Analysis on decomposed time series of Congenital Heart Defects among Children Born to New York State Residents, 1983-2005

Gang Liu[1], Ying Wang[1], Charlotte M Druschel[1], Igor Zurbenko[2]
[1] Congenital Malformations Registry, Bureau of Environmental & Occupational Epidemiology, Center for Environmental Health New York State Department of Health, Empire State Plaza- Corning Tower, Room 1243, Albany, NY 12237.
[2] Department of Epidemiology & Biostatistics, University at Albany, School of Public Health, One University Place, Room 131, Rensselaer, NY 12144-3456

ABSTACT

Background: Several studies have been conducted on examining the seasonality in congenital heart defects (CHDs), but the results were not consistent. The objective of this study was to explain the prevalence of nine selected CHDs using New York State Congenital Malformation Registry (CMR) data from 1983 to 2005.

Methods: Time series of daily prevalence was created on children with nine selected CHDs. Kolmogorov-Zurbenko (KZ) filter was used to decompose the time series. Linear regression was applied to explain the long term trend. Graphical analysis was applied to examine seasonal and weekly patterns. Walter & Elwood test was also applied on the seasonality.

Results: The long term trend could be modeled by two long term trends of the percentage of older maternal age (>35) and the percentage of Hispanic mother. Neither KZ filter nor Walter & Elwood find seasonal pattern in CHDs prevalence, while weekly pattern was found as decreased prevalence in Sunday.

Key Words: Kolmogorov-Zurbenko (KZ) filter, congenital heart defect, approximation, seasonality, weekly pattern.

INTRODUCTION

Congenital heart defects (CHDs), which is the most common birth defect category in United State, has a relatively higher prevalence of 4 to 12 per thousand live births compared to other birth defects. Several studies have been conducted recently on CHDs to examine the seasonality of CHDs using statewide (Eghtesady et al., 2011) or regional (Siffel et al., 2005) population-based birth defects surveillance data in the United State. Population-based studies on seasonality of CHDs have also been conducted in other countries including Swiss (Bosshardt et al., 2005), Malta (Grech, 1999), Finland (Tikkanen et al., 1994; Tikkanen et al., 1993) and Puerto Rico (de la Vega A et al. 2009).

Previous works on seasonality of CHDs only focused on single category, such as hypoplastic left heart syndrome (Eghtesady et al., 2011; Tikkanen et al., 1994); coarctation of aorta (Tikkanen et al., 1993); or several single categories in one paper, such as (Siffel et al., 2005; Grech, 1999). There was no study giving a completing examination on seasonality of CHDs. Most of these studies used the raw monthly number of cases to examine the seasonality. The limitation of this method is that it does not consider the effect of the size of the population at risk, which has a significant contribution in a statistical evaluation of seasonality. Moreover, those studies only focused on seasonality without analyzing long term trend or any short term fluctuations. Most important, when we want to examine the time variables' effect on the time series of CHDs' prevalence, if we use raw prevalence to conduct the analysis, it is hard to find changes in seasonal and short term domain, because the prevalence of CHDs is always fluctuate around a certain number which is much bigger than the change in local changes, so it is necessary to separate the different time scale components in time series of CHDs' prevalence and address the relationship between the prevalence and different time variable using different time scale data. In this study, with the help of Kolmogorov-Zurbenko (KZ) filter, we could conduct analysis on CHDs' prevalence to evaluate the different time variables' affection on CHDs.

In 1982, the Congenital Malformation Registry (CMR) of the New York State Department of Health (NYSDOH) was established as part of the Environmental Disease Surveillance Program. Since the inception of the CMR, birth defects surveillance data have not been used to investigate the effect of the time variable on the prevalence of CHDs among New York State (NYS) children. NYS has its unique geographic location, environment and demographic situation which may cause the prevalence of CHDs having some long term trend, seasonal pattern or short term fluctuations. Because CHDs is the most complicated category among all birth defects, differences in case ascertainment and case inclusion and exclusion criteria among the 23 years cause the prevalence changing a lot in some individual CHDs, which are not available for our study. In the study, we only include nine CHDs, which have relatively consistent prevalence.

The objective of this study was to examine the prevalence patterns of nine major selected CHDs over time using New York State Congenital Malformation Registry (CMR) data from 1983 to 2005. We used a new method to examine the prevalence of selected CHDs for seasonality, long term trend and the weekly pattern using longitudinal birth defects surveillance data.

METHODS

1. Study Cohort.

A cohort was constructed containing children born in 1983-2005 with selected nine major CHDs. ICD-9 codes were used to identify CHDs in this study. The selected CHDs included common truncus (ICD-9: 745.0), pulmonary valve atresia and stenosis (ICD-9: 746.01,746.02), transposition of great arteries (ICD-9: 745.10, 745.11, 745.12, 745.19), tetralogy of fallot (ICD-9: 745.2), Ebstein's anomaly (ICD-9: 746.2), aortic valve stenosis (ICD-9: 746.3), hypoplastic left heart syndrome (ICD-9: 746.7), atrioventricular septal defect (ICD-9: 745.60, 745.61, 745.69), and coarctation of aorta (ICD-9: 747.1). The selected nine CHDs were a subset of all CHDs reportable to CMR.

2. Data match.

Deterministic data linkage methods were used to match CMR cases to birth certificates, in order to obtain various birth variables, and maternal demographic information from birth certificates. In this matching processing, multiple matching variables, such as both child's and mother's names, data of birth, medical record number, and mother's social security number and residential information were used. This matching method resulted in more than 97% of CMR children matched to their birth certificates. Only the CMR children who were matched to their birth certificates were included in the study.

3. Create time series.

Two daily time series and nine monthly time series were created using the study cohort and live births data. The first daily time series was a series of daily live births counts, which was calculated using the live births between 1983 and 2005. The second daily time series was a series of daily prevalence of selected major CHDs using the study cohort as the numerator and the live births series as the denominator. The nine monthly time series were nine series of the monthly prevalence of nine selected CHDs, which were also calculated using the study cohort and live births series.

4. Statistical analyze.

Spectrum analysis was used to explore the cyclical patterns in the time series. It is a nonparametric method with the purpose to decompose a complex time series with cyclical components into a few underlying sinusoidal (sine and cosine) functions of particular wavelengths. As a result of a successful analysis, you might uncover just a few recurring cycles of different lengths in the time series of interest, which at first looked more or less like random noise.

Kolmogorov-Zurbenko (KZ) filter is a low-pass filter that can decompose the time series based on time scale to long term trends, the seasonal variations, short term fluctuations and the white noise. KZ filter is a high resolution filter and its algorithm is simple to understand. Mathematically, KZ filer can be defined as following.

Let X(t), t=… -2, -1, 0,1,2… be a real-valued time series, KZ filter with parameters m and k is:

$$KZ_{m,k}[X(t)] = \sum_{s=-k(m-1)/2}^{k(m-1)/2} \frac{a_s^{m,k}}{m^k} X(t+s)$$

, where $a_s^{m,k}$ are given by the polynomial coefficients of $\left(1 + z + \ldots + z^{m-1}\right)^k$. Using KZ filter, the original time series X(t)= Lt(t)+Sv(t)+St(t), where Lt(t) is the long-term trends, Sv(t) is the seasonal variations, St(t) is short-term fluctuations (weekly pattern) (Wei et al., 2010).

Using KZ filter, we decomposed these two daily time series into long-term trend ($KZ_{365,3}$(X(t)), seasonal variations ($KZ_{29,3}$(X(t)-Lt(t)) and short-term variations (X(t)-Lt(t)-Sv(t)); decomposed these nine monthly time series into long-term trend ($KZ_{12,3}$(X(t)) and seasonal variations (X(t)-Lt(t)).

After decomposed by KZ filter, we use linear regression method to explain the long term trend of CHDs. We selected some risk factors in the population, then withdraw their long term trend using KZ filter. Using these variables’ long-term trend, we modeled the long-term trend of CHDs. Because the noise in the regression variable will lead to underestimation of the regression coefficient, the smoothed date will give an accurate estimation of the regression coefficient (Zurbenko I, Sowizral M (1999), Tsakiri K, Zurbenko I). Multivariate analysis between long term component can face difficulties in direct statistical approach because observations within those curves cannot be treated as independent. We may avoid this difficulty by addressing Gauss Least Square approximation between nonrandom curves. In order to evaluate the goodness fitting of this nonrandom linear regression, we used the method of coefficient of explanation (R2) which is introduced in approximation theory (Selezjev (1999), Maiorov (1996), Ritter (1996), Sard (1963)). The linear regression equation could be expressed as Y=+ε, where Y is the long-term trend of dependent variable, X□ are the long-term trends of independent variables, a□ are coefficients, then the method that least-square approximation was used to estimate the coefficients, that means minimizing the value of (Y-)^2. The content and calculation of coefficient of explanation is almost be the same as its original definition (Nagelkerke, 1991): firstly calculate the total sum of squares (SStot) as , sum of squares of residuals (SSerr) as , then R^2=1-SSerr/SStot. (is each observations of Y, is the mean of and is each observations of Xi). But in this nonrandom regression, coefficient of explanation is not independent of the sample size, adjusted R^2, which equal to 1-(1- R^2)*(n-1)/(n-p-1) (n is the effect sample size, p is the number of independent variable), should be used to evaluate the goodness fitting. Moreover, in this linear regression, moving average was used instead of raw time series, the effect sample size should be equal to N/(effective span of the filter) (N is the real sample size).

In order to examine the seasonality and weekly heterogeneity, we plotted the mean and confidence interval (CI), then examined if the CIs had intersection to conclude if there is any seasonal or weekly pattern in the prevalence of CHDs or in live births data.

Walter-Elwood Test is an extension of Edwards' test, which arranges 12 monthly frequencies as a set of weights on the rim of a circle representing the year. As the most popular method examining seasonality, we used its result as a comparison (Walter et al., 1975; 1977).

The data matching was performed using the Statistical Analysis System (SAS) software package (SAS Institute Inc., Cary, NC, version 9.2) and the statistical analysis was performed using R (programming language) and STATA (general-purpose statistical software package).

RESULTS

Figure 1 presents the period-gram of daily prevalence of CHDs and daily live births series. In the daily time series of CHDs' prevalence, only one frequency value (0.143) was observed. It means a circle with 1/0.143=7 days is present as weekly pattern. In live births time series, three frequencies (0.0027, 0.0055 and 0.143) corresponding to three circles with 1/0.0027=365, 1/0.0055=183 and 1/0.0143=7 days, which means that seasonality and weekly pattern existing.

Figure 2, Figure 3 and Table 1 present the decomposing results of KZ filter. In figure 2, we see a significant increase in long term trend of CHD prevalence from 2.2 per thousand to 2.9 per thousand, but no obvious pattern presenting in seasonality, which has already presented in the period-gram. In figure 3, we see a significant change in long term trend and an obvious pattern in seasonality for live births data. The number of live birth increased since 1983 to the peak at 1991, then decreased every year until 2005. The seasonality of live births presented as increasing births in summer. In table 1, we compare the variances of different components. The variances of short terms are enough large in both CHDs and live births indicating that the fluctuation is very significant in the short term components. This fluctuation might cause by some time variable shorter than month.

Figure 4 and Table 2 present the result of linear regression explaining of CHDs' long-term trend. In the linear regression, long term trends of two variables, percent of older maternal age (>35 years) and percent of Hispanic mothers are selected to estimate the response variable (Prevalence of CHDs). As can be seen in figure 4, that all these variables have significant relationship with the response variable. The adjusted R^2 value is 0.69 indicating that almost 70% of the response variable can be explained by the two variables, while less than 1% could be explained when we used the raw data. In figure 4, we plot the observed long-term trend and the predicted of the long-term trend, the predicted curve fitted the observed curve well except for the period from 1990 to 1997. When we compared it to the long term trend of live births, we found that there was a significant decrease in live births in the same period which might be because of some unknown factors that also had some effect on the prevalence of CHDs.

Figure 5 presents Annual patterns with the mean and 90% confidence interval of CHDs and live births. For CHDs, low prevalence was observed in April and August (90% CIs in these month were less than 0), but no obvious seasonal pattern was observed because the lowest 90% confidence intervals is overlap with the highest one. In the Live births, the seasonality is present with high number of live births in summer time and low numbers in winter time. The number of the highest point is almost 70 more than the lowest point. In figure 6, we also examine monthly prevalence change from the mean for each CHD category and their 95% CI after KZ decomposing. Although we found some high prevalence for some individual CHDs during certain periods such as the month of December for common truncus, August for transposition of great arteries and June for tetralogy of fallot, we could not find any seasonal pattern in individual CHDs.

Figure 7 presents weekly patterns with the mean and 95% confidence interval of CHDs and live births. For CHDs, the weekly pattern presented with decreased prevalence on Sundays. For live births, the weekly pattern presented with decreased number on weekends compared to that on weekdays from Tuesday to Friday.

Table 3 lists the result of Walter & Elwood test results. All p-value are bigger than 0.05, indicating that there is no seasonality in any CHDs or the nine CHDs combine. The results are consistent with that obtained when analyzing only on the seasonal component after KZ decomposing.

DISCUSSIONS

Our findings suggested that the prevalence of selected nine major CHDs had significant change in long-term trend and obvious weekly pattern, but neither KZ filter nor Walter & Elwood test found seasonal pattern in individual CHDs or their combination. The long-term trend of the CHDs' prevalence increased 20% from 1983 to 2005. This increase might be due to the improvement in diagnoses of CHDs over the years. The large increase in prevalence since 2000 may reflect the implementation of web-based, electronic care reporting system by the CMR.

The use of KZ filter to examine the long term trend of daily CHDs prevalence is the strength of our study, because KZ filter could remove the noise in the time series and decrease the variance of the time series, and then make the time series easier to modeling. When we use time series to estimate time series, their noises could cause underestimation of the regression coefficient, so we need to destruct the time series before the regression. Our method is withdrawing the long term component by decomposing the raw time series using KZ filter, the long term component can present over 90% of the raw observation and it is much smoother for regression. To compare the result with conventional linear regression, 70% of time series explained using smooth data, while only about 0.5% could explain using conventional linear regression model.

Previous studies had found that seasonality was present with increased prevalence in winter for valvular heart defects (Bosshardt et al., 2005) and cardiac anomalies (De la Vega et al., 2009). In our study, seasonality didn't present for any of these nine selected major CHDs. This discrepancy between our study and the literature could be explained by the following. A. Heart disease is an acquired disease studied previously, so it might affect by the environmental factors, while CHDs are mostly affected by maternal risk which might have no significant relationship with environmental. B. In the valvular heart defects paper, they used frequency to examine seasonality without considering the frequency of live births and their live births seasonal pattern was significant different from our live births pattern. C. In the cardiac anomalies study, they included more CHD categories that we had a few studies reported that seasonality was found for hypoplastic left heart syndrome but not for other left-sided heart diseases (Eghtesady et al., 2011) and no seasonal patterns were found for coarctation of aorta (Tikkanen et al., 1993) and hypoplastic left heart syndrome (Tikkanen et al., 1994) that are consistent results from our study.

To our knowledge, this is the first study to examine the weekly pattern of CHDs using population-based birth defected surveillance data. The finding of a weekly pattern in live births with decreased frequency on weekends can be explained, in part, that doctors tend to arrange caesarean section and inductions on the weekday (Joshua et al, 2008). The maternal risk, rather than the environmental, factors of CHDs affected the prevalence of CHDs. The finding of a weekly pattern with decreased prevalence on Sundays cannot be confirmed by other reference paper. A guess reason for this finding is if a birth has CHDs, it might be more difficult than the normal birth, so doctors are willing to move this case to week day, when the situation is safer for the birth because of the support of more manpower and resources.

This study had the following limitations. A, the data for some important CHD categories, such as ventricular septal defect, atrial septal defect and patent ductus arteriosus, could not be analyzed because of misdiagnosis or incorrectly coded reporting. B, some variables may affect the prevalence of CHDs, such as maternal obesity (Mills et al. 2010), the family history of the disease and the information of mother's smoke, alcohol or drug usage (Tikkanen et al., 1993; 1994) were not available for the analysis, if these variables are available, the adjusted R^2 could be increased to a higher level.

REFERENCES

Stepanets AI (2005). Methods of Approximation Theory.

Barnett AG, Dobson AJ. 2004. Estimating trends and seasonality in coronary heart disease. Stat Med. 23(22):3505-23.

Bosshardt D, Ajdacic-Gross V, Lang P, et al. 2005. Season of birth in valvular heart disease. Paediatr Perinat Epidemiol 19(3):246-52.

Browne ML, Bell EM, Druschel CM. 2007. Maternal caffeine consumption and risk of cardiovascular malformations. Birth Defects Res A Clin Mol Teratol 79(7):533-43.

De la Vega A, López-Cepero R. 2009. Seasonal variations in the incidence of some congenital anomalies in Puerto Rico based on the timing of conception. P R Health Sci J: 28(2):121-5.
Eghtesady P, Brar A, Hall M. 2011.Seasonality of hypoplastic left heart syndrome in the United States: a 10-year time-series analysis. J Thorac Cardiovasc Surg 141(2): 432-8.

Eskedal LT, Hagemo PS, Eskild A, et al. 2007. A population-based study relevant to seasonal variations in causes of death in children undergoing surgery for congenital cardiac malformations. Cardiol Young17 (4):423-31.

Fellman J, Eriksson WA. 1999. Statistical analysis of the seasonal variation in the twinning rate. Twin Research 2, 22-99.

Grech V. 1999. Trends in presentation of congenital heart disease in a population-based study in Malta. Eur J Epidemiol 15(10):881-7.

Joshua SG, Andrew L. 2008. What Explains the fall in Weekend Births? Journal of Economic Literature Classification Numbers: I11, J13.

Mills JL, Troendle J, Conley MR. 2010. Maternal obesity and congenital heart defects: a population-based study. Am J Clin Nutr 91(6):1543-9.

Siffel C, Alverson CJ, Correa A. 2005. Analysis of Seasonal Variation of Birth Defects in Atlanta. Birth Defects Research (Part A) 73:655-662.

Tikkanen J, Heinonen OP, 1993. Risk factors for coarctation of the aorta. Teratology 47(6): 565-72.

Tikkanen J, Heinonen OP, 1994. Risk factors for hypoplastic left heart syndrome. Teratology 50(2):112-7.

Walter SD, Elwood JM. 1975. A test for seasonality of events with a variable population at risk. British Journal of Preventive and Social Medicine 29: 18-21.

Walter SD. 1977. The power of a test for seasonality. British Journal of Preventive and Social Medicine 31: 137-140.

Selezjev, O. V. (1999). Linear approximation of random processes and random processes and sampling design problems. Prob. Theory and Math. Stat., pp. 665–684

Tsakiri K, Zurbenko IG, 2011, Effect of noise in principal component analysis, Journal of Statistics and Mathematics, ISSN: 0976-8807 & E-ISSN: 0976-8815, Volume 2, Issue 2, 2011, PP-40-48.

Maiorov, V. E. and Wasilkowski, G. W. (1996). Probabilistic and average linear widths in -norm with respect to -fold Wiener measure. J. Approx. Theory 84, 31–40.

Marian Neamtu, Larry Schumaker(2010). Approximation Theory XIII: San Antonio.

Nagelkerke, Nico J.D. (1992) Maximum Likelihood Estimation of Functional Relationships, Pays-Bas, Lecture Notes in Statistics, Volume 69, 110p ISBN 0-387-97721-X.

Radu Păltănea. Approximation Theory Using Positive Linear Operators.

Ritter, K. (1996). Average Case Analysis of Numerical Problems, Habilitationsschrift, Erlangen.

Sard, A. (1963). Linear Approximation, AMS, Providence, Rhode Island.

Theil, Henri (1961). Economic Forecasts and Policy. Holland, Amsterdam: North.

Yang W, Zurbenko IG. 2010(a). Nonstationarity. Wiley interdisciplinary Reviess. In: Computational statistics. Wiley, vol 2, pp 107-115.

Yang W, Zurbenko IG. 2010(b). Kolmogorov-Zurbenko filters. Wiley Interdisciplinary Reviews-Computational Statistics 2(5):340-351.

Zurbenko IG (1989). The spectral analysis of time series. North Holland Series in Statistics and Probability.

Zurbenko IG, Sowizral M (1999). Resolution of the destructive effect of noise on linear regression of two time series. Far East J Theor Stat 3: 139-157.

ATTACHED: Figures and Tables

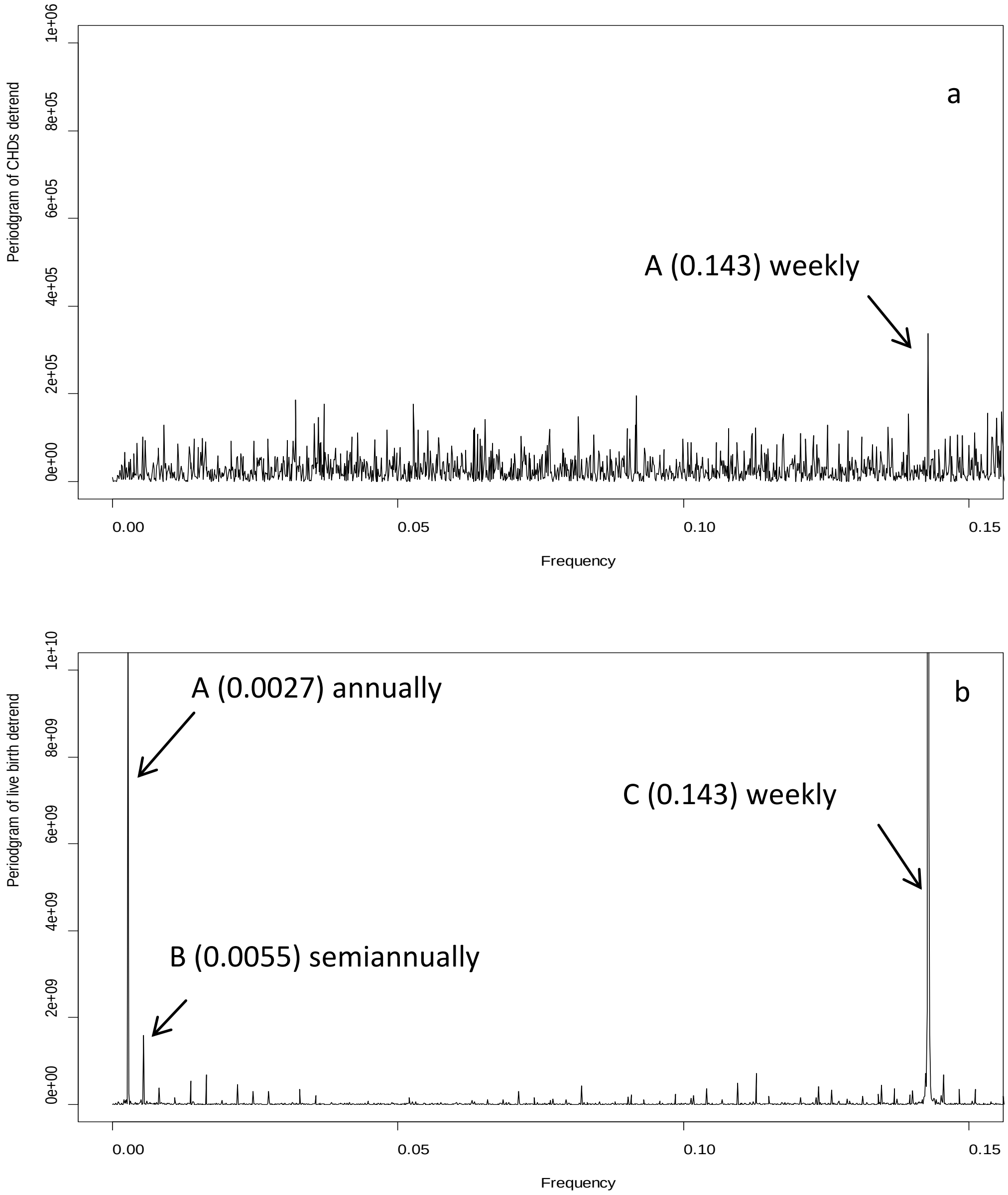


Figure 1. Periodgram of daily prevalence of CHDs (a) and daily live births series (b).

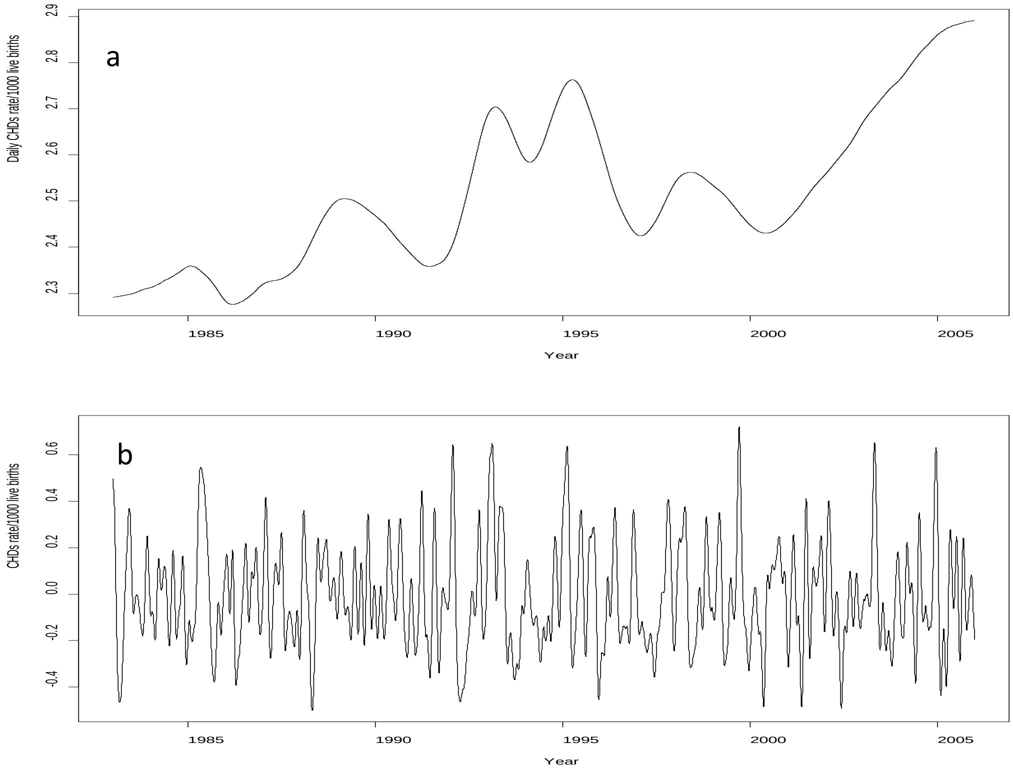


Figure 2. Long term (a) and seasonal (b) component of daily CHDs' prevalence after decomposing.

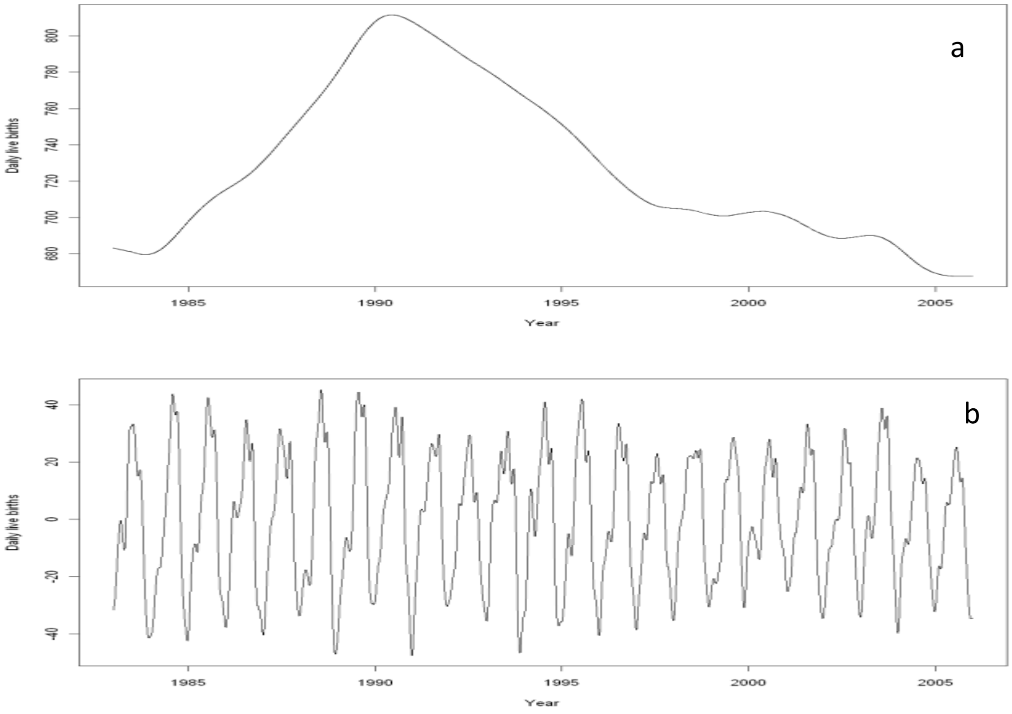


Figure 3. Long term (a) and seasonal (b) component of daily live birth after decomposing.

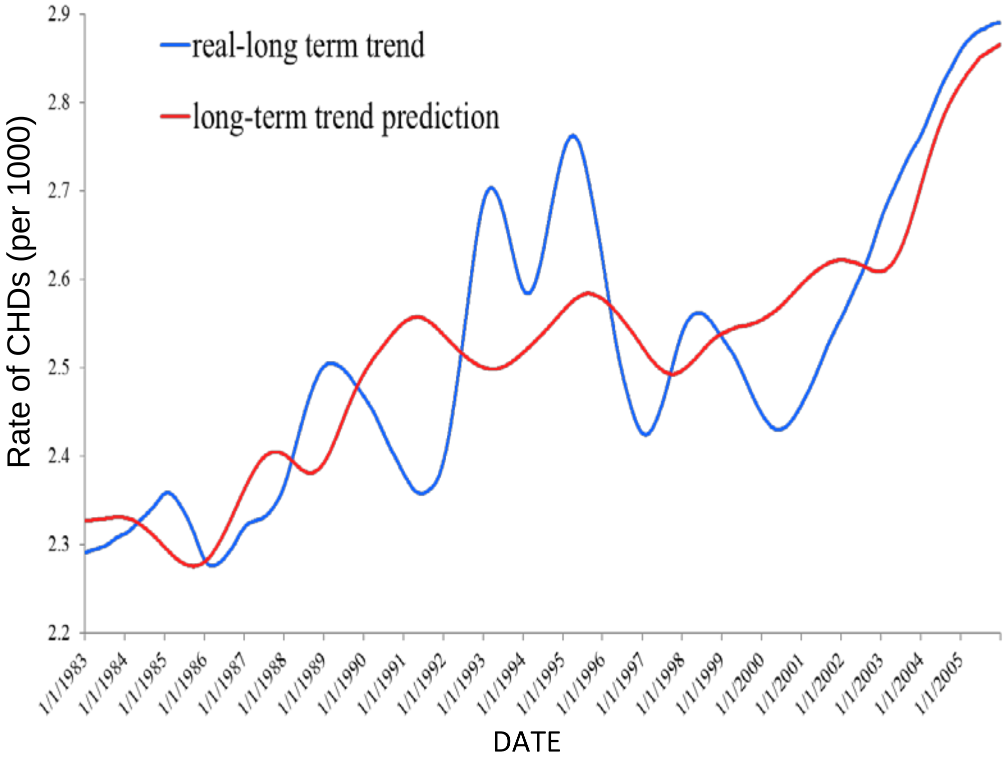


Figure 4. CHDs' real long-term trend vs. predicted result.

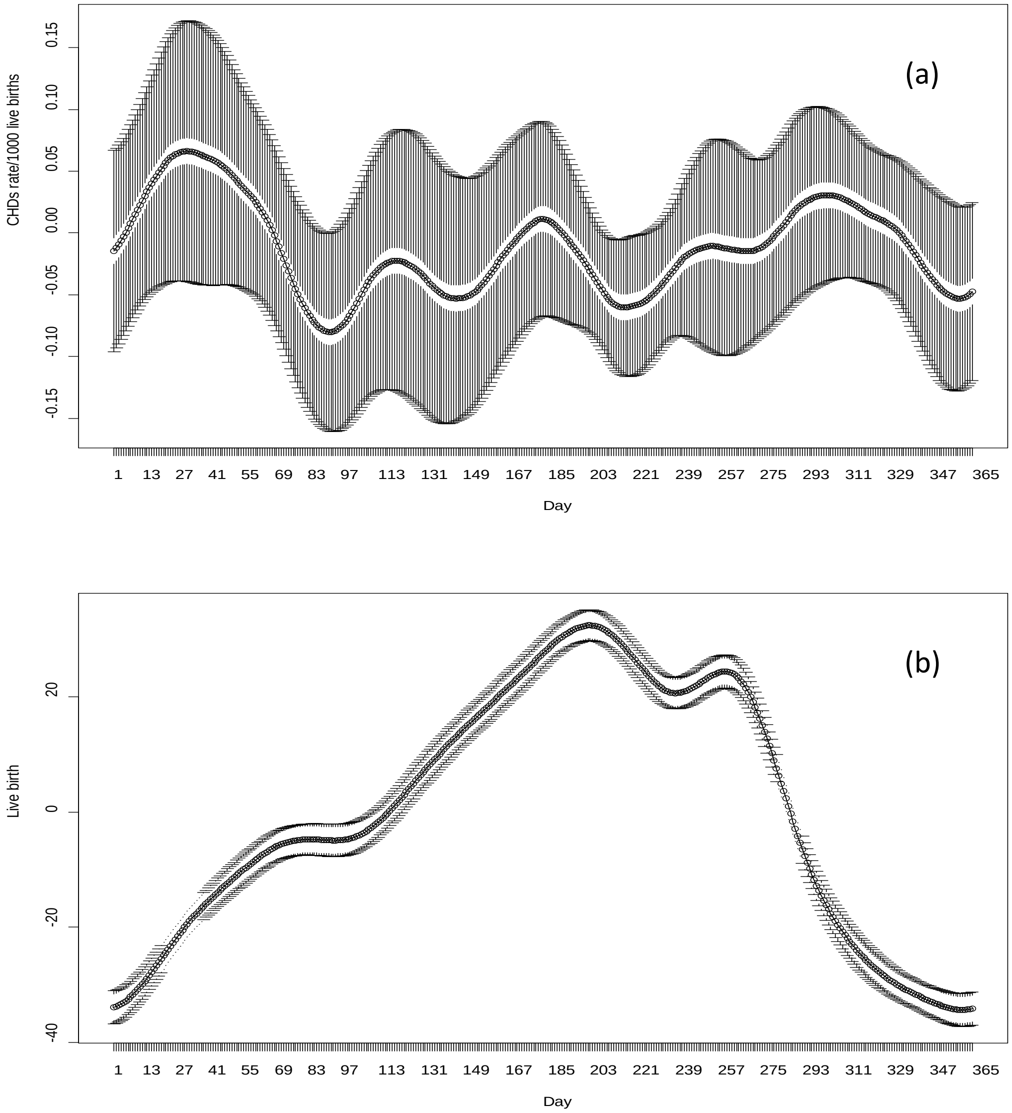
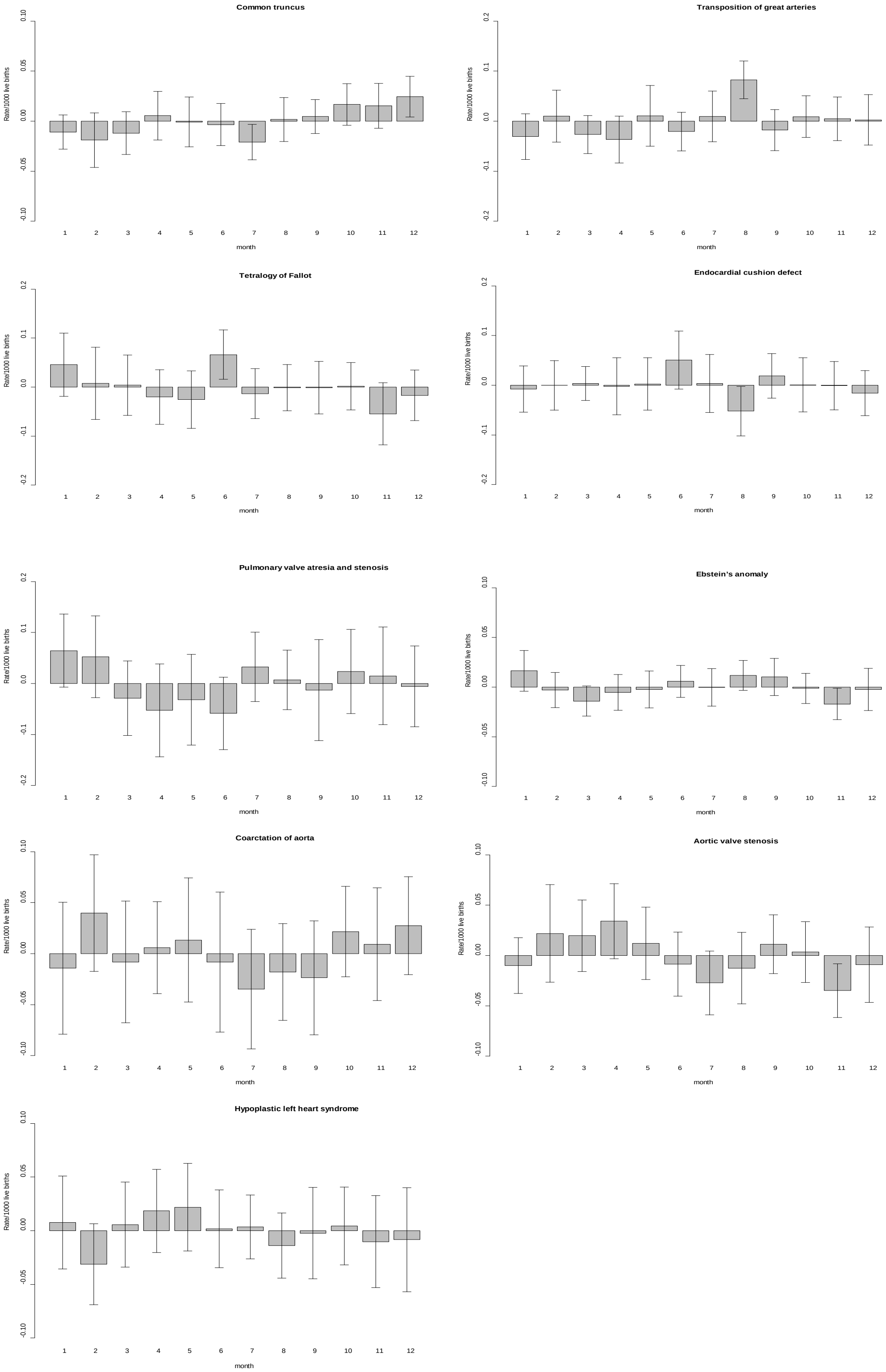


Figure 5. Annual patterns with the mean and 90% confidence interval: (a) the prevalence of selected CHDs; (b) the number of live births.



Common truncus

Transposition of great arteries

Tetralogy of Fallot

Endocardial cushion defect

Pulmonary valve atresia and stenosis

Ebstein's anomaly

Coarctation of aorta

Aortic valve stenosis

Hypoplastic left heart syndrome

Figure 6. Annual patterns with the mean and 95% confidence interval of nine selected CHDs' prevalence.

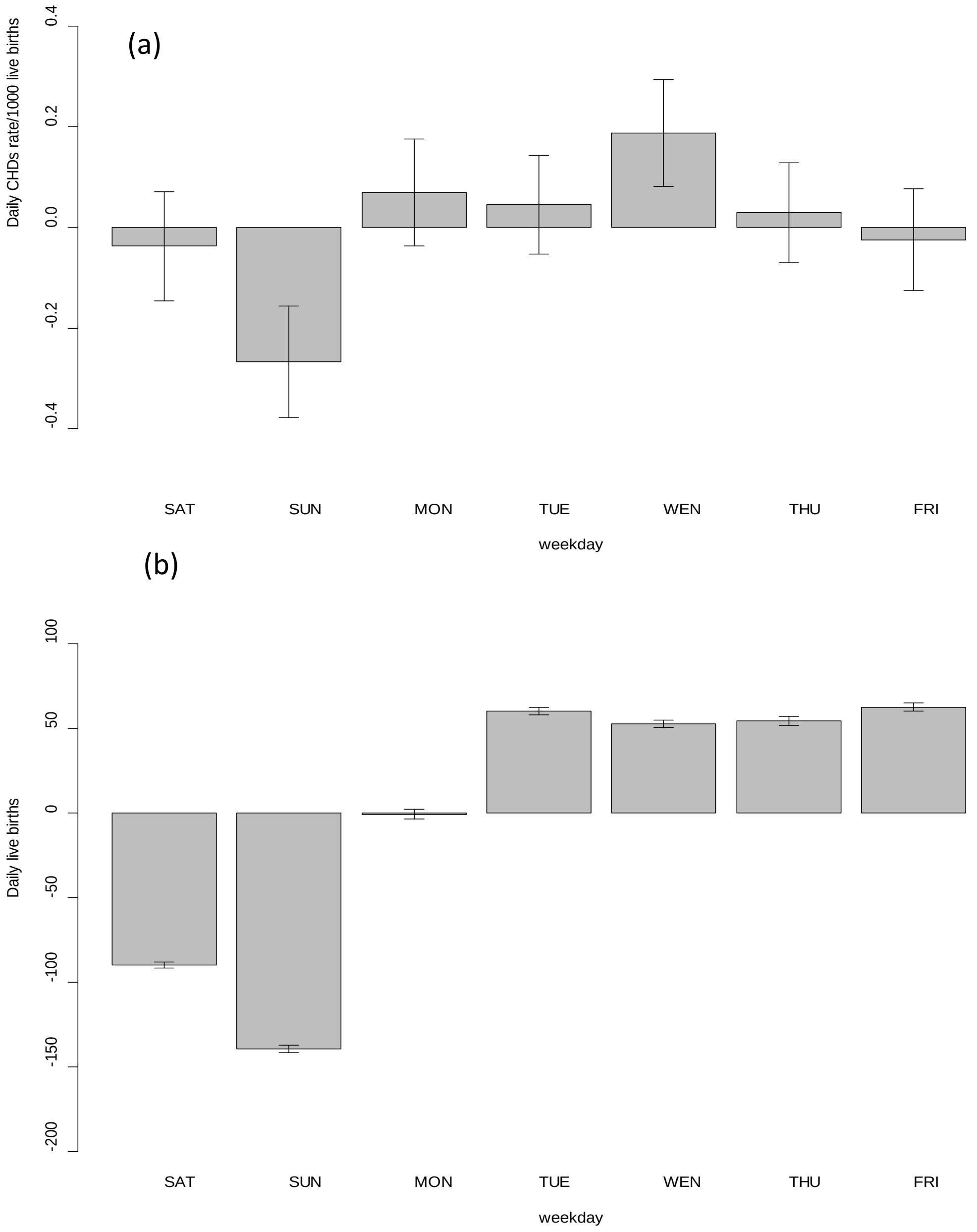


Figure 7. Weekly patterns with the mean and 95% confidence interval: (a) the prevalence CHDs; (b) the number of live births.

	Raw data(Xt)	Long term(Lt)	Seasonality(Se)	Short term(St)
Component	Xt	$Kz_{365,3}$(Xt)	$Kz_{29,3}$(Xt-Lt)	Xt-Lt-Sv
Variance(CHDs)	3.531326	0.02706679	0.04916425	3.412448
Variance(LB)	10114.18	1854.126	513.607	7587.277

Table 1. Decomposition statistics of CHDs and Live births.

Models using same predictor variables	Model P value	Adjusted R-square
Conventional linear regression model	<0.001	<0.01
Long term trend linear regression model	<0.001	0.69

Table 2. Linear regression result of two methods

	Walter & Elwood test		
Category	ϴ□	peak month	p-value
Common truncus	-65.49	October	0.08
Transposition of great arteries	240.22	September	0.12
Tetralogy of fallot	132.55	May	0.82
Atrioventricular septal defect	140.31	May	0.69
Pulmonary valve atresia and stenosis	-34.59	November	0.16
Ebstein’s anomaly	235.14	August	0.48
Aortic valve stenosis	87.14	March	0.08
Hypoplastic left herat syndrome	141.14	May	0.56
Coarctation of aorta	16.32	January	0.30
Total	-13.21	December	0.81

□: ϴ is the degree of the month angle.

Table 3 Results of Walter & Elwood test on seasonality.