastric cancer were reported to Ardabil population-based cancer registry in 2006 and 2008. The pathology source, hospital records and death certificates were reported 439, 361 and 378 new cases of gastric cancer respectively. Of 857 subjects 587 (68.49%) were male. The mean age of participants was 66.95 (±12.16), 67.37 (±11.68) years

DOI:http://dx.doi.org/10.7314/APJCP.2015.16.5.1981

Estimating the Completeness of Gastric Cancer Registry in Ardabil/Iran by Capture-Recapture Method using Population-Based Cancer Registry Data

for men and 66.04 (±13.13) 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 with log-linear model, a model in which two sources of pathology and medical records were mutually interdependent 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 gastric cancer in 2006 and 2008 was 2356.35 (95%CI: 2019.12-2794.04). The completeness of registration for all three sources after removing duplicates was 36.35% (857 cases) and also for pathology reports, hospital records and death certificates were 18.63% (439 cases), 15.32% (361 cases) and 16.04%
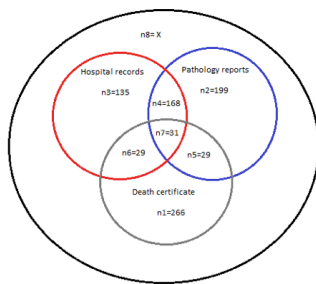
(378 cases), respectively. The completeness of gastric cancer was estimated generally and to the age and sex subgroups based on the estimated new cases of an gastric cancer (by use of selected model in log-linear analysis). (Table 2). The estimated numbers of gastric cancer for 2006 and 2008 were 1209.53 (95% CI: 983.21-1532.85) and 1148 (95% CI: 920.18-1478.14), respectively. Also the estimated completeness for 2006 and 2008 was 36.70% and 35.97%, respectively.

Discussion

In this study, the completeness of gastric cancer was estimated by the capture-recapture method and log-linear models. The mean age of all subjects was 66.95±12.16 years (67.37±11.68 for men and 66.04±13.13 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 was 58-65 years (Biglarian et al., 2009; Khodabakhshi et al., 2009; Mehrabian et al., 2010; Rajaiefard et al., 2011), but reported average age in the study conducted in Tehran metropolitan is consistent with present study (Aghaei et al., 2013). Male to female ratio was 2.17 that is consistent with country reports, as it has been equal to 2.59, 2.61 and 2.58 in country reports of the years of 2005, 2006 and 2007, respectively (Aghaei et al., 2013). But in Finland and the union of Europe this value

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

Table 1. Model Selection in Log-Linear Analysis by AIC, BIC and G2 Statistics.

Model	X***	N****	95% CI for N	DF**	G2*	BIC*	AIC*
†P/C/D	515.80	1372.80	(1288.24-1473.94)	4	191.31	243.06	243.27
PC/D	1500.35	2357.35	(2019.12-2794.04)	5	1.90	55.59	55.86
PD/C	292.50	1149.50	(1089.30-1225.28)	5	121.70	175.39	175.66
CD/P	375.30	1232.30	(1160.50-1321.09)	5	1.60	213.69	213.96
PC/PD	1238.27	2095.27	(1671.60-2739.28)	6	0.71	56.35	56.67
PC/CD	1825.31	2682.31	(2071.16-3601.07)	6	0.28	55.92	56.24
PD/CD	159.91	1016.91	(979.59-1065.59)	6	49.86	105.49	105.82
PC/PD/CD	1567.92	2424.92	(1644.67-3978.07)	7	0.00	57.58	57.96

*Akaike's information criterion/Bayesian information criterion/goodness of fit, **Degree of freedom, ***The estimated number of gastric cancer that were not recorded in any of three sources, ****The estimated total number of gastric cancer in Ardabil province in 2006 and 2008, †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 Number of Gastric Cancer by Log-Linear Model Based on Ardabil Population in 2006 and 2008.

Calendar year	Subgroups		Reported new cases	Estimated new cases	95% CI for estimated new cases	Completeness registration (%)**
	Gender	Male	301	749.6	(598.12 - 978.33)	40.2
2006		Female	143	498.9	(321.17 - 853.94)	28.7
	Age groups	50≤	48	148.8	(79.12 - 374.13)	32.3
		51-64	111	250.5	(177.43 - 363.57)	44.3
		65≥	285	841.5	(640.60 - 1155.88)	33.9
	Total		444	1209.5	(983.21 - 1532.85)	36.7
2008	Gender	Male	286	699.6	(554.22 - 923.86)	40.9
		Female	127	500.3	(303.43 - 916.95)	25.4
	Age groups	50≤	39	135.0	(64.46 - 400.86)	28.9
		51-64	98	279.1	(175.58 - 520.84)	35.1
		65≥	276	732.0	(570.71 - 981.53)	37.7
	Total		413	1148.0	(920.18 - 1478.14)	36.0

*Number of new cases reported by pathology reports, hospital reports and death certificates after removing duplicates

is equal to 1.34 and 1.66, respectively and little difference in male to female ratio was existed (Ferlay et al., 2013).

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 G2 statistics.

The description of this relationship in situations that what happens in society seems also logical, especially about dependencies between pathology reports and hospital records. The reported incidence rate 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 35.32 and 32.50, respectively. Also the registered incidence rate in men and women, after removing the duplicates between sources, in 2006 was 46.6 and 23.3 per 100,000 populations and in 2008 was 43.9 and 20.5 per 100,000 populations, respectively. Babaei et al, reported gastric cancer incidence as 51.8 and 24.4 per 100,000 during 2004 to 2006 years for men and women, respectively (Babaei et al., 2009). This result is consistent with the registered incidence rate in this study. In a study conducted in Tehran, the estimated incidence of gastric cancer in men and women was reported 33.9 and 19.7 per 100,000 populations, respectively (Aghaei et al., 2013).

Also the estimated incidence rate reported by GLOBOCAN in 2008 was reported for man and woman in Iran was 21.9 and 9 per 100,000 populations, respectively. But this estimation is for the country in general, while in Iran, most northern and northwestern areas are at high risk for gastric cancer. Also the central and western provinces are at medium risk and the southern regions are at a low risk (Radmard, 2010). As the Ardabil was located in northwestern of Iran and has the highest incidence of gastric cancer in Iran, with an ASR of 49.1 and 25.4 in men and women, respectively (Radmard, 2010), the higher incidence in this study seems reasonable. In age subgroups with increasing age, the incidence rate was increased. Other studies conducted in Iran and other parts of the world reported the same results (Boyle and Levin, 2008; Matsuda et al., 2008; Babaei et al., 2009; Aghaei et al., 2013).

The estimated completeness in log-linear analysis for 2006 and 2008 was 36.7% and 35.97% respectively. Also the estimated completeness for men and women in 2006 was 40.2% and 28.7% and for 2008 was 40.9% and 25.4%, 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 overall (Robles et al., 1988; Crocetti et al., 2001; Gajalakshmi et al., 2001) and also for gastrointestinal cancers in Canada was reported 95.83% (Robles et al., 1988). The completeness of gastric cancer registry in Tehran was reported 26.6% (Aghaei et al., 2013) and is consistent with results of our study. Thus the results of our study confirmed that the quality of cancer registry in Iran is highly inappropriate and need to more attention to improve it.

References

Aghaei A, Ahmadi-Jouibari T, Baiki O, Mosavi-Jarrahi A. (2013). Estimation of the gastric cancer incidence in Tehran by two-source capture-recapture. *Asian Pac J Cancer Prev*, **14**, 673-7.

Babaei M, Jaafarzadeh H, Sadjadi A, et al (2009). Cancer incidence and mortality in Ardabil: Report of an ongoing population-based cancer registry in Iran, 2004-2006. *Iranian J Public Health*, **38**, 22-26.

Babaei M, Pourfarzi F, Yazdanbod A, et al (2010). Gastric cancer in Ardabil, Iran-a review and update on cancer registry data. *Asian Pac J Cancer Prev,* **11**, 595-9.

Biglarian A, Hajizadeh E, Kazemnejad A, et al (2009). Survival analysis of gastric cancer patients using Cox model: a five year study. *Tehran University Med J*, **67**, 3417-25

Boyle P, Levin B (2008). World Cancer Report, IARC Press, Lyon

Castro CSFC (2011). Estimating completeness of cancer registration-Na application to gastric câncer in the North region registry of portugal. Universidade do Porto.

Crocetti E, Miccinesi G, Paci E, Zappa M (2001). An application of the two-source capture-recapture method to estimate the completeness of the tuscany cancer registry, Italy. *Eur J Cancer Prev*, **10**, 417-23.

Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al (2013). Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*, **49**, 1374-403.

Gajalakshmi V, Swaminathan R, Shanta V (2001). An independent survey to assess completeness of registration: population based cancer registry, Chennai, India. *Asian Pac J Cancer Prev*, **2**, 179-83.

Ghojazadeh M, Mohammadi M, Azami-Aghdash S, et al (2013). Estimation of cancer cases using capture-recapture method in Northwest Iran. *Asian Pac J Cancer Prev*, **14**, 3237-41.

Hook EB, Regal RR (1997). Validity of methods for model selection, weighting for model uncertainty and small sample adjustment in capture-recapture estimation. *Am J Epidemiol*, **145**, 1138-,44.

Hook EB, Regal RR (2000). Accuracy of alternative approaches to capture-recapture estimates of disease frequency: internal validity analysis of data from five sources. *Am J Epidemiol*, **152**, 771-9.

Kamo K, Kaneko S, Satoh K, et al (2007). A mathematical estimation of true cancer incidence using data from population-based cancer registries. *Jpn J Clin Oncol*, **37**, 150-5.

Khodabakhshi R, Gouhari M, Biglarian A, et al (2009). Survival analysis of patients with gastric adenocarconoma. *Kosar Med J*, **12**, 345-55.

Matsuda T, Marugame T, Kamo K, et al (2008). Cancer incidence and incidence rates in Japan in 2002: based on data from 11 population-based cancer registries. *Jpn J Clin Oncol*, **38**, 641-8.

Mehrabian A, Esna-Ashari F, Zham H, et al (2010). Gastric cancer prevalence, according to survival data in Iran (national study-2007). *Iran J Public Health*, **39**, 27-31.

Mosavi-Jarrahi A, Ahmadi-Jouibari T, Najafi F, et al (2013). Estimation of esophageal cancer incidence in Tehran by log-linear method using population-based cancer registry

DOI:http://dx.doi.org/10.7314/APJCP.2015.16.5.1981

Estimating the Completeness of Gastric Cancer Registry in Ardabil/Iran by Capture-Recapture Method using Population-Based Cancer Registry Data

data. *Asian Pac J Cancer Prev*, **14**, 5367-70.

Motevalian A, Holakoeinaeini K, Mahmoodi M, et al (2007). Estimating deaths due to traffic accidents in Kerman using capture-recapture method. *J Health Faculty Health Res*, **5**, 61-72.

Parkin D, Chen V, Ferlay J, et al (1994). Comparability and quality control in cancer registration: IARC technical report no19. Lyon, France.

Parkin DM, Bray F (2009). Evaluation of data quality in the cancer registry: principles and methods part II. completeness. *Eur J Cancer*, **45**, 756-64.

Poorolajal J, Haghdoost AA, Mahmoodi M, et al (2010). Capture-recapture method for assessing publication bias. *J Res Med Sci*, **15**, 107-15.

Radmard AR (2010). Five common cancers in Iran. *Arch Iran Med*, **13**, 143-6.

Rajaiefard A, Moghimi B, Mahmoodi M, et al. (2011). Epidemiological and clinical features of gastric cancer: descriptive study of cancer registry cases of Fars province (2001-2006). ISMJ??? 14, 114-21.

Robles SC, Marrett LD, Clarke EA, Risch HA (1988). An application of capture-recapture methods to the estimation of completeness of cancer registration. *J Clin Epidemiol*, **41**, 495-501.

Schmidtmann I. (2008). Estimating completeness in cancer registries-comparing capture recapture methods in a simulation study. *Biom J*, **50**, 1077-92.

Suwanrungruang K, Sriplung H, Attasara P, et al. (2011). Quality of case ascertainment in cancer registries: a proposal for a virtual three-source capture-recapture technique. *Asian Pac J Cancer Prev*, **12**, 173-8.

van Hest NAH (2007). Capture-recapture methods in surveillance of tuberculosis and other infectious diseases, Erasmus University Rotterdam.