ITALIAN JOURNAL OF PUBLIC HEALTH

Bayesian analysis of esophageal cancer mortality in the presence of misclassification

Mohamad Amin Pourhoseingholi[1], Alireza Abadi[2], Soghrat Faghihzadeh[3], Asma Pourhoseingholi[1], Mohsen Vahedi[1], Bijan Moghimi-Dehkordi[1], Azadeh Safaee[1], Mohammad Reza Zali[1]

[1]Research Center for Gastroenterology and Liver diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [2]Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [3]Department of Biostatistics, Tarbiat Modares University, Tehran, Iran

Correspondence to: Mohamad Amin Pourhoseingholi, Research Center for Gastroenterology and Liver diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Email: Aminphg@gmail.com

Abstract

Background: Esophageal cancer (EC) is one of the most common cancers worldwide. Mortality is a familiar projection that addresses the burden of cancers. With regards to cancer mortality, data are important and used to monitor the effects of screening programs, earlier diagnosis and other prognostic factors. But according to the Iranian death registry, about 20% of death statistics are still recorded in misclassified categories. The aim of this study is to estimate EC mortality in the Iranian population, using a Bayesian approach in order to revise this misclassification.

Methods: We analyzed National death Statistics reported by the Iranian Ministry of Health and Medical Education from 1995 to 2004. ECs [ICD-9; C15] were expressed as annual mortality rates/100,000, overall, by sex, by age group and age standardized rate (ASR). The Bayesian approach was used to correct and account for misclassification effects in Poisson count regression, with a beta prior employed to estimate the mortality rate of EC in age and sex groups.

Results: According to the Bayesian analysis, there were between 20 to 30 percent underreported deaths in mortality records related to EC, and the rate of mortality from EC has increased through recent years.

Conclusions: Our findings suggested a substantial undercount of EC mortality in the Iranian population. So policy makers who determine research and treatment priorities based on reported death rates should notice of this underreported data.

Key words: esophageal cancer, mortality, Bayesian analysis, Iran

Introduction

Cancer is the third most common cause of death in Iran [1] and approximately 50,000 new cases of cancer occur annually in the Iranian population [2]. Gastrointestinal (GI) cancers are the most frequently occurring cancer among Iranian males and second only to breast cancer among females [3-5]. In fact, GI cancers account for nearly half (44.4%) of all cancer related deaths in Iran [1].

Esophageal cancer (EC) is one of the most common cancers worldwide [6]. Its survival rates are very low [7] and advanced EC carries an overall poor prognosis with most patients presenting with incurable disease [8]. The incidence and mortality rates show a wide geographical variation with differences between high- and low-risk areas [6, 9]. EC is a relatively rare form of cancer, but some world areas have a higher incidence than others, like China, Iran, India, Japan, and the region around the Caspian Sea [10].

Mortality is a familiar projection in the assessment of the burden of cancers. With regards to cancer mortality, data are important to monitor the effects of screening programs, earlier diagnosis and other prognostic factors [11]. Data on cancer mortality can be used to guide policy makers in order to setup cancer prevention programs. But this aim needs reliable death registry systems which report death statistics annually and accurately. On the other hand, the analysis of death statistics subject to misclassification is a major problem in epidemiological analysis, often leading to biased estimates, and can therefore cause one to underestimate health risks [12]. The World Health

Organization (WHO) has encouraged member states to introduce systems of death registration involving medical certification of the cause of death. Similar to other developing countries, Iranian mortality information is still incomplete [13]. According to the Iranian death registry, between 15% to 20% of death statistics are recorded in misclassified categories such as septicemia, senility without mention of psychosis symptoms, and other ill-defined conditions, etc [14].

In statistical literature, two approaches are recommended when misclassification occurs. The first using a small validation sample [15] and the second being a Bayesian analysis in which subjective prior information on at least some subset of the parameters is used to re-estimate death statistic [16, 17].

The aim of this study is to re-estimate EC mortality rate for the Iranian population, using a Bayesian approach.

Materials and Methods

Data sources

The National Organization for Civil Registration (NOCR) and the Ministry of Health and Medical Education (MOH&ME) have established death registration systems in Iran [13]. National death Statistics Reported by the MOH&ME from 1995 to 2000 (registered death statistics for Iranian population at the Information Technology and Statistic Management Center, MOH&ME) and from 2001 to 2004 (published by MOH&ME) [1, 14, 18] stratified by age group, sex, and cause of death (coded according to the 9th revision of the International Classification of Diseases [ICD-9]) are included in this analysis. ECs [ICD-9; C15] were expressed as annual mortality rates/100,000, overall, by sex, by age group (<15, 15-49 and ≥50 years of age) and age standardized rate (ASR). The population of Iran in 1995-2004 was estimated, using the census from 1996 conducted by the Statistics Centre of Iran and its estimation was made according to population growth rate for years before and after the national census [19].

Statistical implementation

The Bayesian approach considered here was derived from models proposed by Stamey et al to correct and account for misclassification in a Poisson regression [12]. Stamey's technique extended the model recently proposed to overcome the problem of misclassification in cancer data [16, 17] and Pourhoseingholi et al developed this technique to estimate mortality rate of colorectal cancer [20] and liver cancer [21]. We studied Iranian death statistics in a Bayesian Poisson regression using Stamey's approach to re-estimate mortality rate from EC.

Suppose we have two sample groups for death classification; $y_1 = [y_{11}, y_{21}, ..., y_{r1}]$ and $y_2 = [y_{12}, y_{22}, ..., y_{r2}]$ where r is the covariate pattern, y_1 is the exact cause of death and y_2 is the misclassified group in which the cause of death in the first group was incorrectly labeled, and $y_1 \sim Poisson(P_i \mu_{i1})$ and $y_2 \sim Poisson(P_i \mu_{i2})$ in which μ_i is the observed rate of death mortality for the covariate pattern. Let θ be the probability that an observation from group 1 is incorrectly labeled as belonging to group 2. If the actual rate of death for each group (unknown) is supposed to be as λ_i, the relation between actual rate and observed rate can be written in following form; $\mu_{i1} = \lambda_{i1}(1-\theta)$ and $\mu_{i2} = \lambda_{i2} + \lambda_{i1}\theta$.

The joint distribution of the observable mortality data in this case of misclassification is proportional to;

$$\prod_{i=1}^{r} [\lambda_{i1}(1-\theta)^{y_{i1}}][\lambda_{i2} + \lambda_{i1}\theta]^{y_{i2}} \exp\{-P_i[\lambda_{i1}(1-\theta)] - P_i[\lambda_{i2} + \lambda_{i1}\theta]\}.$$

To perform Bayesian inference, we assume that informative beta prior distribution for the misclassified parameter, i.e. $\theta \sim beta(a,b)$. Because θ is an unknown parameter, we employed a latent variable approach according to Paulino et al. [22, 23], Liu et al. [24] and Stamey et al. [12] to simplify the full conditional models and estimate the posterior distribution using a Gibbs sampling algorithm. In this case, we define $U_i | \beta_1, \beta_2, \theta, y_1, y_2 \sim Binomial(y_{i2}, P_i)$ to be the number of counts from the first group incorrectly labeled as being in the misclassified group. So; $P_i = \frac{\lambda_{i1}\theta}{\lambda_{i1}\theta + \lambda_{i2}}$ and finally the posterior appears in the following form; $\theta | \beta_1, \beta_2, U_{i1}, y_1, y_2 \sim beta(\sum_i U_{i1} + a, \sum_i y_{i2} + b)$.

All analysis were carried out using S-plus. The misclassification probability estimate which is proposed in prior distribution was based on Iranian death registrations which introduced between 15% to 20% of misclassified records into total deaths. We assumed a 20% misclassification rate (as a misclassification parameter, reported for death due to cancer without mentioning the exact name of the kind of cancer in the Iranian death registry) with a beta prior to re-estimate the death statistic related to EC from misclassified groups. All analysis were performed by a Macro and developed in S-Plus.

Results

All death records due to EC from 1995 to 2004 are included in this study. The rate of EC mortality classified by sex and age, and generated

from the original database (Frequentist Rate) as well as their Bayesian corresponding projections (Bayesian Rate) appear in Table 1 and Table 2. According to the Bayesian re-estimate, there were between 20 to 30 percent of underreported deaths in mortality records associated to EC (Figure 1). The age standardized mortality rate due to EC also increased dramatically during these years (Figure 1 and Table1) though a slight decreasing trend was observed from 2002 to 2004. Moreover EC mortality was higher for males (Table 1 and Figure 2) and the mortality increased as age increased (Table 2).

Discussion

Our results indicate that between 20%-30% of mortality due to EC remains underreported and suggest a substantial undercount of EC mortality in the Iranian population. This study also revealed that the trend of EC mortality has dramatically increased in the recent decade. In spite of Iranian trends for EC, data in European countries have shown continuously decreasing trends through recent decades and in the US it has reached a plateau [25]. In Asia, China has shown a decreased incidence of and mortality from EC [17] and in Japan, ASR mortality from EC constantly decreased according to the WHO Mortality Database. In addition, EC related mortality in France and Italy has continuously decreased through recent decades [25]. In this study, EC mortality was higher for males and the rate of EC mortality increased as age increased. Males tend to have higher EC mortality compared with females in western countries too [25], and statistics have indicated that EC affects males more than females [26]. In any case, age is a risk factor for EC [27] and there were similar increasing trends for all age groups in other countries for EC deaths [25].

Response misclassification of counted data for death statistics is still a problem in developing counties. In the Iranian Death Registration System, data on the causes of death are collected from various sources and have been assessed to be about 80% complete [13]. In spite of this, it is thought that there are still up to 20% of undefined death records that are categorized as misclassification.

Recently, the Bayesian approach has received much attention in cases of misclassification. Whittemore and Gong used this approach to estimate cervical cancer mortality rates [16], and Sposto et al. developed this likelihood to assess the effect of diagnostic misclassification on non-cancer and cancer mortality dose–response in A-bomb survivors [17]. Stamey et al. used the Bayesian approach in data consisting of the number of deaths due to cancer and non-cancer among residents of Hiroshima and Nagasaki, which were present during the atomic bombings in August of 1945 [12] and other studies have used this technique to estimate mortality rate in colorectal cancer [20] and liver cancer [21] according to Iranian death statistics.

Though the study revealed interesting facts about mortality from EC during the study decade in Iran, the results of this study must be interpreted after the reader considers some of the following issues in this paper. The coverage

Table 1. Age Standardized Rate (per 100,000) for Esophageal Cancer mortality stratified by sex group before and after adjusting for misclassification by Bayesian model.

Year	*Male*		*Female*		*Total*	
	FR	BR	FR	BR	FR	BR
1995	0.73	0.89	0.68	0.84	0.71	0.87
1996	1.70	2.11	1.10	1.33	1.41	1.71
1997	1.61	1.93	1.11	1.33	1.36	1.63
1998	1.61	1.93	1.30	1.56	1.46	1.75
1999	2.36	2.78	1.59	1.87	1.98	2.34
2000	2.55	2.98	1.78	2.08	2.18	2.55
2001	2.39	2.91	1.81	2.21	2.10	2.56
2002	4.28	5.13	3.13	3.76	3.71	4.45
2003	3.92	4.58	3.22	3.77	3.58	4.19
2004	3.93	4.67	2.74	3.29	3.35	3.98
FR: Frequentist Rate, BR: Bayesian Rate						

Table 2. Age specific rate (per 100,000) for Esophageal Cancer mortality stratified by sex group before and after adjusting for misclassification by Bayesian model.

		5-14 Years		*15-49 Years*		*>=50 Years*	
		Male	***Female***	***Male***	***Female***	***Male***	***Female***
1995	FR	0.02	0.00	0.06	0.09	3.48	3.19
	BR	0.03	0.01	0.07	0.11	4.28	3.92
1996	FR	0.04	0.08	0.25	0.14	7.85	5.01
	BR	0.05	0.11	0.30	0.17	9.50	6.06
1997	FR	0.01	0.04	0.20	0.11	7.52	5.20
	BR	0.02	0.06	0.24	0.13	9.02	6.24
1998	FR	0.05	0.05	0.23	0.21	7.42	5.90
	BR	0.07	0.07	0.27	0.25	8.90	7.08
1999	FR	0.04	0.09	0.38	0.28	10.8	7.10
	BR	0.05	0.11	0.45	0.33	12.74	8.38
2000	FR	0.08	0.06	0.35	0.25	11.76	8.20
	BR	0.11	0.08	0.41	0.29	13.76	9.59
2001	FR	0.03	0.03	0.29	0.20	11.19	8.49
	BR	0.04	0.04	0.35	0.24	13.61	10.36
2002	FR	0.02	0.02	0.60	0.47	19.88	14.46
	BR	0.03	0.03	0.72	0.56	23.86	17.35
2003	FR	0.01	0.00	0.50	0.46	18.38	14.96
	BR	0.02	0.01	0.58	0.54	21.50	17.50
2004	FR	0.02	0.02	0.49	0.46	18.43	12.54
	BR	0.03	0.03	0.58	0.55	21.93	14.92
FR: Frequentist Rate, BR: Bayesian Rate							

and quality of death registration has increased in our country during this timeframe and there has been an observed increase in mortality rates associated with EC probably due to better national registration policies. However, the underestimating of mortality for cancers in Iran due to poor registry is still a problem [14].

In conclusion, this study provides a comprehensive projection of the burden of death due to EC, based on the national Iranian death registry, indicating that the trend of EC mortality has dramatically increased in the recent decade. In addition, there is still a substantial undercount of EC mortality according to the Bayesian model. As such, healthcare policy makers, who determine research and treatment priorities on the basis of death rates as an indicator of the burden of disease, should take notice of this underreported data.

Acknowledgment

This study was sponsored by a grant from the Research Center for Gastrointestinal and Liver Disease (RCGLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Figure 1. Trends of Esophageal Cancer mortality during the period 1995-2004 (Age Standardized Rate per 100,000. *FR: Frequentist Rate, BR: Bayesian Rate*).

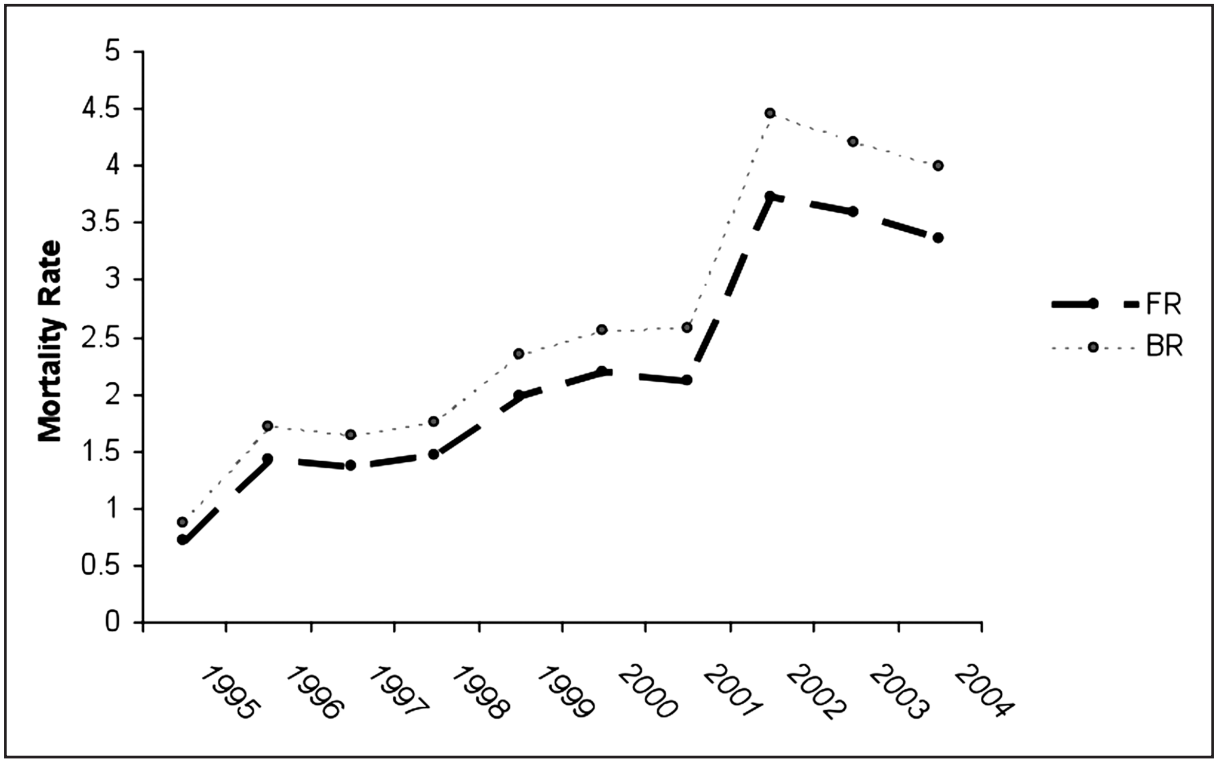
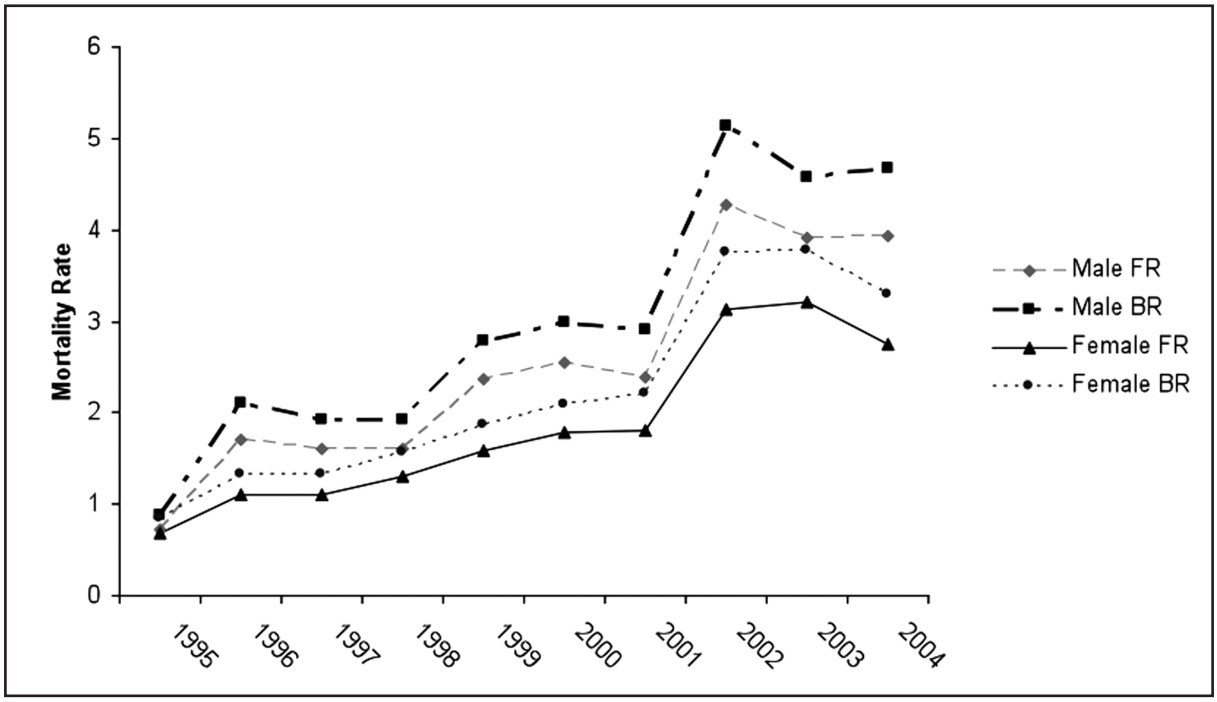


Figure 2. Trends of Esophageal Cancer mortality during the period 1995-2004 by sex groups (Age Standardized Rate per 100,000 *FR: Frequentist Rate, BR: Bayesian Rate*)



References

1) Naghavi M. Death report from 23 provinces in Iran. 1st edition. Tehran, Iran: Ministry of Health and Medical Education, 2003.

2) Mohebbi M, Mahmoodi M, Wolfe R, Nourijelyani K, Mohammad K, Zeraati H, Fotouhi A. Geographical spread of gastrointestinal tract cancer incidence in the Caspian Sea region of Iran: spatial analysis of cancer registry data. BMC Cancer 2008; 4(8):137.

3) Mosavi-Jarrahi A, Mohagheghi MA. Epidemiology of esophageal cancer in the high-risk population of Iran. Asian Pac J Cancer Prev 2006;7:375-80.

4) Pourhoseingholi MA, Vahedi M, Moghimi-Dehkordi B, Pourhoseingholi A, Ghafarnejad F, Maserat E, et al. Burden of hospitalization for gastrointestinal tract cancer patients - Results from a cross-sectional study in Tehran. Asian Pac J Cancer Prev 2009;10:107-10.

5) Pourhoseingholi MA, Vahedi M, Pourhoseingholi A, Moghimi-Dehkordi B, Safaee A, Maserat E, et al. Comparing linear regression and quantile regression to analyze the associated factors of length of hospitalization in patients with gastrointestinal tract cancers. Ital J Public Health 2009;6:32-5.

6) Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74-108.

7) Polednak AP. Trends in survival for both histologic types of esophageal cancer in US surveillance, epidemiology and end results areas. Int J Cancer 2003;105(1):98-100.

8) Qureshi I, Shende M, Luketich JD. Surgical palliation for Barrett's esophagus cancer. Surg Oncol Clin N Am 2009;18:547-60.

9) Ferlay J, Bray F, Pisani P, Parkin DM; GLOBOCAN 2002. Cancer Incidence, Mortality and Prevalence Worldwide. Lyon: IARC Press, 2004.

10) Stewart BW, Kleihues P. World cancer report. Lyon: IARC Press, 2003.

11) Burnet NG, Jefferies SJ, Benson RJ, Hunt DP, Treasure FP. Years of life lost (YLL) from cancer is an important measure of population burden – and should be considered when allocating research funds. Br J Cancer 2005;92:241-5.

12) Stamey JD, Young DM, Seaman Jr JW. A Bayesian approach to adjust for diagnostic misclassification between two mortality causes in Poisson regression. Statist Med 2008;27:2440-52.

13) Khosravi A, Taylor R, Naghavi M, Lopez AD. Mortality in the Islamic Republic of Iran, 1964-2004. Bull World Health Organ 2007;85:607-14.

14) Naghavi M. Death report from 29 provinces in Iran. 1st edition. Tehran, Iran: Ministry of Health and Medical Education, 2004.

15) Lyles RH. A note on estimating crude odds ratios in case–control studies with differentially misclassified exposure. Biometrics 2002;58:1034-6.

16) Whittemore AS, Gong G. Poisson regression with misclassified counts: application to cervical cancer mortality rates. Applied Statistics. 1991;40:81-93.

17) Sposto R, Preston DL, Shimizu Y, Mabuchi K. The effect of diagnostic misclassification on non-cancer and cancer mortality dose–response in A-bomb survivors. Biometrics 1992;48:605-17.

18) Naghavi M. Death report from 18 provinces in Iran. 1st edition. Tehran, Iran : Ministry of Health and Medical Education, 2002.

19) National Statistics Center. Availabel online from: http://amar.sci.org.ir/PlanList.aspx. [Accessd on november 2011].

20.) Pourhoseingholi MA, Faghihzadeh S, Hajizadeh E, Abadi A, Zali MR. Bayesian estimation of colorectal cancer mortality in the presence of misclassification in Iran. Asian Pac J Cancer Prev 2009;10:691-4.

21) Pourhoseingholi MA, Fazeli Z, Zali MR, Alavian SM. Burden of hepatocellular carcinoma in Iran; Bayesian projection and trend analysis. Asian Pac J Cancer Prev 2010;11:859-62.

22) Paulino C, Soares P, Neuhaus J. Binomial regression with misclassification. Biometrics 2003;59(3):670–5.

23) Paulino CD, Silva G, Achcar J. Bayesian analysis of correlated misclassified binary data. Computational Statistics and Data Analysis 2005;49:1120–31.

24) Liu Y, Johnson WO, Gold EB, Lasley BL. Bayesian analysis of risk factors for anovulation. Statistics in Medicine 2004;23(12):1901–19.

25) Qiu D, Kaneko S. Comparison of esophageal cancer mortality in five countries: France, Italy, Japan, UK and USA from the WHO mortality database (1960-2000). Jpn J Clin Oncol 2005;35:564-7.

26) Seitz JF, Dahan L, Jacob J, Artru P, Maingon P, Bedenne L, Triboulet JP. Esophagus cancer. Gastroenterol Clin Biol 2006;30:2S5-2S15.

27) Enzinger PC, Mayer RJ. Esophageal cancer. N Engl J Med 2003;349:2241-52.