

## RESEARCH ARTICLE

# Estimation of the Gastric Cancer Incidence in Tehran by Two-Source Capture-recapture

Abbas Aghaei<sup>1</sup>, Toraj Ahmadi-Jouibari<sup>1</sup>, Omid Baiki<sup>2,3</sup>, Alireza Mosavi-Jarrahi<sup>4\*</sup>

### Abstract

**Introduction:** Capture-recapture methods have been suggested for reducing costs of disease registration as well as reducing bias in incidence estimations. This study aimed to estimate the gastric cancer incidence in the Tehran metropolis population during 2002-2006. **Materials and Methods:** We investigated new cases of gastric cancer reported by three sources; death certificates, pathology reports, and medical records to Tehran population-based cancer registry during 2002-2006. G<sup>2</sup> statistics and the two-source capture-recapture method were used to select the best-fitted log-linear model and to estimate incidence, respectively. EXCEL software version 2007 and SPSS software version 16 were used for this research. **Results:** The number of reported cases was 4,463, with an average age of 68.5 (±12.9) years. We found the model that combined two sources of data including pathology reports and medical records and furthermore complemented by death certificates as the best model. The reported and the estimated incidences were 11.0 and 27.1 per 100,000 respectively. **Conclusions:** The incidence estimated by two-source capture-recapture method is about three times higher than the incidence reported by the sources under investigation. It is recommended to move towards the implementation of population-based cancer registration using various sources of data collection to achieve more accurate data.

**Keywords:** Incidence rate - two sources capture-recapture - gastric cancer - Tehran population - cancer registration

*Asian Pacific J Cancer Prev*, 14 (2), 673-677

### Introduction

Knowing of cancer incidence is the basis of planning, controlling, and promoting regional and national programs of cancer (Parkin et al., 1994; Schouten et al., 1994). There are various methods for estimating the hidden population, but when several incomplete lists are available, capture-recapture methods that are a direct method being applied (Robles et al., 1988; Cochi et al., 1989; Hilsenbeck et al., 1992; Laporte et al., 1992; McCarty et al., 1993). Among the techniques described in International Agency for Research on Cancer (IARC) reports, capture-recapture method has been considered at the same level of the best methods (Schouten et al., 1994). At the beginning, capture-recapture method was used to estimate wildlife population during the second half of the 20<sup>th</sup> century, and later it was applied in epidemiological studies, entitled "Multiple-record system" (Sekar and Deming, 1949; Wittes and Sidel, 1968; IWGDMF, 1995). During the last decades, capture-recapture methods have been increasingly used either in epidemiology or for the promotion of health care methods. Furthermore, they are used in various health-related fields to estimate hidden populations, registrations completeness, incidence, and prevalence of diseases and

special events. Such information is very important for policymakers and health workers (Hook and Regal, 1995; IWGDMF, 1995). Using capture-recapture methods is 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).

Tehran population-based cancer registry center has started since 1998. The main goal of this center is to measure cancer incidence in Tehran metropolis. Since Tehran is a referral center for all around the country, epidemiologists and experts of this center faced with the large amounts of data encountering many problems. Data collection area of this center includes Tehran metropolis plus Islamshahr and shemiranat townships (Mohagheghi et al., 2009; Aghaei et al., 2012; Haidari et al., 2012).

Central areas of Iran including Tehran have the highest incidence of gastric cancer after North and Northeast regions (Kolahdoozan et al., 2010). According to the first Tehran population-based cancer registry report during 1998-2001, malignant gastric cancer have allocated the first and the second ranks of cancers among men and

<sup>1</sup>Clinical Research Development Center, Imam Khomeini Hospital, <sup>3</sup>Department of Epidemiology, School of Public Health, Kermanshah University of Medical Sciences, Kermanshah, <sup>4</sup>Department of Epidemiology, The Cancer Research Center of the Cancer Institute, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran, <sup>2</sup>Department of Epidemiology, Unit of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden \*For correspondence: rmosavi@yahoo.com

women, respectively (Mohagheghi et al., 2009). This study aims to estimate the real incidence of gastric cancer in the Tehran population.

### Materials and Methods

New cases of gastric cancer that have been reported by three sources including death certificates, pathology reports, and medical records to Tehran Metropolitan Area cancer registry center during 2002-2006 years. Since Tehran is a referral center, cases not being among the areas covered by the registration center (Tehran, Islamshahr and Shemiranat townships) according to their address were excluded from data lists: 841 cases (25.6%) of death certificates, 10 cases (0.46%) of pathology reports and 3 cases (0.18%) reported by medical records. Duplicates were identified and removed using EXCEL software. Features such as name, surname, and father's name were used to identify common cases among sources and, if necessary, date of birth and national identification numbers were also used.

To use capture-recapture method, each list is considered as a sample and the names registered in any list are considered as unique characteristic of cases.

Total population is estimated based on "Proportionality argument" in Peterson-Lincoln model (Buckland et al., 2000). Where  $n_1$  includes cases reported by source 1,  $n_2$  includes cases reported by source 2 and  $m_2$  includes cases common between sources 1 and 2.  $N$  is an estimation of total study population.

In 1951, Chapman proposed the following formula to eliminate bias in Peterson's-Lincoln model:

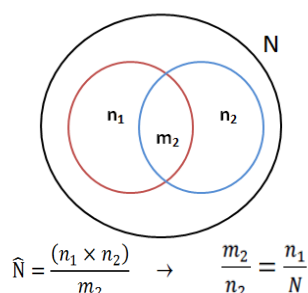
$$N_{cp} = [(n_1 + 1)(n_2 + 1) / (m_2 + 1)] - 1$$

We used the following formula to estimate 95% confidence interval:

$$95\%CI = \hat{N} \pm Z \sqrt{[(n_1 + 1)(n_2 + 1)(n_1 - m_2)(n_2 + m_2) / (m_2 + 1)^2(m_2 + 2)]}$$

Whites et al have suggested the comparison of two pairs of samples when more than two lists exist. When there is no correlation between lists and there is no heterogeneity between members, mutual estimations of samples should be almost similar to each other. If one of the estimations is less than the others, correlation between that sample pair should be suspected. They also recommend the combine of correlated lists and the production of a new sample (Wittes et al., 1974).

After choosing a model based on the suggestion of Whites et al,  $G^2$  statistic was used for model selection



too.  $G^2$  statistic is among the indices used for comparing the models and selecting the best one in the case of using two or more sources (Maydeu-Olivares and Cai., 2006). Therefore, when three sources are available, researcher calculate  $G^2$  statistic for possible choices of two sources capture-recapture method, and select the model with the lowest  $G^2$  statistic as the most appropriate model. We used SPSS package (version 16) to perform statistical analysis.

The incidence rate is product of the new cases of an event (gastric cancer) occurred at a certain time over the number of the population at risk at that time. For calendars year the population of the middle of the same year and for age and gender subgroup we used 5-year population of 2002-2006 as denominator. In this study, all the incidence rates are reported based on the incidence per 100,000 persons.

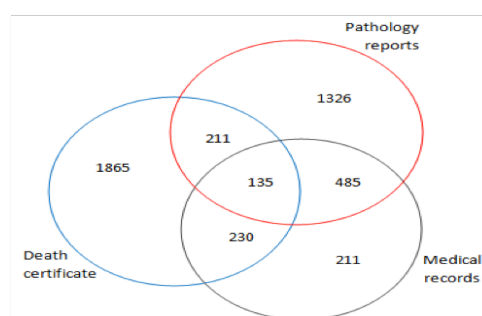
We also estimated the incidence rates by the selected model and by Peterson-Chapman method for the subgroup of study calendar years, gender subgroup, and age subgroups.

### Results

After investigating common cases between sources and removing duplicates, new cases of gastric cancer reported to Tehran population-based cancer registry during 2002-2006 were equal to 4463. The mean age of all cases was 68.5 ( $\pm 12.94$ ) years \_ 68.2 ( $\pm 12.1$ ) years for men (2970 cases) and 67.7 ( $\pm 14.2$ ) years for women (1493 cases). Male to female ratio was 1.99 in men to women.

Estimation of all cases obtained by two sources capture-recapture method for three sources of death certificates, medical records and pathology reports was conducted in two steps. In the first step, we performed the estimation through possible choices of tow sources together. In the second step, we combined the choices indicating high correlation according to the suggestions by Whites et al. That resulted in the estimation from three sources. To compare models selection method, the log-linear model was fitted for all possible choices obtained from three sources (six choices). The model with the lowest value for  $G^2$  statistic was chosen as the best model (Table 1).

Overall incidence rate of gastric cancer and incidence rates for each study subgroups estimated based on the Census of 2006 and based on the population estimation



**Figure 1. New Cases of Gastric Cancer Reported by Three Sources of Death Certificates, Pathology Reports, and Medical Recordsto Tehran Metropolitan Area Cancer Registry Center During 2002-2006**

**Table 1. Results of Sources' Mutual Comparison by Chapman Estimator and G<sup>2</sup> Goodness-of-Fit Statistics**

Sources*	Estimated of the hole cases	Standard error	95%confidence interval	G <sup>2</sup> goodness of fit statistics
1/2	15196	481	(16141 -14252)	9421
1/3	7091	129	(7344 - 6838)	5511
2/3	3692	21	(3734 - 3649)	4754
2/1&3	8142	62	(8264 - 8020)	1054
3/1&2	5310	44	(5395 - 5224)	1084
1/3&2	11004	191	(11379 -10630)	259

\*Source 1=death certificate, source 2=pathology reports, source 3=hospital documentation. In mode 1/2, data of death certificate and pathology reports are compared to each other. In mode 1/3, data of death certificate and medical records are compared to each other. In mode 2/3, data of medical records and pathology reports are compared to each other. In modes 1&2/3, data of two sources of death certificates and pathology reports are combined and compared to medical records data. In modes 1&3/2, data of two sources of death certificates and medical records are combined and compared to pathology reports data. In modes 2&3/1, data of two sources of pathology reports and medical records are combined and compared to death certificates data

**Table 2. Estimating Tehran Population-based Gastric Cancer Incidence during 2002-2006**

subgroups	Reported incidence rate*	Estimate of incidence rate (in 100000)	95%confidence interval of estimation
Calendar (year)	2002	11.8	27.7 (24.5-30.9)
	2003	12.0	29.4 (25.9-32.9)
	2004	10.3	25.5 (22.2-28.8)
	2005	10.1	24.6 (21.4-27.7)
	2006	10.7	27.6 (23.9-31.2)
Sex	Male	14.3	33.9† (31.7-36.2)
	Female	7.5	19.7 (17.7-21.9)
Age group (year)	50≤	2.0	5.2 (4.3- 6.0)
	51-64	26.8	61.9 (55.5-68.4)
	65≥	112.1	266.3 (247.4-285.1)
hole	11.0	27.1	(25.5-28.6)

\*Incidence reported by death certificates, pathology reports and hospital documentation. †Incidence for sex and age subgroups is obtained in terms of the number of the population at risk during 5 years of study

of 2002-2005 for the population of Tehran metropolis (Tehran, Islamshahr and shemiranat townships) (Table 2).

## Discussion

In this study we estimated cancer incidence by capture-recapture method in a population-based cancer registry. There are two important and relevant assumptions when using capture-recapture methods in epidemiological studies. First, if two sources are used, they should be independent of each other. Second, all people should have equal chance of entering to the sources. To reduce the problems raised from the lack of these assumptions, using of more than two sources or using of stratified methods to equalize chances of cases registration in each source are recommended (Hook and Regal, 1995; McClish and Penberthy, 2004). In this study, three sources were used to reduce the effects of these two assumptions. We selected gastric cancer as a high morality tumor to be able to use death certificates as third source for this study. Furthermore, we tried to control the effect of moderator variables by investigating incidence stratified by age, gender and calendar years.

The average age of incidence was 68.5 (68.2 for men

and 67.7 for women). The age of incidence does not show much difference between men and women, but is slightly higher than the average age reported by other studies conducted in the country, so that the average age is reported about 58-65 years in most studies (Hajyan et al., 2003; Biglaryan et al., 2008; Biglaryan et al., 2009; Mehrabian et al., 2010; Rajaeifard et al., 2011). It should be noted that most of the researches have been conducted based on data from pathology reports and death certificates data have not been considered or have not been of much importance. But in the presented study, data are collected based on the population-based registration data in which death certificates are considered as one of the main used sources. This can affect the increase of the average incidence age in registered cases reported only by the source of death certificates to the population-based cancer registry system.

Incidence in men over women was 1.99, while this ratio is different from 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 and also has been equal to 2.33 in Ardabil's population-based cancer registry report during 2004-2006 (Babaei et al., 2009). These ratios show very differences in cancer incidence in male and female subgroups of the population under the coverage of Tehran Metropolitan area Cancer Registration in comparison with other country and regional reports. Based on World Health Organization (WHO) report on cancer in 2008, gastric cancer pattern is similar in women and men, except that the incidence in men is twice women (Boyle and levin., 2008), and this value is consistent with the findings of this research.

The estimation obtained from two sources of pathology reports and medical records by Chapman method is less than other estimations and even less than the reported cases (3692<4463) (Table 1). This shows high dependency of two sources of pathology reports and hospital documentation. So, according to the method of Whites et al, we combine them and consider them as one source. Model selection by Chapman estimator and model fitness by G<sup>2</sup> statistic showed similar results. In both methods, the model in which two sources of pathology reports and medical records were interdependent and consequently combined and investigated as one source along with death certificates was considered as the best model.

According to WHO report, the countries of East Asia, East Europe and some countries of Latin America are among the countries with high gastric cancer incidence and African countries, whites of America, Indians and some countries of west Europe are populations with low incidence (Boyle and levin, 2008). In addition, GLOBOCAN (2008) estimation shows that incidence in East Asia is in the range of 40-62 and 18-24 per one hundred thousand populations in men and women, and incidence in East Europe is about 30 and 15 per one hundred thousand populations in men and women, respectively; And incidence is less than 5 cases per one hundred thousand populations for both men and women in countries such as Egypt, Arabia, India and South Africa. However, in this report, the estimated incidence is 21.9 and 9 per 100,000 population for Iranian men and women,

respectively (GLOBOCAN, 2008), similar to the values related to the countries of East Asia. Of course, it is notable that statistics reported by GLOBOCAN (2008) are about national (country) incidence and there may be populations with more or less incidence in these countries, so that north and north-eastern regions of Iran as well as Ardabil, central regions and southern regions of the country have very high, mean and low incidence, respectively (Kolahdoozan et al., 2010). Babaei et al, reported gastric cancer incidence as 51.8 and 24.4 per 100,000 during three years for men and women, respectively (Babaei et al., 2009).

Incidence estimations among calendar years are almost similar to each other, and most differences are related to the estimations of the years of 2003 (29.4) and 2005 (24.6). This is justifiable regarding the fact that one of the sources used is death certificate and 5-year survival rate of gastric cancer is about 20%; such that cases becoming afflicted by gastric cancer in early years of this study are more likely to be found in the list of gastric cancer fatalities; consequently more cases will be reported by the source of death certificates for early years of study. Therefore, it is expected that the middle years of this study, i.e. the years of 2004 and 2005 show estimates more similar to actual incidence.

Completeness estimation in age subgroups is less than the values reported in the studies conducted in Japan country and Ardabil province in Iran and on the other hand, is higher than most studies conducted throughout the world. Incidence estimation for over 65-year-old age group is similar to the incidence reported by countries with high incidence rates (Boyle and Levin, 2008; Matsuda et al., 2008; Babaei et al., 2009).

As findings of this study show, the estimated incidence rate by capture-recapture method is more than twice the reported values. According to global estimations and studies conducted throughout the world, the value estimated in this study is more similar to reality compared to what is reported by studied sources to Tehran population-based cancer registry system. So more researches should be done regarding this issue. Therefore, it is recommended to move toward the implementation of population-based cancer registration using various sources of data collection to achieve more accurate data.

## References

- Aghaei A, Najafi F, Mosavi-Jarrahi A, Ahmadi-Jouibari T (2012). Coverage, density and completeness of sources used in Tehran metropolitan area cancer registry: according to the data of esophageal cancer, 2003-2007. *Asian Pac J Cancer Prev*, **13**, 3617-9.
- Babaei M, Jaafarzadeh H, Sadjadi AR (2009). Cancer incidence and mortality in ardabil: report of an ongoing population-based cancer registry in Iran, 2004-2006. *Ir J Public Hlth*, **38**, 35-45.
- Biglarian A, Hajizadeh E, Kazemnejad A, Zali M (2009). Survival analysis of gastric cancer patients using cox model: a five year study. *Tehran University Med J (TUMJ)*, **67**, 24-32.
- Biglaryan A, Hajizade E, Gohari M, Khodabakhshi R (2008). Survival analysis of patients with gastric cancer and related factors. *Kowsar Med J*, **12**, 337-47.
- Boyle P, Levin B (2008). World cancer report 2008, IARC Press, Int Agency for Research on Cancer, 308-10.
- Buckland ST, Goudie IBJ, Borchers DL (200). Wildlife population assessment: past developments and future directions. *Biometrics*, **56**, 1-12.
- Cochi SL, Edmonds LE, Dyer K (1989). Congenital rubella syndrome in the United States, 1970-1985. *Am J Epidemiol*, **129**, 349-55.
- GLOBOCAN (2008). Available from: <http://globocan.iarc.fr/factsheets/cancers/stomach.asp>.
- Haidari M, Nikbakht MR, Pasdar Y, Najafi F (2012). Trend analysis of gastric cancer incidence in Iran and its six geographical areas during 2000-2005. *Asian Pac J Cancer Prev*, **13**, 3335-41.
- Hajyan K, Firuzjahi A, Taghikya M (2003). Epidemiological study of cancers in the city of Babylon in the age pattern 2001. *J Med*, **27**, 239-45.
- Hilsenbeck SG, Kurucz C, Duncan RC (1992) Estimation of completeness and adjustment of age-specific and age-standardized incidence rates. *Biometrics*, **48**, 1249-62.
- Hook EB, Regal RR (1995). Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev*, **17**, 243-9.
- International Working Group for Disease Monitoring and Forecasting (1995). Capture-recapture and multiple-record systems estimation II: applications in human diseases. *Am J Epidemiol*, **142**, 1059-68.
- Kolahdoozan S, Sadjadi A, Amir Reza R, Khademi H (2010). Five common cancers in Iran. *Arch Iran Med*, **13**, 143-6.
- Laporte RE, Tull ES, McCarty D (1992). Monitoring the incidence of myocardial infarctions: applications of capture-mark-recapture technology. *Int J Epidemiol*, **21**, 257-63.
- Laporte RE, McCarty DJ, Tull ES, Tajima N (1992). Counting birds, bees, and NCDs. *Lancet*, **339**, 494-9.
- Matsuda T, Marugame T, Kamo K (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.
- Maydeu-Olivares A, Cai L (2006). A cautionary note on using G<sup>2</sup> (dif) to assess relative model fit in categorical data analysis. *Multivariate Behavioral Res*, **41**, 55-64.
- McCarty DJ, Tull ES, Moy CS, Kwok CK, LaPorte RE (1993). Ascertainment corrected rates: applications of capture-recapture methods. *Int J Epidemiol*, **22**, 559-64.
- McClish D, Penberthy L (2004). Using multivariate capture-recapture techniques and statewide hospital discharge data to assess the validity of a cancer registry for epidemiologic use. *Hlth Services and Outcomes Res Methodol*, **5**, 141-52.
- Mehrabian AA, Esna-Ashari F, Zham H (2010). Gastric cancer prevalence, according to survival data in Iran (National Study-2007). *Ir J Public Hlth*, **39**, 27-31.
- Mohagheghi MA, Mosavi-Jarrahi A, Malekzadeh R, Parkin M (2009). Cancer incidence in tehran metropolis: the first report from the tehran population-based cancer registry, 1998-2001. *Arch Ir Med*, **12**, 15-23.
- Parkin DM, Chen VW, Ferlay J, et al (1994). Comparability and Quality Control in Cancer Registration. Lyon: International Agency for Research on Cancer. IARC Technical Report.
- Rajaeifard A, Moghimi-Dehkordi B, Tabatabaei SMR, Safaei A, Tabeieet SZ (2011). Epidemiological and clinical features of gastric cancer: a descriptive study of registered cases of cancer in fars province (1380-85). *Teb-E-Jonoob*, **14**, 114-21.
- Robles SC, Marrett LD, Aileen CE, Risch HA (1988). An application of capture-recapture methods to the estimation of completeness of cancer registration. *J Clinical Epidemiol*, **41**, 495-501.
- Schouten LJ, Straatman H, Kiemeny LALM, Gimbrere CHF,

- Verbeek ALM (1994). The capture-recapture method for estimation of cancer registry completeness: a useful tool? *Int J Epidemiol*, **23**, 1110-4.
- Sekar CC, Deming WE (1949). On a method of estimating birth and death rates and the extent of registration. *J Am Statistical Assoc*, **44**, 101-15.
- Tilling K (2001). Capture-recapture methods—useful or misleading? *Int J Epidemiol*, **30**, 12-4.
- Wittes J, Sidel VW (1968). A generalization of the simple capture-recapture model with applications to epidemiological research\*1. *J Chronic Diseases*, **21**, 287-301.
- Wittes JT, Colton T, Sidel VW (1974). Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *J Chronic Diseases*, **27**, 25-31.