

# Quality Assessment of Fetal Nuchal Translucency Measurements in the First Trimester of Pregnancy

Martin Hynek<sup>1,2,4</sup>, David Stejskal<sup>1</sup>, Jana Zvárová<sup>3,4</sup>

<sup>1</sup>Gennet, Center for Fetal Medicine, Prague, Czech Republic

<sup>2</sup>Department of Gynecology and Obstetrics, Thomayer University Hospital, Prague, Czech Republic

<sup>3</sup>Center of Biomedical Informatics, Institute of Computer Science AS CR, Prague, Czech Republic

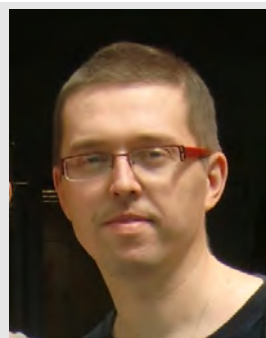
<sup>4</sup>Institute of Hygiene and Epidemiology, First Faculty of Medicine, Charles University, Prague, Czech Republic

## Abstract

**Objectives:** To evaluate and compare the performance of various quality control methods for nuchal translucency (NT) measurements.

**Methods:** Fetal NT measurements performed over a one-year period in a single center were used for the study. The retrospective quality review methods proposed by the Fetal Medicine Foundation (FMF) and the Woman & Infants Hospital of Rhode Island (WIHRI) were assessed in the whole dataset and in sonographer-specific distributions. Further prospective statistic process control (SPC) methods were applied (Shewhart  $\bar{x}$  and  $s$  charts, exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) charts).

**Results:** Three thousand five hundred and seventy eight NT measurements obtained by seven sonographers were eligible for designed analysis. In the assessment of the sonographer-specific NT distributions three of them did not meet due to the underestimation the FMF and one the WIHRI criteria. Using SPC methods, three sonographers presented unsatisfactory performance with underestimation, three sonographers overall satisfactory performance with transient periods of over- and underestimation and one sonographer showed perfect performance.



MUDr. Martin Hynek

**Conclusions:** Assessed SPC methods showed close agreement with the retrospective ones, but with the advantage that they can be applied prospectively allowing the prompt action in case of malperformance. The EWMA and CUSUM methods were regarded as the most suitable.

## Keywords

Nuchal translucency, control chart, statistical process control, cumulative sum, exponentially weighted moving average, prenatal screening

## Correspondence to:

**MUDr. Martin Hynek**  
Gennet, Center for Fetal Medicine  
Address: Kostelní 9, 170 00 Praha 7  
E-mail: martin.hynek@gmail.com

**EJBI 2011; 7(1):22–32**  
received: September 15, 2011  
accepted: October 24, 2011  
published: November 20, 2011

## 1 Introduction

Nuchal translucency (NT) is the sonographic appearance of a collection of fluid under the skin behind the fetal neck and can be identified and measured in the first trimester of pregnancy [1, 2]. It has been shown to be the single most effective marker of trisomy 21 and all other

major chromosomal abnormalities [3, 4]. NT thickness increases with gestational age (GA) and GA is determined by means of crown-rump length (CRL) measurements. In screening for chromosomal aneuploidies patient-specific risk is derived by multiplying the *a priori* maternal age and gestational-related risk by a likelihood ratio, determined from the deviation of the fetal NT measurement

from the normal median for given CRL. One of the usual approaches to quantifying the NT deviation from the normal median, similar to the approach used for laboratory values, is to divide NT measurement by the normal median to produce a multiple of median (MoM) value [5, 6]. In MoM method it is assumed that the distributions of the  $\log_{10}$  transformed MoM values in trisomy 21 and unaffected pregnancies are Gaussian and the ratio of the heights of the distributions at a particular MoM, which is the likelihood ratio for trisomy 21, is used to modify the *a priori* maternal age-related risk to produce a patient-specific risk [5, 7]. Using NT, the detection rate for trisomy 21 for a fixed false-positive rate of 3% in screening by a combination of maternal age and fetal NT reaches about 70%, and in screening by maternal age, fetal NT and biochemical markers free  $\beta$ -human chorionic gonadotropin ( $f\beta$ -hCG) and pregnancy-associated plasma protein-A (PAPP-A) is increased to about 85% [4, 8].

However, NT measurements are displaying the higher variability than biochemical markers due to the lack of automation and significant dependence on the operator [9]. Moreover, it has been reported that even minor deviations in NT measurements cause changes in screening efficacy [10]. Thus, to minimize variability and maintain the satisfactory screening performance, it is of a high importance to have clear international guidelines and ongoing quality review programmes established [2, 11].

International technical guidelines specifying the standardized measurement conditions are provided by the Fetal Medicine Foundation (FMF) which have also set up a training programme with a process of accreditation and ongoing quality control [2]. Basically, quality assurance can be qualitative and quantitative [11, 12, 13]. The former one includes image-scoring system, when individual images are reviewed by an expert, and will not be part of this study. The latter involves the comparison of NT measurements to reference values or their distribution assessment.

The first NT quality review was established by the FMF and was based on annual determination of the proportions above and below certain centiles [2]. In 2008, Palomaki et al. [14] from the Women & Infants Hospital of Rhode Island (WIHRI) proposed to use for the quality review the same three epidemiological parameters that have been proven useful in monitoring biochemical markers, i.e. median NT MoM, logarithmic standard deviation of NT MoM ( $SD \log_{10}(\text{NT MoM})$ ) and percent increase in NT thickness per gestational week.

To look to the issue of quality control more globally, there is a wide range of methods for statistical process control (SPC), originally developed in industry to monitor the quality of manufactured products. Dated back to the year 1926, Walter Shewhart, commissioned by Bell Laboratories to improve the quality of telephones manufactured, developed a simple graphical method [15, 16] - the first of subsequently growing range of SPC charts. Since then, these methods have proven very useful and beneficial in industry [15]. Typical Shewhart control chart is

the chart with center line (CL) representing the average or target value of the quality characteristic and two control lines, upper control limit (UPL) and lower control limit (LCL), representing the interval within which the quality characteristic value should fall with a great probability if the process is 'in control'. In case the values fall outside the control limits, the process is regarded as 'out of control' leading to subsequent investigation of possible causes and corrective actions. The quality characteristic measured by a sample statistic is typically the average  $\bar{x}$  ( $\bar{x}$  chart), standard deviation  $s$  ( $s$  chart), range  $R$  ( $R$  chart) and others [15, 16]. However, these charts are relatively insensitive to small shifts in the process approximately on the order of about  $1.5\sigma$  or less [15]. In this case suitable alternatives represent special types of control charts, such as the cumulative sum control chart (CUSUM) and exponentially weighted moving average chart (EWMA), which are able to detect even small shifts quickly due to the fact that they do not use only information in the last plotted point as Shewhart charts but all the 'historical' information contained in all previous ones.

In the 1970s the first implementation of SPC methods for analysing medical data was reported [17]. Their power to detect the suboptimal clinical performance has been confirmed in various settings - interventional procedures, general, cardiovascular and thoracic surgery, anaesthesia or orthopedics, namely monitoring the success and complication rate of the procedures, mortality as well as infection rates [18, 19, 20]. The SPC methods have the main advantage of being prospective and therefore allow the early detection of deviation from target performance, with prompt feedback and correction. The most commonly SPC method used in medicine is CUSUM [17]. And it was this very method which was first proposed by Biau et al. [17] for the NT quality review, using deviations in millimeters from the expected NT median. Subsequently, Sabria et al. [9] designed a CUSUM models applying NT deviations in MoMs.

The aim of this study is to evaluate the performance of previously applied methods in NT quality control (FMF, WIHRI, CUSUM) with the extension to Shewhart and EWMA charts, using a real dataset of NT measurements from our center. As the screening policy in our center is based on NT deviations in MoMs, all the assessment will regard MoM-based approaches only.

## 2 Patients and Methods

### Patients

The fetal NT measurements performed during a one-year period between July 2010 and June 2011 in the Center for Fetal Medicine Gennet in Prague were used for designed study. From our local database, for each NT measurement we retrieved fetal CRL, date of ultrasound scan and sonographer's name. Only fetuses from singleton pregnancies, without known chromosomal or struc-

tural anomalies and with a CRL between 45 and 84 mm were included in the study. The analysis was restricted to NT measurements between 0.1 and 4.0 mm. Since our center follows the FMF guidelines, as a reference NT median a formula described by Nicolaides et al. [21] was used:  $\log_{10} NT = -0.3599 + 0.0127CRL - 0.000058CRL^2$ ,  $SD \log_{10} (NT \text{ MoM}) = 0.12$ . Each NT measurement was converted into NT MoMs and  $\log_{10} NT \text{ MoMs}$ . Assumption of normality of  $\log_{10} NT \text{ MoMs}$  was assessed using normal probability plot. The statistical analysis was performed using the statistical computing environment R (R Development Core Team 2010) [22] and the additional R packages `qcc` and `iqcc` for quality control charts [23, 24].

### Ultrasound examination

The first-trimester ultrasound scans were performed at GA of 10 weeks<sup>+6 days</sup> - 13 weeks<sup>+6 days</sup> by seven sonographers, labeled A to G, of which sonographers A-F are the FMF accredited and undergoing the regular annual FMF audits and the sonographer G has not been accredited yet. The patients were allocated to the sonographers randomly.

The ultrasound machines Voluson E8 with 4-8-MHz 3D/4D transabdominal probe and Vivid 7 with 7-MHz transabdominal probe (both General Electric Medical Systems, Kretztechnik GmbH & Co, Austria) were used for all examinations. Fetal NT thickness was measured according to the current FMF guidelines [2]:

- the fetal CRL between 45-84 mm,
- the magnification of the image such that fetal head and thorax occupy the whole screen,
- the strict mid-sagittal view of the face, defined by the presence of the echogenic tip of the nose and rectangular shape of the palate anteriorly, the translucent diencephalon in the centre and the nuchal membrane posteriorly,
- the fetus in the neutral position,
- the widest part the translucency measured placing callipers on the inner border of the line that defines the NT.

Example of such an image is presented in Figure 1.

### FMF and WIHRI methods

Two quality review methods based on NT measurement distributions were applied to the whole dataset and to each particular sonographer. The FMF method [2] includes the calculation of the proportion of NT measurements above and below the expected median (expected to be 50%, the acceptable range 40 – 60%), above the expected 95<sup>th</sup> centile and below the expected 5<sup>th</sup> centile (expected to be 5% with the acceptable range 4 – 6%).

WIHRI method [14] examines the median NT MoM (expected to be 1.0, acceptable range 0.90 – 1.10), the

$SD \log_{10} (NT \text{ MoM})$  (expected 0.08 – 0.14) and the rate of increase of NT thickness with advancing gestational age (expected to be 20% per week, acceptable range 15–35%).

### Statistical process control methods

Subsequently, selected SPC methods were applied to our dataset. This was done for each particular sonographer because the interpretation of the whole dataset would be difficult.

The quality characteristic of interest is  $\log_{10} NT \text{ MoMs}$  which is assumed to be normally distributed. For designing of control charts we need to know the target process mean  $\mu$  and standard deviation  $\sigma$  corresponding to the 'in control' process and on the basis of which CL, UCL and LCL are constructed. [15]

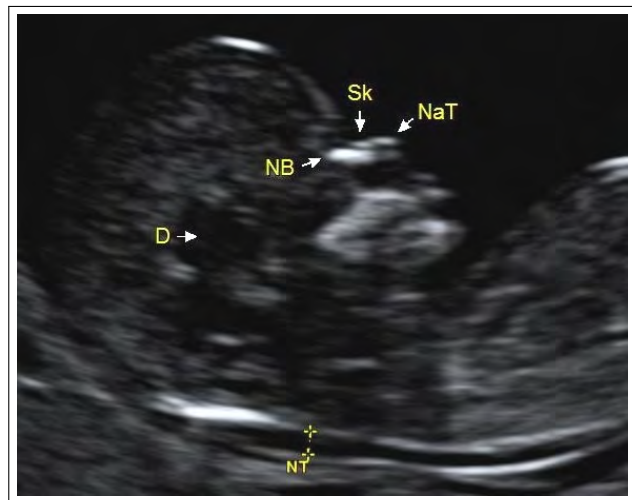


Figure 1: Nuchal translucency measurements according to the Fetal Medicine Foundation protocol. NT, nuchal translucency, Sk, skin, NaT, nasal tip, NB, nasal bone, D, diencephalon.

For this, our target mean is  $\mu = 0$ . The acceptable range outside which the process is considered to be 'out of control' was set to 0.90 – 1.10 NT MoM, the same range used by Palomaki et al. [14] in WIHRI study and by Sabria et al. [9] and based on the extensive knowledge of the impact of inaccuracy on Down syndrome risk estimates using serum markers [25]. The NT MoM interval corresponds to the interval  $\log_{10} NT \text{ MoM} \subset (-0.0458; 0.0414)$ . However, having two different values mean that we are considering differently a process mean shift upward and downward and this is not consistent with the basic principles of SPC [15]. As underestimation is by far more common for NT measurements resulting in a drop in the detection rate of screening test, we decided to set up the acceptable  $\log_{10} NT \text{ MoM}$  interval symmetrically to  $\pm 0.0458$ , corresponding to 0.90 – 1.11 NT MoM, which is entirely clinically acceptable.

The  $SD \log_{10} (NT \text{ MoM}) = 0.086$ , derived from our whole dataset, was markedly lower compared to the 13-year-old Nicolaides's one ( $\sigma = 0.12$ ) [21]. However, if we

used the latter as a reference value, this would lead to control limits which are too loose and do not reflect the real situation. Moreover, if we refer to the literature, we will find that during the years the observed SD is gradually declining: 0.12 in 1998 (dataset  $n = 95476$ ) [21], 0.105 in 2007 ( $n = 23462$ ) [14], 0.079 in 2008 ( $n = 38791$ ) [7]. We believe that the main reason for this fact is the rapid technical improvement of ultrasound equipment with higher resolution and better pre- and postprocessing technologies as well as better training and certification of sonographers participating in regular audits. Therefore, we decided to use as estimate of  $\sigma$  our SD  $\log_{10}$  (NT MoM) = 0.086.

The performance of control charts is evaluated using the average run length (ARL). ARL0 (average run length under the null hypothesis) represents the number of procedures before an extreme value caused by natural variability of the 'in control' process is interpreted as being 'out of control' (corresponds to Type I error). Contrary, ARL1 (average run length under the alternative hypothesis) is the number of procedures which chart shows to be 'in control' despite the process being actually 'out of control' (corresponds to Type II error) [15, 16]. Naturally, the aim is to set control limits to minimize the ARL1 while maximizing the ARL0.

**Shewhart control chart** To control the process mean and variability we used the combination of Shewhart  $\bar{x}$  and  $s$  control charts. Subgroups of  $N$  samples are used to calculate the sample mean  $\bar{x}$  and standard deviation  $s$  and are successively plotted into the diagram together with CL, UCL and LCL. In our case each sample consists of all measurements since the last sample was taken. The use of subgroup samples has the advantage that the variation of sample mean is by  $\sqrt{N}$  lower than the variation in the population from which sample group comes from. Considering  $\bar{x}$  chart, the CL and control limits are placed at [15, 16]

$$UCL = \mu + k \frac{\sigma}{\sqrt{N}} \quad (1)$$

$$CL = \mu \quad (2)$$

$$LCL = \mu - k \frac{\sigma}{\sqrt{N}} \quad (3)$$

where  $k$  is the distance of the control limits from the center line, expressed in standard deviation units. Usually  $k = 3$  is selected for the probability of 0.9973 that 'in control' process will fall within these limits or  $k = 2$  for the probability of 0.9544. Considering our acceptable interval being  $\pm 0.0458$  and  $\sigma = 0.086$ , the minimal necessary sample size  $N$  is from the formula 1:  $0.0458 \geq 0 + k(0.086/\sqrt{N})$ . Thus, for  $k = 3$  is  $N = 32$ , which considering the average number of ultrasound scans would mean too long control interval. A more rational choice is  $N = 15$  for  $k = 2$ , being aware that the Type I error is nearly 5%.

The limits for  $s$  chart are placed at [15]

$$UCL = c_4\sigma + k\sigma\sqrt{1 - c_4^2} \quad (4)$$

$$CL = c_4\sigma \quad (5)$$

$$LCL = c_4\sigma - k\sigma\sqrt{1 - c_4^2} \quad (6)$$

where  $c_4$  is the bias correction constant for the sample standard deviation statistic defined as [16]

$$c_4 = \frac{\Gamma(\frac{N}{2})\sqrt{\frac{2}{N-1}}}{\Gamma(\frac{N-1}{2})} \quad (7)$$

where  $\Gamma(\cdot)$  is the gamma function. For  $N = 15$  is  $c_4 = 0.9823$ . Similarly we selected  $k = 2$ .

To enhance the sensitivity of control charts Shewhart proposed a set of run rules to help to detect nonrandom patterns. In our study we used his rule of seven consecutive points plotted on one side of the center line, having the probability of accidental occurring  $p = 0.5^7 = 0.0078$  [15, 23].

#### Exponentially weighted moving average chart

Contrary to the Shewhart charts where the decision signal obtained depends largely on the last point plotted, using EWMA (sometimes also called 'moving geometric mean') charts the importance of various extent is given to all the previous points. An 'exponentially weighted mean' is calculated each time a new result becomes available [26]:

$$Z_t = \lambda x_t + (1 - \lambda)Z_{t-1} \quad (8)$$

where  $\lambda$  is smoothing coefficient and  $0 < \lambda \leq 1$ , and the starting value of EWMA at time  $t = 0$  is  $Z_0 = \mu$  (the process target). The usually  $\lambda = 0.2$  [26] or  $\lambda = 0.25$  [16] are selected.  $\lambda = 1$  corresponds to Shewhart control chart and the lower the  $\lambda$ , the lower is the reaction of  $Z_t$  to local changes in monitored process and the better is the tendency to emphasize systematic long-term changes [16]. Control limits are placed at [16]:

$$UCL = \bar{x} + k \frac{\sigma}{\sqrt{\lambda/(2 - \lambda)}} \quad (9)$$

$$LCL = \bar{x} - k \frac{\sigma}{\sqrt{\lambda/(2 - \lambda)}} \quad (10)$$

Due to the approximately normal distribution of  $Z_t$ , the choice of  $k$  is similar to the choice in Shewhart charts [16]. For EWMA we used  $\pm 3\sigma$  limit. Since variability of NT measurements is significantly higher than is our 'in control' interval, we set  $\lambda = 0.05$  in order to maximally eliminate local changes and to pick up systematic shift in the process. Using simulated random series of data we verified the eligibility of our settings.

**Cumulative sum chart** Similarly to EWMA, the CUSUM charts utilize all the information contained in-

side previous measurements. The CUSUM test computes, at each time  $t$ , a score  $S_t$  defined by

$$S_t = \max(0; S_{t-1} + W_t) \quad (11)$$

with  $S_0 = 0$  and  $W_t$  the sample weight.  $W_t$  is a measure of the deviation of the observation from the target. At each  $t$  the CUSUM tests the null hypothesis that the process is 'in control' against the alternative one that the process is 'out of control'; practically it happens, if  $S_t$  is equal or greater than a decision limit  $h$ , the null hypothesis is rejected and the process is regarded to be 'out of control'. Until then the null hypothesis cannot be rejected and the process is considered to be 'in control' [17].

A modern CUSUM control charts use the cumulative sum of standardized deviations from target mean  $\mu$ . Consider a standardized variable  $z_t$  [16]:

$$z_t = \frac{x_t - \mu}{\sigma} \quad (12)$$

and two cumulative sums  $S_{H,t}$  to detect positive shift and  $S_{L,t}$  to detect negative shift:

$$S_{H,t} = \max[0; (z_t - K) + S_{H,t-1}], \quad (13)$$

$$S_{L,t} = \max[0; (-z_t - K) + S_{L,t-1}] \quad (14)$$

where starting values  $S_{H,0} = S_{L,0} = 0$ .  $K$  is the reference value defined as  $K = \delta/2$ , where  $\delta$  is the size of the shift one wants to detect in multiples of  $\sigma$ . The usual choice of  $K$  is  $0.5 - 2.0$  ( $1 - 4\sigma$ ) [15, 16, 26].

Control limits are controlled by decision interval  $h$ . Usual choice is  $h = 4$  or  $h = 5$  providing a CUSUM that has good ARL properties against a shift of about  $1\sigma$ . However, for different  $\delta$  Biau et al. [17] recommends that  $h$  is best determined by simulation and compute different ARL0 and ARL1 values by varying  $h$  to obtain acceptable compromise between a very responsive test (short ARL1) and too many false alarms (short ARL0).

For our setting  $\delta = \Delta/\sigma = 0.0458/0.086 = 0.533$ . The decision interval  $h$  was determined using simulation. Thirty series of 20000 random measurements from a normal distribution without any deviation from the expected mean were generated to establish the ARL0 at different  $h$  limits. A further thirty series of 1000 random measurements from normal distribution with the mean located at  $\pm 0.0458$  were simulated to estimate the median ARL1. The  $h$  was then selected to provide the best trade-off between early deviation detection and minimum number of false alarms.

### 3 Results

In total, 3578 NT measurements were eligible for designed analysis. The normal probability plot of  $\log_{10}$  NT MoMs did not violate our assumption of normality. Figure 2 shows the whole data of NT measurements in mm plotted against CRL together with regressed me-

dians expected from the reference distribution and our observed medians. The observed NT median is slightly below the FMF median, suggesting the overall trend towards underestimation in our dataset.

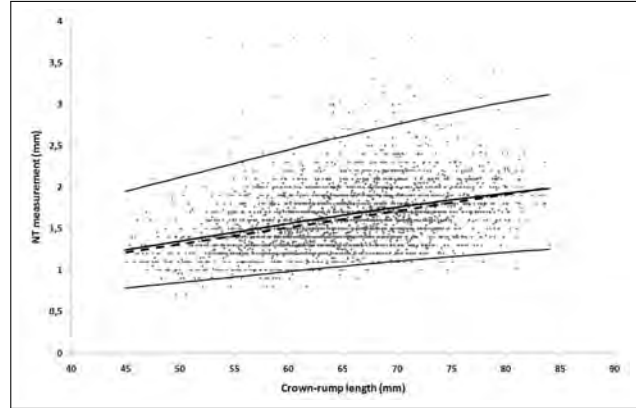


Figure 2: Nuchal translucency (NT) plotted against crown-rump length. ——— expected median and 5<sup>th</sup> and 95<sup>th</sup> centiles of the reference distribution; - - - - - observed median of our measured NT MoMs.

When FMF and WIHRI methods were applied to our whole dataset (see Table 1), the FMF percentage under the median showed underestimation only slightly below the acceptable lower range. However, the proportion below 5<sup>th</sup> and above 95<sup>th</sup> centile were markedly lower than expected, suggesting significantly lower variability within our dataset. All three WIHRI requirements were met. SD  $\log_{10}$  NT MoM from our dataset of 0.086 is indeed markedly lower than  $\sigma$  of 0.12 in the FMF reference serie, thus explaining the results of FMF review method.

Regarding the selection of suitable decision interval  $h$ , the results of simulated random series of measurements are summarized in the Table 2. Value of  $h = 10$  was selected, providing the estimated median number of measurements needed for a false alarm about 1500, and the median number of measurements needed to detect minimum desired deviation at 28.

The results obtained by the five quality review methods - ie. the two retrospective (FMF and WIHRI) and three SPC methods (Shewhart, EWMA and CUSUM charts) - applied to the seven particular sonographers are summarized in the Table 1. The three selected figures with Shewhart, EWMA and CUSUM charts demonstrate three usual performance patterns: unsatisfactory performance (sonographer A in Figure 3), overall good performance with only temporal changes (sonographer C in Figure 4) and finally perfect performance (sonographer F in Figure 5).

Looking more thoroughly to the results, the sonographer A fulfilled the three WIHRI criteria having the median NT MoM at satisfactory 0.95 MoMs suggesting only slight underestimation. However, having 69.8% of measurements below median is slightly below lower FMF acceptable limit. A markedly lower expected number of

Table 1: Results of quality assessment methods of fetal nuchal translucency (NT) measurements.

Method	Whole dataset	Sonographer							Acceptable range
		A	B	C	D	E	F	G	
Number of cases	3578	576	893	641	496	541	320	111	
FMF									
< median (%)	<b>61.2</b>	<b>69.8</b>	56.2	59.8	49.4	<b>72.6</b>	51.6	<b>83.8</b>	40-60
> 95th centile (%)	<b>1.2</b>	<b>0.9</b>	<b>1.0</b>	<b>1.6</b>	<b>2.4</b>	<b>0.9</b>	<b>0.6</b>	<b>0.0</b>	4-6
< 5th centile (%)	<b>1.6</b>	<b>0.9</b>	<b>0.3</b>	<b>1.1</b>	<b>1.4</b>	5.4	<b>0.9</b>	<b>3.6</b>	4-6
WIHRI									
Median NT MoM	0.95	0.93	0.97	0.96	1.01	0.90	0.99	<b>0.83</b>	0.90-1.10
SD log <sub>10</sub> NT MoM	0.086	0.077	0.081	0.084	0.094	0.086	0.087	0.075	0.08-0.13
Weekly increment (%)	19.9	18.4	15.2	26.5	19.0	28.4	21.3	<b>5.6</b>	15-35
Shewhart $\bar{x}$ chart									
Points beyond upper limit (%)	-	0	0	0	9.1	0	0	0	
Upper points violating runs (%)	-	0	0	0	0	0	0	0	
Points beyond lower limit (%)	-	31.6	6.8	2.4	3.0	58.3	4.8	100.0	
Lower points violating runs (%)	-	57.9	11.9	7.1	0.0	69.4	0.0	14.3	
Shewhart $s$ chart									
Points beyond upper limit (%)	-	2.6	3.4	4.8	9.1	2.8	0	0	
Upper points violating runs (%)	-	0	0	0	0	0	0	0	
Points beyond lower limit (%)	-	7.9	3.4	0	3.0	5.6	0	0	
Lower points violating runs (%)	-	7.9	6.8	0	0	0	0	0	
EWMA chart									
Upper limit crossings ( $n$ )	-	0	0	0	1	0	0	0	
Lower limit crossings ( $n$ )	-	20	5	7	4	13	0	1(POL)‡	
Points beyond limit (%)	-	29.0	3.1	3.3	3.6	59.5	0.0	92.8	
CUSUM chart									
Upper limit crossings ( $n$ )	-	0	0	0	1	0	0	0	
Lower limit crossings ( $n$ )	-	1(POL)*	3	5	3	2(POL)†	0	1(POL)‡	
Points beyond limit (%)	-	84.0	7.4	10.0	11.3	67.0	0.0	90.0	

The values lying outside the acceptable range are in bold print. \*Out of limit from 93<sup>rd</sup> measurement. †Out of limit from 177<sup>th</sup> measurement. ‡Out of limit from 9<sup>th</sup> measurement in EWMA and 12<sup>th</sup> measurement in CUSUM.

NT, nuchal translucency, MoM, multiples of median, FMF, Fetal Medicine Foundation method, WIHRI, Women and Infants Hospital of Rhode Island method, EWMA, exponentially weighted moving average, CUSUM, cumulative sum. POL, persistently out of limits.

Table 2: Estimated median number of nuchal translucency (NT) measurements needed until false alarm or deviation detection occurs, for different decision intervals  $h$ , calculated from 30 simulated random series of measurements.

	Decision interval $h$						
	8	9	<b>10</b>	11	12	13	14
First false alarm	323	512	<b>1490</b>	3249	3421	6105	6947
$\pm 0.0458 \log_{10}$ NT MoM mean deviation detection	22	22.5	<b>28</b>	30	31	38.5	48

Our final choice of  $h$  is in bold print. NT, nuchal translucency, MoM, multiples of median.

measurements below the FMF 5<sup>th</sup> and above 95<sup>th</sup> centiles is present in all sonographers with only one exception in the proportion of below 5<sup>th</sup> centile measurements of the sonographer E and will not be further commented. As mentioned previously it is related to the significantly lower variability in our series compared to the FMF one. In relation to the SPC methods (see Figure 3), Shewhart  $\bar{x}$  chart shows underestimation starting from 6<sup>th</sup> sample group onward and manifesting either by points below LCL or points violating the run rule. The 27<sup>th</sup> – 29<sup>th</sup> and further 31<sup>st</sup> and 33<sup>rd</sup> sample groups seem to be 'in control'. The EWMA chart produces a line fluctuating around CL up to the 93<sup>rd</sup> point when it falls below LCL marking the underestimation and staying below until approximately 200<sup>th</sup> measurement. Further course is more or less 'in control' with several temporal periods of underestimations. The CUSUM chart crosses the lower limit at the 93<sup>rd</sup> measurement further gradually bottoming with oc-

casional periods of horizontal course, corresponding very well to the periods in which EWMA shows the 'in control' line. Here it has to be pointed out that the horizontal course of CUSUM line indicates that the observations are scattered around target value although the line is located outside control limits. Shewhart  $s$  chart shows but one point above UCL. As the underestimation in the  $s$  chart representing the lower variability leads to an increase in the screening performance and is thus beneficial, we will not further comment on such results.

In the case of the sonographer B the FMF and WIHRI criteria were fulfilled. The SPC charts displayed the lines oscillating within the limits apart from the final observed period, when  $\bar{x}$  chart presented 'out of control' state from 53<sup>rd</sup> sample group, the EWMA chart from around 817<sup>th</sup> and more profoundly from 872<sup>nd</sup> and the CUSUM from 817<sup>th</sup> observations onward. The other 2 points (12<sup>th</sup>

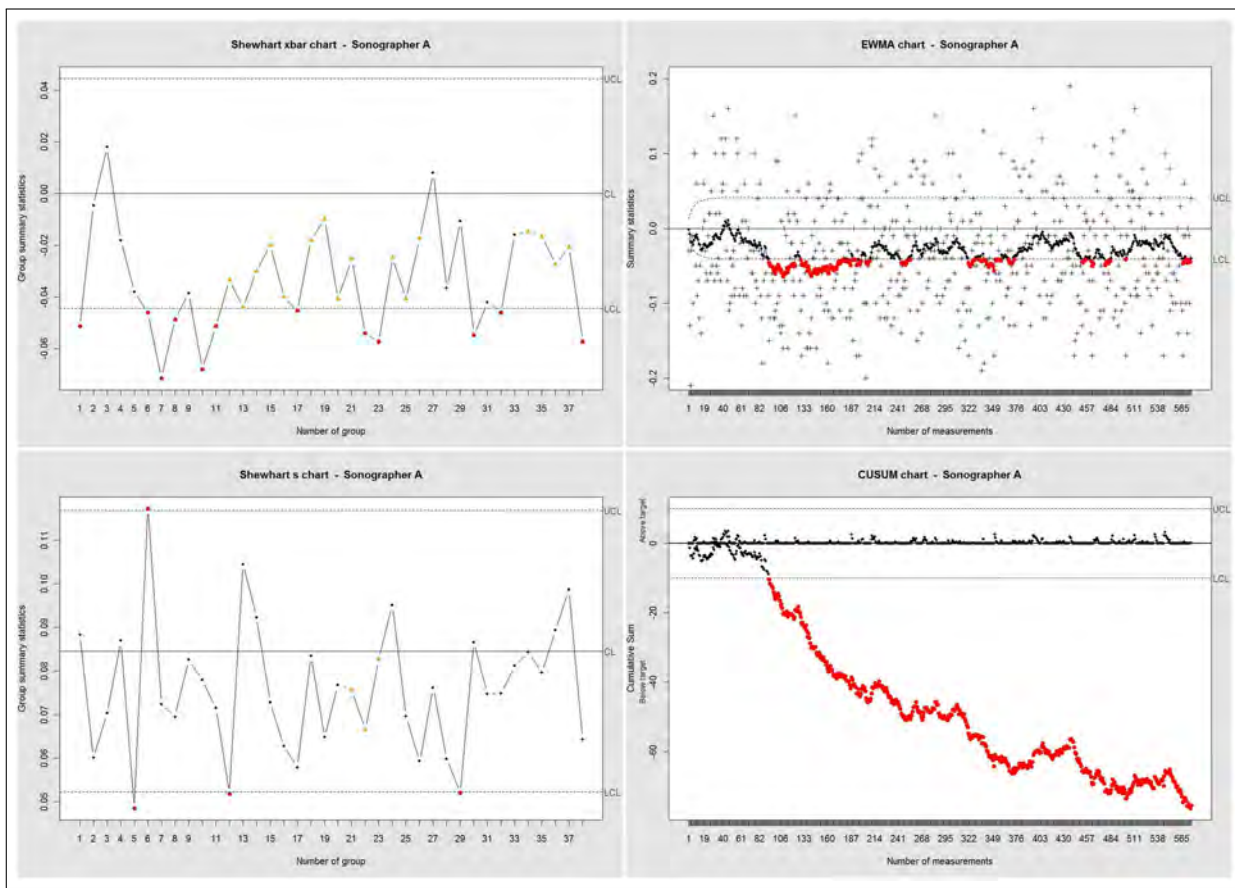


Figure 3: Shewhart  $\bar{x}$  and  $s$  charts, exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) charts for dataset of sonographer A. • point within control limits, • point outside control limits,  $\triangle$  point violating runs. UCL, upper control limit, CL, center line, LCL, lower control limit.

and 26<sup>th</sup>) in  $\bar{x}$  diagram falling below LCL had the corresponding manifestations in EWMA and CUSUM diagrams where the lines almost touched the LCL. The  $s$  chart showed 2 points (3.4%) above the UCL.

Sonographers C (see Figure 4) and sonographer D proved the similar performance. Both satisfactory complied with the FMF and WIHRI limits. However, SPC methods revealed the periods of temporal 'out of control' state. As for the sonographer C the EWMA diagram presented the underestimation of the 38<sup>th</sup> – 51<sup>st</sup>, 129<sup>th</sup> – 135<sup>th</sup> and 305<sup>th</sup> – 308<sup>th</sup> NT measurements. Corresponding response at CUSUM chart had the line sinking below LCL between 40<sup>th</sup> and 75<sup>th</sup>, 128<sup>th</sup> and 142<sup>nd</sup>, 305<sup>th</sup> and 320<sup>th</sup> measurements. The sonographer D overperformed between 236<sup>th</sup> and 246<sup>th</sup> and between 233<sup>rd</sup> and 260<sup>th</sup> measurements looking at EWMA and CUSUM charts respectively and underperformed between 440<sup>th</sup> and 454<sup>th</sup> according to the EWMA and between 435<sup>th</sup> and 468<sup>th</sup> measurements according to the CUSUM charts. In both sonographers the Shewhart  $\bar{x}$  chart displayed in all these mentioned cases of suboptimal performance corresponding points beyond control limits. The proportion of points in  $s$  chart suggesting an increase in variability were 4.8% (2 points) and 9.1% (3 points) regarding the sonographer C and D, respectively.

The sonographer E failed the FMF criteria due to the underestimation, but the WIHRI parameters were within, but close to, the lower limit. Shewhart  $\bar{x}$  chart suggests the underestimation because from 12<sup>th</sup> sample group onward all points are either below LCL or violating the run rule. The same pattern can be seen at the other two diagrams. The EWMA line crossed the LCL at 173<sup>rd</sup> measurement further caterpillaring almost horizontally below LCL with only several very transient returns above LCL. The CUSUM presented a persistent decrement of the lower line from the 177<sup>th</sup> observation, demonstrating similar response to underestimation. There was only one point above UCL in the  $s$  chart.

The sonographer F (see Figure 5) perfectly met the FMF and WIHRI criteria and in concordance with this all SPC charts presented lines rocking between control limits. The only exception was the last 21<sup>st</sup> sample group falling below LCL at Shewhart  $\bar{x}$  chart, which on one hand could be regarded as false positive signal, on the other hand it may be indication of the beginning of 'out of control' as both EWMA and CUSUM diagrams show a line bottoming down to LCL and we know that these charts although more sensitive than Shewhart ones react more slowly.

Contrary to the sonographer F, the sonographer G displayed totally unacceptable performance. There was a

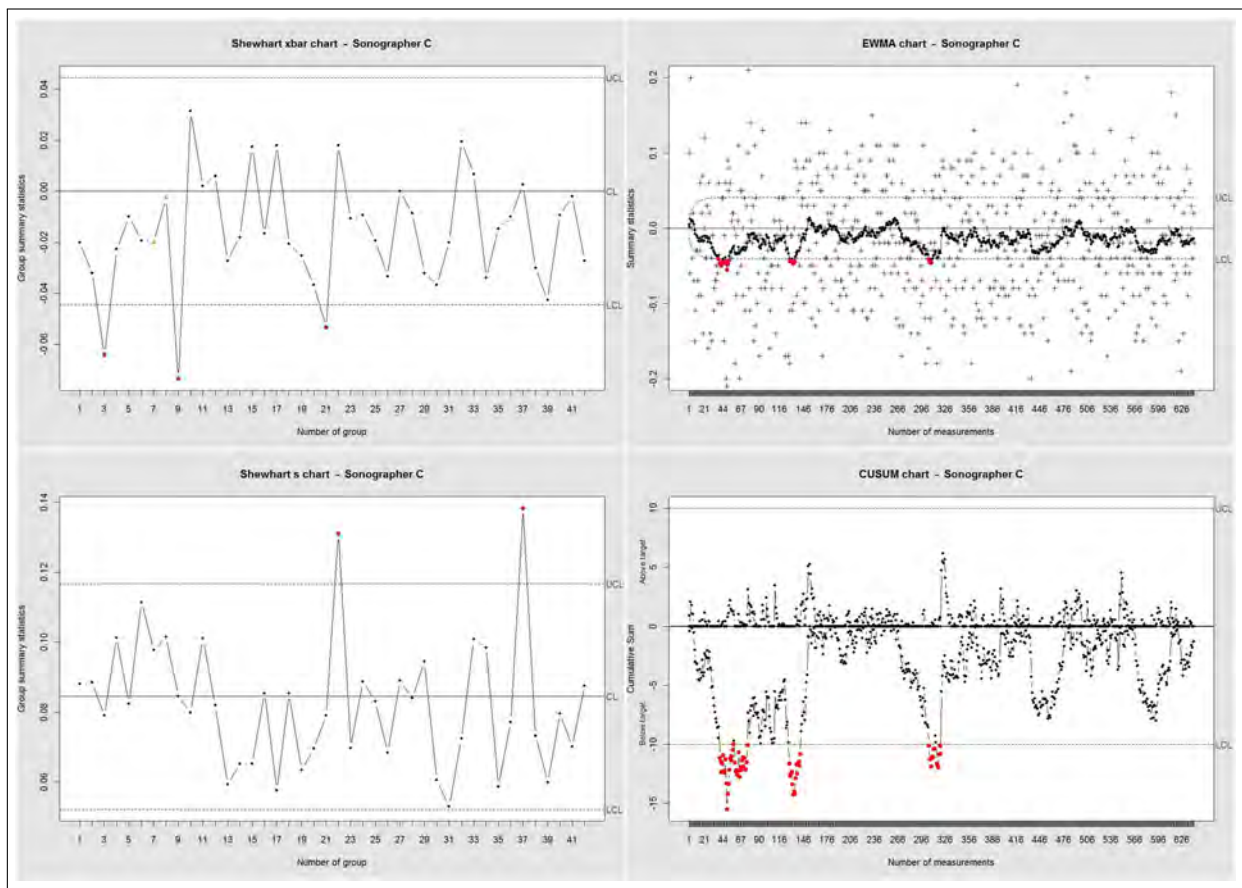


Figure 4: Shewhart  $\bar{x}$  and  $s$  charts, exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) charts for dataset of sonographer C. • point within control limits, • point outside control limits,  $\triangle$  point violating runs. UCL, upper control limit, CL, center line, LCL, lower control limit.

heavy underestimation in FMF and WIHRI criteria. At the same time all  $\bar{x}$ , EWMA and CUSUM charts highlighted the pronounced underestimation, in  $\bar{x}$  diagram from the very 1<sup>st</sup> sample group and from the 9<sup>th</sup> and 12<sup>th</sup> in EWMA and CUSUM charts respectively presenting us how rapidly they are able to expose 'out of control' process.  $S$  chart complied within the limits. It has to be mentioned that the sonographer G has not been accredited by the FMF yet as she is still in the phase of learning the ultrasound scanning.

## 4 Discussion

The contribution of NT as a marker for chromosomal abnormalities is indisputable. Strict quality standards for NT measurements are difficult to follow even for a well-trained and experienced sonographers in tertiary centers [27].

Due to the fact that even minor deviations in NT measurements may have an impact on screening effectiveness (eg. underestimation by 25% may lead to the sensitivity decrease by 1.1%) [10], an establishing of an ongoing audit for NT screening is of a paramount importance [11, 28].

Quality review methods based on distribution parameters (FMF, WIHRI) are easily implemented due to the rather simple methodology. However, the main disadvantages include being retrospective and dealing with all measurements altogether. They are usually performed on the annual basis. Therefore, at the time of audit, a certain number of mothers may have been offered invasive testing unnecessarily (in the case of overestimating) being endangered by the risk of miscarriage which invasive procedures pose or on the other hand some fetuses with Down syndrome may have not been detected prenatally (in case of underestimating). Using the data altogether may lead to undetecting temporal, but significant changes as could be seen in our study for example in the sonographer C (underperforming between 40<sup>th</sup> and 75<sup>th</sup>, 128<sup>th</sup> and 142<sup>nd</sup>, 305<sup>th</sup> and 320<sup>th</sup> measurements) or similarly in the sonographer D (overestimating the 233<sup>rd</sup> – 260<sup>th</sup> measurements, underestimating the 435<sup>th</sup> – 468<sup>th</sup> measurements) but both perfectly fulfilling the criteria of WIHRI. Equally, it has to be pointed out that both methods need a critical minimum number of measurements in order not to be influenced by extreme values.

The presented SPC methods do not have these limitations. They can be used prospectively for ongoing NT audit. Prospectiveness represents an important advantage



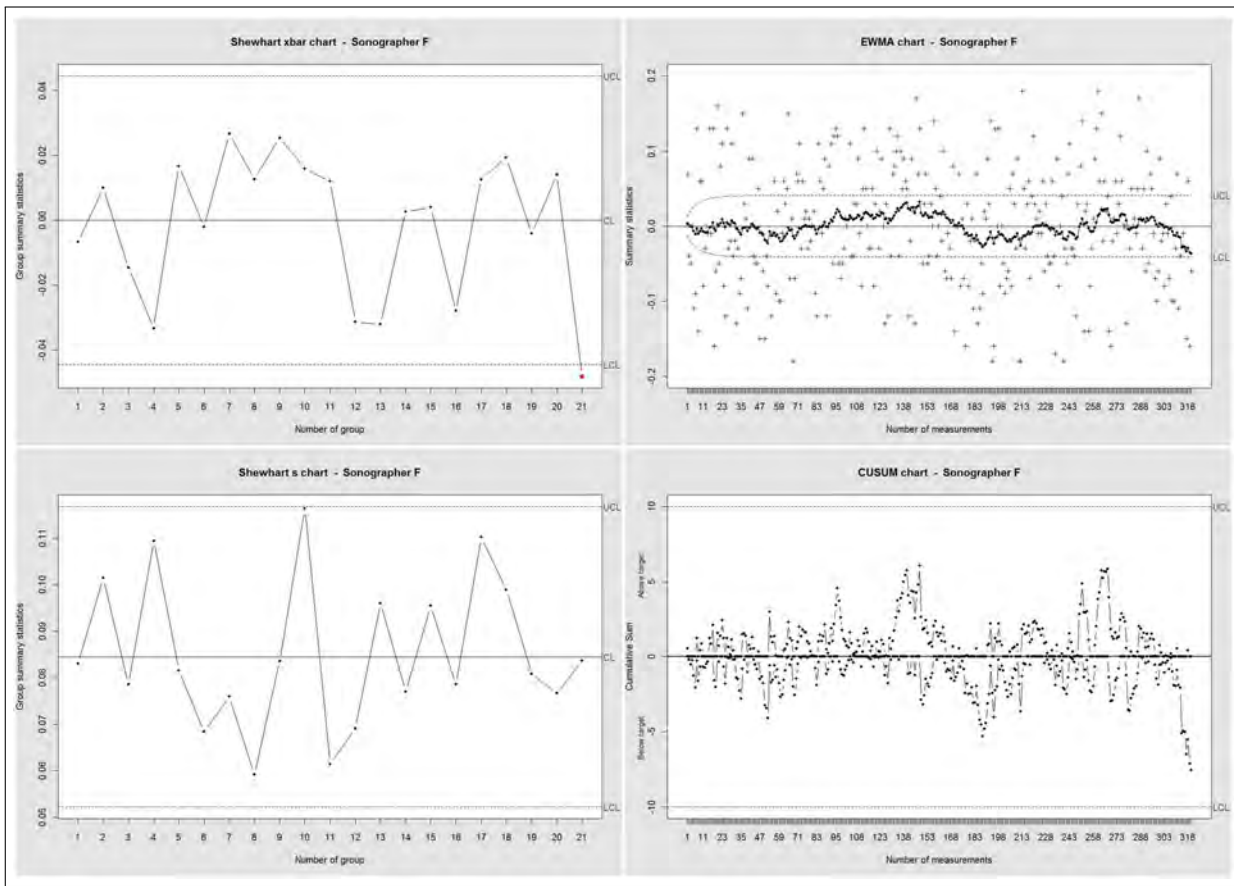


Figure 5: Shewhart  $\bar{x}$  and  $s$  charts, exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) charts for dataset of sonographer F. • point within control limits, • point outside control limits,  $\triangle$  point violating runs. UCL, upper control limit, CL, center line, LCL, lower control limit.

as it allows the early detection of deviation from the target with prompt feedback and correction.

Shewhart control chart have not been to our best knowledge used for NT quality review yet. They are able to control the process accuracy (mean) as well as its precision (variability) showing their change in two separate diagrams. The methodology is simple and straightforward. They are more sensitive to rapid changes within a process [26], however, relatively insensitive to small shifts on the order of about  $1.5\sigma$  or less. To try to overcome these drawbacks, we can implement various kinds of run rules (Shewhart rules, Western Electric rules, etc.), yet making the interpretation more difficult. We observed that implementation of one Shewhart run rule has lead to better ability to detect 'out of control' state.

The other option is to increase the size of sample group. Nonetheless, in our case the desired 'in control' interval is about half the standard deviation and the necessary sample size would be too large considering the average number of ultrasound scans performed by individual sonographer. Increasing the sample size would inevitably mean too long time interval when a subsequent point is added into control diagram. Our suggestion is to lower the control limits to  $\pm 2\sigma$  allowing to have reasonable sample size but at the expense of increasing the risk of false

positive signals at the same time, again making the interpretation more difficult.

EWMA charts have also not been reported to be used for NT quality review. Contrary to the Shewhart charts, EWMA gives the importance of various extent to previous measurements and is able to detect smaller shifts relatively quickly. The amount of weight attributed to the older data and to the last point is controlled by the smoothing parameter  $\lambda$ . We observed that an appropriate  $\lambda$  for NT quality review is 0.05 leading to the low reaction of the chart to local changes and very good ability to emphasize systematic long-term shifts. We have experienced that EWMA charts are capable to reveal the process being 'out of control' very effectively having a very low false positivity rate at the same time (0.27% using  $\pm 3\sigma$  control limits).

Comparing the order of measurements from which the process was considered to be 'out of control' on the basis of EWMA and CUSUM charts the results were almost identical. One of the 'visual' advantage we have noticed is that when the 'out of control' process represented by the curve outside control limits is rectified to 'in control' state the EWMA curve returns back towards center line within the limits in contrast to CUSUM charts where the increasing or declining curve of 'out of control' process switches

to horizontal course remaining in the steady distance from the center line.

CUSUM charts have been repeatedly proved to be advantageous in medicine owing to its simple formulation and very intuitive graphical representation plotting two separate lines, the upper monitoring overestimation and the lower underestimation. It is a very sensitive method for small sustained changes which is able to detect quickly. The CUSUM method has already been proposed as a suitable prospective method for NT quality review; Biau et al. [17] suggested delta-based approach, while Sabria et al. [9] presented the MoM-based one. As most of the fetal medicine centers are using the MoM-based approach for calculating the risk for Down syndrome, the MoM-based CUSUM model is in our opinion more appropriate. Similarly to conclusion of Sabria et al. [9], we have observed that CUSUM represents a very effective and sensitive technique for NT quality review, definitely superior to retrospective FMF and WIHRI methods. It allows to determine the exact time when inaccurate measurements start to occur allowing to find possible causes (eg. change in ultrasound equipment, in scanning routine, etc.) and take corrective actions.

## 5 Conclusion

In our study we have observed that prospective regular NT quality review is of a crude importance. SPC methods represent a powerful tool for this quality process control. They allow a prospective evaluation with a graphically very instructive output. In our opinion the most suitable methods for NT quality control using MoM-based approach are CUSUM as well as EWMA charts as both have the ability to detect the process being 'out of control' very quickly, effectively and with low false positivity rate allowing for prompt correction of the technique when required.

### Acknowledgements

The work was supported by the grant SVV-2011-262514 of Charles University in Prague. Authors would like to thank to Dr. Luca Strucca from Faculty of Economics, University of Perugia, the author of the `qcc` package for R for his advice with the quality control charts and `qcc` package.

## References

- [1] Nicolaides KH, Azar G, Byrne D, Mansur C, Marks K. Fetal nuchal translucency: ultrasound screening for chromosomal defects in first trimester of pregnancy. *BMJ* 1992; 304: 867-869.
- [2] Fetal Medicine Foundation website. URL <http://www.fetalmedicine.com/fmf/training-certification/certificates-of-competence/11-13-week-scan/nuchal/>. [Accesses 1 September 2011]
- [3] Snijders RJ, Noble P, Sebire N, Souka A, Nicolaides KH. UK multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal-translucency thickness at 10-14 weeks of gestation. *Lancet* 1998; 352: 343-346.
- [4] Kagan KO, Wright D, Valencia C, Maiz N, Nicolaides KH. Screening for trisomy 21, 18 and 13 by maternal age, fetal nuchal translucency, fetal heart rate, free  $\beta$ -hCG and pregnancy-associated plasma protein-A. *Hum Reprod* 2008; 19: 1968-1975.
- [5] Spencer K, Bindra R, Nix ABJ, Heath V, Nicolaides KH. Delta-NT or NT MoM: which is the most appropriate method for calculating accurate patient-specific risks for trisomy 21 in the first trimester? *Ultrasound Obstet Gynecol* 2003; 22:142-148.
- [6] Nicolaides KH, Snijders RJ, Cuckle HS. Correct estimation of parameters for ultrasound nuchal translucency screening. *Prenat Diagn* 1998; 18:519-523.
- [7] Wright D, Kagan KO, Molina FS, Gazzoni A, Nicolaides KH. A mixture model of nuchal translucency thickness in screening for chromosomal defects. *Ultrasound Obstet Gynecol* 2008; 31: 376-383.
- [8] Kagan KO, Wright D, Baker A, Sahota D, Nicolaides KH. Screening for trisomy 21 by maternal age, fetal nuchal translucency thickness, free beta-human chorionic gonadotropin and pregnancy-associated plasma protein-A. *Ultrasound Obstet Gynecol* 2008; 31: 618-624.
- [9] Sabria J, Barcelo-Vidal C, Arigita M, Jimenez JM, Puerto B, Borrell A. The CUSUM test applied in prospective nuchal translucency quality review. *Ultrasound Obstet Gynecol* 2001; 37: 582-587.
- [10] Kagan KO, Wright D, Etchegaray A, Zhou Y, Nicolaides KH. Effect of deviation of nuchal translucency measurements on the performance of screening for trisomy 21. *Ultrasound Obstet Gynecol* 2009; 33: 657-664.
- [11] D'Alton ME, Cleary-Goldman J, Lambert-Messerlian G, Ball RH, Nyberg DA, Comstock CH, Bukowski R, Berkowitz RL, Dar P, Dugoff L, Graigo SD, Timor IE, Carr SR, Wolfe HM, Dukes K, Canick JA, Malone FD. Maintaining quality assurance for sonographic nuchal translucency measurement: lessons from the FASTER Trial. *Ultrasound Obstet Gynecol* 2009; 33: 142-146.
- [12] Snijders RJM, Thom EA, Zachary JM et al. First-trimester trisomy screening: nuchal translucency measurement training and quality assurance to correct and unify technique. *Ultrasound Obstet Gynecol* 2002; 19: 353-359.
- [13] Wojdemann KR, Christiansen M, Sundberg K, Larsen SO, Shalmi A, Tabor A. Quality assessment in prospective nuchal translucency screening for Down syndrome. *Ultrasound Obstet Gynecol* 2001; 18: 641-644.
- [14] Palomaki GE, Neveux LM, Donnenfeld A, Lee JE, McDowell G, Canick JA, Summers A, Lambert-Messerlian G, Kellner LH, Yebelman A, Haddow JE. Quality assessment of routine nuchal translucency measurements: A North American laboratory perspective. *Genet Med* 2008; 10: 131-138.
- [15] Montgomery DC, Runger GC. Applied statistics and probability for engineers. 3<sup>rd</sup> ed. United States of America: John Wiley & Sons, Inc.; 2002.
- [16] Meloun M, Militký J. Kompendium statistického zpracování dat. Praha: Academia; 2006.
- [17] Biau DJ, Porcher R, Salomon LJ. CUSUM: a tool for ongoing assessment of performance. *Ultrasound Obstet Gynecol* 2008; 31: 252-255.

- [18] Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard RS, Porcher R. Quality control of surgical and interventional procedures: a review of the CUSUM. *Qual Saf Health Care* 2007; 16: 203-207.
- [19] Noyez L. Control charts, cusum techniques and funnel plots. A review of methods for monitoring performance in healthcare. *Interact Cardiovasc Thorac Surg* 2009; 9(3): 494-9.
- [20] Sibanda T, Sibanda N. The CUSUM chart method as a toll for continuous monitoring of clinical outcomes using routinely collected data. *BMC Med Res Methodol* 2007; 7: 46.
- [21] Nicolaidis KH, Snijders RJM, Cuckle HS. Correct estimation of parameters for ultrasound nuchal translucency screening. *Prenat Diagn* 1998; 18: 519-523.
- [22] R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [23] Scrucca, L. Qcc: an R package for quality control charting and statistical process control. *R News* 2004; 4(1): 11-17.
- [24] Recchia DR, Barbosa EP., Goncalves EJ(2010). IQCC: Improved Quality Control Charts. R package version 0.5. URL <http://CRAN.R-project.org/package=IQCC>. [Accesses 1 September 2011]
- [25] Knight GJ, Palomaki GE. Epidemiologic monitoring of prenatal screenign for neural tube defects and Down syndrome. *Clin Lab med* 2003; 23: 531-551.
- [26] Oakland JS. Statistical process control. 5<sup>th</sup> ed. Cownwall: MPG Books Limited; 2003.
- [27] Ville Y. Semi-automated measurement of nuchal translucency thickness: blasphemy or oblation to quality? *Ultrasound Obstet Gynecol* 2010; 36: 400-403.
- [28] Evans MI, Van Decruyes H, Nicolaidis KH. Nuchal translucency measurements for the first-trimester screening: the 'price' of inaccuracy. *Fetal Diagn Ther* 2007; 22: 401-404.