Application of survival analysis on the prevalence and risk factors of breast cancer in Namibia

Alexandrina Petrus[1]*, Opeoluwa F. Oyedele[1]
[1]University of Namibia, Faculty of Science, Department of Statistics and Population Studies, Main Campus, Private Bag 13301, Windhoek, Namibia.

ARTICLE INFO

Article history:
Received: 23 July 2019
Accepted: 29 June 2020
Published:
Edited by KC Chinsembu

Keywords:
Breast Cancer
Kaplan Meier
Cox Proportional Hazard
Survival Analysis

ABSTRACT

Cancer is a universal disease that drastically affects people regardless of race, sex, socio-economic status and culture. At an approximated population size of 2.3 million people, Namibia is not spared from this disease, with breast cancer diagnosis becoming more rampant in the country. For this reason, this paper was aimed at examining the prevalence and trends for breast cancer patients regardless of patients' sex, as well as establishing the risk factors associated with breast cancer in Namibia. Secondary data obtained from the Cancer Association of Namibia from 2013 to 2016 was used. Survival analysis techniques (Kaplan-Meier and Cox Proportional Hazard) were used to estimate the survive time of the breast cancer patients. Results revealed that Khomas and Oshana regions had the highest percentage of reported breast cancers cases. It also revealed that the survive time of breast cancer was associated with the patient's age group and ethnicity. Furthermore, the survival time for patients aged 41-50 years and 61-70 years were higher, compared to patients who were less than 31 years of age while the survival rate for patients aged 81-90 years was low. Therefore, there is still a need for a greater focus along the breast cancer care pathway in Namibia, with emphases on improving access to early diagnosis at early age.

1 Introduction

Hejmadi (2010) defined cancer as a disease in which a group of abnormal cells grow uncontrollably by disregarding the normal rules of cell division. Its growth can develop in any parts of the human body and the most common parts are breasts, colon, lungs and prostate. Although the lapse of time between awareness of a problem and seeking medical attention can also affect the impact of diagnosis and treatment of cancer, the fear of this disease is very strong that a person may delay examinations and diagnosis hoping that the signs and symptoms will disappear with time. Unfortunately, if not detected on time and treated on time, cancer can make treatment less likely to succeed and reduce the chances of survival. According to the 2014 report of the World Health Organization (WHO), cancer is a leading cause of disease worldwide. In 2012, an estimated 14.1 million new cancer cases were recorded with lung, breast, colorectal and stomach cancers accounting for more than 40% of all diagnosed cases worldwide (WHO, 2014). Lung cancer was the most common cancer diagnosed among the men (16.7% of all new cases), while breast cancer was the most common cancer diagnosed among the women (25.2% of all new cases).

Cancer is a universal disease that affects people regardless of their race, sex, socio-economic status and culture (Iita, 2009). With an approximated population size of 2.3 million people (Namibia Statistics Agency, 2012),

*Corresponding author: Tel.: +264 61 2903144; Fax: +264 61 2902711. *E-mail address:* alexandrinapetrus@gmail.com (A. Petrus)

http://repository.unam.na/handle/11070/790 ISSN: 2026-7673; eISSN: 2026-8653/c 2021 ISTJN.

reported cancer cases in Namibia are elevating. In a research study conducted by the Namib Time in 2013, breast cancer is fast becoming the third most rampant type of cancer in Namibia. In 2008, there were 253 women and 7 men who had breast cancer, while in 2009, there were 278 women and 6 men who developed breast cancer. In 2010, there were 288 cases in total, while in 2011 there were 291 new cases in women and 5 new cases in men (as cited by Mowa (2016) pp.3). With these numbers continuously increasing, awareness in cancer education and research are of utmost importance. Moreover, according to Carrara (2017), breast cancer in the years 2010 to 2014 was recorded as being the most cancer diagnosed among Namibian women, totaling at 1579 cases. Carrara (2017) further indicated that the annual breast cancer incidence has increased with older age group, escalating at 189.1 per 100 000 in women in the 70-74 years of age. This literally means that the number of breast cancer incidences in Namibia is increasing despite the numerous cancer awareness and prevention programmes setups in the country. Furthermore, the burden of cancer is increasing economically in developing countries like Namibia due to cancer mortality, although it is difficult to accurately determine the exact value of cancer burden mortality in these countries. For these reasons, this paper was aimed at examining the prevalence as well as risk factors associated with breast cancer in Namibia. Pazvakawambwa and Embula (2017) did a study on cancer mortality in Namibia. However, their study used a limited cancer data collected for a study period of 2000 to 2015, while in this present paper, a study period of 2013 to 2016 was considered. In addition, the cancer data for 2013 to 2016 had more variables such as region of residence, marital status and occupation (to mention a few), compared to the 2000 to 2015 data used in Pazvakawambwa and Embula (2017).

2 Materials and Methods

Kleinbaum and Klein (2015) define survival analysis as a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. Survival analysis mainly focuses on the survival function, the hazard function and the cumulative hazard function. A survival function is defined as the probability that an individual will survives longer than time (Kleinbaum and Klein, 2012). Let this function be denoted by $S(t)$. Thus,

$$S(t) = P(T > t), \tag{1}$$

where T is a survival time.

$S(t)$ in equation (1) is a monotonically decreasing function of t with the properties:

$$S(t) = \begin{cases} 1 & \text{if } t = 0; \\ 0 & \text{if } t = \infty. \end{cases} \tag{2}$$

In other words, the probability of surviving at least time zero is 1, while at an infinite time, the probability of surviving is 0. The hazard function is defined as the probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval. It is obtained as

$$\lim_{dt \to 0} \frac{Pr\left(T \in [t, t+dt]\, | T \geq t\right)}{dt} \tag{3}$$

The cumulative hazard function is defined as the total number of failures or deaths over an interval of time and it is obtained as

$$H(t) = \int_0^t h(u)\, du, \tag{4}$$

where $h(u)$ is the hazard risk and u is the accumulated risk.

2.1 Kaplan Meier

Kaplan Meier (KM) method is one of the best statistical methods used to measure the survival probability of patients living for a certain period of time after treatment. It involves computing the probabilities of occurrence of event at a certain point of time and these successive probabilities are then multiplied by any earlier computed probability to determine the final estimate.

The general formula for KM estimation at time t when it comes to the survival function is given by:

$$\hat{S}(t) = \prod_{j|t,\, j\leq t} \left(\frac{n_j - d_j}{n_j}\right), \tag{5}$$

where t is the time point, n_j is the number of patient at risk and d_j is the deaths at time t (Etikan et al., 2017).

2.2 Cox Proportional Hazard Model

The Cox Proportional Hazard (CPH) (survival) model is a regression model used for investigating the association between the survival time of patients and one or more predictor variables (LaMorte, 2016). It simultaneously assesses the effects of several risk factors on survival time (LaMorte, 2016). The CPH (survival) model is given as:

$$h_i(t) = h_0(t)\text{exp}(\beta_j X_j) \tag{6}$$

where $h_i(t)$ is the hazard function for patient i, $h_0(t)$ is the hazard function for a patient in the control group (i.e., baseline hazard), exp(β_j) is the hazard ratio that measures the effect of the j^{th} predictor variable on the survival time and X_j is the j^{th} predictor variable, for $j = 1, 2, \ldots, P$.

2.3 Data

In this paper, the study population was all the recorded cancer patients from 1st of January 2013 to 30th of December 2016, obtained from the Namibia National Cancer Registry of the Cancer Association of Namibia. This registry had 11,757 reported diagnosed cancer cases of Namibian nationality, of which 5384 (45.8%) were males, 6360 (54%) were females, and 13 (0.1%) patients were unknown. Data for 2017 and 2018 were not available during the duration of this study. All patients diagnosed with breast cancer in Namibia were considered in this study. Furthermore, the predictor variables in this paper were the patients' sex, age group (in years), occupation, tobacco use, alcohol consumption, marital status, ethnicity, region, treatment, treatment facility and date of diagnosis, while the event variable was the patient's status (alive or dead) measured by their age at diagnosis and their age at death. Treatment facility and region were identified according to the geographical location of the patient's address. Microsoft Excel (2013) was used to clean the study data, while the R software version 3.5.1 (R Core Team, 2017) was utilized to perform the fitted CPH (survival) modelling.

3 Results

Of the 11,757 reported diagnosed cancer cases in Namibia between January 2013 to December 2016, 1148 (10%) were reported as breast cancer. Out of these cases, 178 (16%) cases were reported in 2013, 384 (33%) cases in 2014, 356 (31%) in 2015, while 230 (20%) cases were reported in 2016. Figure 1 shows the reported breast cancer cases across sex. It can be seen that out of the 178 cases reported in 2013, 8 (4.5%) were males while 170 (95.5%) were females. It can be observed that in 2014, 10 (2.6%) were males while 374 (97.4%) were females. In 2015, out of the 356 breast cancer reported cases 12 (3.4%) were males while 344 (96.6%) were females. On the other hand, in 2016, 5 (2.2%) males and 225 (97.8%) females breast cancer cases were recorded as seen in Figure 1.

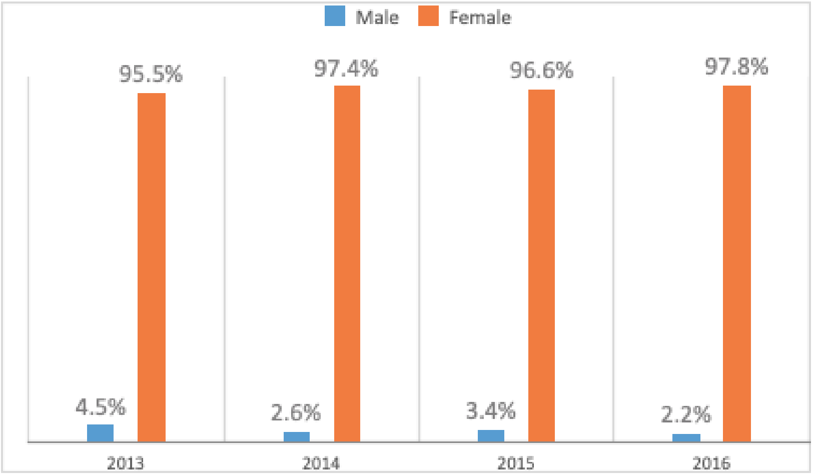

Figure 1: Reported breast cancer case across sex from 2013-2016

Figure 2 shows the percent distribution of reported breast cancer cases across regions in Namibia. From Figure 2, Oshana (45) and Khomas (29) regions had the highest reported breast cancer cases in 2013, followed by Erongo (11), Okavango (11) and Otjozondjupa (11) regions. In 2014 regions that had the highest reported breast cancer cases were Khomas (113) and Oshana (87), followed by Erongo (18), Okavango (18) and Otjozondjupa (18) regions. Khomas (106) and Oshana (49) regions had the highest reported breast cancer cases in 2015 followed by Erongo (26), Otjozondjupa (23) and !Karas (22) regions. However, in 2016 the regions that had the highest reported breast cancer cases were Khomas (38) and Erongo (17), followed by Ohangwena (13) !Karas (11) and Otjozondjupa (9). Overall, Khomas (286) and Oshana (188) regions had the highest number of breast cancer cases reported in 2013-2016, as shown in Figure 2.

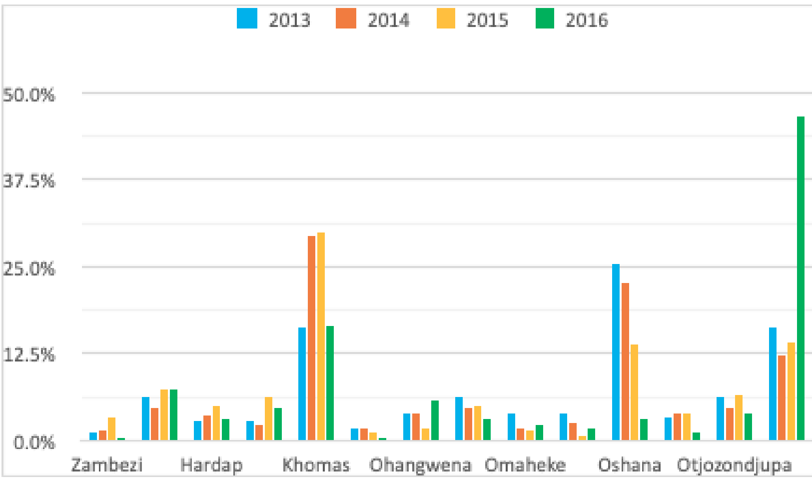

Figure 2: Number of reported breast cancer cases per regions in Namibia from 2013-2016

Looking at the age distributions across the study years, shown in Figure 3, age groups 51-60 (40), 61-70 (35) and 41-50 (33) years had the highest reported breast cancer cases in 2013, followed by age group 31-40 (28) years. The age groups which had the highest breast cancer cases in 2014 were 41-50 (90), 51-60 (85) and 61-70

(60) years, followed by age group 31-40 (56) years. In 2015 age group 51-60 (87) and 41-50 (84) years had the highest reported breast cancer cases, followed by age group 31-40 (55) and 61-70 (47) years. Consequently in 2016 the age group that had the highest breast cancer cases were 41-50 (67) and 61-70 (53) years, followed by 51-60 (41) and 31-40 (35) years. Overall, age groups 41-50 (274) and 51-60 (253) years had the highest numbers of breast cancer cases reported in 2013-2016, as shown in Figure 3.

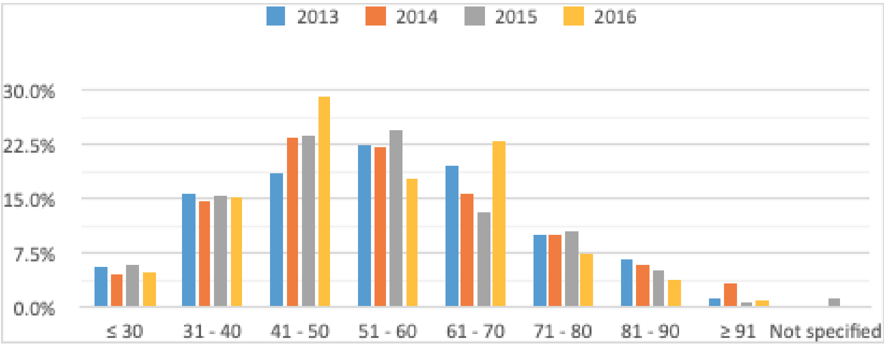

Figure 3: Reported breast cancer cases across age group from 2013-2016

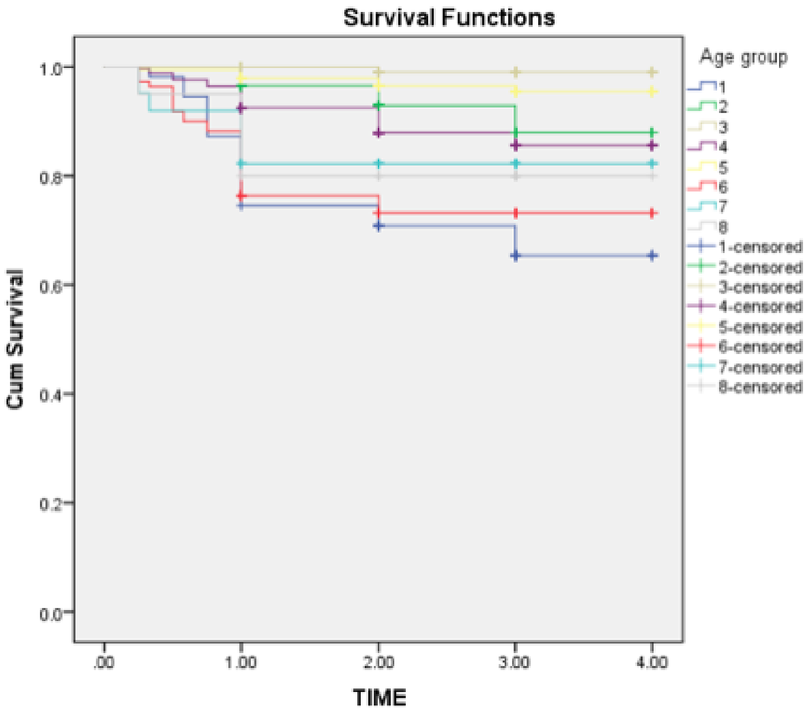

Figure 4: Kaplan Meier curve for age group, where

With the age groups represented by 1 for ≤ 30 years, 2 for 31-40 years, 3 for 41-50 years, 4 for 51-60 years, 5 for 61-70 years, 6 for 71-80 years, 7 for 81-90 years and 8 for ≥ 91 years in Figure 4, it can be observed that the cumulative survival proportion looked to be much higher for the age groups 41-50 years, followed by 61-70 years, and appeared to be much lower for the age groups ≤ 30 and 71-80 years.

Moreover, testing for the equality of survival distribution using the log rank (Mantel-Cox) test, only the patient's age group and ethnicity were significantly associated with their survive time at a 5% significance level, as shown in Table 1.

Table 1: Log rank (Mantel- Cox) test of equality of survival distributions

Variables	Chi-Square Statistics	Degree of Freedom	P-value
Sex	0.035	1	0.852
Age group	106.059	7	0.000***
Occupation	3.320	2	0.190
Tobacco use	0.887	2	0.642
Alcohol consumption	0.808	2	0.668
Marital status	2.352	4	0.671
Ethnicity	36.437	11	0.000***
Region	18.742	14	0.175
Treatment	0.081	1	0.776
Treatment facility	2.997	7	0.885
Date diagnosis	6.346	3	0.096

*** Significant at a 5% level of significance

Furthermore, for the fitted CPH survival model of this paper, the predictor variables that were considered were those that had a probability value (p-value) less than a 5% significance level under the performed (univariate) KM analysis. This elimination scheme was used because not all the predictors were (statistically) relevant in the modelling of survival time. If a predictor variable has a $p-value$ greater than a 5% significance level in the univariate analysis, it is highly unlikely that this predictor variable will significantly contribute to a model which includes other predictor variables.

Table 2 shows the output obtained from the fitted CPH regression model (equation (6)). From this table, it can be observed that there was an association between the patient's survival time of breast cancer and age group 41-50 years ($p-value = 0.0038$) when compared to the age group < 30 years. Likewise, there was an association between the patient's survival time of breast cancer and age group 61-70 years ($p-value = 0.0449$) when compared to the age group < 30 years. Additionally, patients who were aged 41-50 years and 61-70 years were 0.01 and 0.03 times at risk of being diagnosed with breast cancer respectively, compared to patients who were less than 31 years of age. In other words, the survival time for breast cancer patients aged 41-50 years and 61-70 years were higher, compared to patients who were less than 31 years of age.

Table 2: Output from the fitted CPH regression model

Variables	Categories	Hazard ratio (HR)	P-value	95% Confidence Interval for HR	
				Lower	Upper
Age	≤30	Ref			
	31-40	0.00	0.9992	0.00	Inf
	41-50	0.01	0.0038 ***	0.00	0.19
	51-60	0.09	0.1091	0.00	1.73
	61-70	0.03	0.0449 ***	0.00	0.93
	71-80	0.44	0.5344	0.03	5.88
	81-90	3.27	0.4687	0.13	80.64
	>90	1.38	0.8189	0.09	21.31
Ethnicity	White	Ref			
	Baster	0.00	0.9998	0.00	Inf
	San	0.05	0.9999	0.00	Inf
	Caprivian	2.13	0.9999	0.00	Inf
	Damara	5.06	0.2152	0.39	65.63
	Herero	0.00	0.9976	0.00	Inf
	Kavango	0.00	0.9996	0.00	Inf
	Coloured	2.56	0.4529	0.22	29.99
	Nama	4.88	0.1484	0.57	41.97
	Tswana	0.00	0.9999	0.00	Inf
	Ovambo	0.44	0.4325	0.06	3.37

Ref = Reference category, Inf = Infinite, *** Significant at a 5% level of significance

With respect to the patient's ethnicity, although not significant at a 5% level of significance, it can be observed that patients who were Damaras were 5.06 times at risk of being diagnosed with breast cancer compared to patients who were whites as shown in Table 2. Likewise, patients that were Namas were 4.88 times at risk of being diagnosed with breast cancer compared to patients who were whites. Moreover, patients that were in the 81-90 age group were 3.27 times at risk of being diagnosed with breast cancer compared to patients that were less than 31 years old as shown in Table 2. Thus, the older the patient becomes the more likely they were to being diagnosed with breast cancer.

4 Discussion

From this study, it was revealed that the survive time of breast cancer was associated with patient's age group and ethnicity. This finding is in agreement with the study done by Pazvakawambwa and Embula (2017). However their study indicated that region also influenced the cancer survive time. Patients that were Vambos were the most diagnosed with breast cancer followed by Patients that were Whites. Breast cancer patients who were aged 41-50 years and 61-70 years had a high survive time, with HR values of 0.01 and 0.03 respectively, compared to patients who were less than 31 years of age. These findings also concur with the conclusions made under the univariate KM analysis in Figure 4. In addition the hazard ratio of breast cancer was 3.27 times higher for patients aged 81-90 years, compared to patients who were less than 31 years old. This means that the older the patient becomes the more likely they were to experience an event of breast cancer diagnosis. This finding is similar to the observations made in Carrara (2017) and Breyer et al. (2018). Carrara (2017) concluded that the annual breast cancer incidence had increased with older age group, escalating at 189.1 per 100 000 in women aged 70-74 years while Breyer et al. (2018) found out that older age groups were more associated with the development of breast cancer.

5 Conclusion

Breast cancer awareness programs should emphasis on early screening and impart more knowledge to women and men in regions such as Khomas, Oshana, Erongo and Otjozondjupa due to their large number of reported breast cancer diagnosis. Further research should be done on age categories 41-50 and 61-70 years and see other competing risks (such as HIV and cancer related diseases) of breast cancer associated with these age groups. Furthermore, the Ministry of Health and Social Services in Namibia should draft up the country's first national policy on breast cancer diagnosis and management, because at the moment there is none. The breast cancer screening and awareness campaigns should be more prominent in the Khomas, Oshana, Erongo and Otjozondjupa regions since they had the highest percentage of reported breast cancer cases from 2013 to 2016 and this invites for further research on what are the contributing factors for breast cancer prevalence in these regions. Further studies on this topic should consider exploring more predictor variables such as family health history of breast cancer, duration on cancer treatment, types of treatment, adherence to treatment and the use of contraceptives.

Ethical clearance

Ethical clearance was obtained from the Cancer Association of Namibia prior to the commencement of this research study. Records obtained were anonymized to maintain the privacy of those whose records were part of this study.

Acknowledgements

The authors would like to express their appreciation to the Cancer Association of Namibia for given consents prior to the data collection.

References

Breyer, J. Z., Wendland, E. M., Kops, N. L., Caleffi, M., Hammes, L. S., 2018. Assessment of potential risk factors for breast cancer in a population in Southern Brazil. Springer Nature, 169(1), 121–131. https://link.springer.com/article/10.1007/s10549-017-4655-0.

Carrara, H., 2017. Cancer incidences in Namibia 2010–2014. Windhoek: Namibia National Cancer Registry, Accessed date: 15 July 2018.

Etikan, İ., Abubakar, S., Alkassim, R., 2017. The Kaplan Meier Estimate in Survival Analysis. MedCrave - Biometrics & Biostatistics International Journal, 1–5.

Hejmadi, M., 2010. Introduction to Cancer Biology, in Momna, H., How cancer arises. London: Ventus, Accessed date: 7 August 2018.

Iita, S., 2009. Woman's awareness knowledge regarding health promotion on prevention of breast and cervical cancer in Oshakati health district. Masters dissertation. University of Namibia, Namibia.

Kleinbaum, D. G., Klein, M., 2012. Survival analysis, vol.3. Springer-Verlag, New York, USA.

Kleinbaum, G. D., Klein, M., 2015. Survival Analysis - A Self Learning Text, third ed. Springer-Verlag, New York, United States.

LaMorte, W. W., 2016. Cox Proportional Hazards Regression Analysis. School of Public Health. Boston University, USA.

Mowa, K. A., 2016. An exploration of knowledge, attitudes and practices and other breast cancer determinants influencing early detection among women in Windhoek. Masters dissertation. University of Namibia, Namibia.

Namibia Statistics Agency (NSA), National Planning Commission (NPC), 2012. Namibia 2011 Population and Housing Census Preliminary Results. Namibia Statistics Agency, Windhoek, Namibia.

Pazvakawambwa, L., Embula, S.P., 2017. Prevalence, trends and risk factors of breast cancer mortality in Namibia: 2000-2015. International Science and Technology Journal of Namibia, 10, 148–157.

R Core Team, 2017. R: A language and environment for statistical computing. The R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/.

World Health Organisation (WHO), 2014. World cancer report 2012. International Agency for Research on Cancer Report, World Health Organisation, UK.

Prentice, R. L., 1992. Introduction to Cox Regression Models and Life-Tables, in: Kotz, S., Johnson, N.L. (Eds.), Breakthroughs in Statistics. Springer Series in Statistics (Perspectives in Statistics). Springer, New York, USA.