Chapter 7

Influence of cut-off value on the prevalence of short cervical length

M.A. van Os
C.E. Kleinrouweler
E. Schuit
A.J. van der Ven
E. Pajkrt
C.J.M. de Groot
B.W. Mol
M.C. Haak

Submitted
Abstract

Objective: To assess the distribution of cervical length (CL) and explain the low prevalence of short CL ≤ 30 mm measured in a large cohort of women without a history of preterm birth; The Triple P study (NTR-2078).

Methods: CL was measured in a cohort study with the aim to investigate the predictive capacity of cervical length measurement, to identify women at increased risk for preterm birth. Women with short CL were included in a randomized clinical trial evaluating the effect of progesterone compared to placebo. In total 57 centres and 20,234 women participated in this study. Cut off value for a short CL was defined as ≤ 30 mm based on existing literature. The distribution of CL was assessed for each individual centre and measurements were compared between levels of care: primary (29 ultrasound centres), secondary (21 general hospitals) and tertiary care institutions (7 university medical centres). Comparison was also performed for centres with a low, intermediate and high volume of CL measurements. Furthermore the distributions before and after publicly addressing the prevalence of short CL were analysed (12,284 vs.7,950 women).

Results: Between December 2009 and August 2013, 20,234 women had CL measurements of who (1.8%) had a short cervix. Mean cervical length was 44.2 mm (SD 7.8 mm). A ‘dip’ was observed in the 20 – 30 mm CL window, defined by the ratio of CL measurements between the original and expected data being below 50%. The ‘dip’ seemed to be present in the 89% of participating centres. All levels of care and low, intermediate and high-volume centres had a ‘dip’ below 30 mm. A significant difference was found in the distributions over time before and after publicly addressing the low prevalence of short cervix (1.7% and 2.0% of the measurements were 30 mm, respectively (p<0.001)).

Conclusions: A cut-off value was used to include women with a short cervix in a randomised clinical trial which was embedded in a cohort study. We suggest that the use of a pre-defined cut-off value influenced the distribution of the CL measurements. Since the measurement is not blinded, non-blinded assessors might favour the control intervention, inducing selection bias. This might have resulted in too little measurements around the cut-off value. Other trials using similar designs could benefit from this observation and take precaution to avoid selection bias.
Introduction

Preterm birth, defined as birth before 37 weeks of gestation, occurs in 5-13 % of all pregnancies\(^1\)\(^2\). The prevalence of preterm birth in The Netherlands is 7.7\%\(^3\). Spontaneous preterm delivery is an important cause of perinatal mortality. The majority of spontaneous preterm births occur in low-risk women\(^4\). So far, interventions for threatened preterm have shown limited effectiveness\(^5\)-\(^7\). Identification of low-risk women who will deliver prematurely is crucial in the development of preventive strategies. A mid-pregnancy short cervical length (CL), measured by ultrasound, can predict spontaneous preterm birth\(^4\),\(^8\),\(^9\) and is currently the most powerful screening instrument available\(^8\),\(^9\). The shorter the cervix, the higher the risk of preterm birth\(^4\). Heath et al. found that measurement of cervical length was highly reproducible and, on 95\% of occasions, the difference between two measurements by the same observer and by two observers was 3.5 mm or less and 4.2 mm or less, respectively. The variability of measurements was less when the cervical length was shorter\(^10\). Although cervical length is reported to follow a normal distribution, different values for 1\(^{st}\) and 5\(^{th}\) percentile of CL measurement have been mentioned ranging from 11-15 mm and 23-30 mm respectively\(^7\),\(^10\)-\(^13\). Explanations for differences in CL include dynamic changes throughout the pregnancy, gestational age at measurement, parity and measurement techniques\(^14\),\(^15\). Moreover, the prevalence of CL below the 5\(^{th}\) percentile also varies in literature\(^7\),\(^9\),\(^10\)-\(^13\). In light of the above mentioned facts, it seems important to determine the right cut-off value in studies concerning CL in pregnancy. The aim of this study is to analyse the distribution of CL measurements in asymptomatic low-risk women with a singleton pregnancy and explain the low prevalence of short CL ≤ 30 mm measured in a large cohort study of women without a history of preterm birth.
Chapter 7

Methods

This is a secondary analysis of a multicentre cohort study with a randomized clinical trial embedded (Triple P trial; NTR-2078). The aim of the Triple P study was to investigate the predictive capacity of cervical length measurement to identify women with a low risk singleton pregnancy, for an increased risk for preterm birth. Cervical length was measured in asymptomatic low-risk women with singleton pregnancies at the 20 weeks standard anomaly scan\textsuperscript{16}. A short cervix was defined as $\leq$ 30 mm. CL measurements were performed as described by the FMF and comparable to earlier studies on this subject\textsuperscript{7,17-19}. Women with short CL were eligible for a randomized clinical trial to evaluate the effectiveness of vaginal progesterone in reducing adverse neonatal outcome due to preterm birth\textsuperscript{20}. The Triple P study was set within the infrastructure of the Dutch Obstetric Consortium for research in Women’s health and was performed between November 2009 and August 2013. The institutional review board of the Academic Medical Centre, Amsterdam, The Netherlands (MEC AMC 08/374) approved the study. Methods of the cohort study and the embedded randomised clinical trial are described elsewhere\textsuperscript{21}.

In this study 57 ultrasound units participated, of which 29 were ultrasound centres (primary care institutions), 21 were general hospitals (secondary care institutions) and seven were university hospitals (tertiary care institutions). The primary outcome of the prospective cohort study was spontaneous preterm birth before 37 weeks of gestation. Secondary outcomes were preterm birth before 34, 32 and 28 weeks of gestation. The sonographers in the Triple P study were all licensed for the standard anomaly scan. According to the regulations of the national screening program in the Netherlands, it is compulsory for these ultrasonographers to perform at least 150 anomaly scans a year. Additionally, prior to the introduction of the study, all participating sonographers were trained in CL measurements. The training program comprised an e-learning module, specifically designed to teach the cervical length measurement technique\textsuperscript{22}. The ultrasonographers had to pass this e-learning module with a satisfactory grade. Furthermore, they had to prove their ability to perform CL measurement by the submission of five pictures of CL measurements, which had to be approved by the project team. Women were asked to empty their bladder prior to the measurement. The shortest measurement was recorded for further use and entered in a web-based database.
We extracted the CL measurements of this prospective cohort and analysed the distribution of CL measurements for the entire cohort, but also separately for CL measurements performed within different levels of care and for high and low-volume centres. Low volume was defined as centres that measured less than 500 CL's, intermediate volume as centres that measured between 500 and 1000 CL's and high volume was defined as centres that measured >1000 CL's. Centres with a volume of 30 measurements or less were excluded from individual analysis, because a minimum number of CL measurements is necessary to enable the assessment of a normal distribution. Additionally a year by year analysis of the measurements was performed. The CL measurements were plotted in histograms and compared between volume of measurements and the different levels of care; respectively primary care institutions (N=29), secondary care institutions (N=21) and tertiary care institutions (N=7). Histograms were pursued in the course of time during the trial as a continuous monitoring process for quality and progress of the inclusions. In December 2011 the retrieved data of the monitoring process were made public during a national attended symposium in which all sonographers were invited. This symposium was organised for teaching purposes and designed for sonographers, midwives and residents. We compared the distribution before and after the symposium (when respectively 12,284 and 7,950 women were included). The distribution of cervical lengths was tested for normality using the Kolmogorov-Smirnov test. Unpaired Student’s t-tests were used to assess the differences in mean CL between centres. Furthermore, a normal distribution of CL was simulated based on the mean and the SD of the original data in order to compare the observed distribution of the CL measurement. Statistical analysis was performed using IBM SPSS Statistics, Version 21. Distribution of CL was simulated using R, version 3.1.1 (R Foundation for Statistical Computing).

Results

Between December 2009 and August 2013, 20,234 women were screened. Of the 20,234 women, (1.8%) had a cervical length ≤ 30 mm. The mean CL was 44.2 (SD 7.8). The distribution of the total number of measurements (Figure 1) was significantly different from a normal distribution (p< 0.001). Compared to a simulated distribution, a CL ≤ 30 mm was found less in the original distribution (990 (4.9%) vs. 367 (1.8%), p<0.001). The comparison reveals a higher and lower number than expected CL measurement at certain cervical lengths (Figure 2).
Figure 1. Histogram of cervical length (CL) measurements of the entire cohort of 20,234 women.

Figure 2. The number of simulated CL measurements as a percentage of the original CL measurements. The horizontal lines at 100% indicates perfect agreement between both distributions. Cervical lengths for which less than ten measurements were simulated were left out of this figure.

The difference, or ‘dip’, between both distributions was most profound in the 20 – 30 mm CL window as indicated by the ratio of CL measurements between the original and simulated data being below 50% (Figure 3). Additionally, a higher than expected number of CL measurements was found for cervical lengths ≥62 mm (Figure 3).
Influence of cut-off value on the prevalence of short cervical length

**Figure 3.** Simulation of normal distribution of CL measurements, and the difference, between the actual and simulated curve.

A yearly analysis of the CL measurements revealed cumulative distributions of 2009, 2010, 2011, 2012 and 2013 that were all significantly different from a normal distribution (respectively p = 0.005, p< 0.001, p< 0.001, p< 0.001, p< 0.001). The dip below 30 mm was seen from October 2010 onwards. At that time 2360 CL measurements were performed. This dip persisted until the end of the study in August 2013 (Figure 4).

**Figure 4.** Histograms of CL distributions of each cumulative year from 2009 until 2012, the cumulative histogram from 2009 until 2013 is shown in figure 1.
Chapter 7

Hence, showing the preliminary results of the Triple P study on a national attended symposium in December 2011 had no influence on the overall distribution. However, a significant discrepancy ($p<0.001$) of the CL measurements surfaced when the means of the distribution of the study before and after the symposium were compared. The dip was apparent in both distributions, but less prominent in the distribution after the symposium. The percentage of $CL \leq 30$ was 1.7% before and 2.0% after the symposium (Figure 5).

**Figure 5.** Distribution of CL measurements before and after the national symposium.

In a per-centre analysis ($n=57$), 10 centres were excluded because of an insufficient number CL measurement to assess the normality of the distribution. Of the remaining 47 centres, CL was normally distributed in 11 centres. In 42 centres the 20 – 30 mm dip in the measurement was present. Data analysis for primary, secondary and tertiary care institutions did not show a normal distribution for each type of care level ($p < 0.001$). A significant difference in mean CL between primary care institutions and tertiary care institutions ($p<0.001$) and secondary and tertiary care institutions ($p<0.001$) was present. There was no significant difference in CL distribution between primary and secondary care institutions ($p= 0.151$). We found 1.8% of $CL \leq 30$ mm in primary care institutions, 2.4% in secondary care institutions and 2.4% in tertiary care institutions. The dip between 20-30 mm was seen in all three care levels, but was more pronounced in the primary care institutions and the tertiary care institutions (Figure 6).
Influence of cut-off value on the prevalence of short cervical length

Figure 6. Histograms of CL measurements in primary secondary and tertiary care institutions.

If the data were analysed according to a low, intermediate and high volume of the CL measurements, all distributions were significantly different from a normal distribution (p=<0.001). The mean of the CL’s were significantly different between low-volume and intermediate-volume centres (p<0.001) and low-volume and high-volume centres (p<0.001). No difference between the CL distributions of intermediate and high-volume centres (p=0.06) was present. The percentage CL ≤30 mm was 1.8%, 3.0% and 1.4 respectively for low, intermediate and high-volume centres. The dip between 20-30 mm was seen in all volume groups but most prominently in the high-volume centres (Figure 7).
Chapter 7

Discussion

In this secondary analysis of the Triple P study we aimed to analyse the distribution of CL measurements of this large cohort of women without a previous preterm birth. The distribution of CL measurements was significantly different from a normal distribution, with a low prevalence of CL ≤ 30 mm (1.8%) and a high mean CL (44.2 mm) compared to existing literature. The absence of a normal distribution was seen in all levels of care and in high and low-volume centres. We found a typical 'dip' below the cut-off value of 30 mm in the distribution of CL measurements in the entire cohort and the majority (89%) of individual centres. The dip was most pronounced in primary and tertiary care centres and in high-volume centres. The distribution of the CL changed after preliminary results of the study (showing the

Figure 7. Histograms of CL measurements performed in centres including < 500 CLs, 500-1000 CLs, and > 1000 CL measurements.
Influence of cut-off value on the prevalence of short cervical length

low prevalence of CL measurements) were disclosed at a national symposium. Finding a CL above or below the cut-off value, with the consequence to have to offer participation in trial or not, seems to change the CL distribution, creating a 'dip' and reducing the number of short CL's. The mean CL in our data is 6 to 9 mm longer than the previous described 33-38.4 mm\textsuperscript{10-13}. The most probable explanation is the fact that our study population comprises solely of women without a history of preterm birth. Previous histograms were constructed in hospital populations including women with a history of preterm birth\textsuperscript{7,10-13}. Other possible causes might be gestational age at measurement, parity and differences in the ultrasonographic measurement. Our protocol was, however, identical to previous described methods in which the calipers were placed at the distance between the triangular area of echo density at the external os and the V-shaped notch at the internal os\textsuperscript{22}. Gestational age at measurement did not differ from other studies on the subject\textsuperscript{7,10-13}. In our cohort we found a higher number of women with an advanced maternal age and more women from a European origin. These maternal characteristics were found to give longer cervical length, the influence in the whole cohort was, however, only limited\textsuperscript{23}. Therefore the true low-risk population is the most probable explanation for our high mean CL. The second finding of this study was the dip in the distribution below the cut-off value for randomisation. We hypothesize that this can be attributed to the predefined cut-off value due to an observer bias of the ultrasonographers who were not blinded for the measurement they performed. The fact that the gap became less prominent after we addressed this finding at a large national symposium does support our hypothesis that it is a psychological effect of the observer. Additional arguments for this hypothesis are the finding of an overshoot in the number of CL measured above 30 mm compared to a simulated normal curve and the fact that other studies\textsuperscript{10-13} never showed such a gap around this value before. Other studies were, however, never undertaken with an intervention following the measurement. Earlier research showed that observer bias might be important when the outcome assessor has a strong predisposition about the outcome and when the outcome involves personal judgment such as qualitative scores or image recognition\textsuperscript{24}. Non-blinded outcome assessors usually support the experimental intervention. Hrobjartsson et al found that non-blinded outcome assessors exaggerate the hazard ratio by an average of approximately 27%. In certain situations, they support the control interventions, inducing a comparable degree of observer bias in the
reversed direction. The direction of observer bias depends on how the clinical and public contexts have shaped the predispositions of its investigators. It has been shown that it is not unlikely to measure a CL longer than it is. In a study about learning results of ultrasonographers, sonographers were more likely to overestimate a cervical length than to underestimate it. In conclusion, a cut-off point to which an intervention is implemented, may affect the distribution of measurements. More importantly, it effects the selection of women on the basis of CL in a randomised clinical trial and indicates possible problems like under-representation of women with a short CL. A cut-off dip has never been demonstrated so clearly before and observer bias seems to be a possible explanation. Decision making processes of ultrasonographers is an unexplored area and research in this area has never been performed. The decision making process might also effect other fields in which cut off values are important like prenatal screening and diagnosis. Other trials using similar designs could benefit from this observation.

Statement of contribution

MO wrote the first draft of the paper. MO, EK collected data. ES analysed the data. AJ, EP, CG, BM, MH critically revised the manuscript for important intellectual content. All authors approved of the final version of the manuscript to be submitted.
Influence of cut-off value on the prevalence of short cervical length

References

1. Iams JD, Romero R, Culhane JF, Goldenberg RL. Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. Lancet 2008; 371: 164–75.