

RESEARCH COMMUNICATION

Relation between Demographic Factors and Type of Gastrointestinal Cancer using Probit and Logit Regression

Asma Pourhoseingholi, Mohamad Amin Pourhoseingholi, Mohsen Vahedi, Azadeh Safaee, Bijan Moghimi-Dehkordi*, Fatemeh Ghafarnejad, Mohammad Reza Zali

Abstract

Background: Approximately 50,000 new cases of cancer occur each year in the Iranian population of 70.4 million. The organ system involved with more than 38% of all cancers is the gastrointestinal (GI) tract. The objective of this study was to investigate the relation between demographic factors and type of gastrointestinal cancer using probit and logit models. **Methods:** This study was designed as a cross-sectional survey including all consecutive GI cancer patients admitted over a one year period in a randomly selected hospital group located in Tehran in 2006. **Results:** The largest number of cases were colorectal cancers (40.0%), followed by gastric cancers (34.5%) and esophagus cancers (17.1%). There was a significant gender effect in the colorectal, gastric and esophagus cancer also there was a significant association between age and gastrointestinal cancers in both logit and probit regression. The factor of duration was not significant in gastric cancer. **Conclusion:** Men are more likely have colorectal cancer than women. Older people are more likely to have gastric cancer than younger people. In esophagus cancer all factors were significant. Results from probit and logit models were similar, indicating that probit analysis can be employed as a logit model to analyze relationships between demographic factors and cancer type.

Key Words: Logit regression - probit regression - GI tract cancers - hospitalization

Asian Pacific J Cancer Prev, 9, 753-755

Introduction

Cancer is a diverse range of diseases where abnormal cells grow rapidly and generally spread throughout the body in an uncontrolled manner. While some cancers share common causes or risk factors, it is believed that most cancers have a unique set of causes. Cancer is a growing health problem in Iran and according to recent statistics represents the third most common known cause of death after cardiovascular disease and accidents (Naghavi, 2004). the gastrointestinal cancers are the most frequent cancer among Iranian males and second to breast cancer among females (Mosavi-Jarrahi and Mohagheghi, 2006).

Colorectal cancer is the most common problem involving the GI system, with a high mortality rate (Cancer facts figures, 2002). Based on the reports of the World Health Organization (WHO), approximately 87,500 new cases are diagnosed annually (Moghimi et al., 2008). The incidence of the disease is higher in the elderly; more than 90% of cases are older than 55 years of age (Janaut and Kollarova, 2001). Incidence of the disease in younger patients is reported between 0.4% - 35.6% (Janaut and Kollarova, 2001; Turkiewicz et al., 2001). Gastric cancer

remains a major public health issue as the fourth most common cancer and the second leading cause of cancer death worldwide (Crew and Neugut, 2006) and esophageal cancer is one of the least studied cancers worldwide (Enzinger and Mayer, 2003). Iran is one of the known areas with a high incidence of esophageal cancer in Asia (Ghavamzadeh et al., 2001). The most frequent statistical tool to address the relationship among a dichotomous response and other covariates is logistic regression. Logit log linear analysis pertains to a class of log linear models which are used for analyses in which there are one or more categorical dependent variables. Probit regression is an alternative log-linear approach to handling categorical dependent variables. Its assumptions are consistent with having a categorical dependent variable assumed to be a proxy for a true underlying continuous normal (Agresti, 2002). The aim of this study was to compare gastrointestinal cancer with logit and probit models.

Materials and Methods

This study was designed as a cross-sectional study

Research Center of Gastroenterology and Liver Diseases, Shahid Beheshti University, (M.C.), Tehran, Iran *For correspondence: b_moghimi_de@yahoo.com

included all consecutive gastrointestinal cancer patients admitted over one year period in a random selected hospital group located in Tehran metropolitan in 2006. There are 19 hospitals selected randomly from Tehran city and the hospitals that have not any gastrointestinal disease or internal medicine wards and also army hospitals excluded from the study and replaced with another one. Relation between age, sex, type of hospital, length of hospitalization with type of gastrointestinal cancer were analyzed retrospectively using logit and probit. Logistic regression is a technique for making predictions when the dependent variable is a dichotomy, and the independent variables are continuous and/or discrete logit analysis is based on log odds while probit uses the cumulative normal probability distribution and the function used is the inverse of the standard normal cumulative distribution.

For the aim of comparison among models we used Akaike Information Criterion (AIC) proposed as a measure of the goodness of fit of an estimated statistical model (Akaike, 1974).

Results

From all of 2,674 gastrointestinal cancer patients, 40.0% had colorectal cancers, followed by 34.5% for gastric cancer and 17.1% for esophagus cancer. Table 1 shows the results of probit and logit models for all gastrointestinal cancers entered to the analysis. According to the logit model we found a significant gender effect for

colorectal cancer. Men have colorectal cancer more likely than women. Also the relation between age and colorectal cancer was significant in corresponding probit analysis, but other coefficients were negative. In gastric cancer all factors were significant, except the duration of hospitalization. In the probit model the coefficient of age was negative and significant and other factors lead to positive association.

Analysis showed that in esophagus cancer all factors were significant for logit analysis. Also in probit model every coefficient had positive and significant association with esophagus cancer. The results for liver and pancreatic cancer are also given in Table 1.

According to the AIC probit and logit were more similar to each other but in logit model odds ratio estimated and can be used for interpretation of coefficient.

Discussion

Logit and probit models are special cases of general linear models to better treat the case of dichotomous and categorical variables. Although Probit is a variant of Logit modeling based on different data assumptions, results of probit analysis are rarely reported in the original units. Logit is the more commonly used, based on the assumption of equal categories. Probit may be the more appropriate choice when the categories are assumed to reflect an underlying normal distribution of the dependent variable, even if there are just two categories (Agresti, 2002).

Table 1. Relation between Demographic Factors and Cancer Type with Logit and Probit Analyses

Factors	logit				probit			
	AIC	Coefficient	Odds Ratio	Estimate	p-value	AIC	Coefficient	p-value
Colorectal cancer								
intercept	3210.010	1.8638			<0.001	3210.010	1.1419	<0.001
sex		0.2772	1.319		<0.001		0.1722	<0.001
age		-0.0187	0.981		<0.001		-0.0115	<0.001
duration		-0.0219	0.978		<0.001		-0.0126	<0.001
hospital type		-1.2537	0.285		<0.001		-0.7768	<0.001
Gastric cancer								
intercept	3063.350	-1.7660			<0.001	3063.350	-1.0794	<0.001
sex		-0.6337	0.531		<0.001		-0.3842	<0.001
age		0.0129	1.013		<0.001		0.00791	<0.001
duration		0.000371	1.000		0.9522		0.000213	0.9549
hospital type		0.6228	1.864		0.0016		0.3721	<0.001
Esophagus cancer								
intercept	2262.130	-2.9690			<0.001	2262.130	-1.7107	<0.001
sex		0.4032	1.497		<0.001		0.2281	<0.001
age		0.4032	1.006		0.0218		0.00338	0.0138
duration		0.0248	1.025		<0.001		0.0144	<0.001
hospital type		0.7826	2.187		<0.001		0.4261	<0.001
Liver cancer								
intercept	813.319	-4.0315			<0.001	813.319	-2.0287	<0.001
sex		-0.1701	0.844		0.4398		-0.0799	0.4191
age		-0.0124	0.988		0.0790		-0.00554	0.0851
duration		-0.0270	0.973		0.1976		-0.0126	0.1636
hospital type		1.8643	6.451		0.0650		0.7370	0.0397
Pancreatic cancer								
intercept	787.855	-3.2426			0.0031	787.855	-1.6233	<0.001
sex		0.1577	1.171		0.4667		0.0532	0.5894
age		-0.0326	0.968		<0.001		-0.0150	<0.001
duration		0.0171	1.017		0.1254		0.00809	0.1533
hospital type		1.6298	5.103		0.1073		0.6090	0.0850

Approximately 50,000 new cases of cancer occur each year in the Iranian population of 70.4 million. The most common organ system involved with more than 38% of all cancers is the gastrointestinal (GI) gastric, esophagus, and colorectal are the three most common cancers in males; in females, after breast cancer, esophagus, gastric, and colorectal are the major cancers (Sadjadi et al, 2005). Cancer is the third most common cause of death in Iran, accounting for 14% of mortality. Overall, GI cancers account for nearly half (44.4%) of all cancer related deaths in Iran (Naghavi, 2004).

This paper examines relation between gastrointestinal cancer and demographic factors Using logit and probit models. Binary models like yes or no cancer describe with this models respectively. Although probit model is not well-known for researchers in the filed of medical science, there are a few studies using these models to interpret the relation between prognostic factors and a binary response. Costa-Font et al (2005) examined the influence of obesity jointly with other determinants on the prevalence of four chronic conditions, type 2 diabetes, cardiovascular disease, hypertension, and high cholesterol, using probit regression (Costa-Font and Joan, 2005). Shelton Brown et al (2005) investigated the relation between diabetes and employment with probit model (Shelton Brown et al., 2005) and Shahrabani et al (2008) studied the impact of vaccinated in the past perceive influenza to prevent flu infection by probit analysis (Shahrabani et al., 2008). In this study we used probit and logit models and compared this two methods with each other. In reality, this difference was not too important: both models need to have diagnostics done afterwards to check that the assumptions of the model have not been violated and use maximum likelihood. In general, the logit coefficients are larger than the probit coefficients. Because of the probit function is the inverse cumulative distribution function (CDF), or quantile function associated with the standard normal distribution. It has applications in exploratory statistical graphics and specialized regression modeling of binary response variables (Agresti, 2002). In probit model we don't estimate odds ratio and describe models with coefficient directly. Negative and positive coefficients and significance are very important in describe of models. So probit regression can be one of useful statistical method for describe relationship between binary response and explanatory factors.

References

- Agresti (2002). *Categorical Data Analysis*. 2nd edition. John Wiley & Sons, Inc.
- Akaike HA (1974). New look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-23.
- Brown HS 3rd, Pagán JA, Bastida E (2005). The impact of diabetes on employment: genetic IVs in a bivariate probit. *Health Econ*, **14**, 537-44.
- Cancer Control Office of Ministry of Health (2005). Iranian annual cancer registration report: 2003. Tehran: Ministry of Health publication.
- Cancer Facts Figures (2002). New York: The American Cancer Society, Inc. Surveillance Research; 02-250 M-NO
- Costa-Font J, Joan G (2005). Obesity and the incidence of chronic diseases in Spain: A seemingly unrelated probit approach. *Econ Hum Biol*, **3**, 188-214.
- Crew KD, Neugut AI (2006). Epidemiology of gastric cancer. *World J Gastroenterol*, **12**, 354-62.
- Ghavamzadeh A, Moussavi A, Jahani M, Rastgarpanah M, Irvani M (2001). Esophageal cancer in Iran. *Semin Oncol*, **28**, 153-7.
- Moghimi-Dehkordi B, Safaee A, Zali MR (2008). Prognostic factors in 1,138 Iranian colorectal cancer patients. *Int J Colorectal Dis*, **23**, 683-8.
- Mosavi-Jarrahi A, Mohagheghi MA (2006). Epidemiology of esophageal cancer in the high-risk population of Iran. *Asian Pac J Cancer Prev*, **7**, 375-80.
- Naghavi N (2004). Death report from 23 provinces in Iran. 1st edition. Tehran: Ministry of Health.
- Sadjadi A, Nouraie M, Mohagheghi MA, et al (2005). Cancer occurrence in Iran in 2002, an International perspective. *Asian Pacific J Cancer Prev*, **6**, 359-63.
- Shahrabani S, Benzion U, Yom Din G (2008). Factors affecting nurses' decision to get the flu vaccine. *Eur J Health Econ*. Sep 10.
- Turkiewicz D, Miller B, Schache D, Cohen J, Theile D (2001). Young patients with colorectal cancer: how do they fare? *ANZ J Surg*, **71**, 707-10.