Analysis of Birth Defect Monitoring Data and Comparison of Prediction Models in Lanzhou, China

Guangzhuang Jing1§; Yanjun Yang2§; Li'ao Xie1; Jun Zhang2; Qingli Bai1; Li Pan2; Zhilan Li1*
1Institution of Maternal, Child and Adolescent Health, School of Public Health, Lanzhou University, Lanzhou, China
2Maternal and Child Health Hospital of Lanzhou, Lanzhou, China

*Corresponding Author(s): Zhilan Li
Research Institution of Child, Adolescent and Women’s Health Care, School of Public Health, Lanzhou University, 199 West Donggang Road, Chengguan District, Lanzhou, 730000, China
Email: lizhl@lzu.edu.cn

§Equal Contribution: Guangzhuang Jing and Yanjun Yang

Abstract

Background: To analyze the distribution of birth defects in Lanzhou, China from 2012 to 2018, to forecast the prevalence in the future, and to provide evidence for prevention of birth defects.

Materials and Methods: Monitoring datas for birth defects in Lanzhou from 2012 to 2018 were used to describe its epidemiological characteristics, such as the change of the prevalence and trend, urban-rural differences, the varieties and cis-position. Gray modeling, Curve estimation, Exponential smoothing, and ARIMA were used to predict the occurrence for the next year in the future.

Results: The average prevalence rate of total perinatal infants’ birth defects was 104.13 per 10000 perinatal infants in the past 7 years in Lanzhou, with an upward trend increasing by 7.74% on average each year. From 2012 to 2018, the average incidence of birth defects in urban and rural areas of Lanzhou was 105.61/10000 and 106.48/10000, respectively, and the overall trend was upward. The most common birth defects in Lanzhou from 2012 to 2018 were cleft lip ± cleft palate and Congenital heart disease. Exponential smoothing model can better fit the number of birth defects in Lanzhou. The predicted results of exponential smoothing model indicate that birth defects in Lanzhou in 2019 are lower than that in 2018.

Conclusions: The number of birth defects in Lanzhou from 2010 to 2018 is on the rise, and is still in a high incidence in 2019. Congenital heart disease and Polydactylism are the main types of birth defects, which should be strengthened in the future.

Keywords: Birth defects; Perinatal infants; Epidemiological characteristics; Model comparison; Prediction.

Cite this article: Jing G, Yang Y, Xie L, Zhang J, Bai Q, Pan L, Li Z. Analysis of Birth Defect Monitoring Data and Comparison of Prediction Models in Lanzhou, China. A Epidemiol Public Health. 2020; 3(1): 1018.
Introduction

Birth Defects (BDs) consist of a group of conditions, including physical or biochemical abnormalities, present at birth, which may involve malformations, disruptions, deformations [1,2]. The causes of BDs are complex, which can result from genetic defects, non-genetic maternal environmental exposures, or a combination of genetic and non-genetic factors [3-5]. The risk for BDs affects everyone, irrespective of socioeconomic status, race-ethnicity or other demographic characteristics.

The World Health Organization (WHO) estimates that approximately 260,000 deaths (7% of all neonatal deaths) globally were caused by BDs in 2004. The prevalence rates of BDs is estimated reach to 4.7% in the developed countries, 5.6% in the middle-income countries, and 6.4% in the low-income countries [6,7]. As far as we know, China is the most populated developing country in the world, and the prevalence of BDs varies across different socioeconomic groups and areas [8]. China is also a middle-income country with the largest population in the world, which has 16 million births annually and is expected to have 0.9 million BDs each year [9]. According to the most recent surveillance data in 2011, BDs have become the second leading cause of infant deaths in China (the leading cause being premature/low birth weight), which plays a significant role in the main causes of spontaneous abortion, stillbirth, perinatal death, infant death and congenital disability [10]. BDs also affect the child's and the family's quality of life and place an intolerable economic burden to the family and the society, particularly in the setting of the opening of a two-child per family policy in 2015 [11].

Lanzhou is the capital and largest city of GanSu Province. It is located in the geometric center of the geographical map of the motherland and is also an important node city of the “Belt and Road”. Lanzhou governs three counties of Yuzhong County, Gaolan County, Yongdeng County, and five districts of Chengguan District, Qilihe District, Xigu District, Anning District, Honggu District [12]. Lanzhou is located in the northwestern part of China. It's natural and environmental conditions are relatively vague, the level of economic development is backward, the people's living standards are relatively low, and the maternal and child health care facilities are poor. There is still a large gap between the economically developed areas.

The objectives of this study are to estimate the prevalence and types of BDs in Lanzhou, to explore the prevalence of BDs through different predictive models to provide a basis for future prevention and prevention of BDs.

Materials And Methods

Data collection

The surveying population was perinatal infants (including stillbirth, fetal death or live birth) born in the Maternal and Child Health Hospital of Lanzhou from 2012 to 2018. All surveillance data of BDs were collected from Maternal and Child Health Hospital of Lanzhou and confirmed between 28 weeks of gestation through 7 days after birth.

Each delivery that was associated with a BD was reported using a registration card for BDs submitted by physicians in obstetrics and gynecology, pediatric or neonatal medicine through an online hospital-based survey, required by the Chinese government. Each case report card recorded basic maternal information, including ethnicity, residence, family income, education, mother's age, number of antenatal care visits, gestational age, number of reported abortions, pregnancy outcomes, birth information (gestational age, weight of birth), diagnoses of specific BDs, symptoms, early pregnancy disease (fever, virus infection, diabetes), medication use during early pregnancy, and family history. In addition to the case report card, an annual statement for each registered hospital was completed by professional physicians. Each annual statement contained 12 months of data such as the number of perinatal births, maternal age, residence, ethnicity, occupation, pregnancy history, gestational age of birth, gestational age, infant gender, number of BDs, and maternal illness.

Both case report cards and annual statement were reviewed and audited by maternal and child health hospitals and health administrative departments. Periodic quality control measures were in place at the monitored hospitals and occurred quarterly at the county-level and bi-annually at the city-level or province-level to assure reporting accuracy.

Methods

In this study, the Gray model, Curve fitting method, Exponential smoothing model and ARIMA model were used to dynamically predict the trend of BDs in Lanzhou, and the feasibility of four models in the prediction of BDs was discussed.

Gray model (GM)

The GM is to transform the irregular raw data into one or more accumulated data processing methods, transform it into a more regular set of time series, and use discrete series to establish the dynamic model of the differential equation [13]. The GM prediction method is applicable to gray systems with partial information and partial information unknown. It contains both known information and unknown information (or some information is uncertain) [14]. In other words, the prediction of a gray process that changes in a certain azimuth while being time-dependent.

The BDs data are considered as the original time series $X=(x_1, x_2, ..., x_n)$, where $n$ is the length of the time series. GM (1,1) is the main and basic model of grey predictions, that is to say, a single variable first order grey model, which is able to acquire high prediction accuracy despite requiring small sample size (but the sample size must be at least 4). The GM (1,1) model is suitable for sequences that show an obvious exponential pattern and can be used to describe monotonic changes. According to the data characteristics, this study will use the GM (1,1) model for prediction.

The establishment for a GM(1,1) is derived as follows [15].

1. Let non-negative time sequence expressing $X = (x_0, x_1, x_2, ..., x_n)$ be an original time sequence. Where $n$ is the sample size of the data.

2. First-order accumulative generation operation (1-AGO) is used to convert $X_0$ into

$$X^1=(x^1(1), x^1(2), ..., x^1(n)) = (\sum_{t=0}^{n} x^1(t), \sum_{t=1}^{n} x^1(t), ..., \sum_{t=p}^{n} x^1(t)).$$

3. Let $Z=(z(1), z(2), ..., z(n))$ be the sequence generated from $X$ by adjacent neighbor means. That is, $z(t)=0.5(x^1(t)+x^1(t-1))$, $t=2,3, ..., n$. The least square estimate sequence of the grey difference equation of GM(1,1) is defined as $x^1(t)+a(z^1(t)-b)$, where $a$ and $b$ are referred to as the development coefficient and grey action quantity, respectively.
Then, \( [ab]^T = (B_1^T T_i)B_1^T Y_i \)

where

\[
Y_1 = \begin{bmatrix}
 x^2(2) \\
 x^2(3) \\
 x^2(n)
\end{bmatrix}
\]

\[
B_1 = \begin{bmatrix}
 -z^1(2) & 1 \\
 -z^1(3) & 1 \\
 -z^1(n) & 1
\end{bmatrix}
\]

(4) The whitenization equation given by \( dx/dt + ax = b \).

(5) The forecasting model can be obtained by solving the above equation, which is shown as follows:

\[
\hat{X} = \left[ x^2(1) \right] e^{at} + \frac{b}{a}
\]

(6) The predicted value of the primitive data at time point \( (t + 1) \) is extracted:

\[
\hat{X} = \left(1 - e^{-a} \right) \left( x^2(t) - \frac{b}{a} \right) e^{at}
\]

(7) Model accuracy test: The methods of testing include residual test, post-test difference test and small error probability. In this paper, the means of the posterior error ratio and the little probability of error are adopted.

First we calculate the mean square error \( S_0 \) of the original sequence \( x^0 \). It is defined as:

\[
S_0 = \frac{\sum_{n=1}^{n} \left( x(t) - \bar{x} \right)^2}{n-1}
\]

The means of the posterior error ratio: \( c = \frac{S_1}{S_0} \).

Little probability of error:

\[
P = \left\{ \left| e(t) - \hat{e}(t) \right| < 0.6745 \cdot S_0 \right\}
\]

Finally, according to the prediction accuracy classification table, see Table 1, the test results is the prediction accuracy of the model [16].

### Table 1: Prediction accuracy classification table.

| Prediction accuracy | The means of the posterior error ratio \( c \) | Little probability of error \( P \) |
|---------------------|---------------------------------|-----------------|
| Superior            | \(<0.35 \)                      | \( \geq 0.95 \) |
| Eligible            | \(<0.50 \)                      | \( \geq 0.80 \) |
| Reluctantly qualified | \(<0.65 \)                | \( \geq 0.70 \) |
| Failed              | \( >0.65 \)                     | \(< 0.70 \)     |

**Curve Fitting Method**

Curve fitting, also known as regression analysis, was used to find the “best fit” curve for a series of data points in this study. The curve fit often produces an equation that can be used to find points anywhere along the curve. Curve estimation were fitted to the data to quickly estimate regression statistics and produce related plots for 11 different models (linear, logarithmic, inverse, quadratic, cubic, power, compound, S-curve, logistic, growth and exponential models) [17]. F-test was used for selecting the best fitting curve for hypothesis testing. \( p \)-value <0.05 (two sided) was taken statistically significant. Similarly, \( R^2 \) value > 0.80 was taken significant, where \( R^2 \) is the correlation of the contribution of years (independent variables) in predicting BDs cases (dependent variables) [30]. In this investigation, \( x \) stands for the time (year) and \( y \) stands for the BDs cases.

**Exponential smoothing model**

For exponential smoothing method, also known as exponential weighted average method, the major advantage of this model is that it chooses the weighted average of each period as a decreasing exponential sequence, giving greater weight to the historical data closer to the forecast period. It might be suitable for the forecast of time-series data with trends and seasonality [18]. It’s consist of two models of seasonal and non-seasonal. The seasonal models mainly include simple seasonal, winter-addition, and winters multiplication model, and non-seasonal is made up primarily of holt-linear and brown-linear model. Holt-Winter’s method is the extended form of the simple exponential smoothing that include the seasonality in the approach. In this study, SPSS 20.0 statistical software was used to model the data selection model by “Expert Modeler”. In addition, Fitting indicators were analyzed and the optimal model in exponential smoothing model was selected to predict the BDs cases in 2019.

**ARIMA model**

Autoregressive Integrated Moving Average (ARIMA) models, which take into account changing trends, periodic changes, and random disturbances in time series, are very useful in modeling the temporal dependence structure of a time series [19]. In epidemiology, ARIMA models have been successfully applied to predict the incidence of multiple diseases, such as BDs [20]. The Box-Jenkins was used to approach to ARIMA (p, d, q) modeling of time series (Box & Jenkins, 2010). This ARIMA model-building process is adopted to take advantage of associations in the sequentially lagged relationships that usually exist in periodically collected data [21]. The parameters selected for fitting ARIMA model are as follows: p, autoregressive order; d, difference; q, moving average order. The order of Moving Average (MA) and Autoregressive (AR) terms in the ARIMA model is determined by using the Auto Correlation Function (ACF) graph and Partial Auto Correlation Function (PACF) graph. The parameters of the model were estimated by the conditional least squares method. Diagnostic checking including residual analysis and the Akaike Information Criterion (AIC) was used to compare the goodness-of-fit among ARIMA models. The Ljung-Box [22] test was used to measure the ACF of the residuals.

**Statistical Analysis**

SPSS 22.0 was used to analyze the incidence of BDs in Lanzhou from 2012 to 2018, and different models were established to compare the incidence of BDs in Lanzhou in 2019. Using Excel 2013 compiled function to calculate the indicators of dynamic sequence of BDs in Lanzhou from 2012 to 2018. The dynamic series indexes mainly include absolute growth (cumulative absolute growth(CAG) and annual absolute growth(AAG), development speed (base ratio and chain ratio), growth rate (base ratio and chain ratio), average development rate and average growth rate. \( CAG=\left( \sum_{t=1}^{n} x \right) - x_0 \) \( \text{AAG} = x_t - x_{t-1} \) base base
ratio development speed $= \frac{x_t}{x_0} \times 100\%$, chain ratio development speed $= x_t \times 100\%$, base ratio growth rate $= x_t \times 100\%$, and chain ratio development speed $= \frac{x_t}{x_0} \times 100\%$, where $x_t$ and $x_0$ denote observed and fitted values at time point $t$.

In addition, we use the Mean Absolute Percentage Error (MAPE), Error Square Sum (SSE) and the Average Absolute Relative Deviation (AARD) to evaluate the prediction accuracy of the four prediction models. MAPE $= \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_t - x_0}{x_t} \right| \times 100\%$, SSE $= \sum_{i=1}^{n} (x_t - x_0)^2$, AARD $= \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_t - x_0}{x_t} \right| \times 100\%$ where $x_t$ and $x_0$ denote observed and fitted values at time point $t$. A lower MAPE, SSE and AARD value indicates a better fit of the data. Finally, the fitted optimizing model was used for short-term forecasting of BDs cases for 2019 in Lanzhou. All analyses were performed using SPSS 22.0 with a significant level of $P < 0.05$.

**Results**

**Prevalence and trend of total BDs**

A total of 346,729 perinatal infants were monitored from 2012 to 2018. Of the 346,729 perinatal infants, 3653 had BDs, with an average prevalence of 104.13 per 10,000 perinatal infants (PIs). Trend analysis revealed that the annual prevalence rates of total BDs in the 7 years increased linearly ($x^2$ trend $=49.753$, $P<0.001$). The total prevalence rate increased by 7.74% annually from 2012 to 2018 on average (Table 2, Figure 1). In addition, the incidence of urban/rural BDs in Lanzhou is on the rise ($x^2$ urban trend $=11.993$, $P<0.001$; $x^2$ rural trend $=44.029$, $P<0.001$), and it’s higher in rural areas than in urban areas in 2018 (Table 2, Figure 2).

**Dynamic changes of the number of cases of BDs In Lanzhou:**

Table 3 is the dynamic change data of BDs in Lanzhou from 2012 to 2018. As shown in the table 3, the number of cases of BDs in Lanzhou is on the rise as a whole. Perinatal BDs average speed of development is 107.74%, and the Perinatal BDs average speed of increment is 7.74%. Assuming that the number of BDs in Lanzhou will develop at the rate of 107.74% in 2019, it is estimated that the number of BDs in Lanzhou will reach 737 cases in 2019.

| Year | Whole city | Urban | Rural |
|------|------------|-------|-------|
|      | Pls | BDs | Incidence | Pls | BDs | Incidence | Pls | BDs | Incidence |
| 2012 | 41084 | 406 | 98.82 | 19830 | 227 | 114.47 | 21254 | 179 | 84.22 |
| 2013 | 43339 | 303 | 69.91 | 20730 | 135 | 65.12 | 22609 | 168 | 74.31 |
| 2014 | 50056 | 503 | 100.49 | 26455 | 272 | 102.82 | 23601 | 231 | 97.88 |
| 2015 | 44102 | 489 | 110.88 | 22071 | 210 | 95.15 | 22031 | 279 | 126.64 |
| 2016 | 52923 | 601 | 113.56 | 29162 | 332 | 113.85 | 23761 | 269 | 113.21 |
| 2017 | 60602 | 667 | 110.06 | 33974 | 375 | 110.38 | 26628 | 292 | 109.66 |
| 2018 | 54623 | 684 | 125.22 | 31196 | 366 | 117.32 | 23427 | 318 | 135.74 |

$\chi^2$ 49.753 11.993 44.029

$P$ 0.000 0.000 0.000

**Figure 1:** Incidence of Perinatal BDs from 2012 to 2018 in Lanzhou.

**Figure 2:** Incidence of Perinatal birth defects from 2012 to 2018 in Urban/Rural areas of Lanzhou.
Table 3: Dynamic changes in the number of cases of Perinatal birth defects in Lanzhou from 2012 to 2018.

| Year | Birth defects | Absolute increment | Speed of development(%) | Increment speed(%) |
|------|---------------|--------------------|-------------------------|-------------------|
|      |               | CAG                | AAG                     |                   |
|      |               | Base Ratio         | Chain Ratio             | Base Ratio        | Chain Ratio |
| 2012 | 406           | -                  | -                       | 100.00            | 100.00     |
| 2013 | 303           | -103               | -103                    | 74.63             | 74.63      |
| 2014 | 503           | 97                 | 200                     | 123.89            | 166.01     |
| 2015 | 489           | 83                 | -14                     | 120.44            | 97.22      |
| 2016 | 601           | 195                | 112                     | 148.03            | 122.90     |
| 2017 | 667           | 261                | 66                      | 164.29            | 110.98     |
| 2018 | 684           | 278                | 17                      | 168.47            | 102.55     |

Perinatal birth defects average speed of development:
\[ \sqrt{\frac{a}{a_0}} = 107.74\% \]

Perinatal birth defects average speed of increment:
\[ \sqrt{\frac{a}{a_0}} - 100\% = 7.74\% \]

The occurrence and cis-position of all kinds of BDs in Lanzhou

The incidence and the sequence in the first five places were Congenital heart defects, Polydactylism, Cleft lip ± cleft palate, Limb shortening and Ankyloodactylia (Table 4). The most common BDs in Lanzhou from 2012 to 2018 are Cleft lip and palate and Congenital heart disease. The incidence of cleft lip and palate decreased from 19.23/10,000 in 2012 to 5.49/10,000 in 2018, and the incidence of Congenital heart disease increased from 16.06/10,000 in 2012 to 46.87/10,000 in 2018. Moreover, the incidence of Polydactylism increased from 9.98/10,000 in 2012 to 19.77/10,000 in 2018. The change of other defect types is not obvious. (Table 5).

Table 4: The occurrence and cis-position of all kinds of birth defects in Lanzhou from 2012 to 2018 (1/10,000).

| Cis-position | Birth Defects               | N     | Prevalence | Constituent ratio (%) |
|--------------|-----------------------------|-------|------------|-----------------------|
| 1            | Congenital heart defects    | 1153  | 33.25      | 33.40                 |
| 2            | Polydactylism               | 486   | 14.02      | 14.08                 |
| 3            | Cleft lip ± cleft palate    | 448   | 12.92      | 12.98                 |
| 4            | Limb shortening             | 226   | 6.52       | 6.55                  |
| 5            | Ankyloodactylia             | 195   | 5.62       | 5.65                  |
| 6            | Congenital hydrocephalus    | 142   | 4.10       | 4.11                  |
| 7            | Talipes equinovarus         | 125   | 3.61       | 3.62                  |
| 8            | Other malformations of the outer ear | 123 | 3.55 | 3.56 |
| 9            | Neural tube defect          | 110   | 3.17       | 3.19                  |
| 10           | Rectal atresia/stenosis     | 97    | 2.30       | 2.81                  |
| 11           | Hypospadias                 | 81    | 2.34       | 2.35                  |
| 12           | Esophageal atresia/stenosis | 74    | 2.13       | 2.14                  |

Estimation and forecasting

GM(1,1) Model

The results are as follows (Table 6). \( S_0 = 139.27, S_1 = 79.24 \), the means of the posterior error ratio \( c = 0.07 \); little probability of error \( P = 0.857 \). According to the judgment criteria of the GM(1,1) model prediction accuracy (Table 1), the posterior error ratio grade is reluctantly qualified, little probability of error \( P \) grade is qualified, and the accuracy grade is the lower level of the two indicators. It is considered that the fitting accuracy of the model is in a barely qualified \( y = 1808.06e^{-0.18t} - 1402.06 \) Curve fitting method

Table 7 is the model summary and hypothesis test results of 11 models of curve fitting. This study is mainly based on the statistical significance of model hypothesis test. The more \( R^2 \) and variance F are used as the optimal model criteria. A higher \( R^2 \) and lower F value are indicate a better model fit of the data. The fitting results show that the determination coefficient \( R^2 \) is 0.849 and F is 11.249 after quadratic function fitting, and the model has statistical significance. Therefore, the quadratic function model can better reflect the trend of BDs in Lanzhou from 2012 to 2018. The fitting model equation is \( y = 310.714 + 41.952t + 2.167t^2 \) (Figure 3).
### Table 5: Changes of birth defects in Lanzhou from 2012 to 2018 (1/10,000).

| Year | No. 1 | No. 2 | No. 3 | No. 4 | No. 5 |
|------|-------|-------|-------|-------|-------|
|      | Defects | Prevalence (%) | Defects | Prevalence (%) | Defects | Prevalence (%) | Defects | Prevalence (%) |
| 2012 | Cleft lip ± cleft palate | 19.23 | Congenital heart defects | 16.06 | Limb shortening | 10.47 | Polydactylism | 9.98 | Congenital hydrocephalus | 6.33 |
| 2013 | Cleft lip ± cleft palate | 16.38 | Congenital heart defects | 10.84 | Polydactylism | 5.77 | Limb shortening | 5.08 | Congenital hydrocephalus | 3.23 |
| 2014 | Congenital heart defects | 28.17 | Cleft lip ± cleft palate | 13.58 | Polydactylism | 10.99 | Limb shortening | 10.59 | Neural tube defect | 4.59 |
| 2015 | Congenital heart defects | 37.41 | Cleft lip ± cleft palate | 17.91 | Polydactylism | 13.38 | Limb shortening | 6.35 | Ankylopalypia | 5.22 |
| 2016 | Congenital heart defects | 41.95 | Polydactylism | 16.63 | Cleft lip ± cleft palate | 10.58 | Ankylopalypia | 7.37 | Congenital hydrocephalus | 4.35 |
| 2017 | Congenital heart defects | 42.24 | Polydactylism | 18.15 | Cleft lip ± cleft palate | 10.73 | Esophageal atresia/stenosis | 5.12 | Other malformations of the outer ear | 4.62 |
| 2018 | Congenital heart defects | 46.87 | Polydactylism | 19.77 | Ankylopalypia | 9.89 | Limb shortening | 7.14 | Cleft lip ± cleft palate | 5.49 |

### Table 6: Calculation Process of GM (1,1) Model for the Number of Birth Defects in Lanzhou.

| Year | t | Xt | Yt | Zt | Zt² | XtZt | Yt⁻ | Xt⁻ | e | e% |
|------|---|----|----|----|-----|------|-----|-----|---|---|
| 2012 | 1 | 406 | 406 | - | - | - | 406.0 | 406.0 | 0.0000 | 0.00 |
| 2013 | 2 | 303 | 709 | 557.5 | 310806.3 | 214827 | 761.4 | 355.4 | -0.1730 | -17.30 |
| 2014 | 3 | 503 | 1212 | 960.5 | 922560.3 | 609636 | 780.2 | 424.7 | 0.1556 | 15.56 |
| 2015 | 4 | 489 | 1701 | 1456.5 | 2121392.3 | 831789 | 932.3 | 507.6 | -0.0380 | -3.80 |
| 2016 | 5 | 601 | 2302 | 2001.5 | 4006002.3 | 1383502 | 1114.1 | 606.6 | -0.0092 | -0.92 |
| 2017 | 6 | 667 | 2969 | 2635.5 | 6945860.3 | 1980323 | 1331.4 | 724.8 | -0.0867 | -8.67 |
| 2018 | 7 | 684 | 3653 | 3311.0 | 10962721.0 | 2498652 | 1591.0 | 866.2 | -0.2664 | -26.64 |
Table 7: The Main parameters of birth defects curve fitting.

| Curve fitting method | Model statistics | Parameter value |
|-----------------------|------------------|-----------------|
|                       | R^2       | F        | P      | Constant | b_1 | b_2 | b_3 |
| Linear                | 0.846    | 27.394  | 0.003  | 284.714  | 59.286 | -  | -  |
| Logarithmic           | 0.718    | 12.748  | 0.016  | 311.947  | 172.357 | -  | -  |
| Inverse               | 0.486    | 4.729   | 0.082  | 640.496  | -320.293 | -  | -  |
| Quadratic             | 0.849    | 11.249  | 0.023  | 310.714  | 41.952  | 2.167 | -  |
| Cubic                 | 0.898    | 8.759   | 0.054  | 494.714  | -167.603 | 63.500 | -5.111 |
| Compound              | 0.773    | 17.055  | 0.009  | 313.807  | 1.126   | -  | -  |
| Power                 | 0.667    | 10.003  | 0.025  | 330.332  | 0.347   | -  | -  |
| S-curve               | 0.449    | 4.067   | 0.100  | 6.462    | -0.644  | -  | -  |
| Growth                | 0.773    | 17.055  | 0.009  | 5.749    | 0.119   | -  | -  |
| Exponential           | 0.773    | 17.055  | 0.009  | 313.807  | 0.119   | -  | -  |
| Logistic              | 0.773    | 17.055  | 0.009  | 0.003    | 0.888   | -  | -  |

The expert modeler is used to fit the model, and the exponential smoothing model is used to construct the model. The optimal model selection, parameter estimation and prediction are carried out. The statistical results of each model are shown in Table 8. By comparing the statistical results of three seasonal fitting methods of exponential smoothing model, and combining with the selection principle of evaluation index, the winters additive model of exponential smoothing model is finally determined as the best model. The stable R^2, R^2, RMSE and MAPE of the model are the best values in the three models, and the Ljung-Box Q statistic of the model is 0.086, with no significant difference (P>0.05), indicating that the residual is a white noise sequence, suggesting that the Winters additive model has better fitting effect (Figure 5).

ARIMA model

The ARIMA model is suitable for BDs from 2012 to 2018 and is tested by predicting the cases of BDs in 2019. Different ARIMA models were tested to determine the best fitting model. Based on the distribution characteristics, we conducted six models, ARIMA(1,0,1)(1,0,1)_{12}, ARIMA (1,0,1) (1,1,1)_{12}, ARIMA (2,0,1) (1,0,1)_{12}, ARIMA(1,0,1) (2,0,1)_{12}, ARIMA(2,0,1) (2,0,1)_{12}, and ARIMA(2,0,1)(2,1,1)_{12}. Of all the models tested, the ARIMA(2,0,1)(2,0,1)_{12} model was the best fit for the data (Table 9).

Comparison of prediction effect of model:

According to the size of MAE, SSE and AARD of GM(1,1), Curve fitting, Exponential smoothing model and ARIMA, it can be concluded that Winters addition method in exponential smoothing model can better fit the number of BDs in Lanzhou (Table 10). The predicted results of exponential smoothing model indicate that BDs in Lanzhou in 2019 are lower than that in 2018 (Figure 6).
Table 8: Statistical analysis of different model fitting results.

| Model                  | Stable R² | R²     | RMSE | MAPE | Ljung-Box Q | P  |
|------------------------|-----------|--------|------|------|-------------|----|
| Winters additive model | 0.650     | 0.697  | 7.566| 15.616| 22.895      | 0.086|
| Winters multiplication model | 0.538  | 0.614  | 8.544| 17.311| 25.849      | 0.040|
| Simple seasonal model  | 0.643     | 0.689  | 7.624| 15.177| 22.184      | 0.137|

Table 9: Comparisons of tested models.

| Model                  | Stable R² | R²     | RMSE | MAPE | Ljung-Box Q | P  |
|------------------------|-----------|--------|------|------|-------------|----|
| ARIMA(1,0,1)(1,0,1)_{12}  | 0.586     | 0.586  | 8.792| 17.987| 16.625      | 0.677|
| ARIMA(1,0,1)(1,1,1)_{12}  | 0.447     | 0.497  | 9.478| 19.198| 19.261      | 0.155|
| ARIMA(2,0,1)(1,0,1)_{12}  | 0.615     | 0.615  | 8.696| 17.642| 18.798      | 0.129|
| ARIMA(1,0,1)(2,0,1)_{12}  | 0.611     | 0.611  | 8.732| 17.855| 20.973      | 0.073|
| ARIMA(2,0,1)(2,0,1)_{12}  | 0.621     | 0.621  | 8.684| 17.361| 19.558      | 0.076|
| ARIMA(2,0,1)(2,1,1)_{12}  | 0.450     | 0.501  | 9.592| 19.038| 18.933      | 0.095|

Table 10: Comparing results of model fitting accuracy.

| Model                  | SSE      | MAPE  | AARD  |
|------------------------|----------|-------|-------|
| GM(1,1)                | 45673    | 10.41%| 10.81%|
| Curve fitting          | 17520    | 10.31%| 8.26% |
| Exponential smoothing   | 10400    | 5.75% | 5.97% |
| ARIMA(2,0,1)(2,0,1)_{12} | 12721   | 7.36% | 8.07% |

Figure 5: Prediction of the number of birth defects in Lanzhou.

Discussion

BDs are a global problem, but their impact on infant and childhood death and disability is particularly severe in low and middle income countries. Serious BDs can be lethal. For those who survive early childhood with BDs, they can experience lifelong mental, physical, auditory, and visual disabilities that exact harsh human and economic tolls on them, their families, and their communities. Up to 70% of BDs could either be prevented, or with proper care, cured or ameliorated [23,24]. As the largest developing country in the world, a hospital-based surveillance system for monitoring BDs was established in China in 1986 [25]. Many hospitals and institutions have generated a large number of birth defect files and publications. There are large regional differences in birth defect prevalence. Our study performed a comprehensive analysis of differences in the perinatal infant prevalence of BDs in Lanzhou, China over a 7 year time period. The average prevalence rate of total perinatal infants’ BDs was 104.13 per 10000 PIs in the past 7 years in Lanzhou, with an upward trend increasing by 7.74% on average each year. From 2012 to 2018, the average incidence of BDs in urban and rural areas of Lanzhou was 105.61/10000 and 106.48/10000, respectively, and the overall trend was upward. The upward trend in BDs epidemiology is potentially explained by several factors. Firstly, the liberalization of the second-child policy, the increase in the proportion of older pregnant women in Lanzhou, resulting in an increase in the incidence of BDs related; Secondly, Lanzhou has implemented various maternal subsidies projects and improved prenatal screening and prenatal diagnosis technology in grass-roots hospitals, making the detection of perinatal BDs increased.
China is a country with high incidence of BDs. The incidence of BDs is increasing year by year. In 2010, the incidence rate is 149.9/10000, rising to 157.0/10000 in 2014 and 175.5/10000 in 2016 [26]. The average incidence of BDs in Lanzhou (104.68/10000) is lower than the national level. Considering the reasons, it may be that with the promulgation of the National Comprehensive Prevention and Control Program of BDs [27], the grass-roots maternal and child health institutions in Lanzhou can actively respond to the policy call, conscientiously implement the prevention and control program of BDs, take certain measures in the prevention and control of BDs, and reduce the incidence of BDs.

From 2012 to 2018, the occurrence categories of BDs in Lanzhou were Congenital heart disease, Polydactylyism, Cleft lip ± cleft palate, Limb shortening and Ankylopectylyia. According to WHO reports, the most common serious congenital anomalies in the world are Congenital heart disease, Neural tube defects and Down’s syndrome (anomalies). The most common types in the United States are Down’s syndrome, Cleft lip, and Congenital heart disease [28]. Muscle and urinary system defects are the most common in Russia [29]. By above knowable, Congenital heart disease is a BD worthy of attention at present, which is in a high incidence in Lanzhou. The results of cis-positions of BDs from 2012 to 2018 show that before 2015, BDs were mainly Congenital heart disease and Cleft lip and palate. After 2015, Congenital heart disease and Polydactylyism were the main types of BDs. In recent years, the incidence of Congenital heart disease has increased, on the one hand, due to the development of medical and health level, the continuous improvement of prenatal screening and diagnosis technology, such as the popularization and wide application of four-dimensional color Doppler ultrasound, the detection rate of Congenital heart disease has increased. On the other hand, it is not excluded that the incidence of Congenital heart disease increases due to the increase of risk factors related to Congenital heart disease. These results suggest that the prevention and treatment of Congenital heart disease is not optimistic.

The early recognition of distributions is significantly important for BDs disease control and prevention. The effectiveness of statistical models in forecasting future BDs incidence has been proved useful [30]. The surveillance system is a good way to collect and analyze BDs disease data. With high quality surveillance data, BDs disease may be accurately detected and forecasted. The selections of the BDs forecasting techniques is very important. In the present study, we conducted a comparative study of four typical forecasting methods in the forecasting of the BDs diseases, namely four methods GM(1,1), Curve fitting, exponential smoothing and ARIMA model methods. We have also compared the differences among these methods in both observed value and predicted value. By comparing the results of model evaluation index (SSE, MAPE, AARD), it is found that the order of the four models is: exponential smoothing, ARIMA, Curve fitting and GM(1,1).

The four models selected in this study have different prediction theories with different prediction ideas. The GM(1,1) prediction model needs less data and is suitable for the data with good smoothness of the original data. Moreover, the prediction is more accurate, the accuracy is higher and the calculation is simple. However, because the GM(1,1) model relies on historical data and does not take into account the relationship between various factors, the medium-term and long-term prediction error is large [31]. The Curve fitting model is easy to calculate and can save time for short-term prediction, but it is not suitable for medium-term and long-term prediction because of large errors in data prediction with large fluctuations [32]. Exponential smoothing model has been widely used in the prediction of many diseases. Its main idea is to obtain the predicted value by weighting the actual data. According to the characteristics of the near-time data, the predicted value is given more weight, while the far-time data is given less weight, which makes the predicted value less affected by the long-term data, and more affected by the near-term data [33]. ARIMA model is more suitable for time-varying series. It uses the characteristics of data stationarity to predict the future. It is suitable for short-term and medium-term prediction. It has poor long-term prediction effect [34,35]. Moreover, the sequence must satisfy stationarity and non-white noise. Otherwise, it is not appropriate to establish ARIMA model.

This study chooses the best model from the above models to forecast the BDs in Lanzhou in 2019. The prediction results show that the occurrence of BDs in Lanzhou in 2019 is close to that in 2018. It suggests that strengthening the prevention and control of BDs is an important public health problem in Lanzhou in the future years.

Conclusion

In summary, the occurrence of BDs in Lanzhou has shown an upward trend in recent years, and the types of defects are Congenital heart disease and Polydactylyism. In addition, the predicted results show that the number of BDs in 2019 is close to that in 2018. Therefore, in the future, Lanzhou should increase the prevention and control of BDs.

Funding information

The Science and Technology Planning Project of Lanzhou. Project number: 2018-3-8.

Data availability statement

Data available on request from the authors. The data used in this study come from the Maternal and Child Health Hospital of Lanzhou, which does not support sharing because of the hospital itself.

Reference

1. Jenkins KJ, Correa A, Feinstein JA, Botto L, Britt AE, et al. Noninherited risk factors and congenital cardiovascular defects: current knowledge: a scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. Circulation. 2007; 115: 2995-3014.
2. Kirby RS. The prevalence of selected major birth defects in the United States. SeminPerinatol. 2017; 41: 338-344.
3. Carmichael SL. Birth defects epidemiology. Eur J Med Genet. 2014; 57: 355-358.
4. Anderka M, Mitchell AA, Louik C, Werler MM, Hernández-Diaz S, et al. Medications used to treat nausea and vomiting of pregnancy and the risk of selected birth defects. Birth Defects Res A Clin Mol Teratol. 2012; 94: 22-30.
5. Pei L, Kang Y, Cheng Y, Yan H. The Association of Maternal Lifestyle with Birth Defects in Shaanxi Province, Northwest China. PLoS One. 2015; 10: e0139452.
6. Organization WH. The global burden of disease: 2004 update. 2008.
7. Castillo Taucher S. Servicios para la atención y la prevención de defectos congénitos: Síntesis de una reunión de la Organización Mundial de la Salud y la Fundación March of Dimes [Services for the care and prevention of birth defects. Reduced report of a World Health Organization and March of Dimes Foundation meeting]. Rev Med Chil. 2007; 135: 806-813.

8. Zeng L, Cheng Y, Dang S, Yan H, Dibley MJ. Impact of micronutrient supplementation during pregnancy on birth weight, duration of gestation, and perinatal mortality in rural western China: double blind cluster randomised controlled trial. Paper presented at the China academic Congress on Health Statistics. 2007.

9. Desrosiers TA, Herring AH, Shapira SK, Hooiveld M, Luben TJ, et al. Paternal occupation and birth defects: findings from the National Birth Defects Prevention Study. Occup Environ Med. 2012; 69(8): 534-542.

10. Dai L, Zhou GX, Zhu J, Miao L, Wang YP, et al. Impacts of birth defects on perinatal deaths in Chinese population. Chinese Journal of Epidemiology. 2004; 25: 138-141.

11. Wu J, Lu AD, Zhang LP, Zuo YX, Jia YP. Study of Clinical Outcome and Prognosis in Pediatric Core Binding-Factor Acute Myeloid Leukemia. Zhonghua Xue Ye Xue Za Zhi. 2019; 40: 52-57.

12. Information LCB. Lanzhou Municipal People’s Government. Lanzhou City Basic Information. 2010.

13. Liu S, Lin Y. An Introduction to Grey Systems Theory. Understanding Complex Systems. 2011; 68.

14. Lun Li. Mortality prediction based on grey forecasting model for cardio-cerebrovascular disease. Modern Electronics Technique. 2015; 38(442): 107-111.

15. Gan R, Chen X, Yan Y, Huang D. Application of a hybrid method combining grey model and back propagation artificial neural networks to forecast hepatitis B in China. Comput Math Methods Med. 2015; 2015: 328273.

16. Zhao J, Guo XL, Wu BY, Wang S, Wang ZX. The application of GM (1,1) grey forecasting model and ARIMA model in the fitting of the mortality of the cardiovascular disease in Shandong province. 2016; 43: 1732-1749.

17. Tang WR, Fang JY, Wu KS, Shi XJ, Luo JY, et al. Epidemiological characteristics and prediction of esophageal cancer mortality in China from 1991 to 2012. Asian Pac J Cancer Prev. 2014; 15: 6929-6934.

18. Guan P, Wu W, Huang D. Trends of reported human brucellosis cases in mainland China from 2007 to 2017: an exponential smoothing time series analysis. Environ Health Prev Med. 2018; 23: 23.

19. Liu Q, Liu X, Jiang B, Yang W. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. BMC Infect Dis. 2011; 11: 218.

20. Yu M, Liu CY, Zhang SP, Yang HJ, Liu F. Mathematical model for predicting birth defect trend in Xi’an. Journal of Xi’an Jiaotong University (Medical Sciences). 2019; 40: 145-150.

21. Akhtar S, Rozi S. An autoregressive integrated moving average model for short-term prediction of hepatitis C virus seropositivity among male volunteer blood donors in Karachi, Pakistan. World J Gastroenterol. 2009; 15: 1607-1612.

22. Box GEP, Jenkins GM, Reinsel GC. Time series analysis: Forecasting and control. Rev. ed[J]. Journal of Marketing Research, 1994; 14: 238-242.