

## RESEARCH ARTICLE

# Estimation of Cancer Cases Using Capture-Recapture Method in Northwest Iran

Morteza Ghojazadeh<sup>1</sup>, Marziye Mohammadi<sup>2</sup>, Saber Azami-Aghdash<sup>3</sup>, Alireza Sadighi<sup>4</sup>, Reza Piri<sup>5</sup>, Mohammad Naghavi-Behzad<sup>6\*</sup>

### Abstract

**Background:** Under-ascertainment and over-ascertainment are common phenomena in surveillance and registry systems of health-related events. Capture-recapture is one of the methods which is applied to determine the sensitivity of surveillance or registry systems to recognize cancer cases. This study aimed to estimate the number of cancers using data available both in the Cancer Registry Center of Northwestern Iran and in the Population-based Cancer Registry Center of Iran. **Material and Methods:** The studied population consisted of all cancerous cases in the northwest of Iran from 2008 to 2010. All data were extracted from two resources and entered into Microsoft Excel software. After removing common and repeat cases the data were statistically analyzed using a capture-recapture studies' specific software "CARE 1.4". Estimations were calculated by Chapman and Petersen methods with the approximate confidence interval of 95%. **Results:** From 2008 to 2010, the number of all cancer cases was estimated to be 21,652 (CI 95%: 19,863-22,101). Sensitivity rate of all cancer cases was 83.9% and that of Population-based Cancer Registry Center of Iran was 52%. It was 93.1% considering both resources. **Conclusion:** Using two resources and the capture-recapture method rather than a single resource may be a more reliable method to estimate the number of cancer cases.

**Keywords:** Capture-recapture method - coverage rate - sensitivity rate - cancer registry - incidence rate

*Asian Pacific J Cancer Prev*, 14 (5), 3237-3241

### Introduction

Nowadays, having accurate and updated information, especially information about mortality, is one of the most important needs for providing effective medical and health care (Doungchawee et al., 2002; Alireza et al., 2005; Sadjadi et al., 2008). Having information about chronic diseases, especially cancers as the second cause of human mortality following cardiac diseases, has a unique importance (Alireza et al., 2005; Sadjadi et al., 2008; Chen et al., 2012).

Accurate information about mortality rate of cancer is the most essential requirement to control cancer. Considering multi-factorial identity of cancer, different diseases with different etiologies, it is not enough to be just aware of its rate in society. The different incidences of cancer diseases in different geographical areas, races, societies and occupational groups highlight the importance of comprehensive information about it (Ali et al., 2010; Kamsa-ard et al., 2011; Zou et al., 2011; Pandey et al., 2012).

Lack of adequate information regarding the behavior

of cancerous diseases and their incidences from some decades ago in different geographical areas, especially in developing countries like Iran, has been a significant challenge in the path of designing programs to investigate cancer control and prevention. During recent years, a wide range of efforts has been conducted to design "cancer registry" systems (Doungchawee et al., 2002; Mousavi et al., 2008; Sadjadi et al., 2008; Zou et al., 2011). In this direction, to ensure the quality of gathered data, many methods have been developed in the recent years and capture-recapture is one of the most applicable methods (Chao et al., 2001; Behtash et al., 2006; Hajian et al., 2011).

In a capture-recapture study, some presuppositions should be established: i) the under study population should be a "closed" one, i.e. no subject is added or omitted during study, ii) common cases should be found between two or more indexes, iii) the indexes should be independent, i.e. if there is a person in one index, the possibility of his/her existence in another index would not increase or decrease, and iii) probability of persons presence in the indexes should not depend on their characteristics (Ballivet et al.,

<sup>1</sup>Liver and Gastrointestinal Disease Research Center, <sup>2</sup>Nursing and Midwifery Faculty, <sup>3</sup>Health Services Management Research Center, <sup>4</sup>Students' Research Committee, <sup>5</sup>Medical Philosophy and History Research Center, Tabriz University of Medical Sciences, Tabriz, Iran, <sup>6</sup>Department of Neurosurgery, University of Wisconsin-Madison, Madison, WI, USA \*For correspondence: Dr.Naghavii@gmail.com

Cancerous cases of northwestern Iran were the matter of consideration for Cancer Registry Center of Health Research Station of Northwestern Iran, regarding high incidence of cancer at this region (Sadjadi et al., 2003; Yazdizadeh et al., 2005). Cancer registry unit, based on Iran population, started to register cancerous items of years 2006-2007, retrospectively. Since the sensitivity of cancer registry system is of great importance, this study aimed to investigate and estimate the coverage rate of cancerous systems in the above mentioned Iranian centers using capture-recapture methods from 2008 to 2010.

## Materials and Methods

The present study was carried out using two-sample Capture-Recapture method based on the data available in the Cancer Registry Center of Northwestern Iran and Population-based Cancer Registry Center of Iran.

All cancerous cases occurred in the northwestern region of Iran from 2008 to 2010 and data were extracted from two under study resources. Cases registered from causative mortality system at each of the above-mentioned centers entered the study providing that the disease incidence and patient's mortality have occurred in the same year. First, all cancers registered at cancer registry center of northwestern Iran during 2008-2010 were extracted and entered the study, and then cancer data were registered in SPSS15 software into Persian, then gathered data, were transferred to *Microsoft Excel* and changed into English. Afterwards, all cancerous items registered at Iranian cancer registry center during 2008-2010 were extracted and entered the study. Data of the center was extracted from special forms used to register cancers in this unit. After being registered at data collection forms, data were entered *Microsoft Excel* in English format.

To prevent misspelling and non-uniform names considering their spelling, patients' names and surnames were written in a notebook in English alphabetical order. If the same name was repeated in the same or other years, or in another resource, the repeated name was recorded as it was previously typed referring to the mentioned notebook. This is done to prevent misspelled names. Following steps were taken into account when names were entered into computer, also: i) According to the data collection form, the required variables in the study were selected and other variables data were deleted. ii) Names were written using capital letters. iii) In compound names, only the first letter was capitalized and the rest of the name was typed in lowercase letters with no space in between. (e.g. Mohammadreza instead of Mohammad Reza). iv) Letter "ç" was written as "aa". v) The stressed letters were repeated (e.g. Abbas). vi) In compound surnames, the second part was capitalized with a space between two parts (e.g. Mohammad Zadeh instead of Mohammadzadeh). vii) The prefixes such as Seyyed, Seyyedeh, Mir, Haji, Hajieh, Molla, Sheikh, Mirza, Agha, and Karbalaei were omitted from beginning of names. viii) In the names such as Forouzan, the second "o" sound was written in this format "ou". ix) The suffixes such as Zadeh, Pour, Far, Panjeh, Mehr, Nejad, Sani, Nasab, etc. were written

without any space in between.

### Identifying the common cases

To identify common cases following steps in the prepared indexes for excel files data were taken into account: The names were reviewed and misspellings were corrected i) The subjects registered in the index section of all resources were reviewed and the repeated ones were deleted.

Then, the indexes were merged two by two and following actions were taken into account: i) "Sort" order was used to arrange the files by surname, name, father's name, and age, respectively. ii) Cases which were common in three or four characteristic of the mentioned variables were called common cases. Cases which lack one or two of the above-mentioned characteristics but are common in other two or three characteristics, also called common cases. Otherwise, two common characteristics do not indicate to commonness of the cases. iii) If surnames had four or more common letters, the name should be considered. If the name is also same, father's name should be looked for. The case will be regarded as a common one if father's name is also the same.

Two-sample capture-recapture method is demonstrated by Van diagram. The rectangular stands for total population (n) which requires measurement. There are two overlapped circles (L1 & L2) inside the rectangular indicating to resources. Missed circles (m) surround these two circles.

*Chapman* formula is used to calculate total (n) as follows:

$$n = [(L1+1)(L2+1)] / (d+1) - 1 \text{ Formula}$$

Where, L1 and L2 refer to number of cases in each resource and d stands for common cases of two resources. The formula is calculated with the confidence interval of 95% as follows (Chao et al., 2001):

$$CI_{95\%} = n \pm 1.96 \sqrt{\frac{(L1+1)(L2+1)(L1-d)(L2-d)}{(d+1)^2(d+2)}} \text{ Formula}$$

Investigating establishment of capture-recapture assumptions

**Closed population:** The aim of this study is to estimate cancerous items in northwestern Iran. As a result the cancerous items during 2008-2010 at northwestern region of Iran are a set of closed population in a way that closed population assumptions are confirmed in this set. Also of homogeneity issue of geographical coverage, it is possible to say that since all registered cancerous items in each of resources were habitants of the incidence place during recent 10 years, geographical coverage of both sources, was as same as that of the incident place i.e. East Azerbaijan, West Azerbaijan, Ardabil, and Zanjan provinces.

**Possibility of finding mutual cases among two or more indexes:** Since characteristics used to identify common cases (including name, last name, Father's name, and age) are registered in both indexes, the second assumption is also met.

**Independence of resources:** Since each of centers individually registers cancerous items in northwestern

Iran, theoretically resources are completely independent.

#### Dependence of capture to persons' characteristics:

About the fourth assumption, on cancers, sex characteristic had not a significant relationship with presence in each of resources ( $P > 0.05$ ). However for habitat characteristic, there was a significant relationship between habitat and persons' presence in two resources ( $P < 0.05$ ), in a way that presence ratio of urban people to rural people in this resource was 1/3. Also about age, there was not any significant relationship between age and presence of persons in each of resources ( $P > 0.05$ ).

Once cases that were identified only by each of the resources and number of common cases were counted, the data were analyzed using SPSS15 statistical software, Chi-square and McNemar tests, and 1.4 CARE software (Zhou et al., 2007). Peterson and Chapman methods (Chao et al., 2001) were used to calculate the estimations.

## Results

Totally, 22605 cases of various cancer types were extracted from population-based cancer registry center of Iran and cancer registry center of Iran health research station from 2008 to 2010. Out of them, 9315 and 13290 cases were registered by cancer registry center of northwestern Iran and cancer registry center of Iran, respectively. After omitting repeated cases, the cases decreased to 21652 registered in the both above mentioned resources. Again out of 21652 cases, 9163 and 12489 cases were registered by cancer registry center of northwestern Iran and cancer registry center of Iran, respectively. Also of 21652 cases, 75623 cases had been registered in both resources, concurrently.

There was not any statistically meaningful difference between subjects registered in both resources, in terms of patients gender ( $P = 0.24$ ).

Civil proportion to rustic ones is 1:2 in the cancer registry center of northwestern Iran while it is 1:8 in the cancer registry center of Iran. Considering three years of study, there was a statistically meaningful difference between subjects registered in both resources considering residency ( $P < 0.001$ ).

Also, more than 67% of subjects registered at every resource are older than 45 years old. There was not any statistically meaningful difference between the registered subjects considering age ( $P = 0.16$ ).

During the study, there were estimated 6702 (with confidence interval of 95%: 6183-6948) and 14950 (with confidence interval of 95%: 14381-15249) cases of digestive system cancers and other cancers respectively. The analysis was conducted on digestive system and other cancers considering high incidence rate and prevalence of digestive system cancers in these regions.

According to the northwestern Iran Center, categorized estimation for each city of the region indicates the highest rate of sensitivity for cancer registry in Ardabil (96.3%) and Zanjan (88.2%) provinces. Based on reports of northwestern Iran center, the highest and lowest rates of sensitivity for cancers registry relates to Ardabil (53.2%) and West Azerbaijan (38.9%) provinces, respectively. According to common report of both resources, the highest

and lowest rate of sensitivity for cancer registry belong to Ardabil (91.8%) and West Azerbaijan (87.2%) provinces respectively.

Classified gender-based estimation indicates equal sensitivity rates for both genders (92.7%).

According to the northwestern Iran center, classified residency-based estimation demonstrated that sensitivity rate in the registration of cancerous cases in rustics is higher than that of the civil ones while cancer center of Iran refers to high rate in civil rather than rustics. Total rate of sensitivity (sensitivity of both resources) is almost equal for civil and rustics.

According to northwestern Iran center, categorized age-based estimation demonstrated that highest and lowest rates of sensitivity in registering cancerous cases respectively belongs to age group of 5-14 (83.9%) and 0-4 (94.7%) years old while cancer center of Iran refer to the highest and lowest rates in age groups of 5-14 (67%) and 0-4 (8.3%) years age. Generally, the highest rate of total sensitivity (sensitivity of both resources) was seen in age group of 5-14 (91.9%) and the lowest rate of total sensitivity was observed in age group of 0-4 (76.2%) years old. Over three years, starting 2008 up to 2010, the under-ascertainment rate for all cancers is 16.1%, based on the report of Cancer Registry Center of Northwestern Iran, is 48% and based on the report of Population-based Iranian Cancer Registry Center is 6.9% considering both resources.

## Discussion

During the study, under-ascertainment of all cancers were 16.1%, 48%, and 6.9% in accordance to the reports of cancer registry center of northwestern Iran, population-based cancer registry center of Iran, and both resources.

In studies conducted in different places, various amounts of under-ascertainment were reported for cancerous cases (Al-Zahrani et al., 2003; Hafdi-Nejjari et al., 2006; Thompson et al., 2007; Zhou et al., 2007; Zhang et al., 2012). In the study conducted in Linzhou, China (Zhou et al., 2007) using three-sample capture-recapture method for mortalities resulting from malignant tumors, there was an under-ascertainment of 6.6%. It was 0.8% in a study conducted in Rhone Alpes, France (Hafdi-Nejjari et al., 2006) using two-sample capture-recapture method to estimate cancers. Riyadh study (Saudi Arabia) (Al-Zahrani et al., 2003) as a three-sample capture-recapture method estimating cancers in 2003, revealed sensitivity and under-ascertainment rate of 68% and 32%, respectively. Sensitivity and under-ascertainment rate were 91% and 9% in the study conducted at the University of Harvard (Wang et al., 2001) using two-sample capture-recapture method to estimate The number of breast cancers for the next three years, starting 2001.

A study in Tuscany of Italy (Crocetti et al., 2001) using two-sample capture-recapture method to estimate cancers in 2001 indicated 97.4% sensitivity rate and 2.6% under-ascertainment rates. Therefore, results of our study demonstrate that sensitivity rate of both resources are acceptable for cancer cases registration and also there is a low under-ascertainment rate in both resources.. In study

of Dortaj et al. (2011) in the cities of Fars province of Iran on coverage amount of data from cancer-related mortality registry, sensitivity rate has been reported 58%. According to the under-ascertainments in different mentioned countries we can understand that under-ascertainment rate in developed countries is low due to their proper facilities and better information system and also based on the importance they give to accurate statistics and reference-based information. When we compare the results of current study with the results obtained from mentioned studies, we can see that sensitivity rate of two references for registering cancerous cases is acceptable but under-ascertainment rate of two sources is low. Therefore according to the importance of accurate information especially about cancers to provide proper and effective health and medical care services and also considering high cancer prevalence in Iran especially at northern and northwestern regions (Semnani et al., 2006; Sadjadi et al., 2007; Yavari et al., 2008; Suwanrungruang et al., 2011; Ghojzadeh et al., 2012a; 2012b), it is necessary to pay attention to quality and efficiency of data registry systems and centers and policy makers and planners of health and medical system should have special attention to this issue.

Over three years, starting 2008 up to 2010, based on the reports of Cancer Registry Center of Northwestern Iran, the highest sensitivity rate has been reported in the Ardabil province of Iran and the lowest sensitivity rate was reported from the West Azerbaijan province. This shows the necessity of sever attention by authorities of mentioned cancer registry centers to register cancerous cases in West Azerbaijan province.

Also based on reports of both Cancer Registry Center of Northwestern Iran and Population-based Cancer Registry Center, the highest sensitivity rate is related to digestive system cancers. The high sensitivity of these centers registering digestive system cancers is due to high incidence of this type of cancers in northwest of Iran which needs more investigations.

Population-based Cancer Registry Center of Iran was equipped with electronic data bank by which all cases were registered in SPSS software in Persian format. Since the repetitive cases were identifiable in this software, the available data were transferred to *Microsoft Excel* software. Therefore, it is suggested that this pattern may be of value to be used in the cancer registry center of northwestern Iran and other cancer registry centers of the country.

Considering usefulness of capture-recapture analysis and necessity of non-repetitive cases, it is necessary to provide conditions to not register repetitive cases in designing cancer registry software.

Considering necessity of availability of cancer registry data, it is recommended to design a web-based program to observe data cohesion and prevent registry of repetitive cases. All pathology clinics and hospitals should be obliged to register identified cases of cancers in this program.

To prevent registration of repetitive cases and over-ascertainment, it is recommended to register cancerous cases based on full characterization and national ID number of the subjects, if possible.

As a weak point of this study its double-resource nature should be mentioned since in this form it is possible for some errors to appear due to the lack of meeting assumptions of capture-recapture methods in spite of simplicity and high speed of study. In spite of efforts by researchers of this study to avoid such errors, studies with more data registry resources are suggested.

In conclusion, finally, considering the importance of accurate information for planning and providing health and medical care services, and regarding to relatively weak functioning of cancer registry centers and systems, additional attention of authorities is necessary for improving the methods and plans of cancer registration in cancer registry centers.

## References

- Al-Zahrani A, Baomer A, Al-Hamdan N, Mohamed G (2003). Completeness and validity of cancer registration in a major public referral hospital in Saudi Arabia. *Ann Saudi Med*, **23**, 6.
- Ali K, Sutaryo S, Purwanto I, et al (2010). Yogyakarta pediatric cancer registry: an international collaborative project of university gadjah mada, university of saskatchewan, and the saskatchewan cancer agency. *Asian Pac J Cancer Prev*, **11**, 131-6.
- Ballivet S, Salmi LR, Dubourdieu D, et al (2000). Capture-recapture method to determine the best design of a surveillance system. Application to a thyroid cancer registry. *Eur J Epidemiol*, **16**, 147-53.
- Behtash N, Mehrdad N (2006). Cervical cancer: screening and prevention. *Asian Pac J Cancer Prev*, **7**, 683-6.
- Chao A, Tsay PK, Lin SH, Shau WY, Chao DY (2001). The applications of capture-recapture models to epidemiological data. *Stat Med*, **20**, 3123-57.
- Chen SM, Lin W, Liu X, Zhang YZ (2012). Significance of Human Telomerase RNA Gene Amplification Detection for Cervical Cancer Screening. *Asian Pac J Cancer Prev*, **13**, 2063-8.
- 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 Prev*, **10**, 417.
- Dortaj E, Bahrapour A, Haghdoost A, et al (2011). Completeness of fars province deaths registry on cancer death using capture recaptures method. *J North Khorasan University of Med Sciences*, **3**, 33-43.
- Doungchawee G, Y Khoaprasert, et al (2002). Use of Wild Rodents for Environmental Monitoring-Comparison of Rats in Bangkok and Rural Areas of Thailand. *Asian Pac J Cancer Prev*, **3**, 367-8.
- Ghojzadeh M, Azar ZF, Saleh P, Naghavi-Behzad M, Azar NG (2012). Knowledge and attitude of Iranian university students toward human papilloma virus. *Asian Pac J Cancer Prev*, **13**, 6115-9.
- Ghojzadeh M, Naghavi-Behzad M, Azar ZF, et al (2012). Parental knowledge and attitudes about human papilloma virus in Iran. *Asian Pac J Cancer Prev*, **13**, 6169-73.
- Hafdi-Nejjari Z, Couris CM, Schott AM, et al (2006). Role of hospital claims databases from care units for estimating thyroid cancer incidence in the Rhone-Alpes region of France. *Rev Epidemiol Sante Publique*, **54**, 391.
- Hajian S, Vakilian K, Najabadi KM, Hosseini J, Mirzaei HR (2011). Effects of education based on the health belief model on screening behavior in high risk women for breast cancer, tehran, iran. *Asian Pac J Cancer Prev*, **12**, 49-54.

- Kamsa-ard S, Wiangnon S, Suwanrungruang K, et al (2011). Trends in liver cancer incidence between 1985 and 2009, Khon Kaen, Thailand: cholangiocarcinoma. *Asian Pac J Cancer Prev*, **12**, 2209-13.
- Mousavi SM, Somi MH (2009). Gastric cancer in Iran 1966-2006. *Asian Pac J Cancer Prev*, **10**, 407-12
- Pandey S, Mishra CM (2012). Human papillomavirus screening in North Indian women. *Asian Pac J Cancer Prev*, **13**, 2643-6.
- Sadjadi A, Nouraei M, Mohagheghi MA, et al (2005). Cancer occurrence in Iran in 2002, an international perspective. *Asian Pac J Cancer Prev*, **6**, 359.
- Sadjadi A, Malekzadeh R, Derakhshan MH, et al (2003). Cancer occurrence in Ardabil: results of a population based cancer registry from Iran. *Int J Cancer*, **107**, 113-8.
- Etemadi A, Sadjadi A, Semnani S, et al (2008). Cancer registry in Iran: a brief overview. *Arch Iran Med*, **11**, 577.
- Sadjadi A, Zahedi M, Nouraei M, et al (2007). The first population-based cancer survey in Kerman Province of Iran. *Iran J Public Hlth*, **36**, 26-34
- Semnani S, Sadjadi A, Fahimi S, et al (2006). Declining incidence of esophageal cancer in the Turkmen Plain, eastern part of the Caspian Littoral of Iran: a retrospective cancer surveillance. *Cancer Detect Prev*, **30**, 14-9.
- 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.
- Thompson B, Thompson AL, Chan NL, Hislop GT, Taylor VM (2007). Cost effectiveness of cervical cancer screening among Chinese women in North America. *Asian Pac J Cancer Prev*, **8**, 287-93.
- Wang PS, Walker AM, Tsuang MT, et al (2001). Finding incident breast cancer cases through US claims data and a state cancer registry. *Cancer Causes Control*, **12**, 257-65.
- Yavari P, Sadrolhefazi B, Mohagheghi MA, et al (2008). An epidemiological analysis of cancer data in an Iranian hospital during the last three decades. *Asian Pac J Cancer Prev*, **9**, 145-50.
- Yazdizadeh B, Jarrahi AM, Mortazavi H, et al (2005). Time trends in the occurrence of major GI cancers in Iran. *Asian Pac J Cancer Prev*, **6**, 130-4.
- Zhang HY, Tiggelaar SM, Sahasrabudde VV, et al (2012). HPV prevalence and cervical intraepithelial neoplasia among HIV-infected women in Yunnan province, China: a pilot study. *Asian Pac J Cancer Prev*, **13**, 91-6.
- Zhou P, Wang BZ, Quan PL, et al (2007). Application of capture-recapture method in the malignant tumor death registry. *Zhonghua Yu Fang Yi Xue Za Zh*, **41**, 101.
- Zou L, Bao YP, Li N, et al (2011). Life-style and genital human papillomavirus in a cross-sectional survey in Shanxi Province, China. *Asian Pac J Cancer Prev*, **12**, 781-6.