Global Advanced Research Journal of Medicine and Medical Sciences Vol. 1(6) pp. 154-162, July, 2012
Available online http://garj.org/garjmms/index.htm

Full Length Research Paper

Age at cancer diagnosis in Malawi

Humphrey Misiri[1,2] * and Abdi Edriss[3]

[1]Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway.
[2]Department of Community Health, College of Medicine, Blantyre, Malawi.
[3]Bunda College, University of Malawi, Lilongwe, Malawi.

Accepted 28 June, 2012

Cancer which is one of the leading causes of death worldwide is emerging as a serious public health problem in Malawi due to the AIDS pandemic. Research has shown that HIV causes the syndrome of premature aging and accelerates carcinogenesis. The objective of this study was to describe age at cancer diagnosis and to fit the age distribution of childhood and adult cancer diagnosis in Malawi. We therefore fitted the normal, gamma, lognormal and inverse Gaussian probability distributions to the data for the 1996–2005 period from the Malawi National Cancer Registry and selected the model of best fit using the Akaike Information Criterion. Additionally, a finite mixture distribution of lognormals was also fitted to the data. According to the analysis for this study, the median ages at diagnosis are at most 42 years for AIDS-defining cancers and at least 46 years for non-AIDS defining cancers. Furthermore, the ages at childhood and adult cancer diagnosis follow lognormal distributions and the distribution of age at cancer diagnosis (all cancers) is a finite mixture distribution of lognormals with estimated mixing proportions equal to 0.071 and 0.929. The estimated means of the mixture distribution are 5.1 and 45.1 years and the corresponding estimated standard deviations are 1.211 and 2.842 years. This analysis suggests that age at cancer diagnosis in Malawi is relatively low and has a bimodal distribution. Therefore, to achieve maximum impact, cancer prevention and control activities should target the 15-50 year age range.

Keywords: cancer, diagnosis, AIC, finite mixture.

INTRODUCTION

Background

Cancer is one of the leading causes of death worldwide. Although cancer can occur at any age, many cancers are common among persons advanced in age because aging body systems and body cells are more susceptible to mutations that lead to ailments and the malfunctioning of some body parts (Breivik 2007; Gidds 2008). Cancer is a progressive disease and mortality may therefore be prevented if the cancer is detected at an early stage. Early detection is achievable through screening and mass public awareness. For example, regular screening can reduce the risk of breast and cervical cancer mortality. However, in low income countries, cancer screening is rare or non-existent because of financial and

*Corresponding author E-mail: hmisiri@gmail.com, humphrey.misiri@medisin.uio.no; Cell: +265-888-342-864: Fax: +265-1874-700

logistical challenges. In Malawi, cancer screening is rare because of financial challenges and the fact that attention is directed towards the country's huge burden of communicable diseases.

The age at cancer diagnosis is the age at which a person is diagnosed with cancer. The age at diagnosis is "of vital consideration for maximizing the benefits of screening recommendations, prevention initiatives, and treatment strategies" (Karami 2007). Knowledge of the age at cancer diagnosis is helpful to policy makers for initiating measures for the prevention, control and palliation of cancer. For example, if the age at diagnosis of cancer is known, planning for screening is easy because the target age group is known well in advance. Furthermore, formulating cancer prevention strategies and interventions for cancers with known etiology is simplified since the knowledge of the age at diagnosis will help in approximating the best age group to target in order to achieve reasonable impact from cancer interventions. Besides, since more cases of cancer occur at relatively older ages, if the age at diagnosis of such cancers is low, it is easy to imagine the extent of the burden of such cancer because the total burden will be the sum of those cancers occurring prematurely and the cancers occurring at expected older ages.

Research has shown that the age at cancer diagnosis is 10 to 20 times younger among persons with HIV as compared with the general population (Alshafie 1997; Demopoulos 2003; Puoti 2004; Brock 2006 ; Crum-Cianflone 2010). Studies have also shown that HIV infected individuals have an increased risk of some non-AIDS defining cancers (Long 2008; Patel 2008; Shiels 2009) In Malawi, where HIV is spread through heterosexual contact (National Statistical Office (NSO) 2010), the AIDS pandemic has increased the incidence of cancer and the majority of cancers in Blantyre are AIDS-defining such as Kaposi's sarcoma, non-Hodgkin's lymphoma, and eye and cervical cancers.

The aim of this paper, therefore, is to describe the age at cancer diagnosis in Malawi, a low – income country. Additionally, we aim to fit the distribution of age at cancer diagnosis. The distribution of age at cancer diagnosis will be important as a benchmark for fitting serious cancer models in the future.

MATERIALS AND METHODS

Data

The data were extracted from the Malawi National Cancer Registry of Malawi, which is based in Blantyre, Malawi. Founded in 1989, the registry collected national cancer data until 1992. From 1993, the registry has been population-based for Blantyre residents. A person who has lived in Blantyre for at least 6 months is regarded as a resident of Blantyre city. For this paper, we analyse cancer data for Blantyre for the 1996–2005 period.

The response variable for this paper is age at diagnosis. This was computed from the date of birth of the patient and the date of cancer diagnosis. Other variables are year of diagnosis, cancer site and gender. The ages are grouped into discrete single-age categories with corresponding frequencies (see Table 1). Note that cancer site and gender are not shown in the table below.

Fitting the Probability Distribution

A probability distribution for a random variable Y is said to be member of the exponential family if it can be expressed as follows:

$$f\left(y \mid \underset{\sim}{\theta}\right)=\exp\left[\left(\frac{t\left(\underset{\sim}{y}\right)\underset{\sim}{\theta}-b\left(\underset{\sim}{\theta}\right)}{a(\phi)}\right)+c\left(\underset{\sim}{y},\phi\right)\right]$$

where

$\underset{\sim}{\theta}$ *is the canonical parameter,*

ϕ *is the dispersion parameter,*

$b\left(\underset{\sim}{\theta}\right)$ *is the variance function,*

$a(\phi)$ *is the normalizing constant for that distribution, and*

$t\left(\underset{\sim}{y}\right)$ *is a sufficient statistic vector.*

Examples of members of the exponential family are the normal, binomial, Poisson, gamma, lognormal and inverse Gaussian distributions. Different distributions have different numbers of sufficient statistics. The probability distribution of members of the exponential family can be fitted using Poisson log linear models if the random variable can be categorized into a frequency distribution with discrete categories (Lindsey 1992; Lindsey 1995).

Since cancer occurs with different frequencies in different age ranges, cancer cases were divided into two groups: Children were defined as persons aged less than 15 years, and adults as persons aged at least 15 years. For each group, the normal, lognormal, inverse Gaussian and gamma distributions were fitted to the data and compared using the Akaike information criterion (Akaike 1973). The model which provided the best fit was

Table 1. Age at cancer diagnosis in Blantyre, Malawi for 1996-2005

Children		Adults							
age	**count**	**age**	**count**	**age**	**count**	**age**	**count**	**age**	**count**
0	4	15	30	35	289	55	101	75	30
1	49	16	23	36	161	56	67	76	14
2	53	17	22	37	116	57	39	77	12
3	56	18	44	38	164	58	45	78	15
4	60	19	28	39	117	59	24	79	9
5	56	20	84	40	335	60	269	80	33
6	62	21	61	41	91	61	32	81	11
7	50	22	95	42	156	62	45	82	8
8	48	23	84	43	83	63	36	83	9
9	31	24	101	44	78	64	48	84	11
10	38	25	178	45	211	65	101	85	4
11	18	26	119	46	81	66	29	86	2
12	35	27	137	47	79	67	27	87	4
13	26	28	160	48	91	68	35	88	2
14	25	29	136	49	55	69	28	89	4
Total	*611*	30	367	50	331	70	156	90	2
		31	145	51	79	71	15	91	2
		32	273	52	85	72	34	96	1
		33	151	53	68	73	13	98	2
		34	123	54	64	74	14	*Total*	*6428*

selected. The response variable is denoted by *Y*. To fit the probability distribution of age at cancer diagnosis, the data were organized into an age distribution with discrete age categories. Poisson log-linear models were fitted to the frequencies. All sufficient statistics for each candidate probability distribution were treated as covariates in the models. If the coefficient for a covariate was non-significant, this was taken as evidence that the corresponding distribution did not fit the data. The Poisson log-linear models for fitting the normal, gamma, lognormal and inverse Gaussian distributions are shown in Table 2 below.

Two competing distributions were compared by including their sufficient statistics in a model as covariates. If one or all of the sufficient statistics for one distribution were found to be non-significant, the distribution with significant sufficient statistics was chosen as the distribution providing a better fit. Gender was added to the models in interaction with the sufficient statistics. If the coefficient for a sufficient statistic was zero or non-significant for some categories of gender, this was taken as evidence that each gender had different distributions.

About sixty percent of all cancers in Blantyre are AIDS-defining. If we had cancer data from a population-based HIV follow up study conducted in a certain area in Blantyre, that area would have been designated as an AIDS population. We would have proceeded to fit that population's age distribution. The age distribution of the AIDS population would have been compared to the age distribution of the general Blantyre population to see if the two distributions differ. Since we only have cancer data from the whole Blantyre population without the HIV status of each patient, this is not possible. However, since the 30-34 year age range is associated with a high prevalence of HIV in Malawi, we would expect the fitted age distribution to have a mode close to this age range. Since cancer occurs at older ages, if the age distribution of cancer is more skewed to the right, this would imply that cancers are occurring at younger ages. For the current analysis, we assume a constant age distribution for the population of Blantyre for the 1996-2005 period. This assumption is tenable if one closely examines the population forecasts of Blantyre for the same period on the NSO website.

Normal Mixture Model

The most common finite mixture distribution is the finite mixture distribution of univariate normal distributions (Rabe-Hesketh 2007) given by the following equation:

Table 2. Examples of sufficient statistics of selected members of the exponential family and the corresponding Poisson models

Distribution	Sufficient Statistic	Offset	Corresponding Poisson Model
Normal	Y_i, Y_i^2	-	$\log(\mu) = \beta_0 + \beta_1 . y + \beta_2 . y^2$
Lognormal	$\log(Y_i)$, $(\log(Y_i))^2$	-	$\log(\mu) = \beta_0 + \beta_1 . \log(y) + \beta_2 . (\log(y))^2$
Gamma	Y_i, $\log(Y_i)$	-	$\log(\mu) = \beta_0 + \beta_1 . y + \beta_2 . \log(y)$
Inverse Gaussian	Y_i, Y_i^{-1}	-1.5*log(Y_i)	$\log(\mu) = \beta_0 + \beta_1 . y + \beta_2 . \left(\frac{1}{y}\right)$

Source: Lindsey (1995)

$$f\left(y_i, \underset{\sim}{p}, \underset{\sim}{m}, \underset{\sim}{\sigma}\right) = \sum_{j=1}^{k} p_j . g(y_i, \mu_j, \sigma_j)$$

where

$g(y, \mu, \sigma)$ is the normal density with mean μ and standard deviation σ, and

$p_1, p_2, \ldots, p_k$ are mixing proportions whose sum is 1.

A two-component normal mixture model is a special case (where k=2) given by

$$f(y) = p.g(y, \mu, \sigma_1^2) + (1-p).g(y, \mu, \sigma_2^2).$$

Because *a priori* exploratory analyses revealed that the distribution of data in each group was skewed, the data were log-transformed to normality. A finite mixture model was fitted to the log-transformed data using the alldist function in the npmlreg package in R (Einbeck 2009). Two finite mixture models, one with two components, and one with three, were fitted, The components of each fitted mixture distribution were assumed to have a common variance. The difference between the disparities of the two models was computed. A likelihood ratio test of the change in disparity was conducted to determine the number of components to be retained. The level of significance for all statistical tests was 5%.

RESULTS

Descriptive statistics

For the 1996-2005 period there was a total of 7904 cases of cancer of which 862 (10.9%) had missing ages and so were excluded from the analyses. 611 (7.7%) were childhood cancer cases and 7293 (92.3%) were adult cancer cases. The median ages at cancer diagnosis were lower for AIDS-defining cancers than for non AIDS – defining cancers (see Table 3 below). The median ages at diagnosis of all AIDS-defining cancers are at most 40 years for men and at most 42 years for women. For non-AIDS defining cancers, the median ages at diagnosis are at least 46 years for both men and women.

Fitted Distributions for Childhood and Adult Cancers

The distribution of age at cancer diagnosis was skewed and was bimodal (Figure 1). The median ages at childhood and adult cancer diagnosis are 6 and 40 years respectively.

For all the distributions fitted to childhood and adult cancer data, the coefficients for sex were found to be non-significant. The models fitted to age at diagnosis of childhood cancer using the normal, gamma, inverse Gaussian and lognormal distributions had AIC's of 100.1, 97.6, 104.1 and 99.6 respectively. The AIC's of the normal and inverse Gaussian models show that the fits were poor, and the AIC's of the gamma and lognormal distributions show that the fits were good. A comparison of the gamma and lognormal models yielded inconclusive results as the sufficient statistics for both distributions were not significant. Furthermore, the difference between the deviances of the lognormal and gamma models was small (0.552), therefore a Chi-square test was not necessary. Our conclusion is that both the gamma and lognormal distributions provided good fits to the data for age at childhood cancer diagnosis. The data and the lognormal distribution are plotted in the left panel of Figure 2.

Similarly, for all the distributions fitted to age at adult cancer diagnosis, the coefficients for sex were found to be non-significant. The models fitted to age at adult cancer diagnosis using the normal, gamma, inverse Gaussian and lognormal distributions had AIC's of 907.3, 710.2, 674.7 and 661.4 respectively. The data and the lognormal distribution are plotted in the right panel of Figure 2.

Table 3. Median age at cancer diagnosis for selected cancers

Group	Age at diagnosis(years)		
Boys	**Median age**	**Minimum age**	**Maximum age**
AIDS-defining cancers			
Kaposi sarcoma	9	1	14
Eye	2	1	8
NonHodgkin lymphoma	6	0	14
Girls			
AIDS-defining cancers			
Kaposi sarcoma	7.5	1	14
Eye	2	1	12
NonHodgkin lymphoma	6	0	14
Men			
AIDS-defining cancers			
Kaposi sarcoma	35	15	85
Eye	35	15	70
NonHodgkin lymphoma	40	15	83
non_AIDS-defining			
Oesophageal	52	20	91
Others	52	15	98
Women			
AIDS-defining cancers			
Kaposi sarcoma	30	15	77
Eye	35	15	68
NonHodgkin lymphoma	35	15	78
Cervical	42	18	80
non_AIDS-defining			
Oesophageal	50	20	96
Breast	46	15	83
Others	46	15	90
All cancers combined	48	15	96

Figure 2 shows that the fit was good for each model. From the right panel it can be seen that most of the cancers occurred in adults aged between 20 and 50 years. This age range envelopes the 30-34 year age range associated with a high HIV prevalence in Malawi.

Fitted Finite Mixture Model

Two finite mixture models with two and three components were fitted. The disparity of the two-component model was 45806.18 and that of the three-component model was 45806.19. The difference between the disparities of the two models is approximately equal to zero, showing that the two-component mixture model fits the log-transformed data well. The estimated parameters of the two-component normal mixture model are $\hat{\mu} = 1.47, \hat{\sigma} = 0.563$ for the first component and $\hat{\mu} = 3.65, \hat{\sigma} = 0.563$ for the second component. The estimated mixing proportions were $\hat{p} = 0.071$ and $\hat{p}_2 = 0.093$. Thus the estimated finite mixture normal distribution with two components is

$$f\left(y_i, \underset{\sim}{\hat{p}}, \underset{\sim}{\hat{\mu}}, \underset{\sim}{\hat{\sigma}}\right) = \hat{p}_1 . g\left(y_i, \hat{\mu}_1 = 1.47, \hat{\sigma}_1 = 0.563\right) + \hat{p}_2 . g\left(y_i, \hat{\mu}_2 = 3.65, \hat{\sigma}_2 = 0.563\right)$$

where $\hat{p}_1 = 0.087$ and $\hat{p}_2 = 0.913$. On the original scale, the corresponding estimated lognormal mixture distribution is:

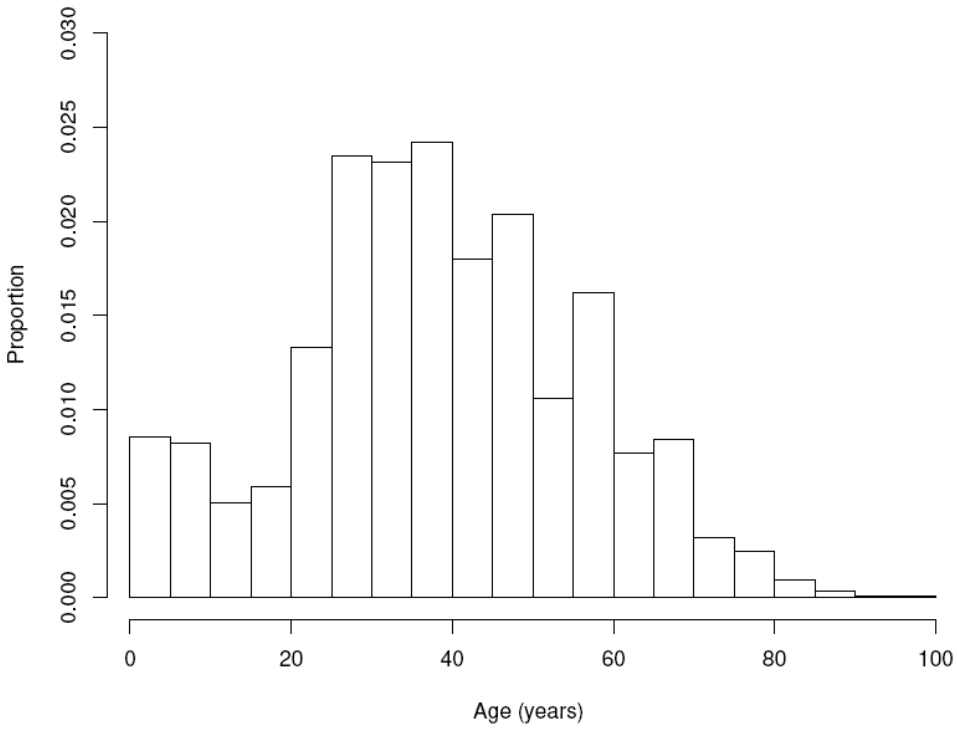

Figure 1. Histogram of all ages at cancer diagnosis

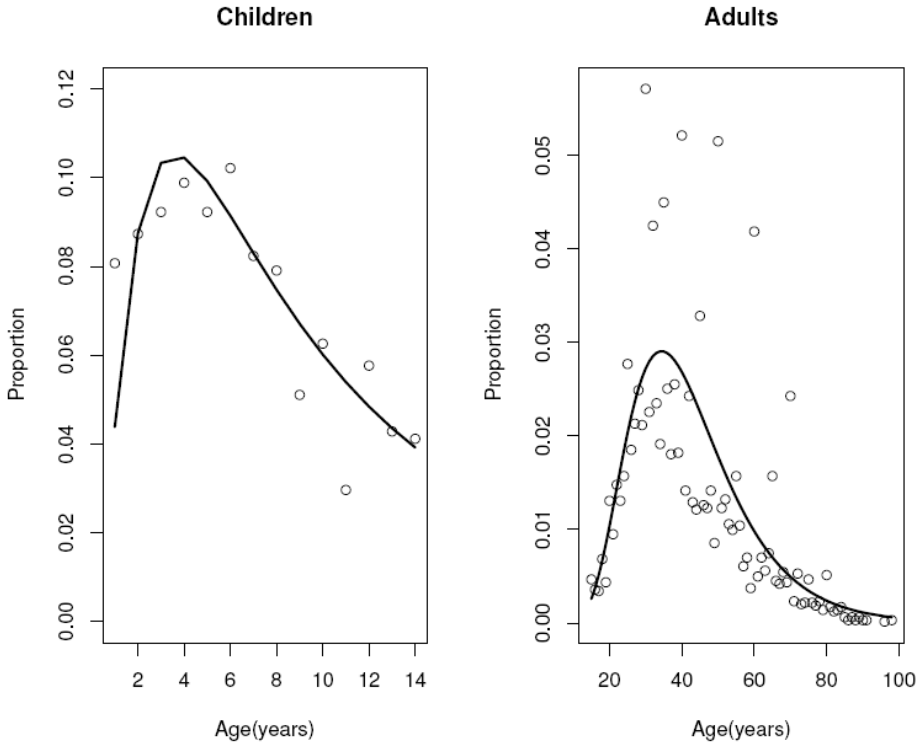

Figure 2. Fitted lognormal distributions

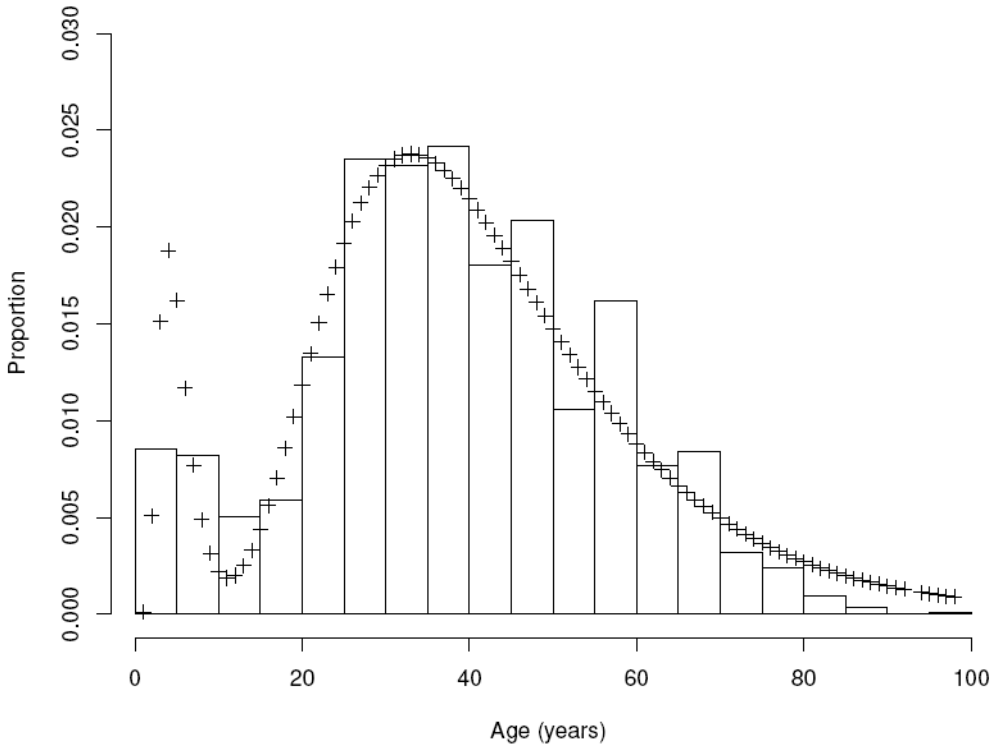

Figure 3. Estimated finite mixture distribution of time to cancer diagnosis

$$f\left(y_i, \underset{\sim}{\hat{p}}, \underset{\sim}{\hat{\mu}}, \underset{\sim}{\hat{\sigma}}\right) = \hat{p}_1 . g\left(y_i, \hat{\mu}_1 = 5.1, \hat{\sigma}_1 = 1.21\right) + \hat{p}_2 . g\left(y_i, \hat{\mu}_2 = 4.51, \hat{\sigma}_2 = 2842\right)$$

where $\hat{p}_1 = 0.071$, $\hat{p}_2 = 0.93$ and

$$g\left(y_i, \mu, \sigma\right) = \left(\frac{1}{\sqrt{2\pi}}\right)\left(\frac{1}{y_i . \sigma}\right) . \exp\left(-\frac{1}{2} . \left(\frac{\log(y) - \mu}{\sigma}\right)^2\right)$$

, in general. The finite lognormal mixture distribution with two components (+++) is super-imposed on the histogram of the data in Figure 3 Above.

DISCUSSION

Our findings show that the time to both childhood and adult cancer diagnosis is skewed and has a lognormal distribution. Furthermore, the age at diagnosis for AIDS-defining cancers is relatively lower than that of non-AIDS defining cancers. The distribution of the age at diagnosis of cancer is a finite mixture of two lognormal components. This is a mixture of the lognormal distributions for the age at diagnosis of paediatric and adult cancers.

Meredith et al. (2010) also fitted age at cancer diagnosis distributions(Meredith 2010) but their analysis is slightly different from ours in that they had cancer cases from a defined AIDS population and so were able to compare the distribution of the age at cancer diagnosis of that AIDS population with that of the general population. They also corrected for differences between the age structures of the AIDS and general populations. We only have cancer data from the Malawi Cancer Registry but we have no information on the HIV status of the cancer patients. Therefore, unlike Meredith et. al.(2010), we don't have another age distribution which would act as a basis for comparison. Our cancer incidence data is from the general Blantyre population. Nevertheless, our assumption of a constant population age structure between 1996 and 2005 is adequate. Besides, our approaches to fitting the age at diagnosis distributions are standard parametric methods and so more accurate for drawing inferences for our purposes than the crude descriptive approaches by Meredith et al. (2010) (Meredith 2010).

The majority of childhood cancers in Malawi are lymphomas, especially Burkitt's lymphoma (Banda 1999), which is very common among children in Malawi. The total number of cases of non-Hodgkin's lymphoma from the childhood cancer case series for 1967–1976, 1985–1993 and 1991–1995 were 187, 472 and 283 respectively, and of these 64.7%, 78.0% and 70.3% were

Burkitt's lymphoma (Parkin 2003). There is an overlap between the modal age range (5-9 years) for childhood cancer as reported by Mukiibi et. al. (1995) (Mukiibi, Banda et al. 1995) and our modal age range (about 2-9 years) inferred from the fitted distribution of age at diagnosis. Besides lymphomas, Kaposi sarcoma is the second most frequent cancer in Malawian children (Banda, Parkin et al. 2001).

The majority of adult cancers in Blantyre are AIDS-defining cancers (Misiri H, Dzamalala,C., Edriss,A.K., Parkin,D.M., Bray,F submitted). The observed median ages of all AIDS-defining cancers were in the neighbourhood of the 30-34 year age group. This coupled with the fact that most adult cancers occurred between the ages of 20 and 50 years support the observation that most of the adult cancers are HIV-related since the 20-50 age range envelopes the 30-34 year age range associated with a high prevalence of HIV in Malawi. Naturally, if human beings live long enough, they are bound to suffer from cancer (Breivik 2007). Furthermore, the life expectancy in Malawi is lower than 55 years. Therefore, only few people live beyond 55 years and a small percentage of these suffers from cancer. This explains why the fitted age distribution had a sharp decline in the proportion of cancer cases from the age of about 50 years up to 100 years. This also explains why the median ages at cancer diagnosis for non-AIDS defining cancers were above 46 years.

CONCLUSION

In Blantyre, age at cancer diagnosis for AIDS-defining cancers is relatively lower and has a bimodal distribution. For the public to optimally benefit from cancer prevention and control activities like screening, mass awareness, treatment and palliation, planning for these activities should take into account the age structure of the population. In particular, cancer prevention and control activities should target the 15-50 year age range.

Limitations

The major limitation of our findings is the quality of the cancer data analysed for this paper. For about 10% of the patients, ages at cancer diagnosis were unknown. Besides, there is a problem of incomplete case ascertainment or morphological verification in Malawi (Banda 1999). This implies that not all recorded cases of cancer are genuine cancer. In addition to this, not all cases of cancer in Blantyre are recorded by the registry due to logistical and financial challenges. For example, currently the registry has no premises of its own. This is a restriction on the amount cancer data the registry can collect because this lack of office space affects data collection and archiving.

ACKNOWLEDGEMENTS

We are very grateful to Dr Charles Dzamalala of the Malawi Cancer Registry for granting us permission to use cancer data. Humphrey Misiri acknowledges support from the Norwegian Programme for Development, Research and Education (NUFU). The study was also partially supported by funds from the Research Council of Norway through contract/grant number: 177401/V50.

REFERENCES

Akaike H (1973). Information theory and an extension of the maximum likelihood principle. Second International Symposium on Inference Theory. -. Budapest,Hungary, Akademikiai Kiado, pp. 267-281.

Alshafie MT, Donaldson B, Oluwole SF (1997). "Human immunodeficiency virus and lung cancer." Br. J. Surg. 84(8): 1068-1071.

Banda LT, Parkin DM, Dzamalala CP, Liomba NG (2001). Cancer incidence in Blantyre, Malawi 1994-1998. Trop. Med. Int. Health. 6(4):296-304.

Banda LTL, Liomba NG (1999). Malawi National Cancer Registry 1991-1995. International Incidence of Childhood Cancer. D. M. Parkin, Kramarova, E., Draper, G.J., Masuyer, E., Michaelis, J., Neglia, J, Qureshi, S., Stiller, C.A. Lyon,France, IARC. 2: 31-34.

Breivik J (2007). "No Solution To Cancer: Have Our Genes Evolved To Turn Against Us?" Science News Retrieved 13/12/2011, 2009, from http://www.sciencedaily.com/releases/2007/04/070416160429.htm.

Brock MV, Hooker CM, Engels EA, Moore RD, Gillison ML, Alberg AJ, Keruly JC, Yang SC, Heitmiller RF, Baylin SB, Herman JG, Brahmer JR (2006). "Delayed diagnosis and elevated mortality in an urban population with HIV and lung cancer: implications for patient care." J. Acquir. Immune Defic. Syndr. 43(1): 47-55.

Crum-Cianflone NF, Hullsiek KH, Marconi VC, Ganesan A, Weintrob A, Barthel RV, Agan BK, Infectious Disease Clinical Research Program HIV Working Group (2010). "Anal cancers among HIV-infected persons: HAART is not slowing rising incidence." AIDS 24(4): 535-543.

Demopoulos BP, Vamvakas E, Ehrlich JE, Demopoulos R (2003). "Non-acquired immunodeficiency syndrome-defining malignancies in patients infected with human immunodeficiency virus." Arch. Pathol. Lab. Med. 127(5): 589-592.

Einbeck J, Darnell R, John Hinde (2009). "npmlreg: Nonparametric maximum likelihood estimation for random effect models."

Gidds WW (2008). "Untangling the Roots of Cancer." Scientific American 18: 30-39.

Karami S, Young HA, Henson DE (2007). "Earlier age at diagnosis: Another dimension in cancer disparity?" Cancer Detection and Prevention 31: 29-34.

Lindsey JK (1995). Modelling Frequency and Count Data. Oxford, Oxford University Press.

Lindsey JK, Mersch G (1992). "Fitting and comparing probability distributions with loglinear models." Computational Statistics and Data Analysis 13: 373-384.

Long JL, Engels EA, Moore RD, Gebo KA (2008). "Incidence and outcomes of malignancy in the HAART era in an urban cohort of HIV-infected individuals." AIDS 22(4): 489-496.

Meredith SS, Pfeiffer RM, Engles EE (2010). "Age at cancer diagnosis among persons with AIDS in the United States." Annals of Inter. Med. 153(7): 452-460.

Misiri H, Dzamalala C, Edriss AK, Parkin DM, Bray F. ”Cancer incidence in Malawi: time trends in Blantyre 1996-2005 and predictions up to 2015.” Submitted.

Mukiibi JM, Banda L, Liomba NG, Sungani FC, Parkin DM (1995). Spectrum of childhood cancers in Malawi 1985-1993. East Afr. Med. J. 72(1):25-29.

National Statistical Office (NSO), ORC Macro (2010). Malawi Demographic and Health Survey 2010. Zomba, NSO and ORC Macro.

Parkin DM, Ferlay J, Hamdi-Chérif M, Sitas F, Thomas JO, Wabinga H, Whelan SJ (2003). Cancer in Africa: Epidemiology and prevention. Lyon,France.

Patel P, Hanson DL, Sullivan PS, Novak RM, Moorman AC, Tong TC, Holmberg SD, Brooks JT (2008). "Incidence of types of cancer among HIV-infected persons compared with the general population in the United States, 1992-2003." Ann. Intern. Med. 148(10): 728-736.

Puoti M, Bruno R, Soriano V, Donato F, Gaeta GB, Quinzan GP, Precone D, Gelatti U, Asensi V, Vaccher E, HIV HCC Cooperative Italian-Spanish Group (2004). "Hepatocellular carcinoma in HIV-infected patients: epidemiological features, clinical presentation and outcome." AIDS 18: 2285-2293.

Rabe-Hesketh S, Everitt BS (2007). A Handbook of Statistical Analyses Using Stata. New York, Chapman and Hall/CRC.

Shiels MS, Cole SR, Kirk GD, Poole C (2009). "A meta-analysis of the incidence of non-AIDS cancers in HIV-infected individuals." J. Acquir. Immune Defic. Syndr. 52(5): 611-622.