Defining the reference range of 8-iso-PG $F_{2\alpha}$ for pregnancy women

Essay in Statistics

Department of Economics and Society, Dalarna University

Author Hao Luo

Supervisor Johan Bring

Date **June, 2007**

Table of Contents

Summary

In recent years, reference ranges are routinely used in clinical research as a screening tool to detect the abnormality. In order to test the value of a chemistry element named 8-iso-prostaglandin $F_{2\alpha}$ which is supposed to show some clinical clues for the etiology of pre-eclampsia in human pregnancy, the homogeneous reference ranges as well as the gestational-age-specific reference ranges will be estimated. During such calculation procedure, both non-parametric method and parametric method will be used. For the non-parametric method, the empirical reference ranges has been estimated for different weeks independently. The linear regression method and the LOESS smoothing method are used when taking the gestational age effect into consideration. For the parametric method, the Box-Cox transformation method towards normality is applied and the LMS method which built up upon this transformation method is introduced as well to get the final estimates. A series of comparison of the estimation results will be given among those methods. And a relatively better method will be chosen.

Keywords: Reference range; 8-iso-PG $F_{2\alpha}$; Homogeneous reference range; Gestational-age-specific reference range; Empirical quantiles; Box-Cox transformation towards normality; LOESS smooth; LMS method

1. Introduction

Reference range is a commonly used tool in detecting the abnormal individuals in many clinical researches or diagnoses. By definition, a univariate *p%* reference range is a pair of numbers (the *reference limits*) that enclose the central *p%* of a sample of observations (the sets of *reference values*) obtained from a specified group of individuals (the *reference subjects*). Thus *(100-p)/2%* of the values lie below the lower limit and the same proportion above the upper limit. By convention, the reference range for whatever is set to cover ninety-five percent (*95%*) of all values from the 'normal', 'healthy' population. Five percent (*5%*) of results consequently fall outside the reference range. The main task of this essay is to estimate the 2.5^{th} and 97.5^{th} percentile of the values of isoprostanes in each gestational age for human uncomplicated pregnancy by each week and each two week. Furthermore, taking the impact of gestational age (GA) into consideration, three smoothing methods will be applied based on the reference limits calculated by each week and each two week respectively.

Many studies have been done on the etiology of pre-eclampsia focusing on the role of oxidative stress in recent years. However, the involvement of oxidative stress in uncomplicated normal human pregnancy has not been taken into some in-depth research. In order to evaluate the role of oxidative stress in normal pregnancy, an index named 8-iso-prostaglandin $F_{2\alpha}$ (8-iso-PG $F_{2\alpha}$) belonging to the family of isoprostanes will be collected from human uncomplicated pregnancies throughout the gestational period by urine samples. By defining the reference ranges of 8-iso-PG $F_{2\alpha}$, women of high values in such index can be identified.

For this specific issue, the reference population mentioned before indicates the 37 women who had pregnancies classified as uncomplicated.

It is common for the distribution of 8-iso-PG $F_{2\alpha}$ to be affected by characteristics such as weight, height and genetic factors. However, due to the limitation of the data we gathered and the difficulty to construct reference intervals for so many subgroups, it is clearly impractical. In addition, another important way to define the normal range should be taken the biologically motivated limits into consideration. One of the good examples to demonstrate such issue is the battle waged in the endocrinology community regarding the so-called "reference range" for the TSH test in recent years. This debate is focusing on whether the reliance on the TSH test-to the exclusion of symptoms and other tests-is medically sound. Unfortunately, although we can predict that defining the reference range only through the firstly cited 95% theory may not be accurate enough, limited by the relating data, we have to make a compromise for such method.

Section 2 briefly introduces the data set and some details regarding how the pre-treatment of the data set is conducted. Section 3 gives a brief description of the methods which have been used in the analysis procedure such as the linear regression method, the LOESS smoothing method and the LMS method used in calculating the gestational-age-specific reference range. Results of the estimated calculated by different methods will be presented in Section 4 and four comparisons are made to give a general overview of the differences between methods. Limitations of this essay are talked about in Section 5 as discussion.

2. Data

2.1 Data set

Thirty-seven pregnant women from one outpatient antenatal clinic in Uppsala City were consecutively recruited into the study during 2003-2004. The variables which have been measured relating to the estimate of the reference intervals are patient number, the date on which the urinary sample have been taken, the value of 8-iso-PG $F_{2\alpha}$, the length of pregnancy when the urinary sample were taken measured by day, week, and each two week respectively. The full day as well as the full week of pregnancy is also included. Table 1 gives an example of the dataset of patient number 100.

Patient No.	date	ISOPgf2	preglength	week	twoweek	Fday	Fweek
100	19-Feb-2004	0.26	79.00	12.00	12.00	284.00	41.00
100	19-Apr-2004	0.35	139.00	20.00	20.00	284.00	41.00
100	1-Jun-2004	0.35	182.00	26.00	26.00	284.00	41.00
100	5-Jul-2004	0.50	216.00	31.00	30.00	284.00	41.00
100	3-Aug-2004	0.38	245.00	35.00	34.00	284.00	41.00
100	9-Sep-2004	0.87	282.00	41.00	40.00	284.00	41.00

Tab. 1 The data set of patient number 100

2.2 Some pre-treatment of the data

Since the samples are gathered from urine, it is quite possible to be gathered every day especially for the days that are close to the expected date of childbirth, such as in week 39, 40 and 41. In such weeks, many patients have provided the urine sample every day, but only one of those 7 values should be selected to represent the value of the whole week. These values were chosen to guarantee that it has an exact 7 days interval both to its pre-week record and the week-after record. Similarly, in order to make a more reasonable estimation, an independent identically distributed sample should be ensured when calculating the reference intervals by each two week. If one woman had more than one sample taken within a two week interval, the values of even number weeks will be inducted since using the mean of the two values will unscientifically reduce the variance of that week. The scatter plots of the values of 8-iso-PG $F_{2\alpha}$ against week and every two week intervals are shown in Fig. 1(a) and 1(b) respectively. And box plots are also shown in Fig.2(a) and 2(b) to give a general overview of the skewness of each sample.

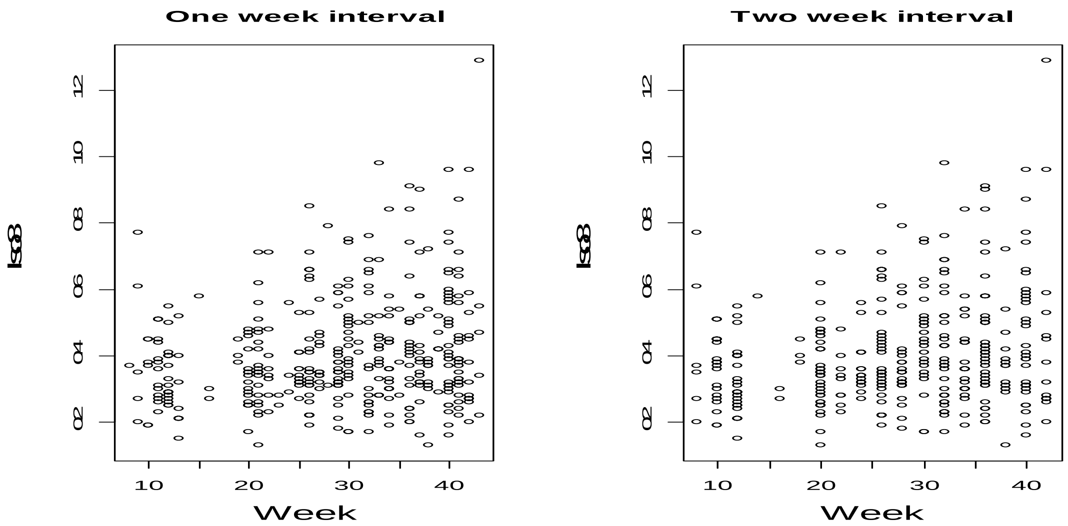


Fig. 1(a) Fig. 1(b)

Fig.1. Scatter plots of ISOPgf2 versus gestational age calculated by one and two week interval respectively from the data of 37 women

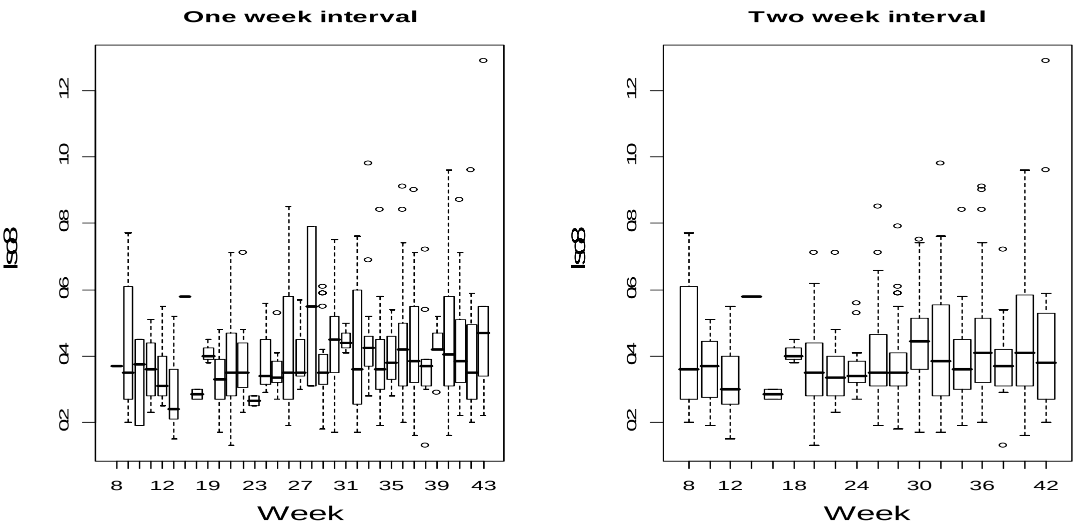


Fig. 2(a) Fig. 2(b)

Fig.2. Box plots of ISOPgf2 versus gestational age calculated by one and two week interval respectively from the data of 37 women

Lots of extreme values have been detected from Figure 2 especially in the last few weeks. However, due to our limited small sample size, any value of the observations may have a big influence on the final estimation result. The extreme values can not be removed in a hurry without taking the in-depth analysis.

To give a general idea of the change for the values of 8-iso-PG $F_{2\alpha}$ for each individual, the values' movement of patient number 100, 101, 104, 105 and 106 are graphed as an example in Figure 3. Figure 3 indicates that there is no clear trend exists for any of those patients, and the values of 8-iso-PG $F_{2\alpha}$ fluctuate intensively for different weeks.

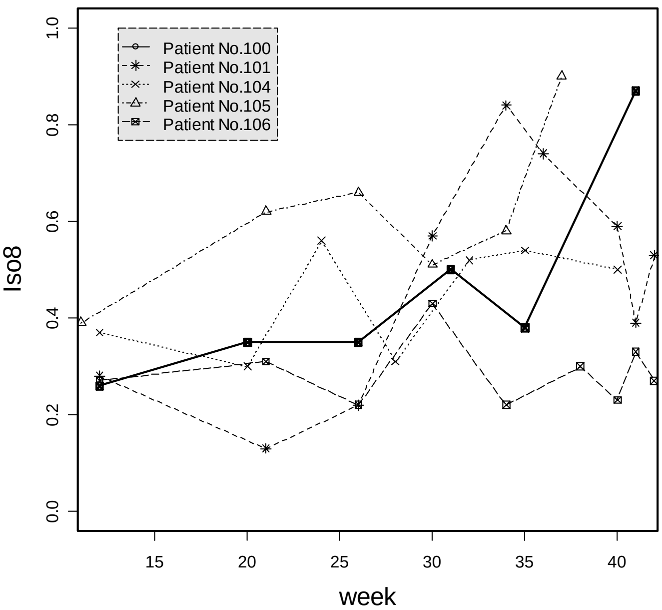


Fig. 3 The changing tendency of the 8-iso-PG $F_{2\alpha}$ values according to different weeks of 5 patients

3. Method

3.1 Method for estimating reference range without gestational age effect

Before going to more complex cases, methods for estimating reference ranges for samples with no gestational age effect are viewed first. This kind of reference range has a formal name-homogeneous reference range-which indicating that the gestational age effect has not been taken into consideration. Generally, there are two ways for estimating the homogeneous reference range: non-parametric method and parametric method. The empirical quantiles which estimated by non-parametric method is easy to be handle and will be introduced first.

3.1.1 Method of estimating the empirical quantiles (Method 1)

Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed continuous random variables with cumulative distribution function F and order statistics $Y_{(1)}<Y_{(2)}< \ldots < Y_{(n)}$. Let p $(0 \le p \le 1)$ be a position on the distribution function of Y and C the corresponding quantile, such that $F(C)=p$. For our case, p equals to 0.025 and 0.975 respectively.

Since $C=F^{-1}(p)$, requires inversion of the distribution function. The simplest estimator of F is the empirical distribution, defined for the ordered sample of observations $Y_{(1)}<Y_{(2)}< \ldots < Y_{(n)}$ by $\widehat{F}(Y_{(i)}) = i/n$, $i=1, 2, \ldots, n$. The $100p$th centile may then be estimated by $\widehat{C} = \widehat{F^{-1}}(p) = Y_{([np])}$ where $[np]$ denotes the nearest integer. When np is not an integer, interpolation between the nearest two order statistics should be used. In R, the one we used here is based on a sum polygon where the ith ranking observation is the $(i-1)/(n-1)$ quantile and intermediate quantiles are obtained by linear interpolation defined as:

$$p(i) = (i-1)/(n-1), \text{ where } p(i) = \text{ model } [F(x_{(i)})] \quad \textbf{(1)}$$

Non-parametric estimation of the reference range is distribution-free and therefore robust, but is usually not efficient enough comparing to the parametric method especially when a suitable distribution is found. According to <Encyclopedia of Biostatistics> volume 4[2], assuming large samples (say, $n \geqslant 100$), that variance of each limit on the normal scale is given approximately by $(3/n)\sigma^2$. For comparison, if these limits had been estimated by the corresponding non-parametric method, for $p=0.025$ or 0.975, there large sample variances would be approximately $Q_p=(7.13/n)\sigma^2$. In other words, on the normal scale, the variance of the quantile estimator of a 95% normal limit would be expected to be 2.44 times larger than that of the normal estimator. As a result, finding a suitable parametric distribution to estimate the reference range is indispensable. And a comparison between those two results will be presented in the result section.

3.1.2 Method of estimate the reference range based on normal distribution (Method 2)

Although it is quite possible that other families of continuous distribution are potentially useful, only the normal distribution will be discussed here for its mathematical convenience and widespread usefulness. In the normal case, the percentile limits of the range are estimated from the sample **mean** and standard deviation **s**.

Define the Z-score for Y_i as $Z_i=\Phi^{-1}\{F(Y_i)\}$ where Φ^{-1} is the inverse normal distribution function. When $F=\Phi$, Y_i is normally distributed and the ith Z-score is simply $(Y_i - \mu)/\sigma$ For most cases, transformation is often needed because the distribution of reference values is skewed.

Basically, there are four most commonly used ways of transformation towards normality:

- Logarithmic transformation: Logarithmic transformation of positive variables may reduce positive skewness and heteroscedasticity, but it may still not be sufficiently normal for accurate inference in the tails of distribution.
- Box-Cox (power) transformation: $(Y^{\lambda}-1)/\lambda$
- Scaled exponential transformation: $[\exp(\lambda Y)-1]/\lambda$ (closely related to Box-Cox)
- Origin-shifted logarithm transformation: $\log(Y-\lambda)$. Estimation of λ by maximum likelihood may present computational difficulties for it is a 'non-regular problem'.

Comparing the pros and cons of those for transformation method, the Box-Cox transformation will be applied. For this method of transformation, $g(Y) = \ln(Y)$ when $\lambda=0$; the distribution of Y is negatively skewed for $\lambda>1$; normal for $\lambda=1$; and positively skewed for $\lambda<1$. Estimates of λ will be obtained by maximum likelihood method.

To test whether the reference population belongs to a normal distribution, the Shapiro-Wilk Normality Test is introduced. The Shapiro-Wilk test calculates a W statistic that tests whether a random sample comes from a normal distribution. Small values of W are evidence of departure from normality. The W statistics is calculated as follows:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \textbf{(2)}$$

where the $x_{(i)}$ are the ordered sample values ($x_{(1)}$ is the smallest) and the $a_{(i)}$ are constants generated from the means, variances and covariances of the order statistics of a sample of size n from a normal distribution

The Shapiro-Wilk test will be taken to each week and each two week to test the distribution of different samples. For the samples which passed the test, a comparison of the W statistics between the original normal population and the transformed normal population will be given. Larger value of W statistics and *p-Value* indicates a better fit towards normality is preferred and will be chosen to estimate the reference limits of each gestational age. For those who haven't pass the normality test, the transformation method will be used, and the reference range will be obtained by back transformed those estimates.

3.2 Modeling gestational age effect

Methods for incorporating gestational age effect into the estimation of reference ranges are now discussed. The basic tasks here are to obtain graphs, tables and functions representing the required reference limits. All the notation defined in Section 3.1 will still be applied here, $f(Y_i, T_i)$ except the measurements of interest will be written as pairs of variables $\{(Y_i, T_i)\}_{i=1}^{n}$, where T_i represents gestational age.

3.2.1 Curves Based on Empirical Estimates using linear regression method (Method 3)

The method discussed in this section, in fact, is an extension of the method discussed in section 3.1.1. When it comes to smoothing or regression method, the first thing that comes to mind might be linear regression or polynomial regression method. As a result, these regression methods will be tried first based on the empirical estimates.

In this adjusting procedure, a significance test will be taken first by assuming a linear regression model first. The model is given by

$$y_i = \alpha + \beta T_i + \varepsilon_i \qquad (3)$$

where y stands for the value of 8-iso-PG $F_{2\alpha}$, and T represents the week number. The significance test is done by checking the p value of the t-Test for the coefficient β. If β is significant, some higher order factor will be introduced step by step as polynomial regression which defined as:

$$y_i = \alpha + \beta_1 T_i + \beta_2 T_i^2 + \cdots + \beta_k T_i^k + \varepsilon_i \qquad (4)$$

The introducing procedure will be stopped when the new added factor is demonstrated as insignificant. The estimates of the reference limits are computed by the fitted value of the best fit model. On the opposite, if β is shown insignificant at the first step, the reference limits should be estimated as a constant reference line: the mean value.

3.2.2 Curves Based on Empirical Estimates using LOESS smoothing method (Method 4)

A big obstacle for linear regression in this specific issue is that the construction of regression function should build on a significant relationship between gestational age and the value of 8-iso-PG $F_{2\alpha}$. Such prerequisite can not be guaranteed especially for the 2.5^{th} percentile and the mean value. As a result, another smoothing method, the LOESS procedure, will be introduced.

LOESS is one of many "modern" modeling methods that build on "classical" method, such as linear and nonlinear least squares regression. By combining the simplicity of linear least squares regression with the flexibility of nonlinear regression, it fits simple models to localized subsets of the data to build up a function that describes the deterministic part of the variation in the data, point by point. The attraction of such method is only to fit segments of the data instead of to specify a global function. It is particularly useful when a theoretical function $f(Y_i, T_i)$ is complicated or do not exist. Furthermore, it is a technique which is an improvement over least squares smoothing when the data is not equally spaced. And this characteristic fits our data very well for there is a lack of research record for earlier gestational age.

LOESS is more descriptively known as "locally weighted polynomial regression". The basic idea of the LOESS method applied in this specific empirical reference ranges smoothing procedure is briefly discussed below:

Take percentile 2.5% for example, at each point in this data set, a low-degree polynomial is fit to a subset of the data with the values of variable representing gestational age near the point whose response is being estimated. The polynomial is fit using weighted least squares, giving more weight to points near the point whose response is being estimated and less weight to points further away. The value of the regression function for the point is then obtained by evaluating the local polynomial using the gestational age variable values for the data point. The LOESS fit procedure is complete after each regression function value of the ***n*** data points has been computed. In such procedure, the only two parameters that we should define first are the "*smoothing parameter*" and "*the degree of local polynomials*".

The choices of those two parameters are flexible and subjective. The range of choices for such parameters will be discussed next.

▪ Choice of "*smoothing parameter*":

The smoothing parameter ***q*** or "bandwidth" determines how much of the data is used to fit each local polynomial. It is a number between ***(d+1)/n*** and ***1***, with ***d*** denoting the degree of the local polynomial. The value of ***q*** indicates the proportion of data used in each fit. Large values of ***q*** produce the smoother functions that wiggle less in response to fluctuations in the data. The smaller ***q*** is, the closer the regression function will conform to the data. In most LOESS applications, the smoothing parameters typically lie in the range 0.25 to 0.5.

However, since our gestational age is recorded by week and the sample size is very small, larger value of the smoothing parameter is acceptable. As a result, ***q=0.5*** and ***q=0.75*** will be applied in estimation to give a comparison in our data set.

▪ Choice of "*degree of local polynomial*"
Generally, the local polynomials fit to each subset of the data are almost always of first or second degree which stands for either locally linear (in the straight line sense) or locally quadratic. Higher-degree polynomials are theoretically workable, but the over-fitted and unstable models are not yield in the spirit of LOESS. Under the spirit that any function can be well approximated in a small neighborhood by a low-order polynomial and that simple models can be fit to data easily, ***d=1*** and ***d=2*** will be applied in our data during the LOESS smoothing procedure.

▪ Weight function
As LOESS gives different weights for different points, there should be a function defining such weights. Traditionally, the weight function used for LOESS is the tri-cube weight function,

$$w(\alpha) = \begin{cases} (1-|\alpha|^3)^3 & \text{for } |\alpha| < 1 \\ 0 & \text{for } |\alpha| \geq 1 \end{cases} \qquad (5)$$

For the fit at point x, the fit is made using points in a neighborhood of x, weighted by their distance from x. The size of the neighborhood is controlled by α (set by span) which is just the *smoothing parameter* ***q*** that described above. For $\alpha<1$, the neighborhood includes proportion α of the points. For $\alpha \geqslant 1$, all points are used.

3.2.3 Smoothing the reference limits by the LMS method (Method 5)

The LMS method provides a general method for fitting smooth centile curves to reference data. As we discussed in Section 3.1.2, among the four kinds of transformation methods, Box-Cox transformation method has its own superiority and has been widely used in recent years. The LMS method is just utilizes the Box-Cox transformation to allow the skewness of the measurement distribution, as well as the median and variability, to vary with age or some other time variables (gestational age for our data set). As we mentioned above, the Box-Cox transformation method works by choosing an optimal power at a given gestational age to completely remove skewness in the distribution. As skewness change with gestational age, the calculated power also changes. Slightly different from the expression defined for the Box-Cox transformation formula above, the transformed variable ***x*** is defined as:

Given a variable of interest ***y*** with median ***μ*** and a power transformation so that $\boldsymbol{y^{\lambda}}$ (or ***log(y)*** is ***λ =0***) is normally distributed, we consider the transformed variable

$$x = \begin{cases} \dfrac{(y-\mu)^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(\dfrac{y}{\mu}) & \text{if } \lambda = 0 \end{cases} \qquad (6)$$

based on the Box-Cox transformation. For λ=1, the SD of $\boldsymbol{x}$ is the coefficient of variation (CV) of $\boldsymbol{y}$, and this remains approximately true for all moderate values of $\boldsymbol{\lambda}$. The optimal value of $\boldsymbol{\lambda}$ minimizes the SD of $\boldsymbol{x}$. Although the $\boldsymbol{\lambda}$ defined here is not exactly the same as the $\boldsymbol{\lambda}$ described in section 3.1.2, it can more or less give us a clue when choosing the parameter for the estimation of the $\boldsymbol{\lambda}$ in formula (6).

Furthermore, the Z-score of $\boldsymbol{x}$ is given by:

$$z = \frac{x}{\sigma} = \begin{cases} \dfrac{(y-\mu)^{\lambda}-1}{\lambda\sigma} & \text{if } \lambda \neq 0 \\ \dfrac{\log\left(\dfrac{y}{\mu}\right)}{\sigma} & \text{if } \lambda = 0 \end{cases} \tag{7}$$

and is assumed to take a standard normal distribution.

Assume that the distribution of y varies with gestational age-*t,* and that *λ, μ* and *σ* at t are read off smooth curves *L(t), M(t)* and *S(t)*. Then

$$z = \begin{cases} \dfrac{\left(\dfrac{y}{M(t)}\right)^{L(t)}-1}{L(t)S(t)} & \text{if } L(t) \neq 0 \\ \dfrac{\log\left(\dfrac{y}{M(t)}\right)}{S(t)} & \text{if } L(t) = 0 \end{cases} \tag{8}$$

By some simple rearrangement, the centile 100_{α} of $\boldsymbol{y}$ at $\boldsymbol{t}$ is given by

$$C_{100\alpha}(t) = \begin{cases} M(t)[1+L(t)S(t)z_{\alpha}]^{\frac{1}{L(t)}} & \text{if } L(t) \neq 0 \\ M(t)\exp(S(t)z_{\alpha}) & \text{if } L(t) = 0 \end{cases} \tag{9}$$

where z_{α} is the normal equivalent deviate of size *α*. This shows that if *L, M* and *S* are smooth, then so are the centile curves.

From equation (8), the introducers Cole and Green[9] of the LMS method developed a penalized likelihood function with three integrals providing roughness penalties for the curves *L(t), M(t)* and *S(t).* The smoothness of the curves is controlled only by specifying such three smoothing parameters in the model fitting procedure. Further, by calculating each fitted curve as a function of these smoothing parameter, 'equivalent degrees of freedom' (EDFs) have generated and give a more usable measure of the extent of the smoothing.

4. Result

4.1 Result of the estimated empirical quantile (Method 1)

The results of homogeneous reference ranges calculated by each week as well as each two week are listed below in Table 2(a) and Table 2(b). Further more, Fig. 4(a) and Fig. 4(b) will give us a general overview of the changing tendency of different sample from week 12 to week 42.

From the last line of Table 2(b), we can see that compared to Table 2(a), the upper reference limit changed a lot to 1.191. It could be happened when outliers were included. Traced back to the original data set, it shows that the value of 8-iso-PG $F_{2\alpha}$ for the patient number 122 on week 43 is abnormally high to 1.29. This extreme value has severely affected the computation of the upper reference limit. From Fig. 1, we can see that all the other values of 8-iso-PG $F_{2\alpha}$ are less than 1 without any exception. As a result, defining the value 1.29 as an outlier and remove it is reasonable. Remove that value and recalculate the reference limit again, we got the result of the adjusted reference limits of week 42 listed in Table 3.

Week	***Freq***	***Empirical Quantile***		***Median***
		Q-0.025	***Q-0.975***	
11	13	**0.24**	**0.51**	0.36
12	17	**0.25**	**0.53**	0.31
20	16	**0.20**	**0.48**	0.33
21	21	**0.18**	**0.67**	0.35
25	12	**0.28**	**0.50**	0.34
26	24	**0.21**	**0.77**	0.35
27	11	**0.31**	**0.55**	0.35
29	23	**0.20**	**0.60**	0.35
30	21	**0.17**	**0.75**	0.45
32	19	**0.19**	**0.73**	0.36
33	14	**0.28**	**0.89**	0.43
34	18	**0.20**	**0.73**	0.36
36	21	**0.20**	**0.88**	0.42
37	16	**0.20**	**0.83**	0.39
38	10	**0.17**	**0.68**	0.37
40	30	**0.18**	**0.82**	0.41
41	24	**0.23**	**0.78**	0.39
42	12	**0.22**	**0.86**	0.35

Tab. 2(a) Homogeneous reference ranges
calculated by each week using non-parametric method

Week	*Freq*	*Empirical Quantile*		*Median*
		Q-0.025	***Q-0.975***	
10	19	**0.19**	**0.51**	0.37
12	24	**0.19**	**0.53**	0.30
20	37	**0.17**	**0.63**	0.35
24	15	**0.28**	**0.55**	0.34
26	35	**0.22**	**0.73**	0.35
28	25	**0.20**	**0.68**	0.35
30	24	**0.17**	**0.74**	0.45
32	32	**0.21**	**0.81**	0.39
34	21	**0.21**	**0.71**	0.36
36	36	**0.20**	**0.90**	0.41
38	13	**0.18**	**0.67**	0.37
40	31	**0.18**	**0.89**	0.41
42	13	**0.22**	**1.19**	0.54

Tab. 2(b) Homogeneous reference ranges

Calculated by each two week using non-parametric method

Week	*Freq*	*Q- 0.025*	*Q- 0.975*	*Median*
42	12	**0.22**	**0.86**	0.35

Tab.3 The adjusted homogeneous reference range of week 42 calculated by each two week

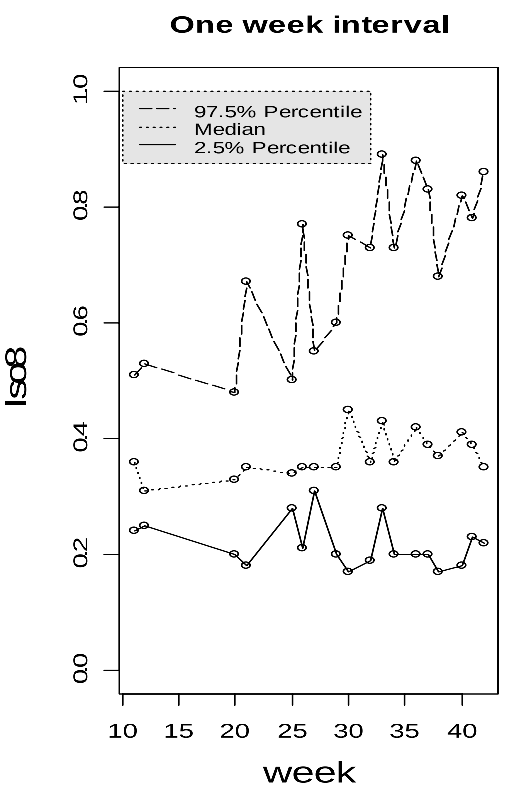


Fig. 4(a) Empirical reference limits calculated by one week interval

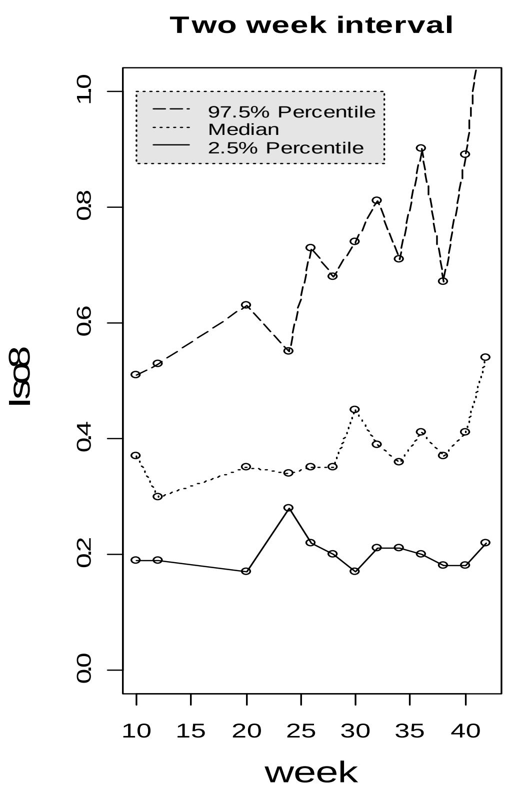


Fig. 4(b) Empirical reference limits calculated by two week interval

4.2 Result of the reference range based on normal distribution

According to the method discussed in section 3.1.2, the general steps for estimating the homogeneous reference ranges based on normal distribution are listed below:

a) Test each sample by the Shapiro-Wilk test to get a first step *W* statistic *W1* and a *p value p1* which indicating the probability for obtaining such *W* statistic.
b) Basing on the real data of each sample, calculate the power of each Box-Cox transformation formula by maximum likelihood estimation method. In another word, the value of λ in the formula $(Y^{\lambda}-1)/\lambda$ will be calculated by different samples.
c) Putting the estimated λ values back to the formula to get new transformed samples.
d) Using the Shapiro-Wilk test again to get a new *W* statistic *W2* and *p value p2*.
e) Comparing the values of *W1* and *W2* as well as *p1* and *p2*, the sample which has larger values of *W* statistic and *p* value will be used to the forthcoming estimation.
f) If *W1* is larger, the *2.5%* centile and *97.5%* centile can be directly estimated by the original sample using the equation $C=F^{-1}(p)$ where *p* equals to *0.025* and *0.975* respectively and *F* is the cumulative normal distribution. Otherwise, if *W2* is larger, the transformed sample should be used. The estimates of each reference limits will be calculated by back transforming the Box-Cox transformation formula.

The estimates of the reference limits for each week and each two week which has observations number greater than 10 are listed in Table 4(a) and Table 4(b).

Week	Freq	W1	p1	λ	W2	p2	Q-0.025	Q-0.975	Median
11	13	0.9313	0.36	−0.08	0.9473	0.56	0.21	0.59	0.35
12	17	0.8805	0.03	−1.48	0.9352	0.27	0.23	0.63	0.32
20	16	0.9537	0.55	0.62	0.9594	0.65	0.17	0.52	0.33
21	21	0.9706	0.75	0.44	0.9596	0.65	0.10	0.66	0.38
25	12	0.8495	0.04	−2.00	0.9590	0.77	0.27	0.54	0.34
26	24	0.8928	0.02	−0.34	0.9493	0.26	0.17	0.98	0.36
27	11	0.8987	0.18	−1.49	0.9354	0.47	0.28	0.64	0.38
29	23	0.8991	0.02	0.01	0.9213	0.33	0.26	0.58	0.39
30	21	0.9732	0.80	0.81	0.9040	0.21	0.14	0.76	0.45
32	19	0.8852	0.03	−0.11	0.9144	0.09	0.15	0.99	0.37
33	14	0.7753	0.00	−1.10	0.9563	0.66	0.27	0.99	0.41
34	18	0.8901	0.04	−0.30	0.9828	0.97	0.19	0.80	0.36
36	21	0.9112	0.06	−0.02	0.9551	0.42	0.17	0.98	0.40
37	16	0.9108	0.12	0.06	0.9715	0.86	0.18	0.89	0.40
38	10	0.8927	0.18	0.50	0.9141	0.31	0.13	0.73	0.37
40	30	0.9547	0.23	0.19	0.9864	0.96	0.17	0.92	0.42
41	24	0.8944	0.02	−0.56	0.9803	0.90	0.22	0.90	0.39
42	12	0.8303	0.02	−0.69	0.9708	0.92	0.19	1.12	0.36

Tab. 4(a) The homogeneous reference ranges calculated by each week basing on normal distribution

Week	Freq	W1	p1	λ	W2	p2	Q-0.025	Q-0.975	Median
10	19	0.9503	0.40	0.96	0.9501	0.40	0.16	0.55	0.35
12	24	0.9496	0.27	0.19	0.9372	0.24	0.13	0.53	0.33
20	37	0.9593	0.19	0.32	0.9861	0.92	0.16	0.64	0.35
24	15	0.8200	0.01	−2.32	0.9658	0.79	0.27	0.63	0.34
26	35	0.9114	0.01	−0.26	0.9708	0.47	0.19	0.84	0.37
28	25	0.8727	0.00	−0.31	0.9563	0.35	0.20	0.74	0.36
30	24	0.9720	0.72	0.85	0.9723	0.72	0.17	0.75	0.45
32	32	0.9272	0.03	−0.11	0.9799	0.80	0.18	0.95	0.39
34	21	0.8982	0.03	−0.28	0.9842	0.97	0.20	0.78	0.37
36	36	0.9093	0.01	−0.14	0.9761	0.62	0.19	0.94	0.41
38	13	0.9239	0.28	0.57	0.9395	0.45	0.15	0.69	0.38
40	31	0.9522	0.18	0.18	0.9851	0.93	0.17	0.96	0.43
42	12	0.8303	0.02	−0.69	0.9708	0.92	0.19	1.12	0.36

Tab. 4(b) The homogeneous reference ranges calculated by each two week basing on normal distribution

The graphs indicating the change of different reference limits are described in Figure 5(a) and Figure 5(b). From the two graphs which indicating the changing tendency for different values of reference limits, it can be seen that a clear trend exists. For the 97.5th upper reference limit, the value is becoming larger and larger as the time increased. Oppositely, the 2.5th lower reference limit has a descending trend as the time increased although it is not very significant. It demonstrates that the gestational age is an influential factor for the reference ranges. And the value of 8-iso-PG $F_{2\alpha}$ tends to become more and more instable as the gestational age increases. Advanced analysis is needed for smoothing the reference limits and a regression model of the reference referring to gestational ages as a factor will be fitted in the forthcoming section.

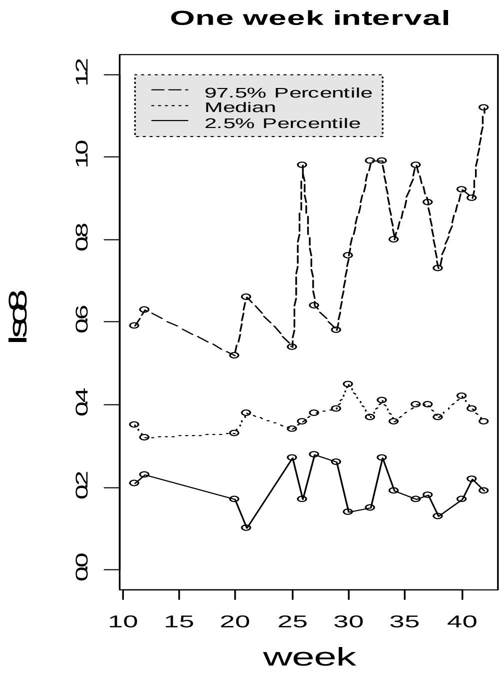


Fig. 5(a) Reference limits calculated by each one week basing on normal distribution

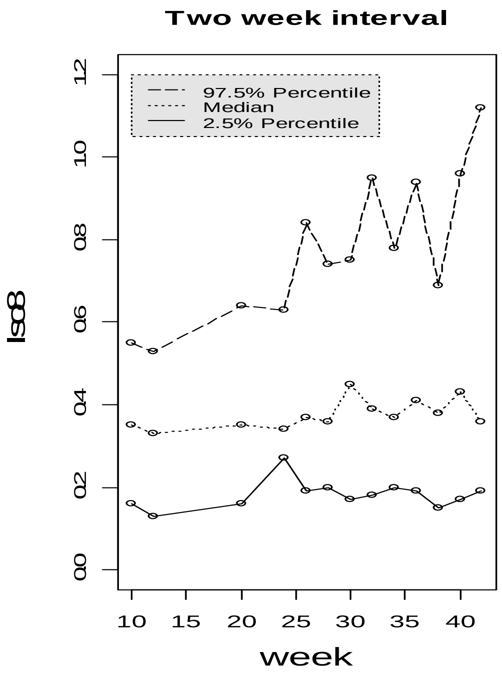


Fig. 5(a) Reference limits calculated by each two week basing on normal distribution

4.3 Result of the regression method based on the empirical quantiles (Method 3)

Applying the method described in section 3.2.1, the result shows that the gestational age effect for both the estimates of 2.5% percentiles which calculated by each one week and each two week interval are not significant. The mean value of the 2.5% empirical quantiles will be used as a constant reference limit. For the 97.5% percentile, the simple linear regression method has been chosen as the best fit model since the higher order factors are not significant at all. The fitted values are listed in Table 5 and the fitted regression models as well as the residuals are graphed in Figure 6.

Week	Q-2.5%	Q-97.5%
11		0.480
12		0.492
20		0.585
21		0.597
25		0.643
26		0.655
27		0.667
29		0.690
30	0.217	0.702
32		0.725
33		0.737
34		0.748
36		0.772
37		0.783
38		0.795
40		0.818
41		0.830
42		0.841

Tab. 5(a) Estimates for reference limits by linear regression calculated by each one week interval

Week	Q-2.5%	Q-97.5%
10		0.459
12		0.489
20		0.607
24		0.666
26		0.695
28		0.725
30	0.202	0.754
32		0.784
34		0.813
36		0.843
38		0.872
40		0.902
42		0.931

Tab. 5(b) Estimates for reference limits by linear regression calculated by each two week interval

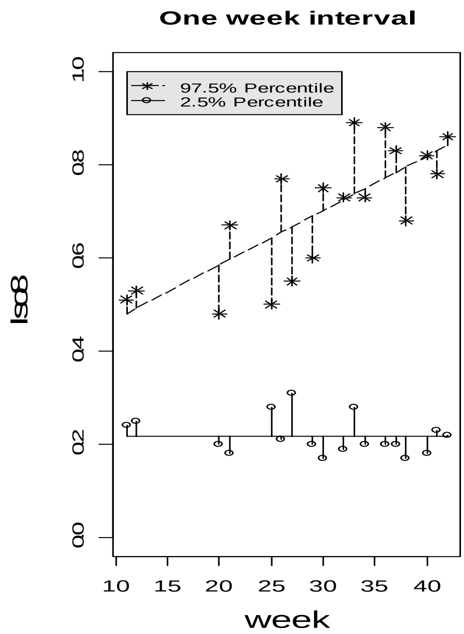


Fig. 6(a) Reference limits calculated by each one week basing on linear regression model

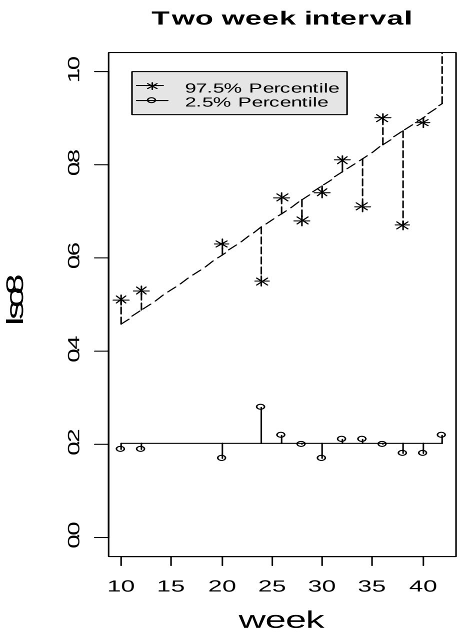


Fig. 6(a) Reference limits calculated by each two week basing on linear regression model

4.4 Result of smoothing the empirical quantiles by LOESS procedure (Method 4)

As discussed in section 3.2.2, ***d=1*** and ***d=2*** as well as ***q=0.5*** and ***q=0.75*** should generate four combinations of the extent of smoothing. A large value of the degree of local polynomial accompanied by a small value of smoothing parameter can contribute to a better fit comply with the real observations. On the opposite, Small values of the degree of local polynomial given large values of smoothing parameter can result in a smoother curve. As a result, only two combinations will be chosen here to represent two contrary extent of smoothing. They are ***d=2*** given ***q=0.5*** and ***d=1*** given ***q=0.75***. By setting the smoothing parameter and degree of local polynomial in advance, the graph of the regression curves for different parameters calculated by each week and each two week respectively are shown in Figure 7. And the regression values of reference limits calculated each one week interval and two week interval are listed in Table 6 in the next page after the more appropriate parameters have been chosen.

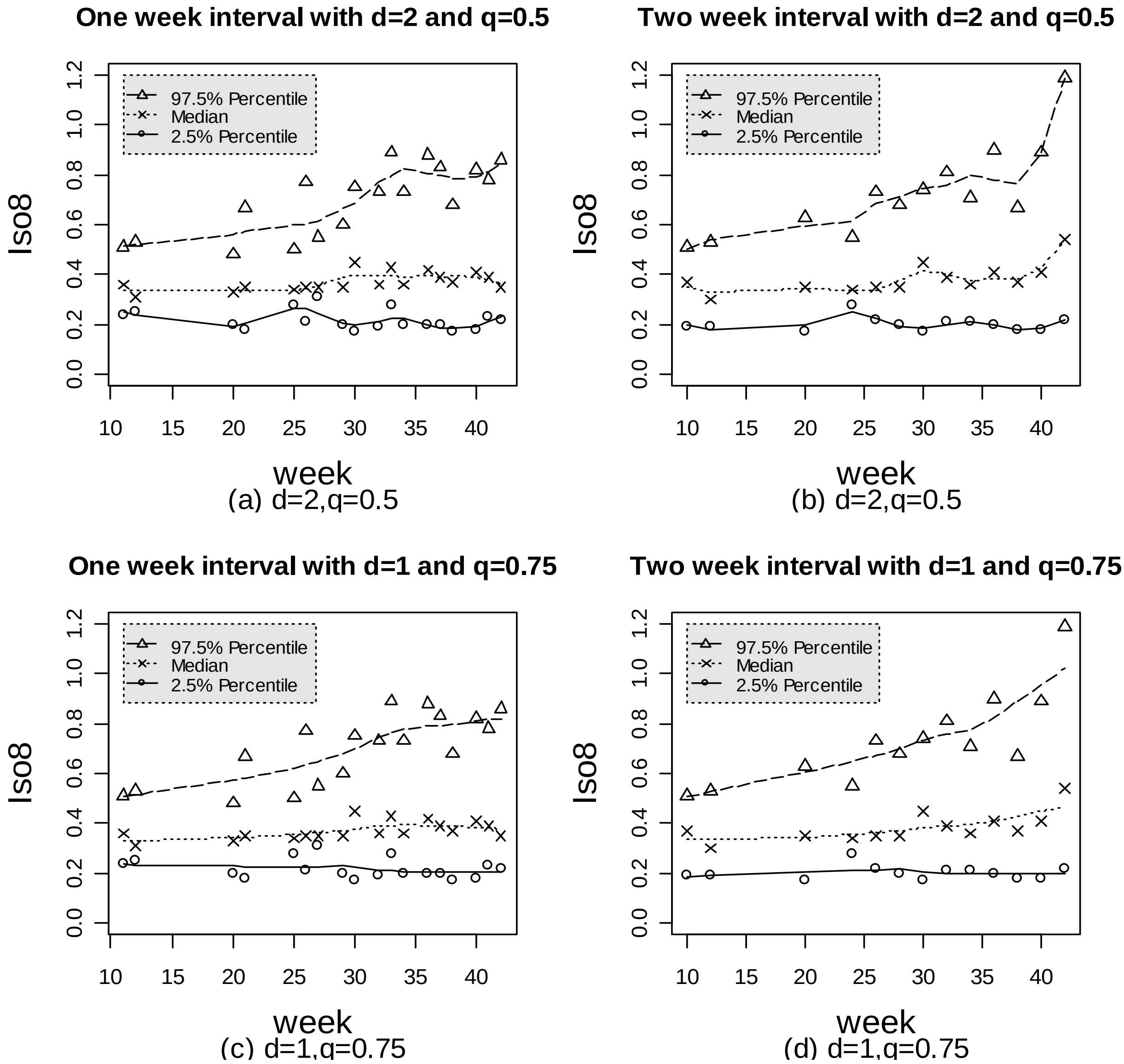


Fig. 7 The regression curves based on different smooth parameters and degrees of local polynomial by LOESS procedure using the estimates of empirical reference limits

Comparing those two groups of regression curves, the last two graphs clearly give better smoother curves when ***d=1*** and ***q=0.75***. The regression values for such setting of parameters are listed in Table 6 below.

From the last line of Table 6(b), we can see that the regression value of the 97.5% percentile is extremely high comparing to other values of 97.5% reference limit. Trace it back to figure 7(d), the nearest point for the empirical estimate of week 40 is comparatively small which produced a sharp increasing trend with the empirical estimate of week 42. Due to the basic idea of LOESS procedure, the nearest point gives the most weight, it is quite reasonable that the regression value has been seriously affected by the empirical value of week 40. It can be viewed as a drawback of the LOESS smoothing method for the regression value can be easily influenced by the nearest value even if it is abnormal to some extent.

One week interval			
week	*Q-2.5%*	*Median*	*Q-97.5%*
11	0.235	0.330	0.506
12	0.234	0.332	0.513
20	0.228	0.344	0.574
21	0.227	0.345	0.582
25	0.222	0.354	0.617
26	0.225	0.360	0.633
27	0.226	0.364	0.649
29	0.228	0.373	0.682
30	0.226	0.380	0.701
32	0.213	0.389	0.744
33	0.208	0.392	0.762
34	0.205	0.393	0.777
36	0.204	0.391	0.790
37	0.205	0.389	0.795
38	0.204	0.387	0.800
40	0.204	0.383	0.811
41	0.204	0.381	0.815
42	0.204	0.378	0.819

Tab. 6(a) Regression values calculated by one week interval by LOESS procedure

Two week interval			
week	*Q-2.5%*	*Median*	*Q-97.5%*
10	0.185	0.336	0.511
12	0.190	0.338	0.529
20	0.206	0.344	0.605
24	0.213	0.354	0.644
26	0.213	0.364	0.671
28	0.215	0.372	0.697
30	0.208	0.380	0.729
32	0.198	0.391	0.760
34	0.196	0.393	0.775
36	0.196	0.406	0.823
38	0.197	0.426	0.888
40	0.198	0.450	0.954
42	0.199	0.471	1.024

Tab. 6(b) Regression values calculated by two week interval by LOESS procedure

4.5 Result of the reference range estimated by the LMS method (Method 5)

4.5.1 The procedure description of estimating the reference range using the software 'LmsChartMaker Light'

As it mentions above in section 3.2.3, the EDFs is defined as 'equivalent degrees of freedom'. It can be better used for setting the extent of smoothing in advance. And the main difficulty for using the software to estimate the centiles is about how to choose the EDFs. Here are some chain rules for choosing EDF:

By default, the EDF for L, M and S are 3, 5 and 3 respectively. The EDF of each curve is a measure of its complexity. 2 EDF corresponds to a straight line, 3 EDF gives a simple curve like a quadratic, and 4 or more EDF indicates progressively more complex curve shapes. For small datasets, the default EDF of 5 for the M curve may be adequate. Fix 4 for M curve, set both the initial values as 1 for S and L curves, and optimize the curves by increasing of decreasing the EDF by 1 until the change in deviance is sufficiently small.

Where the deviance is set to be:

*Deviance (D(y,θ))= -2 * penalized log likelihood*

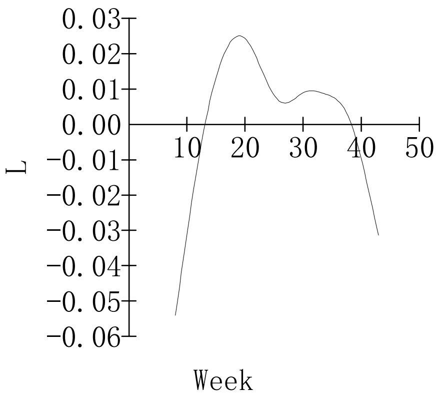
And $\boldsymbol{\theta}$ is a parameter vector stands for the estimates of parameter λ, μ and σ. If $\boldsymbol{\theta}$ is the true parameter, the deviance will follow a chi-square distribution. In our procedure for applying the LMS method, assessing the goodness of fit is realized by viewing the decrease in deviance for a given increase in EDF as χ^2 distributed. One of the most important rules is that a change in deviance less than 4 per unit change in EDF is modest. A change more than 4 can be viewed as significant. In general, M edf ≥ S edf ≥ L edf. And if in doubt, smaller values of EDF are preferred.

4.5.2 The final result of the reference ranges calculated by LMS method

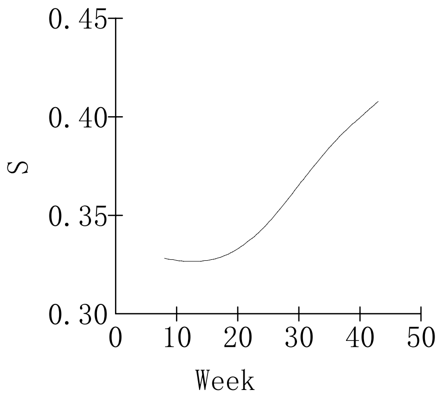
For the LMS method is built up upon the Box-Cox transformation method, the value of $\boldsymbol{\lambda}$ in Table 4 can more or less give some guidance for choosing the EDF for the L curve. Since it shows that the value of $\boldsymbol{\lambda}$ differs a lot for different gestational ages, we can conclude that a quadratic or cubit form of L curve might be appropriate. After some adjustment of the EDF by comparing the value of deviances, the EDF for L, M and S curves are 3, 4 and 3 respectively.

The result of the reference ranges calculated by each week is listed in Figure 8 and Table 7. And the result of the reference ranges calculated by each two week is shown in Figure 9 and Table 8 in the following four pages.

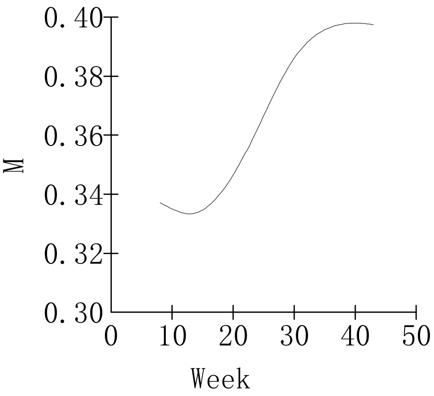
(a)



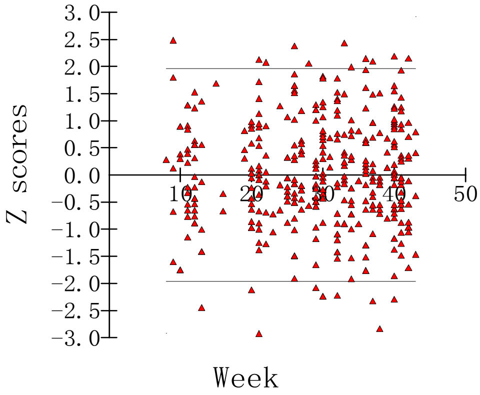
(b)



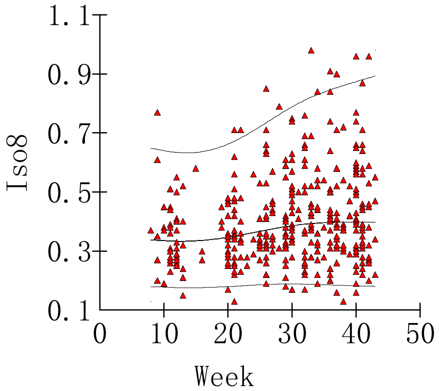
(c)



(d)



(e)



(f)

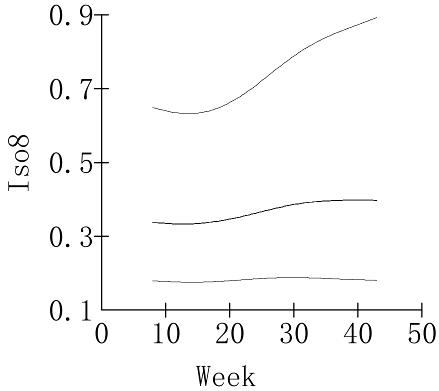
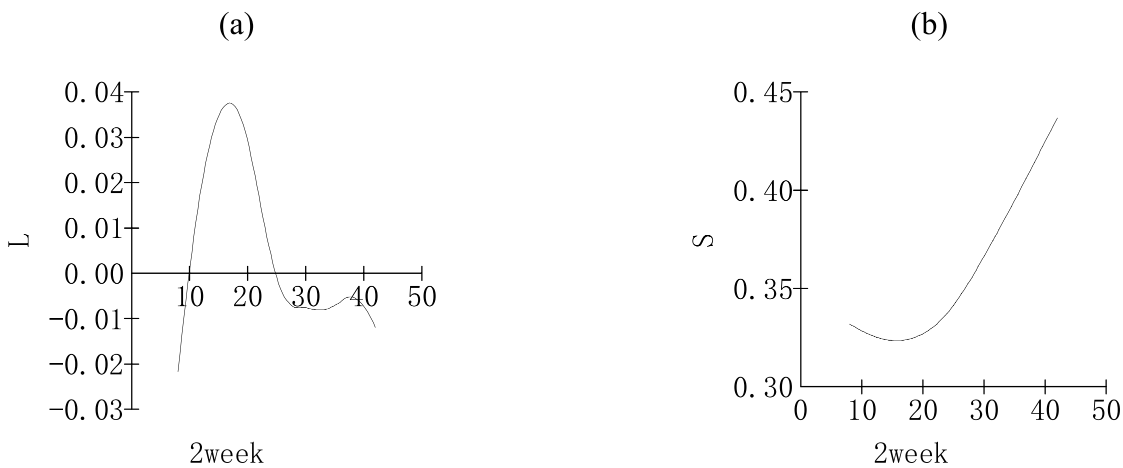


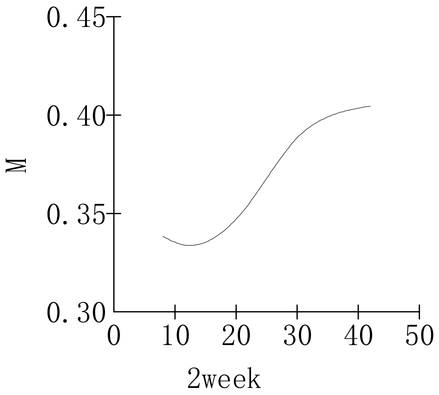
Fig. 8 Graphs showing the result of indexes of LMS method calculated by each week- (a) The value of L curve in different gestational age; (b) The value of S curve in different gestational age; (c) The value of M curve in different gestational age; (d) Z-scores after transformation with the centile curves; (e) The real centile curves with scatter plots; (f) The final centile curves of Quantile 2.5%, mean value, and Quantile 97.5%.

week	L	M	S	Q-2.5%	Median	Q-97.5%
8	-5.40E-02	0.337	0.328	0.179	0.337	0.649
9	-4.25E-02	0.336	0.328	0.178	0.336	0.645
10	-3.10E-02	0.335	0.327	0.178	0.335	0.640
11	-2.00E-02	0.334	0.327	0.177	0.334	0.637
12	-9.85E-03	0.334	0.327	0.176	0.334	0.634
13	-7.66E-04	0.333	0.327	0.176	0.333	0.632
14	6.94E-03	0.334	0.327	0.176	0.334	0.632
15	1.33E-02	0.335	0.327	0.176	0.335	0.634
16	1.83E-02	0.336	0.328	0.176	0.336	0.637
17	0.0219466	0.338	0.329	0.177	0.338	0.641
18	2.42E-02	0.340	0.330	0.177	0.340	0.646
19	2.50E-02	0.343	0.331	0.178	0.343	0.653
20	2.43E-02	0.346	0.333	0.179	0.346	0.662
21	2.21E-02	0.350	0.335	0.181	0.350	0.672
22	1.87E-02	0.354	0.337	0.182	0.354	0.683
23	1.49E-02	0.358	0.340	0.183	0.358	0.695
24	1.12E-02	0.362	0.343	0.184	0.362	0.708
25	8.26E-03	0.367	0.346	0.186	0.367	0.721
26	6.45E-03	0.371	0.350	0.187	0.371	0.735
27	6.00E-03	0.375	0.353	0.187	0.375	0.749
28	6.59E-03	0.379	0.357	0.188	0.379	0.762
29	7.79E-03	0.383	0.361	0.188	0.383	0.775
30	8.93E-03	0.386	0.365	0.188	0.386	0.788
31	9.48E-03	0.389	0.369	0.188	0.389	0.800
32	9.42E-03	0.391	0.373	0.188	0.391	0.811
33	9.01E-03	0.393	0.377	0.187	0.393	0.821
34	8.50E-03	0.395	0.381	0.187	0.395	0.830
35	7.82E-03	0.396	0.384	0.186	0.396	0.839
36	6.66E-03	0.397	0.388	0.185	0.397	0.846
37	4.67E-03	0.397	0.391	0.184	0.397	0.853
38	1.40E-03	0.398	0.394	0.184	0.398	0.860
39	-3.47E-03	0.398	0.397	0.183	0.398	0.867
40	-9.67E-03	0.398	0.400	0.182	0.398	0.874
41	-1.67E-02	0.398	0.402	0.182	0.398	0.880
42	-2.41E-02	0.398	0.405	0.181	0.398	0.887
43	-3.14E-02	0.397	0.408	0.180	0.397	0.893

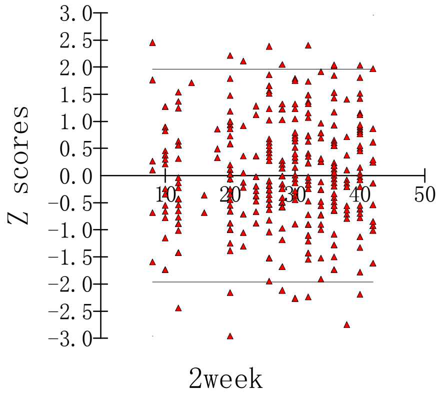
Tab. 7 The result showing the value of L, M, S curves as well as the final estimates of Quantile 2.5%, mean value, and Quantile 97.5% each week.



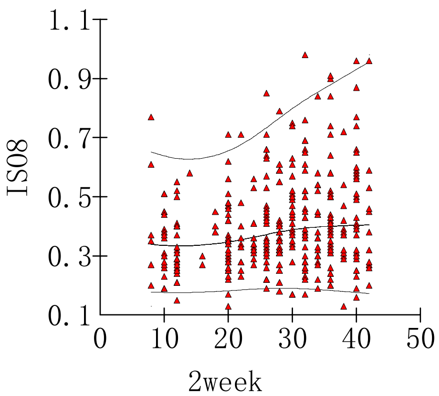
(c)



(d)



(e)



(f)

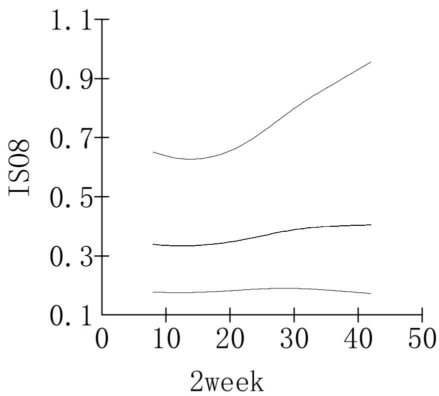


Fig. 9 Graphs showing the result of indexes of LMS method calculated by each two week- (a) The value of L curve in different gestational age; (b) The value of S curve in different gestational age; (c) The value of M curve in different gestational age; (d) Z-scores after transformation with the centile curves; (e) The real centile curves with scatter plots; (f) The final centile curves of Quantile 2.5%, mean value, and Quantile 97.5%.

week	L	M	S	Q-2.5%	Median	Q-97.5%
8	-2.31E-02	0.3384	0.3322	0.1770	0.3380	0.6520
10	-4.17E-04	0.3355	0.3287	0.1760	0.3360	0.6390
12	1.80E-02	0.3338	0.3258	0.1760	0.3340	0.6300
14	3.02E-02	0.3344	0.3240	0.1760	0.3340	0.6270
16	3.62E-02	0.3369	0.3234	0.1770	0.3370	0.6300
18	3.60E-02	0.3412	0.3243	0.1790	0.3410	0.6400
20	2.90E-02	0.3470	0.3269	0.1820	0.3470	0.6550
22	1.66E-02	0.3545	0.3311	0.1850	0.3540	0.6760
24	4.08E-03	0.3632	0.3374	0.1870	0.3630	0.7030
26	-4.48E-03	0.3723	0.3457	0.1890	0.3720	0.7340
28	-7.48E-03	0.3809	0.3554	0.1900	0.3810	0.7660
30	-7.49E-03	0.3884	0.3662	0.1900	0.3880	0.7980
32	-7.94E-03	0.3939	0.3775	0.1880	0.3940	0.8270
34	-7.73E-03	0.3976	0.3889	0.1860	0.3980	0.8540
36	-6.35E-03	0.4003	0.4006	0.1830	0.4000	0.8800
38	-5.17E-03	0.4022	0.4126	0.1790	0.4020	0.9040
40	-7.24E-03	0.4036	0.4247	0.1760	0.4040	0.9300
42	-1.19E-02	0.4045	0.4369	0.1730	0.4050	0.9570

Tab. 8 The result showing the value of L, M, S curves as well as the final estimates of Quantile 2.5%, mean value, and Quantile 97.5% each two week.

4.6 Comparison among different methods

From all the analysis above, we can see that many approaches have been used in such procedure. To sum up, the methods which have been used are listed in Table 8.

Estimation Method	Non-parametic (Empirical quantile)	Parametric (Normal distribution)	GA effect 3 Linear regression	GA effect 2 LOESS	GA effect 3 LMS
1	Yes				
2		Yes			
3	Yes		Yes		
4	Yes			Yes	
5		Yes			Yes

Tab. 9 A summary of methods which have been used

There are many combinations for comparison, but since some comparisons between different methods are meaningless, only 4 combinations of the methods will be chosen to make a comparison. They are method 1 versus 2, method 2 versus 5, method 3 versus 4, and method 4 versus 5.

4.6.1 Comparison of method 1 versus 2 without gestational age effect

As Table 9 shows, both estimates carried out by method 1 and 2 are estimation methods without taking the gestational age effect into consideration. The main different is whether parametric method has been used or not. Graph 10 gives a clear comparison between the results of two methods calculated by both one week interval and two week interval. In order to give an easy to be distinguished view, the estimates of median have been removed.

Graph 10 gives us a legible view that the estimates based on normal distribution that represented by dotted lines have wider reference ranges than the empirical quantiles. And graph 11 shows the exact differences between different estimates. It shows distinguishable evidence that as sample size grows larger, the differences between the two methods tend to getting smaller. Estimates calculated by each two weeks provide closer estimates of those two methods than estimates calculated by each one week.

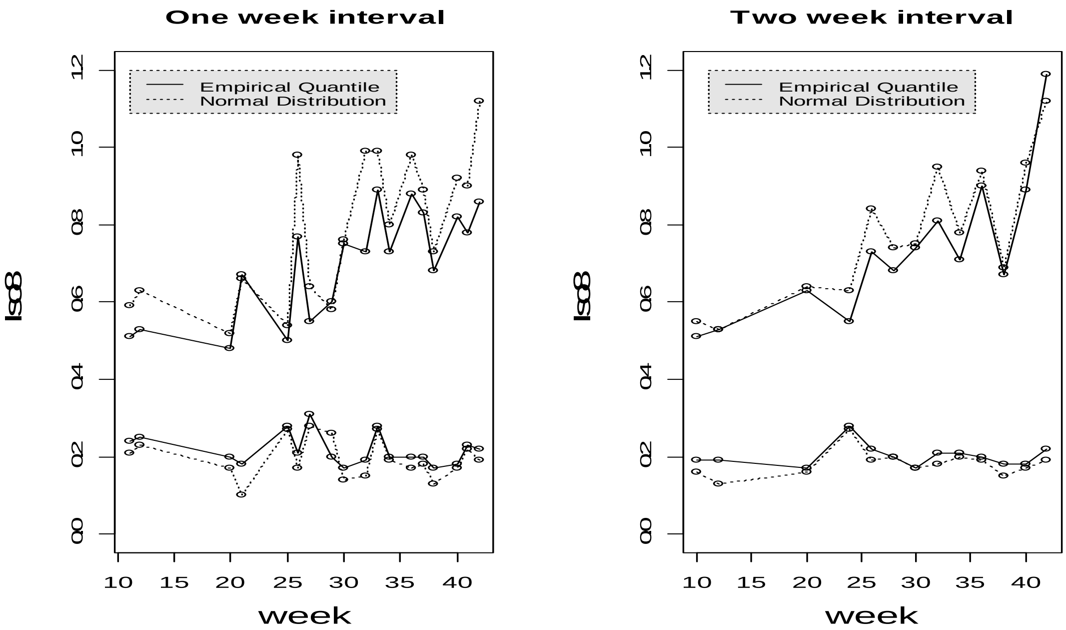


Fig. 10 A comparison of the estimation result calculated by non-parametric method and parametric method without gestational age effect

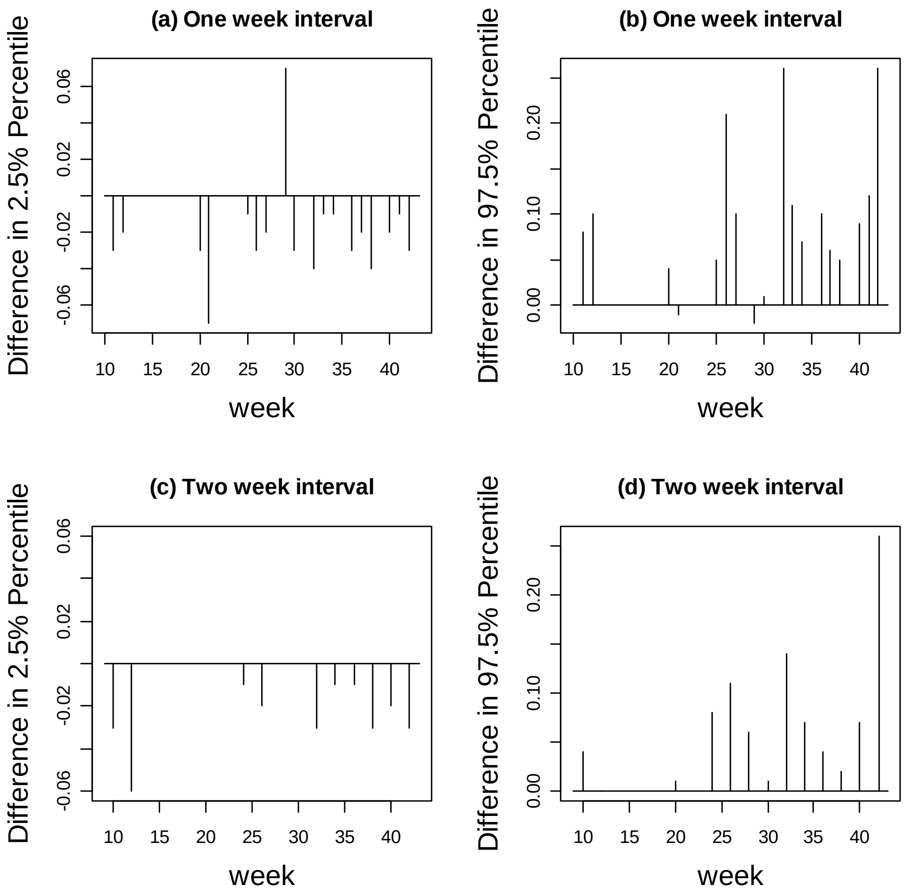


Fig. 11 The exact differences between method 1 and method 2

For graph 11, the exact differences are calculated by the formula below:

Exact differences= (Estimates of Method 2)−*(Estimates of Method 1)*
=(Estimates based on Normal Distribution)−*(Estimates of empirical quantiles)*

Although it is not clear enough to see from Fig. 11(a) and Fig. 11(c) for the differences of the 2.5% percentile (since the fluctuation is comparatively small for the lower percentile), it can be seen that, for the 97.5% percentile, the exact differences between the two methods have an explicit trend of getting smaller as the sample size getting larger. And the estimates computed basing on normal distribution are always larger than the empirical quantiles for the upper reference limits.

4.6.2 Comparison of method 2 versus method 5

Since method 5 is an extension of method 2 and both of them are carried out upon the Box-Cox transformation, the main difference of them is that whether the gestational age effect has been taken into consideration. Graph 12 gives a compendious comparison between those two methods.

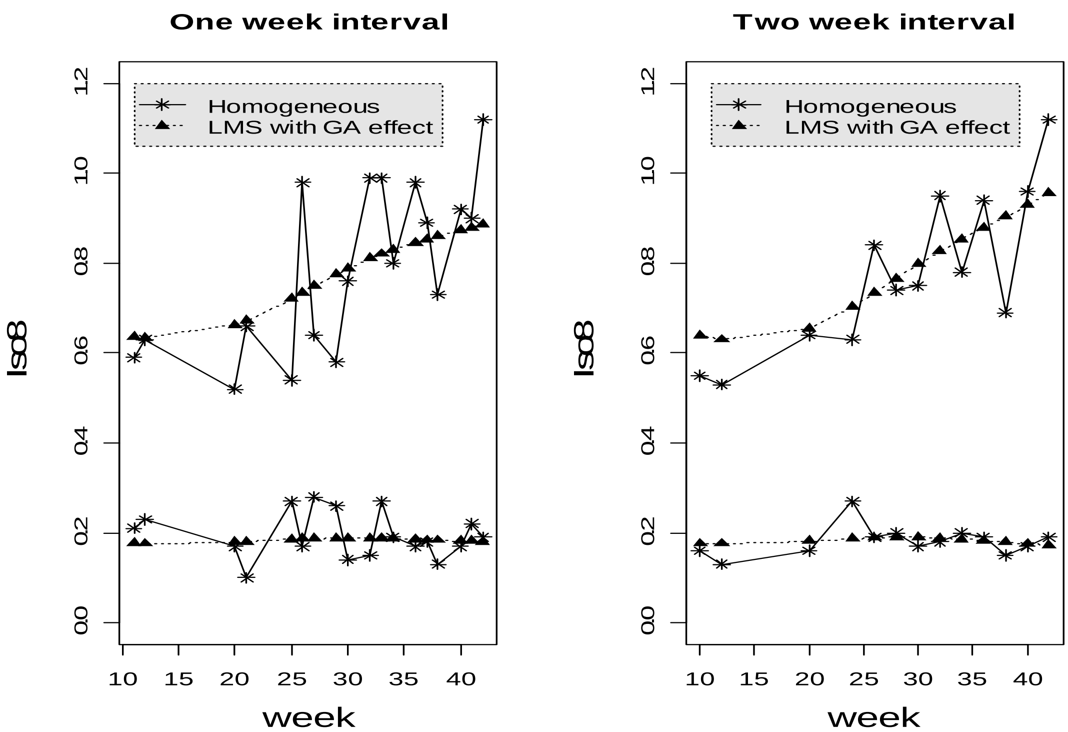


Fig. 12 A comparison of the homogeneous reference limits basing on normal distribution and the reference limits calculated by LMS method taking gestational age effect into consideration

This graph gives us evidence that the LMS method has smoothed the homogeneous reference limits which based on normal distribution a lot. Furthermore, as the time units changed from one week interval to two week interval, the fluctuation of the estimates have been weaken since the sample sizes are becoming larger and larger.

4.6.3 Comparison of method 3 versus method 4 based on empirical quantiles with gestational age effect

Both of method 3 and method 4 are estimations based on the empirical quantiles. The only different is the smoothing method they used. For method 3, the smoothing is done by fitting a linear regression model. And for method 4, the LOESS procedure is applied. Graph 13 gives a comparison between these two methods.

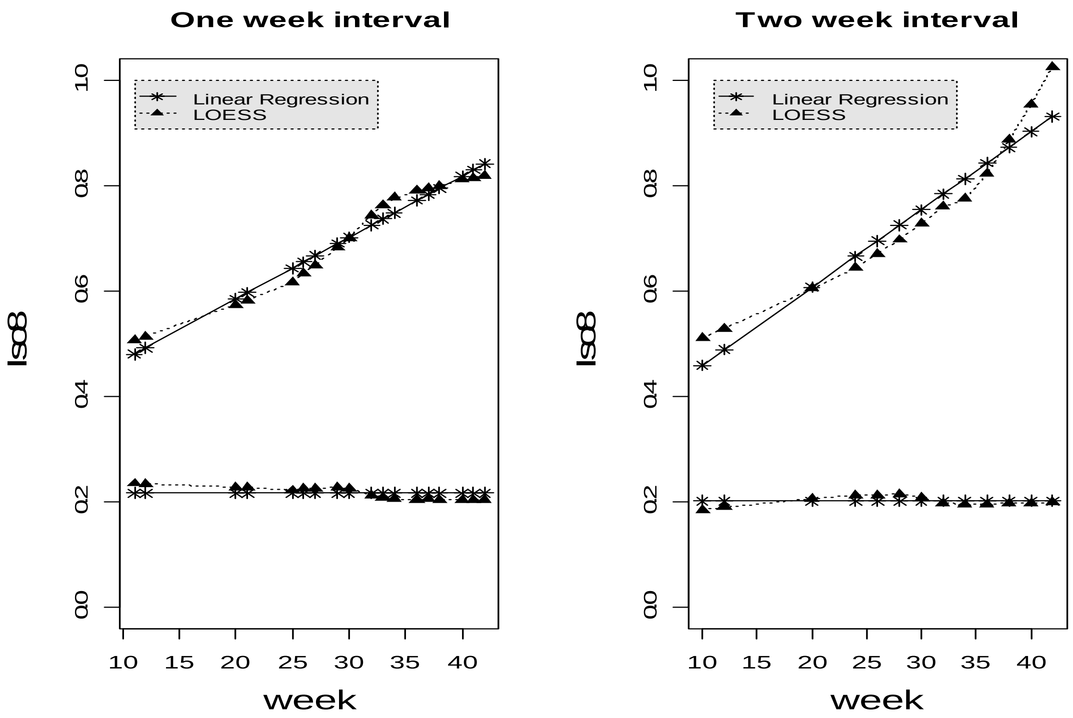


Fig. 13 A comparison of the smoothed reference limits between the linear regression line and the LOESS smoothing line basing on empirical quantiles

Fig. 13 gives us a general view that the estimates of these two methods are roughly the same especially for the 2.5% percentile. In order to get a more accurate comparison between those estimates, one way is to calculate how many percent of the total sample falling outside the reference ranges. The better method should have 5% observations outside the reference ranges.

Take the gestational age effect variable as the two week interval for example, it has been computed that for method 3, under the linear regression method, there are 9.5% of the observations falling outside the estimated reference ranges. On the other hand, by method 4, the LOESS procedure, only 7.8% of the observations falling outside the estimated reference limit. Although none of them have the percentage exactly equal to 5%, but method 4 has made an improvement for 1.7% comparing to method 3. As a result, the LOESS method is superior to the linear regression in this specific case.

4.6.4 Comparison of method 4 versus method 5

Since method 4 has been chosen as the better method between method 3 and method 4, the comparison will be taken between method 4 and method 5. These two methods are differed in many ways. From the first step analysis (non-parametric method and parametric method) to the advanced smoothing method (LOESS procedure and LMS method), they are quite dissimilar in many aspects. It is impossible to make accurate conclusion about which method system is superior to the other due to our limited sample size and a lack of convincible goodness-of-fit tests. However, such comparison in graph 14 can more or less give us an impression of the difference between these two totally different methods.

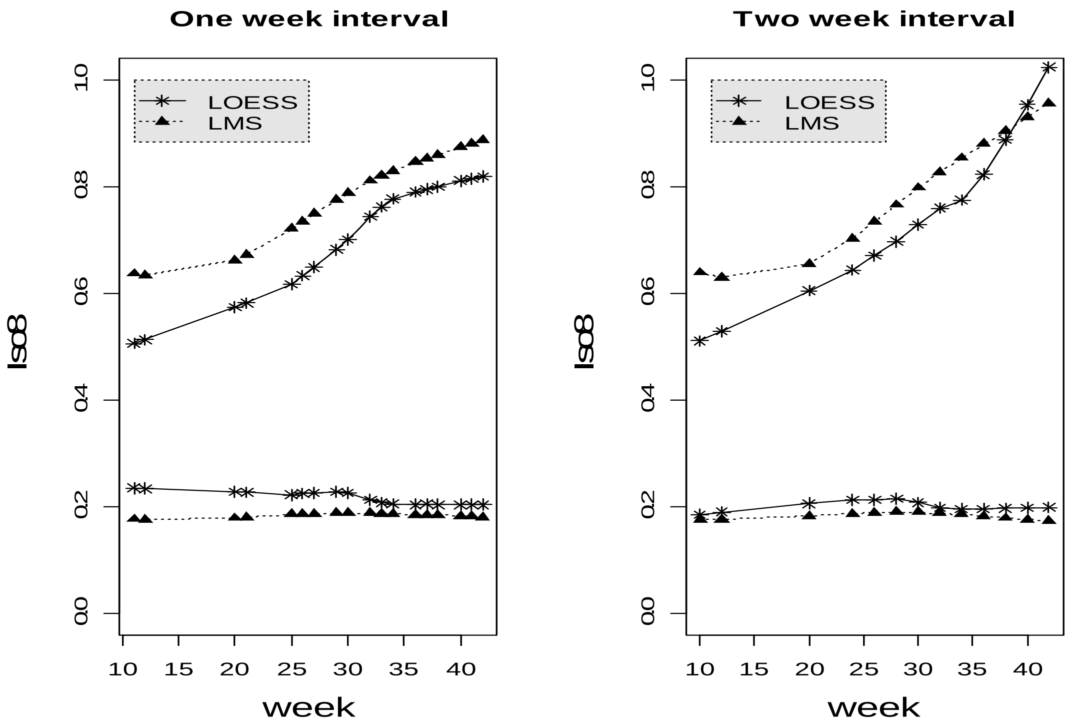


Fig. 14 A comparison of the empirical quantiles smoothed by LOESS procedure and the reference limits smoothed by LMS method taking gestational age effect into consideration

Figure 14 gives us a very clear impression that the reference ranges calculated by the LMS method are always wider than that estimated by LOESS procedure basing on the empirical quantiles. It is quite reasonable because this comparison is actually basing on the comparison between method 1 and method 2 illustrated in section 4.6.1. Both of the LOESS method and the LMS method are smoothing technique built up upon method 1 and method 2 respectively. In addition, Figure 14 also shows that the differences in the 97.5% percentile estimates are bigger than the differences in the 2.5% percentile estimates due to the original stable situation of the lower percentile. The cross part of the 97.5% percentile estimates in the two week interval graph is generated because of the big weight that has been given on the nearby points as discussed in section 4.4. Similar as the comparison of method 1 versus method 2, the differences for the estimates of two different methods are becoming smaller as the time units changing from one week interval to two week interval. More evidence will be given in figure 15 below. The exact differences are calculated by the formula on the next page:

Exact differences= (Estimates of Method 5)−(Estimates of Method 4)
= (Estimates based on LMS method)−(Estimates based on LOESS procedure)

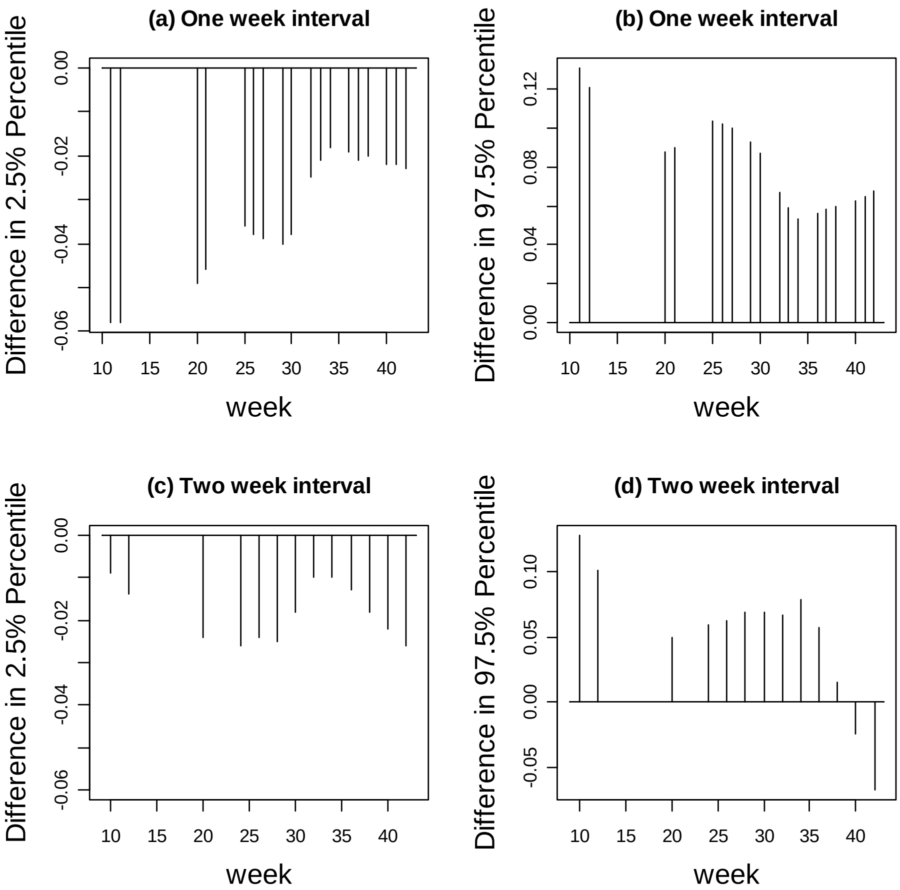


Fig. 15 The exact differences between method 3 and method 4

To get a more accurate comparison result, similar action will be taken as section 4.6.3. As it has been already calculated that method 4 has 7.8% observations outside the estimated reference ranges, we should compute the corresponding percentage of method 5. After some calculation, it is shown that there are 4.9% observations falling outside the reference ranges calculated by the LMS method. Although we can not simply generalize that the LMS method is always a mature method that can be used in all the estimation of reference ranges. At least, we can conclude that for our specific issue, this method is the best one comparing to the other methods. Table 10 gives a summary of the accurate comparison showing the percentage of outside observations by all the three methods which have taken the gestational age effect into consideration.

	Method 3 (Linear regression)	**Method 4 (LOESS procedure)**	**Method 5 LMS method**
Percentage of the observatsions outside the reference ranges	**9.5%**	**7.8%**	**4.9%**

Tab. 10 The comparison of the percentage of the observations outside the estimated reference ranges calculated by method 3,4 and 5 basing on the two week interval

5. Discussion

5.1 About the sample size

Although there is not a clear requirement of how many reference subjects should be sampled to provide the data for calculating a reference range, there is a common sense that a sample size that $n>100$ is more evidential to draw a convincible conclusion especially when the parametric method is being used.

Unfortunately, due to a difficulty of getting the measurement of 8-iso-PG $F_{2\alpha}$ (since the laboratory work is quite complicated to measure the value), our sample size is quite small to draw an exact conclusion. As a result, the confidence interval for each estimate should be introduced in for the further estimation.

5.2 About the LOESS smoothing procedure

There are lots of methods for smoothing the existing data, the reason why the LOESS procedure is used here has been discussed in Section 3.3.2. But one of the drawbacks of such method is that the choice of the smoothing parameter is quite subjective, the estimates changed a lot for different views of the choice of such parameters. Another disadvantage of LOESS is that it does not produce a regression function that is easily represented by a mathematical formula. This can make it difficult to transfer the results of an analysis to other people.

5.3 About the use of LMS method

Actually, one of the requirements for applying the LMS method is that the data set is from a cross-sectional study where there is only one observation on each individual. This requirement is clearly can not be fulfilled since many patients have been tested for many different weeks in order to obtain the measurement value. However, this data set can not be simply viewed as a longitudinal data for the patients have not been tested every week and different weeks have different combination of patients' measurements. From Fig. 3, the twisted lines between different patients demonstrate that there does not exist a clear trend or character for each patient. In another word, although in fact, the measurements of different weeks for the same patients should be dependent, treat them as independent value is acceptable due to its seemingly random distribution.

5.4 The use of the smoothing method

It might be doubtful about whether the smoothing of the homogeneous reference ranges is necessary or not. There are two reasons for the smoothing. First, as discussed in section 4.2, since the value of 8-iso-PG $F_{2\alpha}$ tends to become increasingly instable as the gestational age increased, it gives us a hint that the gestational age might be a factor of the change of each reference limit. And from the analysis in 4.3, the result of linear regression analysis, more evidence has been built since the estimates of the 97.5% percentile are strongly related to the change of gestational age. Secondly, traced it back to the real life, it is nearly impossible that the estimated reference limit jumping low and high as a zigzag when the gestational age increased. The reason why the homogeneous reference limits appearing like a zigzag in this case is because of our limited sample size.

It should become smoother as the sample size grows larger. As a result, it also demonstrates that the smoothing of the homogeneous reference ranges is indispensable.

5.5 About the goodness-of-fit test

For most of the cases, the smoothing procedures need some tests for the goodness-of-fit to compare which model is the best fitted one. Take the LMS method for example, the Q test as well as the worm plot are the most commonly used test methods for the goodness-of-fit. Specified to our issue, due to our limited sample size and the severe fluctuation among the estimates of different weeks, the object for getting a best fitted model may have a big contradiction against the original intention of smoothing. As a result, none of the goodness-of-fit tests have been illustrated in this essay.

In conclusion, because of the limitation of the data set, many other methods for calculating the homogeneous as well as the gestational-age-specific reference ranges can not be applied to give a comparison between methods. And the test for the goodness-of-fit should also be included to make the estimation procedure completed. Some other experiments should be applied to this data set to make some further estimation. However, the result of table 10 gives us strong evidence that the LMS method is good enough for this specific issue.

References

1. Peter Armitage, Theodore Colton. Encyclopedia of Biostatistics, Volume 4 (1998)
2. Patrick Royston and Eileen M. Wright. A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomial and exponential transformation. *J. R. Statist. Soc.* A (1998) 161, Part 1, pp. 79-101
3. H. Pan and T. J. Cole. A comparison of goodness of fit tests for age-related reference ranges. *Statist. Med.* 2004; 1749-1765 (DOI: 10.1002/sim.1692)
4. Eileen M. Wright and Patrick Royston. A comparison of statistical methods for age-related reference interval. *J. R. Statist. Soc.* A (1997) 160, Part 1, pp. 47-69
5. Huiqi Pan and Tim Cole. User's guide to lmsChartMaker.
6. Roger Koenker and Kevin F. Halllock. Quantile Regression. *Journal of Economics Perspecitves*-Volume 15, Number 4-Fall 2001-Pages 143-156.
7. Patrick Royston. Algorithm AS 181: The W test for normality. *Applied statistics,* 31, pp. 176-180.
8. Maria Palm, Ove Axelsson, Lisa Wernroth and Samar Basu. Oxidative stress is associated with normal human pregnancy: A hallmark of free radical mediated lipid peroxidation in physiology. (manuscript)
9. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992; 11: 1305-1319.