RESEARCH ARTICLE

Estimating Completeness of Cancer Registration in Iran with Capture-Recapture Methods

Gohar Mohammadi[1], Mohammad Esmaeil Akbari[1]*, Yadolah Mehrabi[2], Ali Ghanbari Motlagh[3], Elham Partovi Pour[3], Gholamreza Roshandel[4], Ardasheir Khosravi[5]

Abstract

Completeness is an important indicator of data quality in cancer registry programs. This study aimed to estimate the completeness of registered cases in a population based cancer registry program implemented in five provinces of Iran. Capture-recapture methods were used to estimate the number of cases that may have been missed and to estimate rates of completeness for different categories of age, year, and sex. The data used for this study were obtained from three sources: 1) National Pathology Database; 2) National Hospital Discharge Database; and 3) National Death Registry Database. The three sources were linked and duplicates were identified based on first name, last name, father's names, and date of birth, ICD code, and case's residency address using Microsoft Excel. Removing duplicates, the three sources reported a total of 35,643 cases from March 2008 to March 2011. Running many different multivariate models of capture-recapture and controlling for source dependencies revealed an overall under-reporting of 49% in all five registries combined. The estimated completeness differed based on age, sex, and year. The overall completeness was higher for males than females (71.2% for males and 59.9% for females). Younger age had lower rates of completeness compared to older age (38.1% for <40 years, 55.4% for 40-60 years, and 76.7 for >60 years). The results of this study indicated a moderate to severe (depending on the age, sex and year) degree of completeness in the population based cancer registration of Iran.

Keywords: Cancer - registration - completeness - capture-recapture - Iran

Introduction

Cancer is a leading cause of death in both more and less economically developed countries. Based on GLOBOCAN estimates, about 14.1 million new cancer cases and 8.2 million deaths occurred in 2012 worldwide (Torre et al., 2015). Cancer is the third leading cause of death in Iran, (Moh and ME, 2009) and it accounted for 12% of all deaths (Organization, 2014). Moreover, it is estimated that more than 70,000 new cases of cancer occur in Iran annually (Moh and ME, 2009). Knowing of cancer incidence is essential like to plan, control, and promote regional and national cancer control programs (Schouten et al., 1994; Kamo et al., 2007). The cancer registry is an organization set up for the systematic collection, storage, analysis, interpretation and reporting of data on subjects with cancer, using the available data result in saving time and cost (Hearst and Hulley, 1988). Cancer registries originated in the first half of the twentieth century, and have expanded in the last 20 years (Parkin, 2006). In Iran, the first activities that were performed in order to organize cancer reporting were started in 1956 when Cancer Society in Tehran University was founded (Habibi, 1984).

The report was based on the data collected from pathologic centers. It included cancer data from 1945 – 1956 with an incidence rate of 28/100,000 in the south and 42/100,000 in the north of Iran (Habibi, 1984; Mohagheghi and Mosavi-Jarrahi, 2010). At the moment, in Iran the incidence of cancer is estimated by the data obtained from the pathology-based cancer registry which is unique.

The value of cancer registry and its ability to carry out such activities rely heavily on the quality of the data (Bray and Parkin, 2009). The completeness level of cancer registration is one of the main parts of quality control in such registration (Mosavi-Jarrahi et al., 2013). Completeness of registration is the proportion of all incident cases in the registry's population that have been included in the registry database. Incidence rates and survival proportions will be close to their true value if maximum completeness can be achieved. Cancer registry completeness can be evaluated by

[1]Cancer Research Center, [2]Department of Epidemiology, School of Health, Shahid Beheshti University of Medical Sciences, Tehran, [3] Cancer Prevention Department, [5]Director of Information and Statistic Group of Health, MOHME, Tehran ,[4]Golestan Research Center of Gastroenterology and Hepatology, Golestan University of Medical Sciences ,Gorgan, Iran. *For correspondence: drakbari.drakbari@gmail.com

independent case ascertainment, capture-recapture, or death-certificate methods (Shin et al., 2007; Bray and Parkin, 2009). Among the techniques described in the reports of the International Agency for Research on Cancer (IARC), capture-recapture method was considered to be at the same level of the best methods (Schouten et al., 1994). This is important enough since a significant part of quality control in cancer registration reported by the International Association for Research on Cancer (IARC) is based on the estimated number of cases in the community, that is, not based on the registered cases (Schouten et al., 1994; Ghojazadeh et al., 2013). Capture-recapture method is used in various health related fields to estimate hidden populations, completeness of registrations, incidence, and prevalence of diseases and special events (Hook and Regal, 1995). Also, capture-recapture methods are recommended for reducing the costs of disease registration as well as reducing bias in incidence estimations and for comparing population subgroups. Modeling the effect of intervening variables presents better estimations of population size and therefore solves many problems of the estimation of population size (Tilling, 2001). Although passive cancer data registration was started since around 1999 in the Center for Disease Control report of cancer / Ministry of Health and Medical Education, as it did not cover all pathology labs and other departments that had related data about cancerous patients all over the country, the first results of national reports were not accurate for national estimation. For example ,the first report had only 18% coverage (Mohagheghi and Mosavi-Jarrahi, 2010), thereafter, pathologic based registration continued until the last national reports in 2009 that claimed coverage of more than 86% (Moh and ME, 2009). Nevertheless, these reports are based on the data obtained from the majority of pathologic labs, but not all of the centers. There are also some independent scientific departments like research centers that have domestic data about cancer cases, which are population based and reported separately (Modirian et al., 2014). Since most cancer registries employ more than one data source for case finding, the capture-recapture method may be used to estimate the number of incident cases in the population and hence to assess the completeness of case ascertainment (Robles et al., 1988). Evaluation of completeness is important for all registries (Parkin and Bray, 2009). It helps public health professionals in programming and implementing policies to control burden of cancers more effectively (Modirian et al., 2014).

Therefore, completeness of registration is used as one of the measures of determining the quality of a cancer registry (Schmidtmann, 2008). This study aims to estimate the completeness of GI cancer from three sources using the capture-recapture method and the log linear model.

Materials and Methods

Sources of data

The data used for this cross-sectional study were obtained from three sources including the national report cancer registry (pathology report), hospital records and national death registry in five provinces of Iran namely Esfahan, Golestan, Semnan, Bushehr and Kermanshah, during March 2008 to March 2011.

Statistical methods

A capture-recapture analysis was performed to estimate the number of patients that may have been missed. Capture–recapture has been advocated for use in estimating completeness of disease registers, (Hook and Regal, 1995) and it has been applied several times to estimate the completeness of cancer registry data (Brenner et al., 1994; Schouten et al., 1994).

To use the capture-recapture method, two main assumptions should be considered, that is, sources of information should be independent and everybody that contributed in the process of gathering data should be given the opportunity of partaking in the study (19, 22). The capture-recapture and log-linear model was used to estimate the completeness and accurate incidence rate of GI cancer in the selected provinces from three sources.

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 "nteraction parameter. For three registers,

A with i levels, B with j levels, and C with k levels, the natural logarithm (ln or log$_e$) of the expected frequency Fijk for cell ijk, ln Fijk, can be denoted as:

$$InF_{ijk} = \theta + \lambda_i^A + \lambda_j^B + \lambda_K^C + \lambda_{ij}^{AB} + \lambda_{ik}^{Ac} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABc}$$

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.

To assess how the various log-linear models fit the data (model fitting) and select the best model, the log likelihood-ratio test, also known as G^2 or deviance, was used, as well as Akaike's information criterion (AIC) and Bayesian information criteria (BIC) which can be expressed as:

$G^2 = -2\Sigma$ Obs$_j$In[Obs$_j$/Exp$_{ji}$]

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

AIC = G^2−2df

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

Another information criterion is the bayesian Information Criterion (BIC), which can be expressed as: BIC = G^2− [lnN$_{obs}$][df]

N$_{obs}$ is the total number of observed individuals (Agresti and Kateri, 2011)

AIC is the more appropriate criteria which is used by

Table 1. Source Dependence was Modeled by Adding the Interaction Term to the Mode

Model Number	Formula	Explanation
Model 1	$\mu_{ijk} = \lambda + \lambda_i^{sP} + \lambda_j^{sH} + \lambda_K^{sD}$	All sources are independent in model
Model 2	$\mu_{ijk} = \lambda + \lambda_i^{sP} + \lambda_j^{sH} + \lambda_K^{sD} + \lambda_{ij}^{sPsH}$	A model where sP and sH are dependent and independent of the sD
Model 3	$\mu_{ijk} = \lambda + \lambda_i^{sP} + \lambda_j^{sH} + \lambda_K^{sD} + \lambda_{ik}^{sPsD}$	A model where sP and sD are dependent and independent of the source H
Model 4	$\mu_{ijk} = \lambda + \lambda_i^{sP} + \lambda_j^{sH} + \lambda_K^{sD} + \lambda_{jk}^{sHsD}$	A model where sH and sD are dependent and independent of the sP
Model 5	$\mu_{ijk} = \lambda + \lambda_i^{sP} + \lambda_j^{sH} + \lambda_K^{sD} + \lambda_{ij}^{sPsH} + \lambda_{ik}^{sPsD}$	A model where sP andsH and also sP &sD are mutually interdependent and sH and sD are independent
Model 6	$\mu_{ijk} = \lambda + \lambda_i^{sP} + \lambda_j^{sH} + \lambda_K^{sD} + \lambda_{ij}^{sPsH} + \lambda_{jk}^{sHsD}$	A model where sP and sH and also sH&s D are mutually interdependent and sP and s D are independent
Model 7	$\mu_{ijk} = \lambda + \lambda_i^{sP} + \lambda_j^{sH} + \lambda_K^{sD} + \lambda_{ik}^{sPsD} + \lambda_{jk}^{sHsD}$	A model where sP and sD and also sH & s D are mutually interdependent and sP and sH are independent
Model 8	$\mu_{ijk} = \lambda + \lambda_i^{sP} + \lambda_j^{sH} + \lambda_K^{sD} + \lambda_{ij}^{sPsH} + \lambda_{ik}^{sPsD} + \lambda_{jk}^{sHsD}$	A model where all two-way interaction between reources are exist

sD, death certificates; sP, pathology reports; sH, hospital records

researchers for model selection (Hook and Regal, 1997 (Hook and Regal, 1997; Motevalian et al., 2007)). To select the best-fit model, Akaike statistics, which is very common in scientific researches and analyses was used (Hook and Regal, 1995). However, the model is best fit when the value of G2, AIC and BIC is lower.

Therefore, these criteria should be used for evaluating the goodness of fit. The total number of cases that might have been present in the cancer registries were estimated from the best fitting model. The estimated percentage of completeness (EPC) is the proportion of observed to estimated cases. In all registries, the percentage of completeness cannot exceed 100%. The percentage difference can be calculated by 100 - EPC. The estimates of total number of cancer cases in 3 years and annual cases were calculated. Also the completeness was calculated by age groups, gender and calendar year respectively. All registrations were checked for duplication based on full name, father's names, date of birth, ICD codes and address and the duplicates were removed , using Excel software.

All calculations for the capture-recapture method were made using STATA software, version 13 (StataCorp, Texas, USA). The confidentiality of all data was ensured in all stages of the study from extraction of the data from the cancer registry to analysis, reporting and preservation of the backup data. The Ethics Committee of Shahid Beheshti University of Medical Sciences approved this study.

Results

The goal of this study was to evaluate the completeness of registration in 5 provinces (Golestan, Isfahan, Kermanshah, Semnan and Busheher) for cancer cases, of patrons resident in the district of the provinces at the date of diagnosis. GI cancers were chosen in this study because it is, for both sexes, one of the most common and lethal

Table 2. Model Selection in Log-Linear Analysis by AIC, BIC and G2 Statistics (Gastrointestinal Tract)

Model number	X***	N****	95% CI for N		DF**	G²*	AIC*	BIC*
			lower	Upper				
Model 1	6,395.4	15,969.4	15,616.4	16,343.0	4.0	662.0	729.8	729.6
Model 2	5,615.9	15,189.9	14,852.8	15,548.5	5.0	558.9	628.8	628.5
Model 3	3,358.6	12,932.6	12,601.6	13,299.8	5.0	421.0	490. 83	490.6
Model 4	9,227.5	18,801.5	18,217.0	19,425.5	5.0	51.2	120. 95	120.7
Model 5	1,959.1	11,533.1	11,322.2	11,769.6	6.0	72.7	144.4	144.1
Model 6	8,388.8	17,962.8	17,368.8	18,601.9	6.0	22.3	94.1	93.8
Model 7	8,855.8	18,429.8	17,093.3	20,003.9	6.0	50.9	122.7	122.3
Model 8	4,888.7	14,462.7	13,434.9	15,764.2	7.0	0.0	73.8	73.4

*Akaike's Information Criterion/ Bayesian Information Criterion/ Goodness of fit, ** Degree of freedom, *** The estimated number of GI cancer that were not recorded in any of three sources in 2008 - 2010, **** The estimated total number of GI cancer in 2008 - 2010, Model1: All sources are independent in model; Model 2: A model where sP & sH are dependent and independent of the sD; Model 3: A model where sP & sD are dependent and independent of the source H; Model 4: A model where sH & sD are dependent and independent of the sP; Model 5: A model where sP &sH and also sP &sD are mutually interdependent and sH&sD are independent; Model 6: A model where sP & sH and also sH&s D are mutually interdependent and sP&s D are independent; Model 7: A model where sP & sD and also sH & s D are mutually interdependent and sP &sH are independent; Model 8: A model where all two-way interaction between resources are exist

Table 3. Estimated Number of GI Cancer Types and Completeness of Registration by Log-Linear Model Based on Three Sources Data in Iran 2008 - 2010

Subgroups		Reported new cases	Estimated new* cases	95% CI for Estimated cases		Completeness** of registration (%)
				Lower	Upper	
Gender	Male	1,679.0	3,452.4	2,598.6	5,098.9	48.6
	female	1,320.0	4,077.4	2,478.7	7,881.8	32.4
	Total	2,999.0	7,204.2	5,496.6	10,079.4	41.6
Gender	Male	1,931.0	2,467.0	2,253.2	2,822.5	78.3
	female	1,400.0	2,294.2	1,887.2	3,041.0	61.0
	Total	3,331.0	4,731.0	4,284.8	5,385.8	70.4
Gender	Male	1,724.0	2,129.2	1,969.2	2,393.8	81.0
	female	1,520.0	2,214.7	1,904.2	2,775.9	68.6
	Total	3,244.0	4,267.8	3,942.9	4,743.7	76.0
Gender	Male	5,452.0	7,664.7	7,080.2	8,459.0	71.1
	female	4,122.0	6,876.8	6,017.1	8,126.6	59.9
	Total	9,574.0	1,4462.7	13,434.9	15,764.2	66.2
Age group	Age1	944.0	2,477.9	1,491.5	5,241.4	38.1
	Age2	2,635.0	4,758.6	3,991.5	5,959.3	55.4
	Age3	5,970.0	7,794.2	7,330.2	8,416.5	76.6

cancers in Iran, especially in the northern region, so it is of great interest to evaluate the completeness of registration of GI tumors (gastric, colorectal and liver). It was observed that a total of 53398 cancer cases was recorded (pathology reports: 24941, death certificates: 20468, hospital records: 7989) for all cancers in 5 provinces from March 2008 to March 2011. After removing duplicated registrations and linkage of the data from the three data sources provided, 35643 cases were observed to be statistical (44.7% female and 55.3% male) but the number of new cases for GI cancers were 9,574 (43% female and 57% male). The mean age at diagnosis (for GI cancer) was 62.9 ±16.9 years generally and 64.8 ±16.1 and 60.6 ±17.7 years for men and women respectively. 20 patients were excluded because they were of unknown age. The majority of the patients were 60 years or older (62.36%). The men to women ratio for all cancer cases was 1.26 (year 2010) and for GI cancer it was 1.13 (year 2010).

Figure 1 shows the percentage of share of the final data from each data sources.

Based on the results of eight models that estimated the true number of cancer registry from the three data sources,
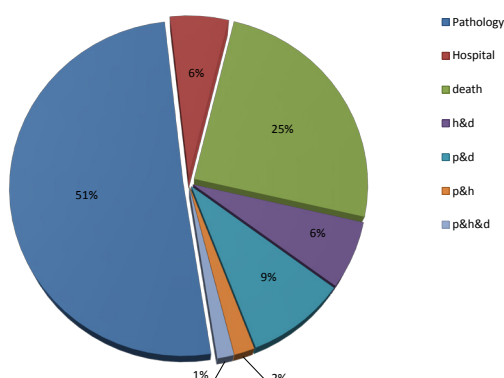
in which the 8th model was assumed to be the optimal model in this data that showed the lowest AIC (73.8) and BIC (73.3) measurements, it was estimated that 14,462.7 (95% CI: 13,434.9-15,764.2) GI new cases cancer have occurred in 5 provinces between 2008 and 2010 (Table 2).

The completeness of GI cancers was estimated generally and to the age, sex and year subgroups based on the estimated new cases (by use of a selected model in log-linear analysis) as shown in Table 3. The overall completeness for GI cancers (three sources) was 66.2% and it increased from 41.5% in March 2008 to 76% in March 2011.

Figure 2 shows that the number of cancer cases increased from 2,999 in year 2008 to 3,244 in year 2010.

The completeness shows the difference between men and women, and it was observed that the completeness of the women is higher than that of the men.

The highest completeness was observed in age group 3 (60 years and above) which was 76.6% and the lowest was observed in age group 1 (under 40 years) which was 31.8%.

The highest completeness is related to stomach cancer

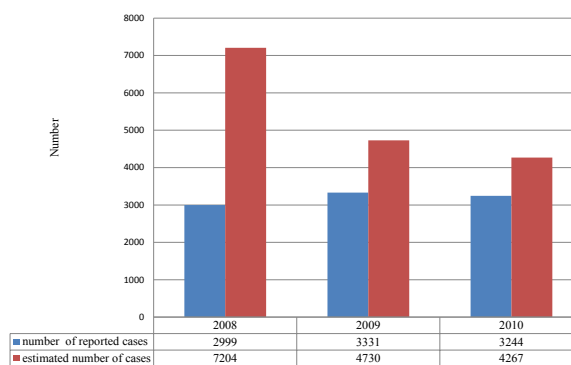Figure 1. Reporting Sources for Incident Cases in Iran, 2008-2010

Figure 2. Number of Reported and Estimated of GI Cancers in 2008-2010 in 5 Provinces of Iran

with an underestimation rate of 14.7% and the lowest is given to liver cancer with an underestimation rate of 43%.

Completeness of stomach cancer increased from 46.8% in 2008 to 85.3% in 2011; during these years, it has increased by 38.5%.

The percentages of the completeness of the cancer registration in colorectal cancers were 78.7% in 2010. Because of the low sample, completeness of liver and pancreases were not calculated in 2008.

Discussion

The findings in this study show that the mean age of the onset of cancer is different by gender and cancer incidence in men is more than in women. The majority of the patients were 60 years old or above (62.4%). The overall mean age was 62.9 years (64.8 for men and 60.6 for women) in 2010 and the mean age of women was lower than that of the men. In most studies conducted in other parts of Iran, average age was reported to be about 58-68 years (Biglarian et al., 2009; Rajaiefard et al., 2011; Aghaei et al., 2013; Khodadost et al., 2014). The men to women ratio for all cancer cases was 1.3 and for GI cancer it was 1.1 for year 2010.

This finding is consistent with country report (1.3) (Moh and Me, 2009) and regional reports such as: Mosavai-jarahi (1.4), (Mosavi-Jarrahi et al., 2013) and Aghaei (1.9), (Aghaei et al., 2013).

Also, in Finland and the union of Europe, this value is equal to 1.4 and 1.7, respectively and small difference in male to female ratio was observed (Bray et al., 2013), though this ratio was different from that of Ardabil's population-based cancer registry report of 2008 (2.2) (Khodadost et al., 2014). In this study, the capture-recapture method and log-linear models were used for estimating cancer registry and evaluating the dependence between pairs of sources, after which the most dependent sources were then grouped. Before estimating the missing data, the log-linear method was handled. In this case, the best option is to use a model with all possible interactions between sources (Regal and Hook, 1991), as such the best model was selected using AIC, BIC and G^2 statistics and it was discovered that the underestimation rate of cancer registry for all cancers were 49%.

The results of some studies such as Suwanrungruang et al., 2011,(Parkin et al., 2001; Dimitrova and Parkin, 2015) and (Lang et al., 2003) were higher than the results of this study . The estimated completeness was in the range of 70-99% in most studies.

The reasons may vary from differences between societies and structural and management differences in health and treatment systems of these countries.

Based on the results of some studies in Iran, sensitivity of cancer registration was lower or higher in other studies than in this study such as: Ghojazadeh in West Azerbaijan province (38.9%) (Ghojazadeh et al., 2013) and Lankarani in Fars province was above 100% (Lankarani et al., 2013). Although completeness of converge varied by cancer site, improvements in the completeness of converge in some places may have resulted from better communication with reporting centers such as pathology laboratories and

clinical centers, which were sent directives and guidelines regarding the registry as well as information regarding the obligation to report cancer cases (Lankarani et al., 2013).

Overall, sensitivity of cancer registries for all cancers in Ardabil province was 53.2% (Ghojazadeh et al., 2013) and that for all the selected hospitals of Shiraz was 58.6% (Sharifian et al., 2015). Also, sensitivity in Gambia was 50.3% (Shimakawa et al., 2013). The results are consistent with results of this study. This finding confirms the need for planning in order to improve the completeness of national cancer registry.

Because of the high fatality, the gastro intestinal tract of cancers was chosen for analysis. The log-linear method was used for statistical analyses. Sub-analysis revealed differences by site, sex and age groups. The overall completeness for GI cancers (three years) shows the difference between men and women respectively. Overall, the sensitivity of the system of this study in the estimation of the completeness of GI cancer was 66.2%, and it was considerably lower than the rate reported in Canada (95.83%) (Robles et al., 1988). Also total percentages of the completeness of cancer registration in Shiraz showed that the registration of females (67.6%) was higher than that of the males (41.4%) respectively (Sharifian et al., 2015), but in the study of Ardabil, completeness for men (68.5%) was higher than for women (Khodadost et al., 2014). Also, Matsuda's and et al (2014) research in Japan showed that the incidence percentage for males (58.4%) was higher than for females (41.6%) .

Studies carried out in Ardabil and Japan were similar to the present research, but they had lower registration percentage when compared with this study.

In this study, with increasing age, the completeness estimation was increased, and the highest and lowest completeness was observed in age group 60 and above (76.6%) and age group of under 40(31.8%) respectively. Other studies conducted in Iran and other parts of the world reported similar results such as: estimation for the old age group is similar to the incidence reported, among males and females in Japan (Matsuda et al., 2014). Also the percentages of completeness of cancer registration in ages between 60 to 69 and above 69 years old were 54% and 51.8% respectively, but the age group of under 20 years old was the lowest for cancer registration (Sharifian et al., 2015).

The percentages of the completeness of the cancer registration in stomach and colorectal cancers were 85.3% and 78.7% in 2010. The highest completeness is related to stomach cancer with an underestimation rate of 14.7%, and the lowest is given to liver cancer with an underestimation rate of 43% in this study. But in the study of Shimakawa (39), completeness for liver was estimated at 46.5% using the capture–recapture method and lower completeness was found in the liver, but they had lower registration percentage when compared with this study. It is possible that some of the metastatic liver could have been erroneously registered as primary tumors of liver in death certificates.

Completeness of gastric cancer registry in this study was considerably higher than the rate reported in Ardabil (35.9%) and Shiraz (58.7%) (Sharifian et al., 2015), but

lower than that of other countries such as Canada (95.9%) (Robles et al., 1988) and South Korea (93.2%) (Im et al., 2000) . This difference may be because of the time of that research, so this comparison seems not to be logical.

The results of the present research show that the registration for stomach cancer is higher than the other types of GI cancer, but it still needs to be improved.

The results of some studies showed that the percentage of the completeness of colorectal cancer was 38.9% for Sharifian (Sharifian et al., 2015), 66% for Zendehdel in Iran (Zendehdel, 2015) and approximately 60% for McClish in Virginia, (McClish and Penberthy, 2004), which is lower than that of the present study.

In the study conducted by Larsen, completeness of colorectal and liver cancers for the period of 2001–2005, estimated by the capture/recapture method, was approximately 99% and 98.6% respectively also indicating that the registration in those researches was higher than the registration done in this study .

The reasons may be as a result of the difference of the spread of colorectal cancer in developed and developing countries like Iran due to the differences in life styles, diets, and intake of alcohol in those countries and also the ability to diagnose the disease based on its different risk factors (Sharifian et al., 2015).

Cancer registry data make it possible to describe the size and dynamics of the cancer burden, to study cancer etiology, evaluate the effects of primary and secondary prevention and to plan health services. Thus, they have great relevance for public health and must be at all times as complete and valid as possible (Sigurdardottir et al., 2012). As a result, the ideal percentage for cancer registration is about 90% to 100% (Kroll et al., 2011).

The results of this study confirmed the underestimation rate in the cancer registry data. The extensive effort to collect, improve coverage and validity of information of the cancer registry are necessary. Considering the importance of subject, additional attention of authorities is necessary for improving the methods and plans of cancer registration in cancer registry centers. If the application conditions of the capture-recapture method are carefully adhered to, it becomes possible, to produce a correct estimate of the number of missing cases.

Acknowledgments

References

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

Agresti A, Kateri M (2011). Categorical data analysis, *Springer,* **104**, 182–7.

Biglarian A, Hajizadeh E, Kazemnejad A, et al (2009). Survival analysis of gastric cancer patients using Cox model: a five year study. *TUMJ,* **67**, 317-25.

Bray F, Parkin DM (2009). Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *Eur J Cancer,* **45**, 747-55.

Bray F, Ren JS, Masuyer E, et al (2013). Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int J Cancer,* **132,** 1133-45.

Brenner H, Stegmaier C, Ziegler H (1994). Estimating completeness of cancer registration in Saarland/Germany with capture-recapture methods. *Eur J Cancer,* **30**, 1659-63.

Dimitrova N, Parkin DM (2015). Data quality at the bulgarian national cancer registry: An overview of comparability, completeness, validity and timeliness. *Cancer Epidemiol Biomarkers Prevent,* **39**, 405-13.

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.

Habibi A (1984). Epidemiological aspects of cancer in Iran. *Int J Surg,* **70**, 105-8.

Hearst N, Hulley S (1988). Using secondary data. Designing clinical research, Williams and Wilkins, Baltimore, 53-62.

Hook EB, Regal RR (1995). Capture-recapture methods in epidemiology: methods and limitations. *Epidemiologic Reviews,* **17,** 243-64.

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.

Im JS, Kweon SS, Kim SY, et al (2000). Completeness estimation of the population-based cancer registration with capture-recapture methods. *Korean J Prev Med,* **33**, 31-5.

Kamo K-i, 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.

Khodadost M, Yavari P, Babaei M, et al (2014). Estimating the completeness of gastric cancer registration in Ardabil/Iran by a capture-recapture method using population-based cancer registry data. *Asian Pac J Cancer Prev,* **16**, 1981-6.

Kroll M, Murphy M, Carpenter L, et al (2011). Childhood cancer registration in Britain: capture-recapture estimates of completeness of ascertainment. *Br J Cancer,* **104**, 1227-33.

Lang K, Mägi M, Aareleid T (2003). Study of completeness of registration at the estonian cancer registry. *Eur J Cancer Prev,* **12**, 153-6.

Lankarani KB, Khosravizadegan Z, Rezaianzadeh A, et al (2013). Data coverage of a cancer registry in southern Iran before and after implementation of a population-based reporting system: a 10-year trend study. *Health Serv Res,* **13**, 169.

Matsuda A, Matsuda T, Shibata A, et al (2014). Cancer incidence and incidence rates in Japan in 2008: a study of 25 population-based cancer registries for the Monitoring of Cancer Incidence in Japan (MCIJ) project. *Jpn J Clin Oncol,* **38,** 490-5.

McClish D, Penberthy L (2004). Using Medicare data to estimate the number of cases missed by a cancer registry: a 3-source capture-recapture model. *Med Care,* **42**, 1111-6.

Modirian M, Kompani F, Rezaei N, et al (2014). Quality evaluation of national cancer registry system in iran: study protocol. *Arch Iran Med,* **17**, 193-8.

Moh , Me (2009). Iranian annual of national cancer registration report 2009, ministry of health and medical education, health deputy enter for disease control, Noncommunicable deputy, Cancer Control Office.

Mohagheghi MA, Mosavi-Jarrahi A (2010). Review of cancer

registration and cancer data in Iran, a historical prospect. *Asian Pac J Cancer Prev,* **11**, 1155-7.

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

Motevalian S, Mahmoodi M, Majdzadeh R, et al (2007). Estimation of death due to road traffic injuries in Kerman district: application of capture-recapture method. *J Sch Public Health Inst Public Health Res*, **5**, 61-72.

Organization WHO (2014). Global status report on noncommunicable diseases. Geneva: 2010. World Health Organization.[acesso em 2012 Fev 28]. Disponível em: http://www. who. int/nmh/publications/ncd_report2010/en.

Parkin DM (2006). The evolution of the population-based cancer registry. *Nat Rev Cancer*, **6**, 603-12.

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.

Parkin DM, Wabinga H, Nambooze S (2001). Completeness in an african cancer registry. *Cancer Causes Control*, **12**, 147-52.

Rajaiefard A, Moghimi B, Tabatabaie SH, 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.

Regal RR, Hook EB (1991). The effects of model selection on confidence intervals for the size of a closed population. *Stat Med*, **10,** 717-21.

Robles SC, Marrett LD, Clarke EA, et al (1988). An application of capture-recapture methods to the estimation of completeness of cancer registration. *JCE,* **41**, 495-501.

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

Schouten LJ, Straatman H, Kiemeney LA, et al (1994). The capture-recapture method for estimation of cancer registry completeness: a useful tool?. *Int J Epidemiol,* **23**, 1111-6.

Sharifian R, SedaghatNia MH, Nematolahi M, et al (2015). Estimation of completeness of cancer registration for patients referred to Shiraz selected centers through a two source capture re-capture method, 2009 data. *Asian Pac J Cancer Prev,***16**, 5549.

Shimakawa Y, Bah E, Wild CP, et al (2013). Evaluation of data quality at the Gambia national cancer registry. *Int J Cancer,* **132**, 658-65.

Shin H, Curado M, Ferlay J, et al (2007). Comparability and quality of data. Cancer Incidence in Five Continents, IARC, France, **9**, 752-9

Sigurdardottir LG, Jonasson JG, Stefansdottir S, et al (2012). Data quality at the icelandic cancer registry: comparability, validity, timeliness and completeness. *Acta Oncol*, **51**, 880-9.

Tilling K (2001). Capture-recapture methods-useful or misleading?. *Int J Epidemiol,* **30**, 12-4.

Torre LA, Bray F, Siegel RL, et al (2015). Global cancer statistics, 2012. *CA Cancer J Clin,* **65**, 87-108.

Zendehdel K (2015). Completeness and underestimation of cancer mortality rate in Iran: a report from Fars province in southern Iran. *Arch Iran Med*, **18,** 160.