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

Mahmoud Khodadost¹, Parvin Yavari¹²³*, Masoud Babaei⁴⁵, Alireza Mosavi-Jarrahi², Fatemeh Sarvi²⁴, Kamyar Mansori⁸, Behnam Khodadost⁹

Abstract

**Background:** 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. **Materials and Methods:** 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.**Results:** 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. **Conclusions:** 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
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 was established in Ardabil in 2003. The Ardabil cancer registry (ACR) actively collects information of cancer incidence from pathology-based, hospital-based and death certificates. The main goal of the ACR is to measure cancer incidence and mortality in residents of Ardabil province (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 kept confidential. We used STATA individual’s information such as name, surname and other calendar time respectively. In all stages of this study the completeness was calculated by age groups and incidence per one hundred thou and populations. Also the incidences are reported based on 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.
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 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 was 1.58.

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

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

*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. C: hospital records; Model P/C/D: A model where all available resources are independent; Model PD/C: A model where sources P and C are dependent and independent of the source D; Model P/D/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/CD: A model where two sources P and C and also two sources C and D are mutually independent and two sources P and D are dependent; Model PC/D: A model where two sources P and D and also two sources C and D are mutually dependent 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.**

<table>
<thead>
<tr>
<th>Calendar year</th>
<th>Subgroups</th>
<th>Reported new cases</th>
<th>Estimated new cases</th>
<th>95% CI for estimated new cases</th>
<th>Completeness registration (%)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006 Gender</td>
<td>Male</td>
<td>301</td>
<td>749.6</td>
<td>(598.12 - 978.33)</td>
<td>40.2</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>143</td>
<td>498.9</td>
<td>(321.17 - 853.94)</td>
<td>28.7</td>
</tr>
<tr>
<td>Age groups</td>
<td>50s</td>
<td>48</td>
<td>148.8</td>
<td>(79.12 - 274.13)</td>
<td>32.3</td>
</tr>
<tr>
<td></td>
<td>51-64</td>
<td>111</td>
<td>250.5</td>
<td>(177.43 - 363.57)</td>
<td>44.3</td>
</tr>
<tr>
<td></td>
<td>65+</td>
<td>285</td>
<td>841.5</td>
<td>(640.60 - 1155.88)</td>
<td>33.9</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>444</td>
<td>1209.5</td>
<td>(983.21 - 1532.85)</td>
<td>36.7</td>
</tr>
<tr>
<td>2008 Gender</td>
<td>Male</td>
<td>286</td>
<td>699.6</td>
<td>(554.22 - 923.86)</td>
<td>40.9</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>127</td>
<td>500.3</td>
<td>(303.43 - 916.95)</td>
<td>25.4</td>
</tr>
<tr>
<td>Age groups</td>
<td>50s</td>
<td>39</td>
<td>135.0</td>
<td>(64.46 - 400.86)</td>
<td>28.9</td>
</tr>
<tr>
<td></td>
<td>51-64</td>
<td>98</td>
<td>279.1</td>
<td>(175.58 - 520.84)</td>
<td>35.1</td>
</tr>
<tr>
<td></td>
<td>65+</td>
<td>276</td>
<td>732.0</td>
<td>(570.71 - 981.53)</td>
<td>37.7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>413</td>
<td>1148.0</td>
<td>(920.18 - 1478.14)</td>
<td>36.0</td>
</tr>
</tbody>
</table>

*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. Babaeei 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 (Babaeei 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; Babaeei 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.

Acknowledgements
The authors wish to express their sincere thanks to all staffs of the Ardabil cancer registry center, especially Dr. Babaei who sincerely helped the data collection.

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


