Modelling count data with an excess of zero values applied to childhood bone tumour incidence in Iraq

Muzahem Mohammed Al-Hashimi,[1] Hasmek Antranik Warttan[2]

[1]Department of Statistics and Informatics, College of Computer Science & Mathematics, University of Mosul; [2]Department of Business Management Techniques, Administrative Technical College, Northern Technical University, Mosul, Iraq

Abstract

Bone tumours are rarely found in children and adolescents (0-19 years old), but there are reports from some provinces in Iraq indicating possible increases in the incidence of childhood bone cancer. Since counts are very low and often zero, or near zero, we fitted zero-inflated Poisson, zero-inflated negative binomial, Poisson hurdle, and negative binomial hurdle regression models to investigate these changes. We used data covering the 2000-2015 period taking age, gender and province into account with the aim of identifying potential health disparities. The results indicate that the zero-inflated Poisson is the most appropriate approach. We also found that, the incidence rate ratio of bone tumours for age groups of 5-9, 10-14 and 15-19 years were 134%, 490% and 723% higher, respectively, compared to the 0-4 year olds. The incidence rate was higher by 49% higher in males compared to females. Compared to 2000-2004, the rate was higher during 2005-2009 and 2010-2015 by 23% and 50%, respectively. In addition, the provinces Al-Muthana and Al-Diwaniyah in the South were found to have a higher incidence rate than other provinces. Join point analysis showed that the age-adjusted incidence rate had a significant, increasing trend, with an average percentage change of 3.1% during 2000-2015. The study suggests that further research into childhood tumours, bone tumours in particular, is needed. Reference to the effect of environmental factors in this group of medical disorders would be of special interest.

Introduction

Bone tumours are a varied group of malignant and benign tumours. Worldwide, osteosarcoma is the most common primary malignant bone tumour in children and adolescents (Meyers and Gorlick, 1997; Damron *et al*., 2007; PosthumaDeBoer *et al*., 2010), with a peak incidence at age 15-19 years old (Pruksakorn *et al*., 2016), accounting for 30-80% of all primary sarcomas (Evola *et al*., 2017). Ewing's sarcoma is the second most common primary malignant bone tumour in children (Grimer *et al*., 2010), with a peak incidence in the 0-9 years olds (Arora *et al*., 2012). Globally, bone tumours in childhood and adolescence are more frequent in males than females (Stiller *et al*., 2006; Mirabello *et al*., 2009; Mohammed *et al*., 2010; Özkan *et al*., 2015; Ghert *et al*., 2019). The incidence of this bone tumour also peaks at 15-19 years (Katchy *et al*., 2005; Mirabello *et al*., 2009; Rhutso *et al*., 2013; Wilhelm *et al*., 2014; Ramdass *et al*., 2015; Pruksakorn *et al*., 2016).

In childhood, malignant tumours are rare events constituting 4.7% of all cancers in children aged 0-14 years, and 7.8% in those aged 15-19 years (Steliarova-Foucher *et al*., 2017). The main diagnostic groups are osteosarcomas and Ewing's sarcoma with related tumours of the bone (Kaatsch *et al*., 2016). However, when they occur, they bring with them many social, psychological and ethical issues, both in terms of the effect of the illness on those involved, the patient and the family, and matters related to the treatment and its consequences. This is in addition to our current lack of knowledge regarding the nature of these effects, their severity and long-term effects, particularly in low- and middle-income countries with limited access to the necessary health facilities, as exemplified in the case of Iraq.

In Iraq, bone tumours comprise 1.7% of all cancer diagnoses during the period from 2000-2015 in Iraq with exception of three Kurdish provinces (Erbil, Duhok and Al-Sulaymaniyah), for which data are currently unavailable. For all ages, neoplasm is the most common bone tumour (morphology code 8000 - unclassified tumour) (31.1%), while osteosarcoma is the most common malignant bone tumour, comprising 21.5% of these, followed by Ewing sarcoma (16.0%), and chondrosarcoma (4.5%). Less common,

comprising 7.5% of all bone tumours, are chondroblast osteosarcoma, fibroblastic osteosarcoma, fibrosarcoma, giant cell tumour, multiple myeloma, synovial sarcoma, and osteoblastoma. Among the benign bone tumours, the most common histological type is the soft tissue tumour (2.8%), which are most frequently diagnosed in those aged 15-19 years and comprise 15.7% of all benign tumours. In comparison, only 1.9% of the benign tumours are diagnosed in those aged 75-79 years (Iraqi Cancer Board, 2018). The age adjusted incidence rate of bone tumours per 100,000 population for all ages per year, from 2000 to 2015, is 1.016 in females and 1.447 in males, while it is 1.221 in both males and females standardized to the world population.

Count data are in common use across a variety of disciplines, including medicine, ecology, public health and epidemiology. However, when applied in the study of childhood cancers, their rarity introduces the problem of excess number of zero values. Zero outcomes can be categorized as either random zeros or structural zeros (Pittman et al., 2018). In practice, however, count data are often over-dispersed or zero-inflated (ZI). Standard regression models are inappropriate for analysing count data which do not fit a normal distribution (Du et al., 2012).The most commonly applied statistical models, Poisson log-linear models and logistic regression, frequently fail to adequately account for over-dispersion in observed data (Martin and Hall, 2017), where the variance tends to be considerably greater than would be expected for the Poisson model (Pittman et al., 2018).

The problem of over-dispersion with zero inflation came to the fore when we looked at the occurrence of bone tumours in the 0-19 years age group across fifteen provinces in Iraq. This was done to assess changes in the incidence and prevalence of this kind of disease over time with a view to identifying potential health disparities. The bone tumour incidence rate ratio (IRR), together with their 95% confidence interval (CI), was estimated. However, inspection of the data revealed over-dispersion, as well as a high proportion of zero counts.

Materials and methods

Study area

With exception of three Kurdish provinces (Erbil, Duhok and Al-Sulaymaniyah), the work presented here includes all of Iraq, a country in the south-western Asia comprising a population of over 40 million (World Population Review) in an area of 437,072 km^2 (Ibp, 2012). It is divided into 18 provinces, three of which are situated in the Kurdish autonomous region.

Data

For this study, the International Classification of Diseases for Oncology, third edition (ICD-O-3) was used to define the bone tumours. This is a dual classification containing a code for the site of the neoplasm and a histological one for each neoplasm. In general, it uses the same 3- or 4-character codes as used in ICD-10 for malignant neoplasms. The topography codes C40.0-C41.9 which signify bones, joints, and articular cartilage for the period of 2000-2015 was used and the patients included people aged between 0-19 years. The data for the study were provided by the Iraqi Cancer Board of the Iraqi Ministry of Health, and covers all provinces in the country, with the exception of the three provinces in the Kurdish autonomous region mentioned above. Population data for the same period come from the Iraqi Central Statistics Organization with incidence rates age-adjusted according to the standard method of the World Health organization (WHO). The incidence of bone tumours was considered as a dependent variable related to the explanatory variables described in Table 1.

Models for data counting

The count data were estimated by four models: zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), Poisson hurdle (PH), and negative binomial hurdle (NBH) regression models with a log link function. We applied these models, all with explanatory variables, determining the performance of these four

Table 1. Explanation of variables used in the analysis of bone cancer incidence data.

Variable	Level	Reference	Variable	Level	Reference
Age group	0-4 5-9 10-14 15-19	0-4	Gender	Female Male	Female
Period	2000-2004 2005-2009 2010-2015	2000-2004			
Province	Al-Anbar Babil Baghdad Al-Basrah Diyala Al-Diwaniyah Kerbala	Al-Basrah	Province (continued)	Kirkuk Maysan Muthanna Al-Najaf Ninewa Salah Al-Din Thi-Qar Wassit	

The logarithm for the population of each age group in each province was used as an offset. A log-linear relationship between the mean and the factors: age group, gender, time-period and province specified by the log link function. The reference factors were Al-Basrah Province (having the highest incidence of cancer of all provinces in Iraq and used as a benchmark), female, the 0-4 age group and the period 2000-2004.

models and investigating changes in the prevalence of bone tumours in age groups from birth until 19 years of age over time by province to identify potential health disparities. Goodness-of-fit was compared among the four models using Akaike's information criterion (AIC) and Sawa's Bayesian information criterion (BIC). Having applied the four methods to evaluate the relationship between predictors and bone tumours, the IRRs and their P-values were used to assess the efficacy of each approach. All statistical analyses were conducted in SAS version 9.4 (SAS Institute, Cary, NC, USA) and the Join point Regression Program, version 4.8.0.1. All maps were created using ArcGIS 10.6 (ESRI, Redlands, CA, USA).

In recent years, ZI and hurdle models of count data with extra zeros have gained popularity in the research literature. These two types of model can be viewed as finite mixture models involving a degenerate distribution whose mass condenses at zero (Cameron and Trivedi, 2013). Data with excess zeros arise in many disciplines, such as epidemiology, economics, insurance, agriculture, general industry and traffic safety. An excess of zeros refers to the event where the occurrence of zeros in the observed data exceeds the number of zeros predicted by standard models for count data (Diallo and Dupuy, 2019).

Hurdle count models are popular for modelling count data when over-dispersion and excess zeros exist. The hurdle models, like the ZI models, use a two-part model. The first part is a binary model, which includes all zeros, and the second part is a truncated-at-zero count model, which includes only the positive counts (Mullahy, 1986). The hurdle model can be specified as a PH model or NBH model.

The Poisson regression model is a popular analytical tool for count data. However, this model assumes equi-dispersion. Unfortunately, this attribute is frequently missing in real datasets (Mouatassim and Ezzahid, 2012). The negative binomial (NB) distribution can be used for count data that are over-dispersed, *i.e.*, when the variance exceeds the mean. However, the NB model may not be an appropriate choice to handle over-dispersion, when the data have long tails or extreme skewness with an excess of zeros due to the rare appearance of non-zero events (Klakattawi *et al.*, 2018). Both Poisson and NB models are inadequate in the case of excess of zeros (Rose *et al.*, 2006). Lambert (1992) developed the ZIP model as an alternative to model counts data with an excess of zero values.

The zero-inflated Poisson model

ZIP regression, an extension of the Poisson model, represents an useful approach to data counting that can handle over-dispersion and excess zeros (Lambert, 1992). ZIP regression actually combines two models: logistic regression and Poisson regression, where the former predicts the structured zero-only count, and the latter predicts the extent of the count outcome, which may include zero counts (Xie *et al.*, 2013).

In the ZIP regression according to Lambert (1992), the counts Y equals zero ($Y = 0$) and follow a Poisson distribution of mean θ_i with probability ($1 - p_i$) where $i = 0,1,\ldots,n$, ZIP can be defined as follows

$$P(Y) = \begin{cases} p_i + (1 - p_i) \exp(-\theta_i), & Y = 0 \\ (1 - p_i)\dfrac{\exp(-\theta_i)\,\theta_i^{\,Y}}{Y!}, & Y = 1,2,\ldots;\ 0 \le p_i \le 1 \end{cases} \quad (1)$$

The zero-inflated negative binomial model

ZINB, an extension of the NB model, also combines two types of regression models: logistic and NB where the former predicts the structured zero-only count, and the latter predicts both non-negative counts and the random zeros defined by the NB distribution. The ZINB distribution is given by:

$$P(Y) = \begin{cases} p + (1 - p)\left(\dfrac{1}{1+k\theta}\right)^{\frac{1}{k}}, Y = 0 \\ (1 - p)\dfrac{\Gamma\left(Y+\frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(Y+1)}\left(\dfrac{1}{1+k\theta}\right)^{\frac{1}{k}}\left(\dfrac{k\theta}{1+k\theta}\right)^{Y}, Y = 1,2,\ldots \end{cases} \quad (2)$$

The Poisson hurdle model

The positive integer numbers based on truncated count data are called the PH model when they are modelled using the Poisson distribution (Yesilova *et al.*, 2010). The PH model can be expressed as follows:

$$P(Y) = \begin{cases} p_i & ;\ Y = 0 \\ (1 - p_i)\dfrac{\exp(-\theta_i)\,\theta_i^{\,Y}}{(1 - \exp(-\theta_i))\Gamma(Y+1)} & ;\ Y = 1,2,\ldots;\ 0 \le p \le 1 \end{cases} \quad (3)$$

The negative binomial hurdle model

The positive integer numbers based on truncated count data are called the NPH model when modelled using the negative binomial distribution. The NPH model can be expressed as follows

$$P(Y) = \begin{cases} p\ ;\ Y = 0 \\ (1 - p)\dfrac{\Gamma\left(Y+\frac{1}{k}\right)^{p}}{\left[1 - \Gamma\left(\frac{1}{1+k\theta}\right)^{\frac{1}{k}}\right]\Gamma(Y+1)\Gamma\left(\frac{1}{k}\right)}\left(\dfrac{1}{1+k\theta}\right)^{\frac{1}{k}}\left(\dfrac{k\theta}{1+k\theta}\right)^{Y}\ ;\ Y = 1,2,\ldots;\ 0 \le p \le 1 \end{cases} \quad (4)$$

Results

The count data contains 73 zeroes, from 360 observations, which amounts to 20.3% of the data. Zero counts represent no incidence of bone tumours in many areas in Iraq during the three study periods. The total number of registered cases for ages between 0-19 years during the period 2000 to 2015 was 1474 (comprising 32.7% of all bone tumours diagnosed). The sub-divided data are shown in Table 2. For the age group 0-14 years, the male: female ratio was 1.41, and for the 15-19 years age group 1.59. The most common tumours for ages 0-14 and 0-19 years were in bone without further specification, *i.e.* not otherwise specific (NOS) (C41.9) followed by long bones of the lower limbs and associated joints (C40.2). The most common histological picture for this group was that of osteosarcoma NOS (coded 9180/3), followed by Ewing's sarcoma (coded 9260/3). The subdivisions of occurrence in the various age groups are shown in Table 2.

The provincial numbers and incidence of bone cancer in the age groups and the division of the population between these same groups for the years 2000-2015 are summarized in Table 3. The age-adjusted (AAIR) incidence rate per 100,000 person-years for the part of the population aged 0-19 years was 0.76 and that of the 0-14 age group 0.54. The rate was 0.605 in females and 0.905 in

males in the 0-19 age group during the study period.

The geographical distribution of bone cancer incidence for the three periods 2000-2004, 2005-2009 and 2010-2015 at the province level is shown in Figure 1. Bone cancer incidence varied by years. The highest incidence rates were found in Al-Al-Diwaniyah, Al-Muthanna, Al-Basrah, Al-Najaf and Kirkuk, while the lowest appeared in Diyala.

Temporal trends

The Join point analysis indicated that the AAIR had a significant increasing trend during 2000-2015, with an average percentage change of 3.1% with a 95% CI of 0.9%-5.3% (Figure 2A). Furthermore, the temporal trend can be divided into five periods of variable AAIR changes (Figure 2B), i.e. from 2000 to 2002 when it was increasing with an average percentage change (APC) of 6.09% followed by a period of rapid growth (2002-2005) with an APC of 10.17% and a third period from 2005 to 2008 when it was decreasing with an APC of 7.34%. The fourth period was from 2008 to 2011, during which there was a 14.8% APC growth finally followed by a slow decrease between 2011 and 2015 with an APC of 4.5%.

Model fitting and selection

Figure 3 displays the histogram for the number of cancer incidence cases during the study period. The distribution of the data is highly skewed to the right, with a spike at zero and a long right tail. The response variable, i.e. the cancer incidence cases, ranged from 0 to 54. The observed mean and variance of the response variable were 4.094 and 33.178, respectively. The variance was about 8 times the mean, so the variance/mean ratio was larger than unity, indicating an over-dispersion the bone tumour data, implying that the Poisson regression model did not adequately fit the data, and that the ZI and hurdle models would be more appropriate.

The study applied the four models by using four factors in order to assess which one performed the best. Table 4 shows the AIC, BIC, and Log-likelihood for the four models. For AIC and BIC models, a lower value was comparatively better. The table

shows that the ZIP model fitted the childhood and adolescent bone tumour data better than other models, because it had the smallest AIC and BIC, while the NBH had the largest.

The four models do well at the numbers of bone cancer cases (Figure 4). Compared to the youngest age group (0-4 years) the IRR of bone tumours for age groups 5-9, 10-14 and 15-19 years were higher by 134.4% (IRR=2.344, 95% CI: 1.583, 2.9441), 490.1% (IRR=5.9008, 95% CI: 4.0213, 6.9092) and 722.9% (IRR=8.2292, 95% CI: 6.4094, 9.7651), respectively. With respect to gender, the incidence rate was higher for the males by 49.1% (IRR=1.4912, 95% CI: 1.1880, 3.0073). Compared to the period 2000-2004, the overall rate was also higher during 2005-2009 and 2010-2015 by 22.9% (IRR=1.2286, 95% CI: 0.9603, 1.4244) and 49.67% (IRR=1.4967, 95% CI: 1.169, 1.8656). respectively (Table 5). Regarding the variation of incidence between provinces, Figure 5

Table 2. Distribution of childhood bone cancer cases by age, gender and bone sarcomas in Iraq in the period 2000-2015.

Variable	Description	No. of cases n (%)
Age	0-4	112 (7.6)
	5-9	229 (15.5)
	10-14	508 (34.5)
	15-19	625 (42.4)
Gender	Male	903 (61.26)
	Female	57 (38.73)
Osteosarcoma	0-4	7 (1.3)
	5-9	47 (9.0)
	10-14	197 (37.5)
	15-19	274 (52.2)
Ewing's sarcoma	0-4	31 (6.7)
	5-9	101 (22.0)
	10-14	180 (39.1)
	15-19	148 (32.2)
Number	Total	1474 (32.7)

Table 3. Childhood bone cancer incidence and population divisions in Iraq during the period 2000-2015.

Province	Incidence (2000-2015)				Population (2000-2015)			
	0-4	5-9	10-14	15-19	0-4	5-9	10-14	15-19
Baghdad	47	65	141	163	17,064,259	15,033,024	13,385,137	11,937,447
Basrah	12	22	51	62	4,645,219	4,077,465	3,615,963	3,208,966
Ninewa	8	25	54	64	7,254,084	6,329,052	5,513,404	4,845,594
Maysan	1	7	21	14	2,054,977	1,793,834	1,581,007	1,392,247
Al-Diwaniyah	4	17	26	30	2,531,285	2,198,408	1,926,473	1,684,216
Diyala	5	11	12	30	4,067,623	3,518,565	3,069,345	2,667,725
Al-Anbar	4	11	24	28	3,799,449	3,299,434	2,891,017	2,527,090
Babil	4	11	27	42	4,260,348	3,693,008	3,229,235	2,815,321
Kerbala	2	6	14	30	2,214,148	1,932,716	1,703,346	1,499,958
Kirkuk	4	11	25	25	2,283,725	1,997,130	1,760,512	1,556,522
Wassit	0	8	24	24	2,706,194	2,352,954	2,062,759	1,828,557
Thi-Qar	8	7	25	39	4,180,932	3,639,864	3,192,513	2,804,758
Al-Muthanna	5	4	7	21	1,595,214	1,381,250	1,206,255	1,049,907
Salah Al-Din	6	16	16	26	3,078,987	2,667,964	2,331,921	2,031,903
Al-Najaf	2	8	41	27	2,676,443	2,339,872	2,065,793	1,823,068

presents maps of IRRs for each of the methods used. According to our findings, there was no difference between ZIP and ZINB models, indicating that both ZIP and ZINB have the same prediction and they are successful with the AIC and BIC results, and ZIP was more suitable than the ZINB model to handle the over-dispersion of the bone tumours data. While PH and NBH models suggest there were some provinces which disagree with the results obtained from ZIP and ZINB models, because these two models unsuccessfully with the AIC and BIC results. According to ZIP results, Al-Muthanna and Al-Diwaniyah had a higher incidence rate than the other provinces (Figure 5A); compared to Al-Basrah Province, their incidence was

Table 4. Model selection criteria for childhood bone tumour data of the fitted models for childhood bone tumour data.

Model	AIC (smaller is better)	BIC (smaller is better)	Log-likelihood
ZIP	1381	1467	1337
ZINB	1383	1472	1337
PH	1390	1553	1306
NBH	1394	1557	1310

AIC, Akaike information criterion; BIC, Bayesian information criteria; ZIP, zero-inflated Poisson; ZINB, zero-inflated negative binomial; PH, Poisson hurdle; NBH, negative binomial hurdle.
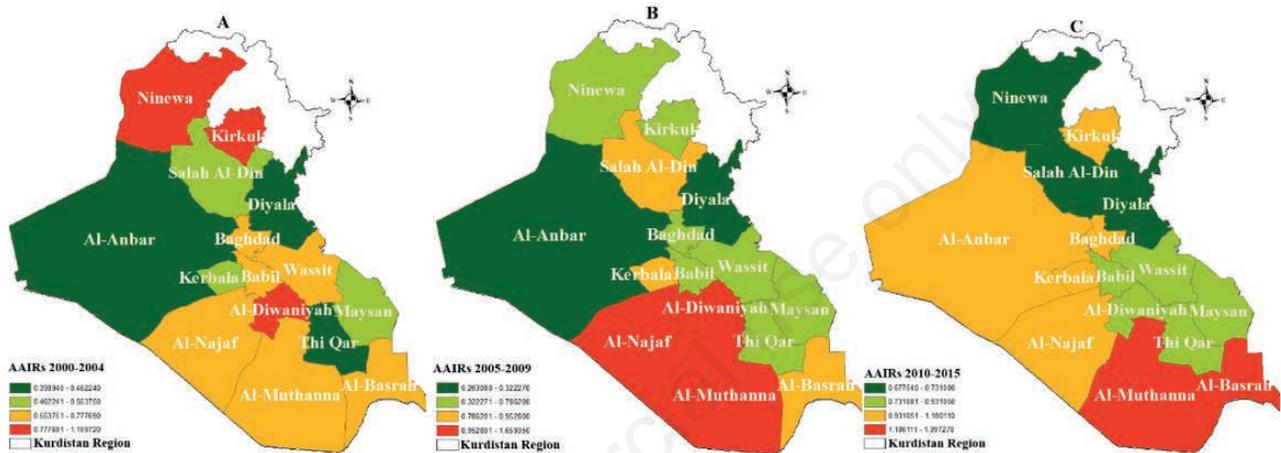
Figure 1. Variation of the geographical distribution of childhood bone cancer incidence between 2000 and 2015. The age-adjusted incidence rate (AAIR) for A) for 2000-2004; B) for 2005-2009; and C) for 2010-2015.
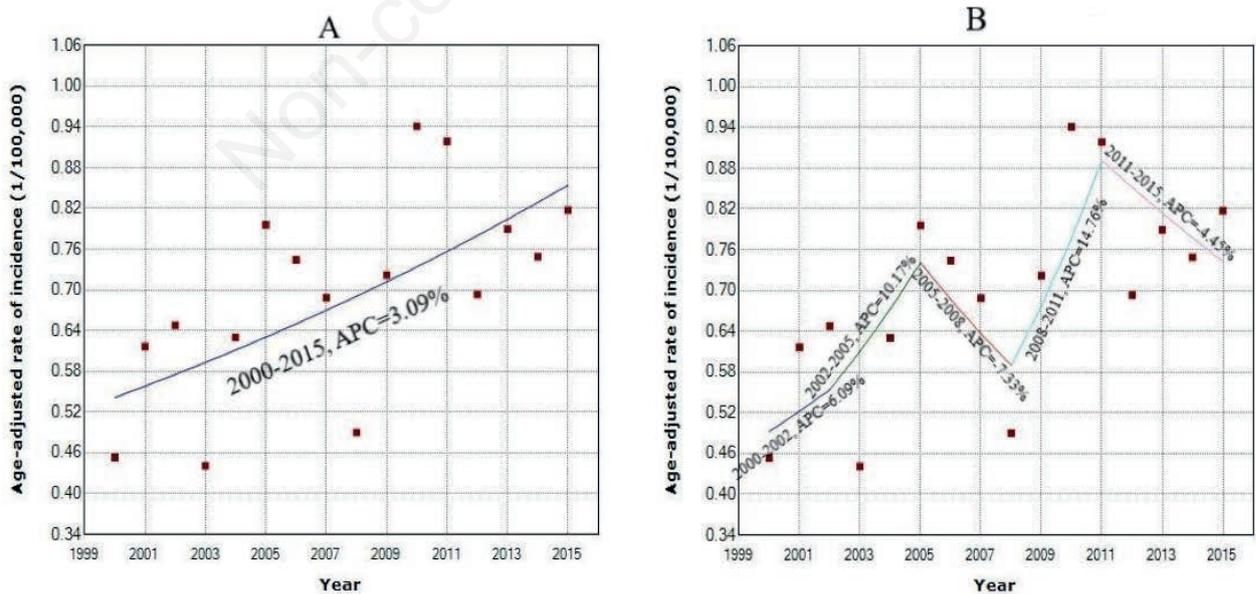
Figure 2. Join point analysis of the age-adjusted incidence rate of childhood bone cancer in Iraq between 2000 and 2015. A) The general temporal trend of the average percentage change (APC) of the age-adjusted incidence rates over the whole period; B) The short-term trend changes in five sub-divisions of the time.

higher, although it was statistically insignificant by 1.7% (IRR=1.0168, 95% CI: 0.8121, 1.4311) and 2.0% (IRR=1.0204, 95% CI: 0.8184, 1.4510), respectively. The incidence rate was lower among the remaining provinces; the lowest one being Diyala, situated in the East, at 51.7% (IRR=0.4829, 95% CI: 0.2934, 0.7952); AL-Anbar, 40.8% (IRR=0.5917, 95% CI: 0.3691, 0.7842) in the West; Thi-Qar, 36.9% (IRR=0.63610, 95% CI: 0.4846, 0.8903) in the South; Babil, 32.76% (IRR=0.6724, 95% CI: 0.4794, 0.8242) in central Iraq and Ninewa 31.5% (IRR=0.6852, 95% CI: 0.4214, 0.8976) in the North (Table 5 and Figure 5A).

Discussion

This is the first study of bone tumours in 0-19 years old in Iraq, with incidence distributed into age groups, year-periods and gender across fifteen provincial areas, covering the entirety of the country with the exception of three Kurdish province, for which data were not available. Four models were applied to analyse the data over a 15-year period. The data showed over-dispersion with excessive zeros where the ZIP performed better than the other regression models because it handled the extra zeroes and the over-dispersion in a superior way.

The situation in Iraq was similar with respect to the worldwide data on osteosarcoma and Ewing's sarcoma, but Ewing's sarcoma
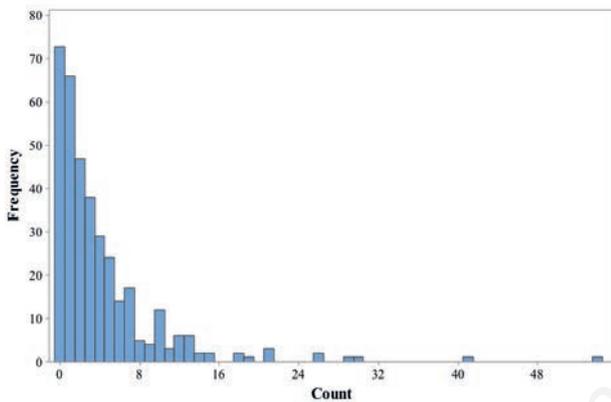
Figure 3. Bone tumour frequency counts from 2000 to 2015.
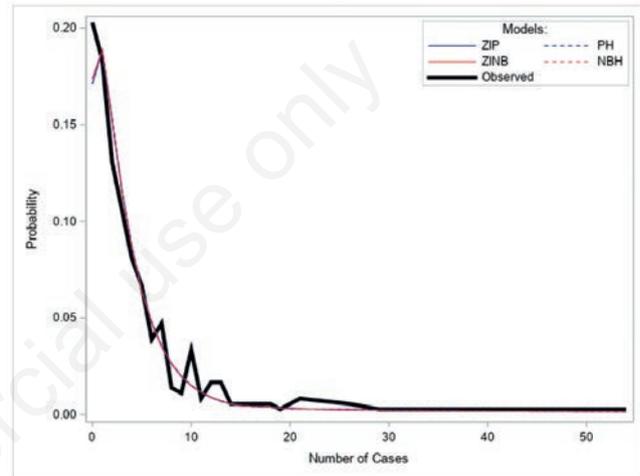
Figure 4. Average predicted count probability.

Table 5. Parameter estimates of the zero-inflated Poisson model fitted to childhood bone tumour data.

Variable	Level	IRR	P-value	95% CI	
Age group	5-9	2.3438	<0.0001	1.5830	2.9441
	10-14	5.9008	<0.0001	4.0213	6.9092
	15-19	8.2292	<0.0001	6.4094	9.7651
Gender	Male	1.4912	<0.0001	1.1880	3.0073
Time-period	2005-2009	1.2286	0.0055	0.9603	1.4244
	2010-2015	1.4967	<0.0001	1.1690	1.8656
Province	Baghdad	0.7792	0.0104	0.6296	0.8911
	Ninewa	0.6852	0.0013	0.4214	0.8976
	Maysan	0.6935	0.0355	0.4903	0.8275
	Al-Diwaniyah	1.0204	0.8862	0.8184	1.4510
	Diyala	0.4829	<0.0001	0.2934	0.7952
	Al-Anbar	0.5917	0.0004	0.3691	0.7842
	Babil	0.6724	0.0038	0.47941	0.8242
	Kerbala	0.793422	0.1505	0.6803	0.9092
	Kirkuk	0.925584	0.6075	0.6474	1.0102
	Wassit	0.689285	0.0184	0.4923	0.8099
	Thi-Qar	0.636100	0.0010	0.4846	0.8903
	Al-Muthanna	1.01679	0.9206	0.8121	1.4311
	Salah Al-Din	0.701173	0.0184	0.5232	0.8930
	Al-Najaf	0.974715	0.8552	0.6742	1.0188

IRR, incidence rate ratio; CI, confidence interval.

in Iraq was found to be most common in the 10-14 age group. In accordance with studies carried out in Africa (Mohammed *et al.*, 2010; Ghert *et al.*, 2019), USA (Mirabello *et al.*, 2009) and Europe (Stiller *et al.*, 2006), it was noted that bone tumours in childhood and adolescence were more frequent in males than females. Our results that bone tumour incidence peaks in the 15-19 years olds are also similar to findings by others (Katchy *et al.*, 2005; Rhutso *et al.*, 2013; Wilhelm *et al.*, 2014; Ramdass *et al.*, 2015; Pruksakorn *et al.*, 2016). The incidence rates had a fluctuating pattern, generally showing an increasing annual trend, with peaks in 2010 and relatively small peak in 2011.

The distribution of the incidence of bone tumours in children and adolescents is different among the provinces in Iraq. However, very little is known about the aetiology of bone tumours in children; there are few papers that have implicated a role of environment. A Spanish study (García-Pérez *et al.*, 2017) have suggested a probable risk of bone tumours in children living near metal processing plants, cement and combustion plants, waste management, energy generating sector and plantations using pesticides. Eyre *et al.* (2009) and Ottaviani and Jaffe (2009) have put forward the possible role of genetic/environment interaction. These reports support the idea of agricultural, industrial and conflict-related contam-
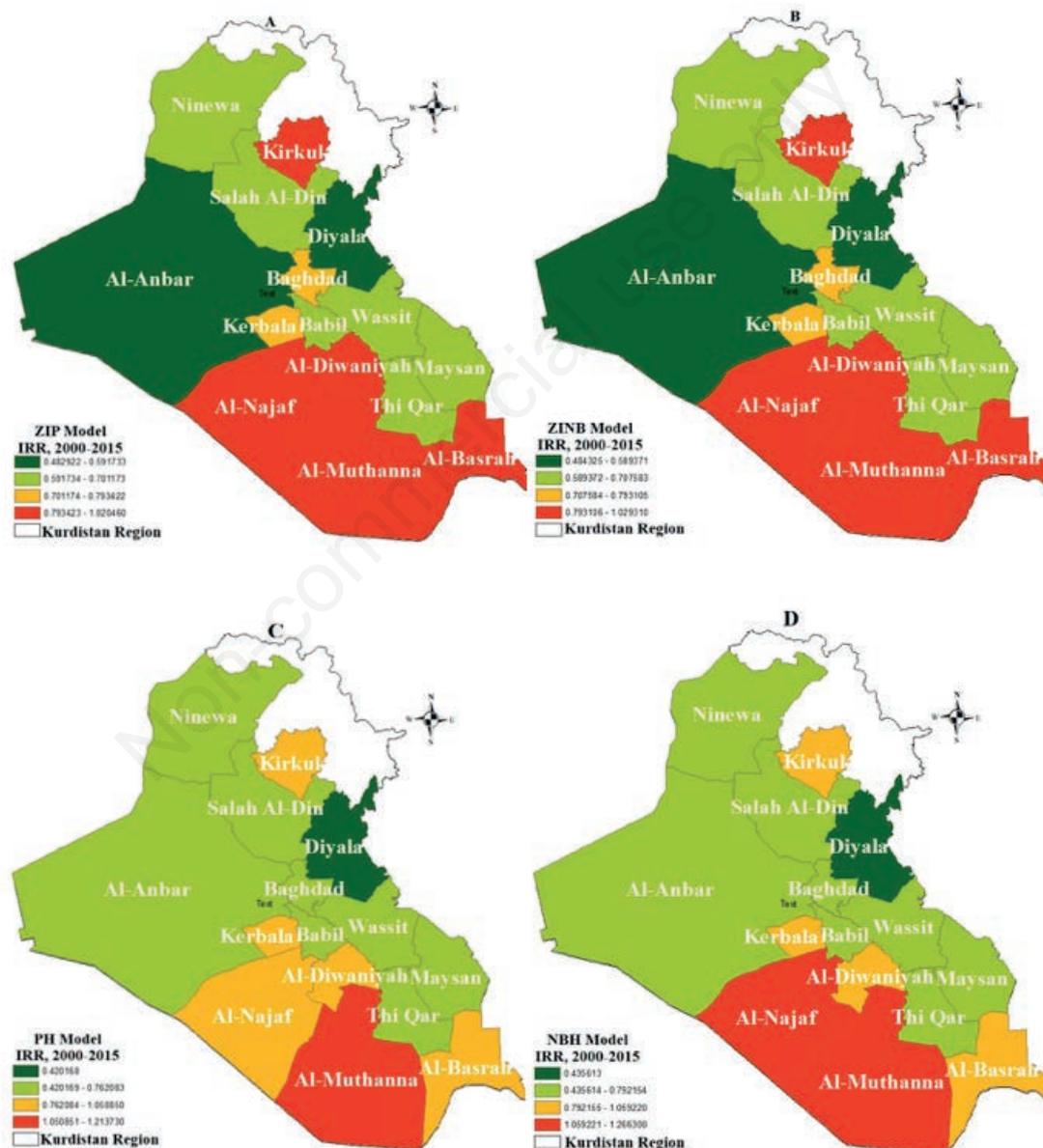
Figure 5. Incidence rate ratio (IRR) of bone cancer, childhood using four models between 2000 and 2015, using A) the zero-inflated Poisson (ZIP) model; B) the zero-inflated negative binomial (ZINB) model; C) the Poisson hurdle (PH) model; D) the negative binomial hurdle (NBH) model.

ination that environmental pollutants could be a factor contributing to the differences of bone tumour incidence among Iraqi provinces.

Al-Diwaniyah Province is a rural environment and is heavily cultivated, increasing the probable risk factors such as exposure to pesticides and various agrochemicals. Many articles have reported that Al-Diwaniyah Province suffers from environmental pollution. Walli (2015) found highly pollution of surface water in many sites in the Al-Dalmaj Marshes, Al-Diwaniyah province. The river is neither suitable for drinking nor swimming due to the high-level of pollution (Hussein *et al*., 2019). Polynuclear aromatic hydrocarbons, considered a soil contaminant, have been discovered in Al-Diwaniyah city centre (Afak, Al-Shamiya and Al-Hamza) (Karam & Kadhim, 2014). Another study found chemical pollutants in different types of food (Karam, 2015).

In Al-Muthanna Province, cement and brick factories use black oil, which is usually not efficiently combusted leading to air pollution. Through the southernmost provinces, there are multiple sites where raw sewage is discharged directly into the rivers which flow through them. The use of pesticides in agriculture is unregulated and rife. These are all polluting factors in Iraq, particularly in Al-Muthanna, Al-Diwaniyah and Al-Basrah provinces (Al-Daghir, 2013).

Iraq that is heavily contaminated with pollutants related to the wars (Al-Shammari, 2016). Many researchers have written about environmental pollution levels in Al-Basrah (Al-Aasadi *et al*., 2015; Al-Hassen *et al*., 2015; Al-Shammari, 2016). All these sources confirm air pollution in Al-Basrah during the years 1980 to 2003 when it suffered wars and their aftermath. Subsequently, there was an expansion of industrial and petrochemical plants accompanied by increased traffic density and air pollutants caused by daily activities (Douabul *et al*., 2013).

Iraq's unique geographical situation with two major rivers that cross the country transporting waste of various kinds flow from their catchment areas that include large parts of Turkey, eastern Syria and western Iran. For this reason, the water in Iraq, in particular in the southern part, is loaded with the results of industrial and natural activities in these areas. This element has the potential to affect the health of the population, especially children and young adults. Not surprisingly, the highest incidence was observed is in the southernmost provinces of Al-Muthanna, Al-Diwaniyah and Al-Basrah.

Although understanding the reasons for the variation of the incidence of bone tumours across Iraqi governorates is not the primary aim of this research, the detection of higher incidence rates in different areas is suggestive of the need to provide services for diagnosis and treatment in the locations with higher incidence. It also highlights the need for more research into the causes behind such variation to develop the necessary public health strategies including diagnostic and therapeutic services, besides the necessary policies regarding any contributory environmental factors.

There is no clear evidence for the relationship between cases and geographical and temporal differences, that may be traced to the development of inconsistent criteria for bone tumour registration across the country, because of variation in the accuracy of registration from one governorate to another. This research is a positive endeavour in that direction, which may strongly suggest the role of environmental factors such as exposure to radiation, chemical agents, and pollution. Further research in that area is well worth pursuing.

Conclusions

This study establishes that ZIP model is the preferred model over others. Osteosarcoma and Ewing Sarcoma are the most common type of bone cancer in children and adolescents in Iraq. Bone tumours showed incidence patterns variation among age groups, with peak in the 15-19 age group. The distribution of the incidence of bone tumours in children and adolescents is different between Iraqi provinces. The highest incidence was observed in Al-Muthanna, Al-Diwaniyah, and Al-Basrah provinces in the south of Iraq. The annual incidence of bone tumours from 2000-2015 showed a fluctuating pattern over the period of study, with significant increasing trend.

The study suggests further research into childhood tumours, and especially bone tumours, is needed to look at the effect of environmental factors in this group of disorders beside establishing the impact of availability of specialist local services on the reported incidence of this group of disorders.

References

Al-Aasadi KA, Alwaeli AA, Kazem HA, 2015. Assessment of air pollution caused by oil investments in Basra Province-Iraq. J Novel Appl Sci 4:82-6.

Al-Daghir HA, 2013. The environmental influences on the incidence of cancer in AL-Muthana governate April 2010. Almuthanna J Pure Sci (MJPS) 1:1.

Al-Hassen SI, Sultan AW, Ateek AA, Al-Saad HT, Mahdi S, Alhello AA, 2015. Spatial analysis on the concentrations of air pollutants in Basra province (southern Iraq). Open J Air Pollut 4:139.

Al-Shammari AM, 2016. Environmental pollutions associated to conflicts in Iraq and related health problems. Rev Environ Health 31:245-50.

Arora RS, Alston RD, Eden TO, Geraci M, Birch JM, 2012. The contrasting age□incidence patterns of bone tumours in teenagers and young adults: Implications for aetiology. Int J Cancer 131:1678-85.

Cameron AC, Trivedi PK, 2013. Regression analysis of count data. Cambridge University Press, Cambridge, UK, pp 112.

Damron TA, Ward WG, Stewart A, 2007. Osteosarcoma, chondrosarcoma, and Ewing's sarcoma: national cancer data base report. Clin Orthop Relat Res 459:40-7.

Diallo AO, Diop A, Dupuy JF, 2019. Estimation in zero-inflated binomial regression with missing covariates. Statistics 53:839-65.

Douabul AA, Maarofi SS, Al-Saad HT, Al-Hassen S, 2013. Gaseous pollutants in Basra city, Iraq. Air Soil Water Res ASWR-S10835.

Du J, Park YT, Theera-Ampornpunt N, McCullough JS, Speedie SM, 2012. The use of count data models in biomedical informatics evaluation research. J Am Med Inform Assoc 19:39-44.

Evola FR, Costarella L, Pavone V, Caff G, Cannavò L, Sessa A, Avondo S, Sessa G, 2017. Biomarkers of osteosarcoma, chondrosarcoma, and Ewing sarcoma. Front Pharmacol 8:150.

Eyre R, Feltbower RG, Mubwandarikwa E, Eden TO, McNally RJ, 2009. Epidemiology of bone tumours in children and young adults. Pediatr Blood Cancer 53:941-52.

García-Pérez J, Morales-Piga A, Gómez-Barroso D, Tamayo-Uria I, Romaguera EP, López-Abente G, Ramis R, 2017. Risk of

bone tumours in children and residential proximity to industrial and urban areas: New findings from a case-control study. Sci Total Environ 579:1333-42.

Ghert M, Mwita W, Mandari FN, 2019. Primary bone tumors in children and adolescents treated at a referral center in Northern Tanzania. J Am Acad Orthop Surg Glob Res Rev 3:3.

Grimer R, Athanasou N, Gerrand C, Judson I, Lewis I, Morland B, Peake D, Seddon B, Whelan J, 2010. UK guidelines for the management of bone sarcomas. Sarcoma 1:2010.

Hussein KM, Fadhil SA, Salih SA, 2019. Assessing water quality for Al-Diwaniyah River, Iraq using GIS technique. Eng Technol J (ETJ) 37:256-64.

Ibp I, 2012. Iraq country study guide, Volume 1. Strategic information and developments. International Business Publications, Inc., Alexandria, VA, USA.

Iraqi Cancer Board, 2018. Annual Report: Iraqi Cancer Registry 2016. Ministry of Health and Environment, Republic of Iraq.

Kaatsch P, Strothotte J, Becker C, Bielack S, Dirksen U, Blettner M, 2016. Pediatric bone tumours in Germany from 1987 to 2011: incidence rates, time trends and survival. Acta Oncol 55:1145-51.

Karam FF, 2015. The concentrations, distribution and sources of Benzo (a) pyrene in vegetables from Al-Diwaniya Province, Iraq. J Al-Qadisiyah Pure Sci Q 2:86-91.

Karam FF, Kadhim MI, 2014. Determination of Chrysene by Gas Chromatography in Selected Soils from Al-Diwaniya Province, Iraq, 2014. Asian J Chem 26:S259-61.

Katchy KC, Ziad F, Alexander S, Gad H, Mota'al MA, 2005. Malignant bone tumours in Kuwait. Int Orthop 29:406-11.

Klakattawi HS, Vinciotti V, Yu K, 2018. A simple and adaptive dispersion regression model for count data. Entropy 20:142.

Lambert D, 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34:1-4.

Martin J, Hall DB, 2017. Marginal zero-inflated regression models for count data. J Appl Stat 44:1807-26.

Meyers PA, Gorlick R, 1997. Osteosarcoma. Pediatr Clin North Am 44:973-89.

Mirabello L, Troisi RJ, Savage SA, 2009. Osteosarcoma incidence and survival rates from 1973 to 2004: data from the Surveillance, Epidemiology, and End Results Program. Cancer 115:1531-43.

Mohammed A, Sani MA, Hezekiah IA, Enoch AA, 2010. Primary bone tumours and tumour-like lesions in children in Zaria, Nigeria. Afr J Paediatr Surg 7:16.

Mouatassim Y, Ezzahid EH, 2012. Poisson regression and Zero-inflated Poisson regression: application to private health insurance data. Eur. Actuar. J. 2(2):187-204.

Mullahy J, 1986. Specification and testing of some modified count data models. J Econom 33:341-65.

Ottaviani G, Jaffe N, 2009. The etiology of osteosarcoma. In: Jaffe N, Bruland OS, Bielack S, eds. Pediatric and adolescent osteosarcoma. Springer, Boston, MA, USA, pp 15-32.

Özkan EA, Göret CC, Özdemir ZT, Yanık S, Doğan M, Gönültaş A, Akkoca AN, 2015. Pattern of primary tumours and tumor-like lesions of bone in children: retrospective survey of biopsy results. Int J Clin Exp Patho 8:11543.

Pittman B, Buta E, Krishnan-Sarin S, O'Malley SS, Liss T, Gueorguieva R, 2018. Models for analyzing zero-inflated and overdispersed count data: an application to cigarette and marijuana use. Nicotine Tob Res 22:1390-8.

PosthumaDeBoer J, Graat HC, Bras J, Saouti R, 2010. Small cell osteosarcoma of a toe phalanx: a case report and review of literature. J Orthop Surg Res 5:36.

Pruksakorn D, Phanphaisarn A, Pongnikorn D, Daoprasert K, Teeyakasem P, Chaiyawat P, Katruang N, Settakorn J, 2016. Age Standardized incidence rates and survival of osteosarcoma in northern Thailand. Asian Pac J Cancer Prev 17:3455-8.

Ramdass MJ, Mooteeram J, Beharry A, Mencia M, Barrow S, 2015. An 8-YEAR analysis of bone tumours in a Caribbean island. Ann Med Surg 4:414-6.

Rhutso Y, Laishram RS, Sharma LD, Debnath K, 2013. Histopathological evaluation of bone tumours in a tertiary care hospital in Manipur, India. J. Med. Soc 27:135.

Rose CE, Martin SW, Wannemuehler KA, Plikaytis BD, 2006. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. J Biopharm Stat 16:463-81.

Steliarova-Foucher E, Colombet M, Ries LA, Moreno F, Dolya A, Bray F, Hesseling P, Shin HY, Stiller CA, Bouzbid S, Hamdi-Cherif M, 2017. International incidence of childhood cancer, 2001-10: a population-based registry study. Lancet Oncol 18:719-31.

Stiller CA, Bielack SS, Jundt G, Steliarova-Foucher E. Bone tumours in European children and adolescents, 1978-1997, 2006. Report from the Automated Childhood Cancer Information System project. Eur J Cancer 42:2124-35.

Walli HA, 2015. Determination of PAHs in Surface Water of Al-Dalmaj Marsh, Al-Diwaniya Province, Iraq. J Enviro Earth Sci (JEES) 5:67-72.

Wilhelm M, Dirksen U, Bielack SS, Whelan JS, Lewis IJ, Jürgens H, Ferrari S, Sundby Hall K, Cleton-Jansen AM, Stark D, 2014. ENCCA WP17-WP7 consensus paper on teenagers and young adults (TYA) with bone sarcomas. Ann Oncol 25:1500-5.

Xie H, Tao J, McHugo GJ, Drake RE, 2013. Comparing statistical methods for analyzing skewed longitudinal count data with many zeros: An example of smoking cessation. J Subst Abuse Treat 45:99-108.

Yesilova A, Kaydan MB, Kaya Y, 2010. Modeling insect-egg data with excess zeros using zero-inflated regression models. Hacettepe J Math Stat 39:273-82.