Teaching Mathematics and Statistics in Sciences - IPA HU-SRB/0901/221/088
Interesting Mathematical Problems in Sciences and Everyday Life - 2011

Investigating the effect of seasonality of birth for acute lymphoblastic leukaemia

Tibor Nyári
Department of Medical Physics and Informatics, University of Szeged, Hungary
nyari@dmi.u-szeged.hu

Summary: The investigation of the seasonality of a disease is an important objective in many epidemiological studies. Two statistical methods (Edwards' and Walter-Elwood methods) are described here to detect cyclic trends in time of birth of children diagnosed with acute lymphoblastic leukaemia (ALL) aged 0 - 4 year . We found evidence of seasonality related to month of birth with peaks in February and August. Both peaks correspond to previous findings and could reflect the seasonality of infectious diseases in temperate climates: respiratory virus infections (for example, influenza) show marked seasonality occurring in the winter months, and gastrointestinal infections peaks in the summer months.

Keywords: seasonality, Edwards' test, Walter - Elwood test, logistic regression, acute lymphoblastic leukaemia

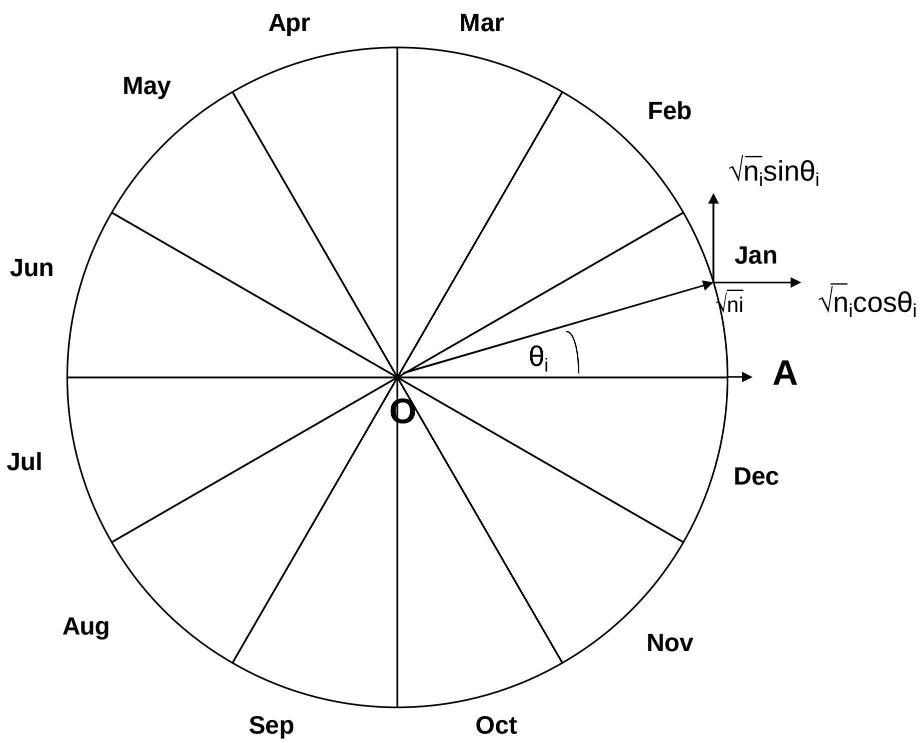
1. INTRODUCTION

Infection has long been suspected as a possible factor in the etiology of acute lymphoblastic leukaemia (ALL)[1]. A cyclic trend in occurrence could be interpreted as supportive evidence of an etiology linked to exposure to infection. The investigation of seasonality, or cyclic variation of a disease frequency throughout a year has been of great interest to researchers in various medical disciplines. Various methods of estimating and evaluating cyclic trends have been reported. The methodology for studying seasonal variation ranges from simple graphical methods to more advanced techniques which are more efficient and mathematically sophisticated. The simplest method for the detection of seasonal variation in disease risk is the Pearson chi-square test, defined as the sum of square of the deviations divided by the expectation, which has $n - 1$ degrees of freedom (d.f.). This is a poor test for detecting a cyclic trend as of the $n - 1$ degrees of freedom only one or two are likely to be necessary to specify any biologically meaningful type of trend (H_o was uniform distribution throughout the year) Instead of the Pearson chi-square test the Edwards' test has been used by epidemiologist [2]. This test was generalized by Walter and Elwood in 1975 [7]. These two statistical methods and their application to detect cyclic trends in time of birth of children diagnosed with acute lymphoblastic leukaemia (ALL) aged 0 - 4 year are described here using the original Edwards' and Walter - Elwoods' methodology. Both methods are special types of chi-square test and compare the nullhypothesis of no seasonality against the alternative hypothesis of an existence of a cyclic trend.

2. HARMONIC ANALYSIS

2.1. Edwards' test

The Edwards' test is one of the earliest of the more sophisticated methods developed as a specific test of cyclic variation and is here reintroduced using definitions and formulas of the original paper [2]. A simple test for cyclic trend in independent events is presented in the form of the rim of a unit circle divided into equal sectors (Figure 1.), corresponding to time intervals (eg.: months), and a number in each rim sector specifying the number of events observed.



Edwards' model where θ_i = (2i - 1) π /12 for i =1, 2, …, k

Figure 1

The expected centre of gravity of these masses will be at the circle's centre in the absence of any seasonal trend and any excess of deficit in neighboring sectors will have a consistent effect on the position of the centre of gravity. Thus, the distance from the centre of which will have an uniform probability distribution on the null hypothesis, and the direction will indicate the position of maximum or minimum liability, or both. This χ^2 −test compares the hypothesis of a uniform distribution throughout the year (H_o) with the assumption that the frequencies follow a seasonal fluctuation over the year (H_A).

Let's consider N number of ALL cases which are distributed over k equal sectors (e.g.: k = 12 months). Let n_i denote the number of events in i^{th}sector. The square root of n_i are taken as weights and then placed around a unit circle at points corresponding to the sector midpoints at angles θ_i to an arbitrary diameter. The moment about this diameter is $\sqrt{n_i}\,\sin\theta_i$ and $\sqrt{n_i}\,\cos\theta_i$ where θ_i = (2i - 1)π /12 for $i = 1, 2, \ldots, k$.

Adding all these sectors ($i = 1, 2, ..., k$), the expected moment is zero. The position of the centre of gravity of these points ($\bar{x}$; $\bar{y}$) where

$$\bar{x}=\frac{\sum_{i=1}^{k}\sqrt{n_i}\,\cos\theta_i}{\sum_{i=1}^{k}\sqrt{n_i}} \text{ and } \bar{y}=\frac{\sum_{i=1}^{k}\sqrt{n_i}\,\sin\theta_i}{\sum_{i=1}^{k}\sqrt{n_i}}\ . \qquad (1)$$

The distance from the origin is

$$\mathbf{d} = \sqrt{\bar{x}^2 + \bar{y}^2}\ . \qquad (2)$$

The square root of an observed number has an expected sampling variance of slightly less than $\frac{1}{4}$ (was proved by C.A.B. Smiths[2]), thus the sampling variance is approximately:

$$\sigma^2=\sum_{i=1}^{k}\frac{1}{4}\sin^2\theta_i=\frac{k}{8}\text{ or }\frac{1}{8N}\tag{3}$$

units of moments squared, and the test statistics:

$$\chi^2=8N(\bar{x}^2+\bar{y}^2)\text{ with 2 d.f.}\tag{4}$$

In case of a simple harmonic trend the frequency is proportional to

$$f_i=1+\alpha\sin(\theta_i-\theta^*)\tag{5}$$

From (5) the estimate of α =4d and $\theta^*=\arctan\left(\frac{\bar{y}}{\bar{x}}\right)$, where θ^* gives the direction of the maximum rate.

It should be noted that such a curve has just one peak and one trough during the year. A simple extension of this method to test for the occurrence of two peaks and troughs in the year would be through the hypothesis that the frequencies follow a sinusoidal curve of period six months [3]. The two peaks are exactly six months apart and of the same height in this curve. In mathematical terms the hypothesis is the same as expression (5) - test against uniform distribution (no seasonal variation) - except that θ_i' is substituted for θ_i where $\theta_i'=(2i\text{-}1)\pi/6$ for $i = 1, 2, \ldots, k$. The calculation and test statistic are the same as (4) above replacing θ_i' instead of θ_i in the expressions of quantities $\bar{x}$ and $\bar{y}$ in equation (1).
The goodnes-of fit of the fitted harmonic curve might be tested using a further χ^2-test statistic with k - 1 d.f. under the usual assumptions:

$$\chi^2=\sum_{i=1}^{k}\frac{(n_i-n_i')^2}{n_i'}\text{ where }n_i'=\frac{Nf_i}{\sum_{i=1}^{k}f_i}.\tag{6}$$

2.2. Walter – Elwood method

Edwards' test is widespreadly used but it ignores any possibility of variation in the population at risk which might be responsible for the cyclic variation of the disease incidence and suffers from the lack of power for small sample sizes [4-5]. Occasional extreme values in the data as well as cyclic variations of other forms than a simple harmonic curve have also an influence on test statistic [6], such as the assumption of the equally spaced time intervals may not be fulfilled in practice, e.g. due to the unequal lengths of the months. Although various modifications of this test have later been employed [3-7], only the Walter-Elwood's test [7] use data of population at risk. Thus, this test is a generalization of Edwards' test. Following Edwards, within a certain time span (e.g.: a year) there are k sectors (e.g.: 12 months) and that in i^{th} sector there are n_i events from a population at risk of size m_i (e.g.: total births during that month). The total number of events and the total number of population at risk are $N=\sum_{i=1}^{k}n_i$ and $M=\sum_{i=1}^{k}m_i$, respectively.

The data also weighted by $\sqrt{n_i}$ and placed around the unit circle at sector midpoints at angles θ_i to an arbitary diameter. The equation (1) is also the position of the centre of gravity of these points ($\bar{x}$; $\bar{y}$). The population at risk is respected to calculate the expected values of $\bar{x}$ and $\bar{y}$, by replacing n_i by its expectation $E(n_i)=Nm_i/M$ everywhere in (1) and obtain after symplifying:

$$E(\bar{x})=\mu_{\bar{x}}==\frac{\sum_{i=1}^{k}\sqrt{m_i}\cos\theta_i}{\sum_{i=1}^{k}\sqrt{m_i}}\text{ and }E(\bar{y})=\mu_{\bar{y}}=\frac{\sum_{i=1}^{k}\sqrt{m_i}\sin\theta_i}{\sum_{i=1}^{k}\sqrt{m_i}}.\tag{7}$$

Walter and Elwood estimated the variance of $\sqrt{n_i}$ as approximately 0.25, thus the variances of the sample centre of gravity's coordinates are:

$$\sigma_{\bar{x}}^2=\frac{\sum_{i=1}^{k}\frac{1}{4}\cos^2\theta_i}{\left(\sum_{i=1}^{k}\sqrt{Nm_i/M}\right)^2}\text{ and }\sigma_{\bar{y}}^2=\frac{\sum_{i=1}^{k}\frac{1}{4}\sin^2\theta_i}{\left(\sum_{i=1}^{k}\sqrt{Nm_i/M}\right)^2}.\tag{8}$$

The test statistic assuming $\bar{x}$ and $\bar{y}$ are normally distributed is:

$$\chi^2=\left(\frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)^2+\left(\frac{\bar{y}-\mu_{\bar{y}}}{\sigma_{\bar{y}}}\right)^2 \text{ with 2 d.f.} \quad (9)$$

The distance (d) of the sample centre of gravity from its null expectation is given by:

$$\mathbf{d} = \sqrt{\left(\bar{x}-\mu_{\bar{x}}\right)^2+\left(\bar{y}-\mu_{\bar{y}}\right)^2} \quad (10)$$

Equation (5) became:

$$c_i = m_i[1+\alpha\cos(\theta_i-\theta^*)] \quad (11)$$

where α is a measure of the cyclic variation and θ^* gives the discretion of the maximum rate and α and θ^* are

estimated

$$\alpha=\frac{2\left[d\sqrt{\mathrm{kM}}-\sum_{i=1}^{k}\sqrt{m_i}\cos(\theta_i-\theta^*)\right]}{\sum_{i=1}^{k}\sqrt{m_i}\cos(\theta_i-\theta^*)} \text{ and } \theta^*=\mathrm{Tan}^{-1}\left(\frac{\bar{y}-\mu_{\bar{y}}}{\bar{x}-\mu_{\bar{x}}}\right)\text{, respectively.} \quad (12)$$

In the case $\theta_i = 2\pi i/k$, $m_i = M/k$, $i = 1, 2, \ldots, k$ it may be shown that $\mu_{\bar{x}} = 0$ and $\mu_{\bar{y}} = 0$ and that (9), (10) and (12) reduce to the corresponding Edwards' test statistic. Here again, the adequacy of the description of the data afforded by a simple harmonic curve may be evaluated by a goodness-of-fit test using a further χ^2 statistics:

$$\chi^2=\sum_{i=1}^{k}\frac{(n_i-n_i^{'})^2}{n_i^{'}} \text{ with 2 d.f., and the expected frequencies } n_i^{'} \text{ are calculated by } n_i^{'}=\frac{\mathrm{N}c_i}{\sum_{i=1}^{k}c_i}. \quad (13)$$

The fitted cyclic curve has also just one peak and one trough during the year. The extension of this method to test for the occurrence of two peaks and troughs in the year is similar to described above for Edwards' test when equal month lengths are used. In mathematical terms $\theta_i^{'}$ is substituted for θ_i to calculate c_i, where $\theta_i^{'} = (2i - 1)\pi/6$ for $i = 1,2,\ldots,k$. The calculation and test statistic are also the same as above using $\theta_i^{'}$ instead of θ_i in the expressions of quantities $\bar{x}$ and $\bar{y}$. The refinement of using exact month lengths, in practice, makes unremarkable difference in results.

3. ILLUSTRATIVE EXAMPLE

3.1. Study population

Children born during 1981-1997 in South Hungary were considered. Data from the Central Demographic Agency were available on the number of births for each month over the study period but by gender only for each year [11]. The number of births in each month for each gender was estimated assuming no monthly variation in gender ratio within any year. Registrations of first malignancies for children, born and diagnosed with ALL under age five years in Hungary before the end of 2002 were obtained from the Hungarian Pediatric Oncology Group (HPOG). Month was used as the time unit of the analysis. The pattern of annual cyclical variation was studied for month of birth. Three methods were applied for the detection of seasonal variations with both twelwe and six month periods in both Edwards' and Walter-Elwood methods.

3.2. Main findings

The total number of childhood ALL cases diagnosed under age five years for the study was 121. There were 481,984 live births in the study area, during the 17 year-interval of 1981-1997. The monthly distribution of the cases and births is shown in Table 1.

Table 1.The monthly number of cases and births in the period 1981-1997 in South Hungary.

Month	*No of ALL cases*	*No of births*
1	14	41120
2	14	37979
3	11	41345
4	7	38272
5	3	39914
6	8	39923
7	14	43619
8	12	42061
9	10	41433
10	11	39394
11	11	37863
12	6	39061
Total	121	481984

Firstly, we tested whether there is a seasonal pattern with one maximum level and one minimum level per year. The results of both tests are summarized in Table 2. There was no evidence of seasonality of birth for ALL in children for twelwe month period. However, evidence of seasonality related to month of birth with peaks in February and August was found in the six month period analysis.

Table 2. Results of the test for seasonal variations (N=121, M=481984)

	Edwards'	*Walter-Elwood*	*Walter-Elwood exact month length*
Twelve month period			
Amplitude	0.19	0.20	0.20
angle of maximum rate	-48.7	-39.9	-39.9
p-value	0.34	0.29	0.29
Six month period			
Amplitude	0.38	0.34	0.34
angle of maximum rate	85.5	87.0	86.9
p-value	0.014	0.027	0.028
Peaks	February, August	February, August	February, August
Troughs	April, October	April, October	April, October
Goodness-of-fit p-value	0.68	0.65	0.65

4. COMMENTS AND CONCLUSIONS

We have presented two methods to analyze cyclic variation of ALL in children diagnosed under five year.

An investigation of seasonality stands as an important component in the etiological description of diseases, particularly for certain cancers of childhood and infectious diseases [12]. The test proposed by Edwards has occupied a central place in analyses of seasonality of epidemiological data having been applied by a large number of investigators [5,13-14]. The only data required are the number of cases of the disorder. Therefore, this test does not consider variation in the population at risk. It has also been criticized for being sensitive to occasional extreme values in the data [6]. Walter and Elwood [7] modified and generalized the Edwards' test to avoid these difficulties, to allow for an arbitrary pattern of variation in the population at risk and also for unequal lengths of time intervals. Different from Edwards' test, the expected number of events is calculated to be proportional to the population at risk. Both Edwards' and Walter-Elwood tests were originally applied on cyclic data with one peak and one trough. Edwards' test was extended by Cave and Freedman [3] to analyze the occurrence of double peaks and troughs, and a similar extension of Walter-Elwood test has been applied in our analysis.

Seasonal variation related to time of birth would indicate exposure to an infection in utero or around the time of birth. We found evidence of seasonality related to month of birth with peaks in February and August. Both peaks correspond to previous findings and could reflect the seasonality of infectious diseases in temperate climates: respiratory virus infections (for example, influenza) show marked seasonality occurring in the winter months, and gastrointestinal infections peaks in the summer months [15]. There are hypothesis about the role of infection in the aetiology of leukemia. Kinlen proposed that childhood leukemia could be a rare response to infectious exposure. Our study provided some evidence of an effect of exposures to infections around the time of birth. This data were further analysed in gender effect which was supported by Bolyai fellowship and published [16]. There were a significant difference between boys and girls in the pattern of seasonal trends which corresponds to the findings of our previous study [17].

5. Acknowledgement

Research supported by the Hungarian National Foundation for Scientific Research Grant No.T 049516, and cofinanced by the European Union through the Hungary - Serbia Cross - border Cooperation programme in the frame of the project IPA HU - SRB/0901/221/088.

6. REFERENCES

[1] Kinlen LJ. Epidemiological evidence for an infective basis in childhood leukaemia. Br J Cancer 1995; 71:1-5.

[2] Edwards JH. The recognition and estimation of cyclic trends. Ann Hum Genet 1961;25:83-86.

[3] Cave DR and Freedman LS. Seasonal variations in the clinical presentation of Chron’s diseae and ulcerative colittis. Int J Epidemiol 1975; 4: 317-320.

[4] Hewitt D, Milner J, Cisma A and Pakula A. On Edwards' criterion of seasonality and a non - parametric alternative. Brit J Prev Soc Med 1971; 25: 174 -176

[5] Roger JH. Significance test for cyclic trends in incidence data. Biometrika 1977; 64 : 152 - 155

[6] Wehrung DA and Hay S. A study of seasonal incidence of congenital malformations in the United States. Br J Prev Soc Med 1970; 24 : 24 - 32.

[7] Walter SD, Elwood JM. A test for seasonality of events with a variable population at risk. Br J Prev Soc Med.1975; 29(1): 18 - 21.

[8] Hallberg F, Johnson E, Nelson W, Runge W, Sothern R. Autorhythmometryprocedures for physiologic self - measurements and their analysis. Physiol Teach 1972; 1 : 1 - 11.

[9] Fleiss JL, Williams JBW and Dubro AF. The logistic regression analysis of psychiatric data. J Psychiatr Res 1986;20: 145 - 209.

[10] Stolwijk AM,Straatman H and Zielhuis GA. Studying seasonality by using sine and cosine functions in regression analysis. J Epidemiol Community Health 1999;53:235-238.

[11] Demographic Yearbook.Budapest:Central Demographic Agency;1982-1998

[12] Le CT, Liu P, Lindgren BR, Daly KA,Scott Giebink G. Some statistical methods for investigating the date of birth as a disease indicator. Stat Med 2003;22:2127-35.

[13] Marrero O. The performance of several statistical test for seasonality in Monthly data. J Statist Comput Simul 1983;17:275-296.

[14] Jones RH,Ford PM& Hamman RF. Seasonality comparisons among groups using incidence data. Biometrics 1988;44:1131-1144.

[15] Lopman BA,Reacher M,Gallimore C,Adak GK,Gray JJ,rown DW. A summertime peak of “winter vomiting disease”:surveillance of noroviruses in England and Wales,1995 to 2002. BMC Public Health 2003;3:13

[16] Nyári TA, Kajtár P, Bartyik K, Thurzó L, McNally R, Parker L. Seasonal Variation of Childhood Acute Lymphoblastic Leukaemia is Different Between Girls and Boys. Pathol Oncol Res 2008;14:423-428.

[17] Nyari TA, Kajtar P, Bartyik K, Thurzo L, Parker L. Childhood acute lymphoblastic leukaemia in relation to population mixing around the time of birth in South Hungary. Pediatr Blood Cancer 2006;47:944 - 948.