Ultrasound Obstet Gynecol 2017; **49**: 330–336
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.15967

Influence of cut-off value on 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. J. MOL¶ and M. C. HAAK**

**Department of Obstetrics and Gynaecology, VU University Medical Center Amsterdam, Amsterdam, The Netherlands; †Department of Obstetrics and Gynaecology, Academic Medical Center Amsterdam, Amsterdam, The Netherlands; ‡Department of Epidemiology Julius Center, University Medical Center Utrecht, Utrecht, The Netherlands; §Stanford Prevention Research Center, Stanford University, Stanford, CA, USA; ¶The Robinson Institute, School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, Australia; **Department of Obstetrics, Leiden University Medical Center, Leiden, The Netherlands*

KEYWORDS: cervical length; prevalence; ultrasound

ABSTRACT

Objective *To assess the distribution of cervical length (CL) in a large cohort of asymptomatic low-risk women with singleton pregnancy and no previous preterm birth and to explain the low prevalence of short CL ≤ 30 mm in this cohort.*

Methods *This was a secondary analysis of a multicenter cohort study with an embedded randomized controlled trial (Triple P trial; NTR-2078) on the prevention of preterm birth with progesterone. In the cohort study, CL was measured in asymptomatic low-risk women with singleton pregnancy to investigate its predictive capacity to identify those at increased risk for preterm birth. A short CL was defined by a cut-off value of ≤ 30 mm, based on existing literature. Women with a short CL were subsequently included in a randomized controlled trial evaluating the effect of progesterone, compared with placebo, on preterm birth. In total, 57 centers and 20 234 women participated in the study. Normal distributions for CL were simulated based on the mean and SD of the original data. The distribution of CL was assessed for each individual center and measurements were compared between levels of care: primary (29 ultrasound centers), secondary (21 general hospitals) and tertiary (seven university medical centers) care institutions. Comparison was also performed between centers with low, intermediate and high volume of CL measurements. CL distributions before (*n = *12 284 women) and after (*n = *7950 women) a national symposium, at which the prevalence of short CL measurements was addressed publicly, were analyzed.*

Results *Between November 2009 and August 2013, 20 234 women had CL measurements, of whom 367 (1.8%) had a short CL. Mean ± SD CL was 44.2 ± 7.8 mm. A 'dip' in the distribution of CL measurements between 20 and 30 mm was observed, defined by a ratio of < 50% when comparing the number of measurements in observed and simulated normal distributions. The dip was present in 89% of participating centers. All centers showed a dip in the distribution of measurements ≤ 30 mm when analyzed according to the level of care and volume of measurements. A significant difference was found when comparing the distribution before and after publicly addressing the low prevalence of short CL (1.7% vs 2.0% of measurements were ≤ 30 mm, respectively; P < 0.001).*

Conclusions *A cut-off value of 30 mm for CL was used to include women in a randomized clinical trial that was embedded in a cohort study. We suggest that the use of a predefined cut-off value for a short cervix influences the distribution of the CL measurements. Since the measurement is not blinded, preference of assessors for the control or intervention arms may have introduced selection bias. This might have resulted in fewer measurements around the cut-off value. Other trials using similar designs could benefit from this observation and take precautions to avoid selection bias. Copyright © 2016 ISUOG. Published by John Wiley & Sons Ltd.*

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

Correspondence to: Dr M. A. van Os, VU University Medical Center, Department of Obstetrics and Gynaecology, 8 F Boelelaan 1117, 1084 HV Amsterdam, The Netherlands (e-mail: m.vanos@vumc.nl)

Accepted: 29 April 2016

preterm births occur in low-risk pregnancies[4]. So far, interventions for threatened preterm birth have shown limited effectiveness[5–7].

Identification of low-risk women who will deliver preterm 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.*[10] found that measurement of CL was highly reproducible and, on 95% of occasions, the difference between two measurements by the same observer and by two different observers was ≤ 3.5 and ≤ 4.2 mm, respectively. The variability of measurements was less when the CL was shorter[10]. Although CL is reported to follow a normal distribution, different values for the 1st and 5th percentiles of CL measurement have been reported, ranging from 11 to 15 mm and 23 to 30 mm, respectively[7,10–13]. Explanations for these 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 5th percentile also varies in the literature[7,9–13]. In light of the abovementioned facts, it seems important to determine an appropriate cut-off value in studies concerning CL in pregnancy.

The aim of this study was to analyze the distribution of CL measurements in asymptomatic low-risk women with a singleton pregnancy and without a history of preterm birth and to explain the low prevalence of short CL ≤ 30 mm measured in this large cohort of women.

METHODS

This was a secondary analysis of a multicenter cohort study with an embedded randomized clinical trial (Triple P trial; NTR-2078). The aim of the Triple P study was to investigate the predictive capacity of CL measurement to identify women with a low-risk singleton pregnancy who are at an increased risk for preterm birth. CL was measured in asymptomatic low-risk women with a singleton pregnancy at the time of their 20-week standard anomaly scan[16]. A short cervix was defined as CL ≤ 30 mm. CL measurements were performed as described by The Fetal Medicine Foundation and were comparable to earlier studies on this subject[7,9,17–19]. Women with a short CL were eligible for inclusion in a randomized clinical trial to evaluate the effectiveness of vaginal progesterone in reducing adverse neonatal outcome due to preterm birth[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 Center, Amsterdam, The Netherlands (MEC AMC 08/374) approved the study. Methods of the cohort study and the embedded randomized clinical trial are described elsewhere[21].

In this study 57 ultrasound units participated, of which 29 were ultrasound centers (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 to carry out the standard anomaly scan. According to the regulations of the national screening program in The Netherlands, it is compulsory for these sonographers to perform at least 150 anomaly scans per year. Additionally, prior to commencement of the study, all participating sonographers were trained in measuring CL. The training program comprised an e-learning module, specifically designed to teach the CL measurement technique[22]. The sonographers were required to pass this e-learning module with a satisfactory grade. Furthermore, they had to prove their ability to perform CL measurement by submitting five images of CL measurement, which had to be approved by the project team. CL was measured during the standard anomaly scan at 18–22 weeks' gestation using transvaginal ultrasound. Women were asked to empty their bladder prior to measurement. The ultrasound image had to display an empty bladder, full length of the endocervical mucosa in an exact mid-sagittal plane of the cervix, and an equal thickness of anterior and posterior cervical walls. The calipers were placed between the triangular area of echodensity at the external os and the V-shaped notch at the internal os. When there was a curved aspect of the endocervical mucosa, calipers were placed on the external and internal ora as described previously and a straight or curved line made between the calipers. In the case of funneling, the funnel itself was not included in the measurement. Sonographers were instructed to measure CL approximately 3 min after insertion of the vaginal probe and to document the shortest measurement, which was entered in an online database.

We extracted the CL measurements of this prospective cohort and analyzed the distribution of measurements for the entire cohort, and separately for CL measurements performed at centers with different levels of care and for centers with high and low volume of measurements. Low, intermediate and high volumes were defined as centers that measured a total of < 500 CLs, 500–1000 CLs and > 1000 CLs in the study period, respectively. Centers with a volume of ≤ 30 measurements 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.

CL measurements were plotted on histograms and compared between centers according to volume of measurements and the different levels of care: primary ($n = 29$), secondary ($n = 21$) and tertiary ($n = 7$) care institutions. Histograms were analyzed throughout the trial as a continuous monitoring process for quality

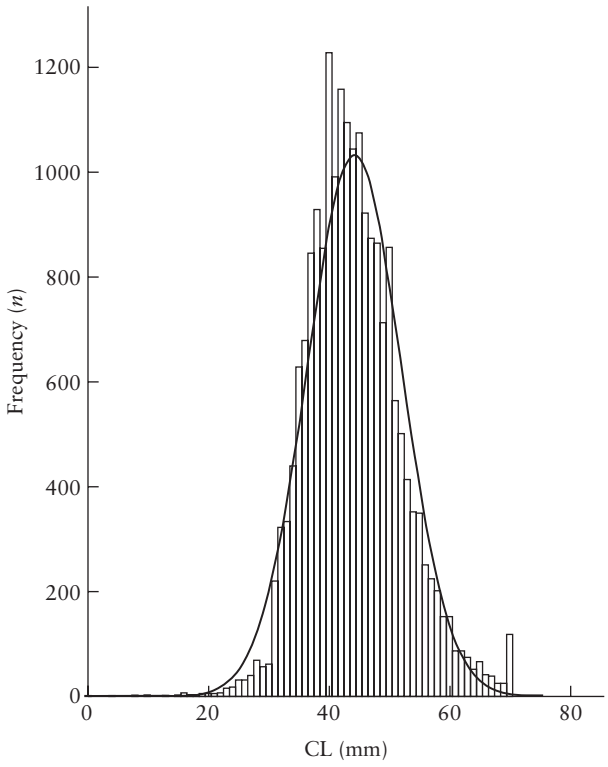

Figure 1 Distribution of cervical length (CL) measurements in 20 234 asymptomatic low-risk women with singleton pregnancy between November 2009 and August 2013 as part of the Triple P study. Mean ± SD CL = 44.18 ± 7.812 mm. The simulated normal distribution is also shown.

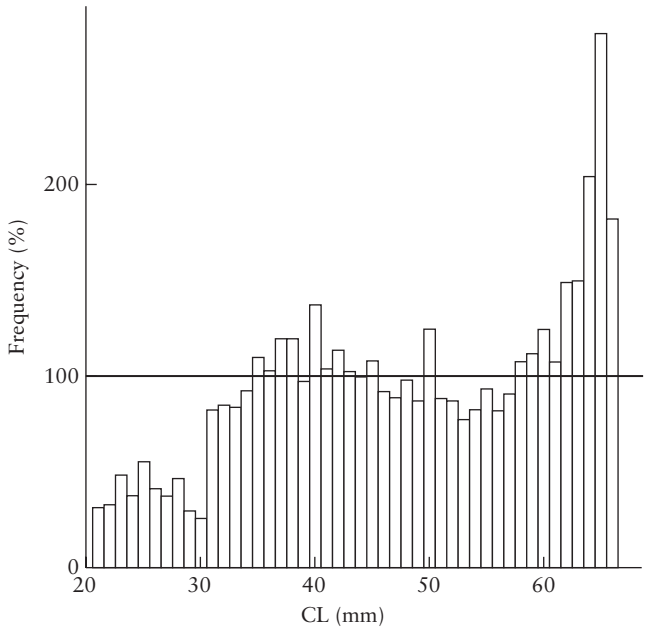

Figure 2 Number of cervical length (CL) measurements as a percentage of number of simulated CL measurements. Horizontal line at 100% indicates perfect agreement between both distributions. CLs for which < 10 measurements were simulated have been excluded.

and progress of the inclusions. In December 2011, the retrieved data of this monitoring process were made public during a national symposium to which all sonographers were invited. The symposium was organized for teaching purposes and designed for sonographers, midwives and residents. We compared the distribution of CL measurements before and after the symposium, for which, respectively, 12 284 and 7950 women were included.

Statistical analysis

The distribution of CL measurements was tested for normality using the Kolmogorov–Smirnov test. Student's unpaired *t*-test was used to assess the differences in mean CL between centers. A normal distribution of CL was simulated based on the mean and SD of the original data in order to compare the observed distribution with a normal distribution. Statistical analysis was performed using SPSS Statistics, v. 21 (IBM Corp., Armonk, NY, USA). Distribution of CL was simulated using R, v. 3.1.1 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Between November 2009 and August 2013, 20 234 women with low-risk singleton pregnancy were screened. Mean ± SD CL was 44.2 ± 7.8 mm at a median gestational age of 20 + 2 weeks. Of the total cohort of women, 367 (1.8%) had a CL ≤ 30 mm. The observed distribution of all CL measurements was compared with the simulated normal distribution and was found to be significantly different ($P < 0.001$; Figure 1). A CL measurement of ≤ 30 mm was less frequent in the original distribution compared with the simulated normal distribution ($n = 367$ (1.8%) *vs* 990 (4.9%), $P < 0.001$). The comparison revealed higher and lower numbers of measurements at certain CLs than expected (Figure 2). The difference, or 'dip', between both distributions was most profound for a CL of 20–30 mm, as indicated by a ratio of < 50% for the number of CL measurements within this range in the original compared with simulated data. Additionally, a higher than expected number of CL measurements was found for CL ≥ 62 mm.

A yearly analysis of CL measurements revealed cumulative distributions for 2009, 2010, 2011, 2012 and 2013, which were all significantly different from a normal distribution ($P < 0.005$, < 0.001, < 0.001, < 0.001 and < 0.001, respectively). The dip for CL measurements ≤ 30 mm was observed from October 2010 onwards. At that time, 2360 CL measurements had been performed. This dip persisted until the end of the study in August 2013. Hence, showing the preliminary results of the Triple P study at a national symposium in December 2011 had no influence on the overall distribution. However, a significant discrepancy ($P < 0.001$) in CL measurements was found when the means of the distribution before and after the symposium were compared. The dip for CL ≤ 30 mm was apparent in both

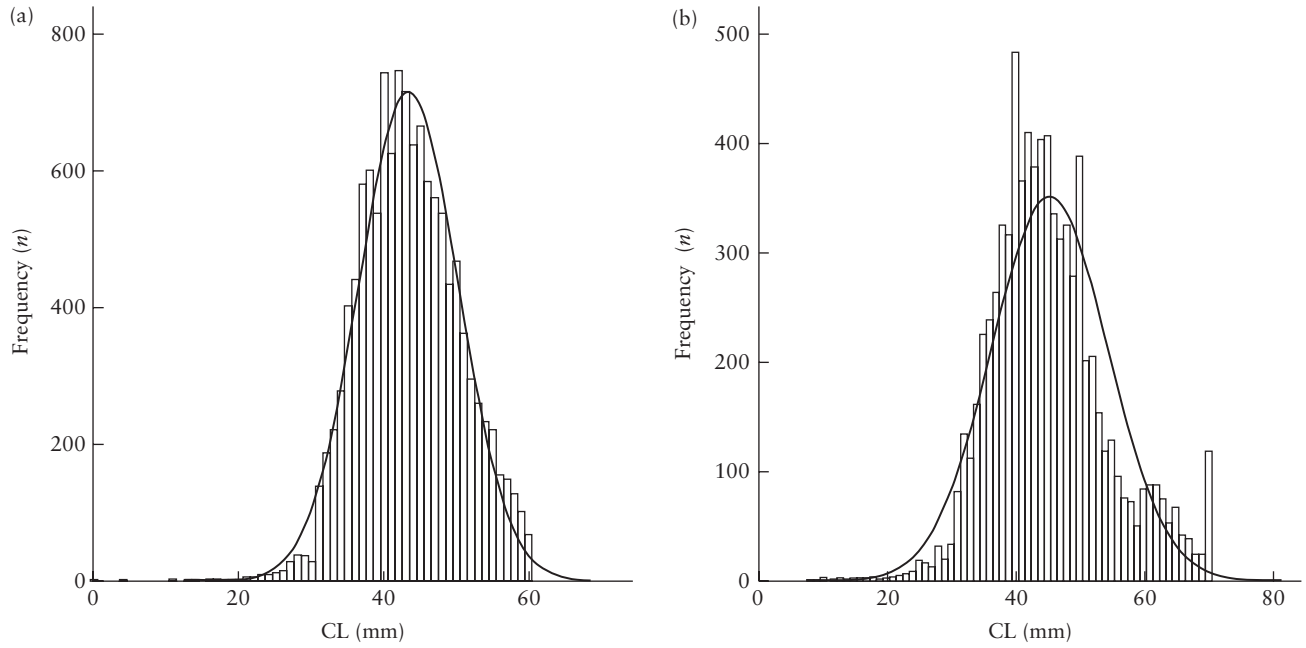

Figure 3 Distributions of cervical length (CL) measurements obtained between November 2009 and August 2013, with respect to timing of a national symposium in December 2011, at which the low prevalence of CL measurements ≤ 30 mm was addressed. (a) Before symposium: mean ± SD CL = 43.49 ± 6.84 mm, $n = 12\,284$. (b) After symposium: mean ± SD CL = 45.26 ± 9.007 mm, $n = 7950$.

sets of data, but was less prominent in the distribution of measurements obtained after the symposium. The proportion of CL measurements ≤ 30 mm was 1.7% before and 2.0% after the symposium (Figure 3).

In a per-center analysis ($n = 57$), 10 centers were excluded because of an insufficient number of CL measurements to assess the normality of the distribution. Of the remaining 47 centers, CL was normally distributed in 11. In all 36 centers in which CL was not normally distributed, the dip in CL measurements at 20–30 mm was visible. Of the 11 centers with a normal distribution, six also showed the dip at 20–30 mm.

Data analysis for primary, secondary and tertiary care institutions did not show a normal distribution for any level of care ($P < 0.001$). A significant difference in mean CL between primary and tertiary care institutions ($P < 0.001$) and between secondary and tertiary care institutions ($P < 0.001$) was found; however, there was no significant difference in CL distribution between primary and secondary care institutions ($P = 0.151$). The incidence of CL measurements ≤ 30 mm was 1.8%, 2.4% and 2.4% in primary, secondary and tertiary care institutions, respectively. The dip in CL measurements between 20 and 30 mm was seen in all three care levels, but was more pronounced in primary and tertiary care institutions (Figure 4).

When data were analyzed separately according to a low, intermediate and high volume of CL measurements, all distributions were significantly different from a normal distribution ($P \leq 0.001$). Mean CL was significantly different between low- and intermediate-volume centers ($P < 0.001$) and between low- and high-volume centers ($P < 0.001$); however, no difference was found between intermediate- and high-volume centers ($P = 0.06$). The proportion of CL measurements ≤ 30 mm was 1.8%, 3.0% and 1.4% for low-, intermediate- and high-volume centers, respectively. The dip in CL measurements at 20–30 mm was seen in all centers when analyzed separately according to the volume of measurements, but was most prominent in high-volume centers (Figure 5).

DISCUSSION

In this secondary analysis of the Triple P study, we aimed to analyze the distribution of CL measurements in a large cohort of women without previous preterm birth. The distribution of CL measurements was significantly different from a normal distribution, with a low prevalence of measurements ≤ 30 mm (1.8%) and a high mean CL of 44.2 mm, when compared with findings in the existing literature[7,10–13]. The absence of a normal distribution was seen in all levels of care and in high- and low-volume centers. We found a typical 'dip' in the distribution of CL measurements below the cut-off value of 30 mm in the entire cohort and in the majority (89%) of participating centers. The dip was most pronounced in primary and tertiary care centers and in high-volume centers. The distribution of CL measurements changed after preliminary results of the study (a low prevalence of short CL measurements) were disclosed at a national symposium. Finding a CL above or below the cut-off value, with the consequence of offering participation in a trial or not, seems to change the CL distribution, reducing the number of short CL measurements and creating a dip.

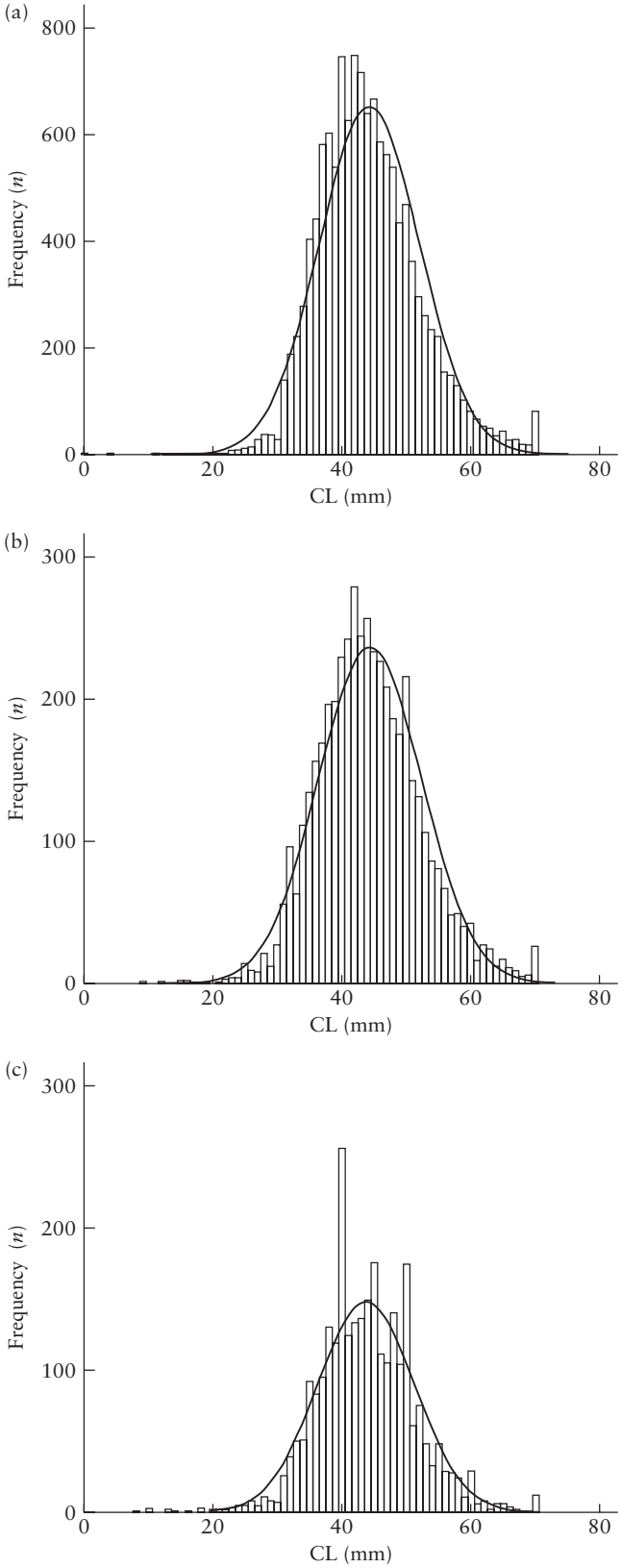

Figure 4 Distributions of cervical length (CL) measurements obtained between November 2009 and August 2013, according to level of care of center. (a) Primary-care institutions: mean ± SD CL = 44.22 ± 7.797 mm, $n = 12\,714$. (b) Secondary-care institutions: mean ± SD CL = 44.35 ± 7.999 mm, $n = 4728$. (c) Tertiary-care institutions: mean ± SD CL = 43.73 ± 7.543 mm, $n = 2792$.

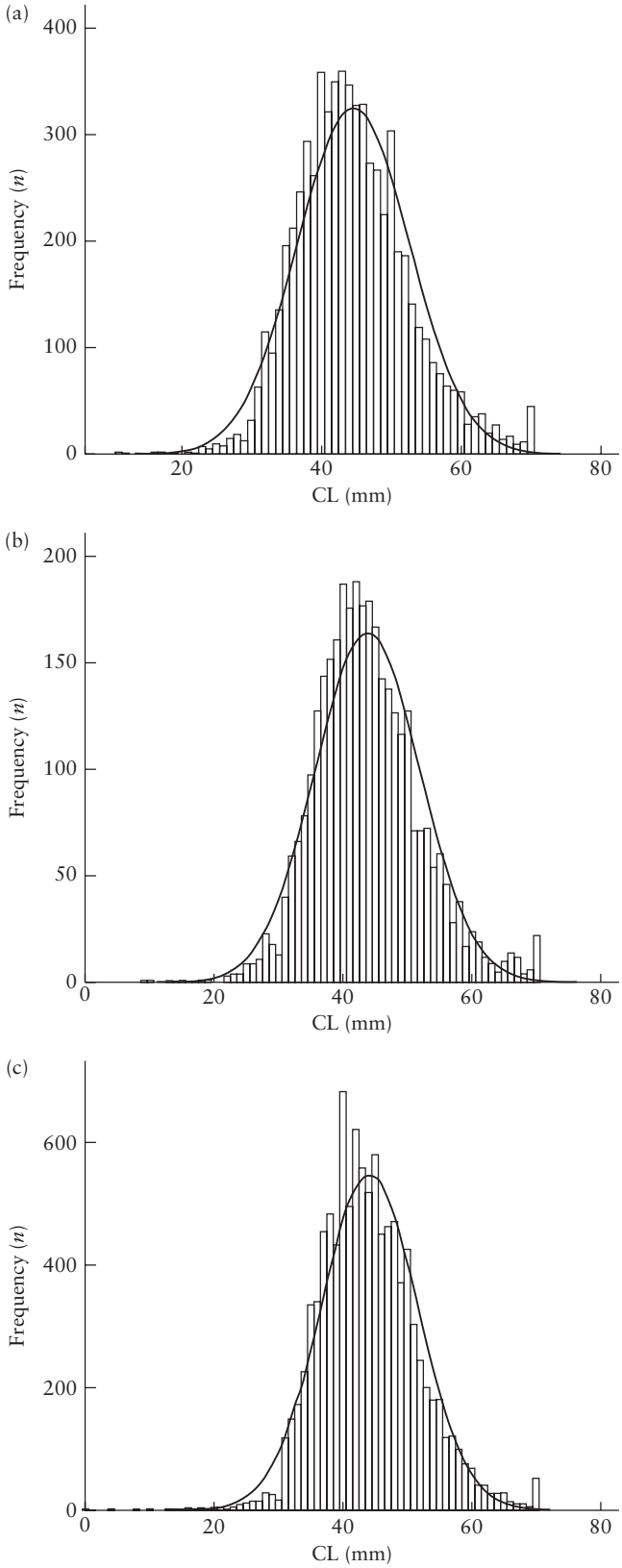

Figure 5 Distributions of cervical length (CL) measurements obtained between November 2009 and August 2013, according to volume of measurements performed in centers. (a) Centers performing < 500 measurements: mean ± SD CL = 44.58 ± 8.022 mm, $n = 6529$. (b) Centers performing 500–1000 measurements: mean ± SD CL = 43.77 ± 8.142 mm, $n = 3331$. (c) Centers performing > 1000 measurements: mean ± SD CL= 44.06 ± 7.555 mm, $n = 10\,374$.

The mean CL in our cohort is 6–9 mm longer than the previously described 33–38 mm[10–13]. The most probable explanation is the fact that our study population is composed solely of women without a history of preterm birth. Previous histograms were constructed in hospital populations including women with a history of preterm birth[7,10–13]. Other possible causes might be gestational age at measurement, parity and differences in the sonographic measurement technique. Our protocol was, however, identical to previously described methods in which the calipers were placed between the triangular area of echodensity at the external os and the V-shaped notch at the internal os[22]. Gestational age at measurement did not differ from other studies on this subject[7,10–13]. The Triple P study cohort contained a relatively high number of women with advanced maternal age and European origin. These maternal characteristics have been found to contribute to a slightly longer mean CL[16,23]. The influence in the entire cohort was, however, limited. 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 CL distribution below the cut-off value used for randomization into the clinical trial. We hypothesize that this can be attributed to the predefined cut-off value creating observer bias amongst the sonographers who were not blinded to the measurement. The fact that the difference in distributions became less prominent after we addressed this finding at a large national symposium supports our theory that it is a psychological effect of the observer. Additional arguments for this hypothesis are the finding of an overshoot in the number of CLs measured above 30 mm compared with a simulated normal curve, and the fact that other studies[10–13] never showed such a gap around this value. In other studies, however, intervention was not given following the CL 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 subjective judgment, such as qualitative scores or image recognition[24]. Non-blinded outcome assessors usually support the experimental intervention. Hróbjartsson *et al.* found that non-blinded outcome assessors exaggerate the hazard ratio by an average of approximately 27%[25]. In certain situations, however, they support the control interventions, inducing a comparable degree of observer bias in the opposite direction. The direction of observer bias depends on how the clinical and public contexts have shaped the predispositions of its investigators[25]. It has been shown previously that it is not unlikely that a CL will be measured longer than it actually is. In a study about the learning curve of sonographers, the sonographers were more likely to overestimate a CL than to underestimate it[22]. We hypothesize that the sonographers underestimate a CL below 30 mm in order to prevent referral, which may cause distress to the pregnant woman, and therefore they are selecting towards a longer CL.

In conclusion, a predefined cut-off point for which an intervention is implemented may affect the distribution of measurements. More importantly, it affects the selection of women for a randomized clinical trial on the basis of CL and indicates possible problems such as the under-representation of women with a short CL. A cut-off dip has not been demonstrated so clearly previously, and observer bias seems to be a possible explanation. The fact that we were not able to eliminate completely the dip after addressing this psychological effect of the observer at a national symposium indicates that even education cannot completely prevent this bias. Decision-making processes of sonographers are an area in which research has not yet been performed. The decision-making process might also affect other fields in which cut-off values are important, such as prenatal screening and diagnosis. To avoid the effect of using a cut-off value, as we describe in this study, it would be advisable to apply strict monitoring in the starting phase of a trial in order to reduce this bias quickly by habituation of the outcome assessor. Other trials using similar designs could benefit from this observation.

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**: 164–175.
2. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet* 2008; **371**: 75–84.
3. Schaaf JM, Mol BW, Abu-Hanna A, Ravelli AC. Trends in preterm birth: singleton and multiple pregnancies in The Netherlands, 2000–2007. *BJOG* 2011; **118**: 1196–1204.
4. Iams JD, Goldenberg RL, Meis PJ, Mercer BM, Moawad A, Das A, Thom E, McNellis D, Copper RL, Johnson F, Roberts JM. The length of the cervix and the risk of spontaneous premature delivery. *N Engl J Med* 1996; **334**: 567–572.
5. Smith V, Devane D, Begley CM, Clarke M, Higgins SA. Systematic review and quality assessment of systematic reviews of randomised trials of interventions for preventing and treating preterm birth. *Eur J Obstet Gynecol Reprod Biol* 2009; **142**: 3–11.
6. Sprague AE, O'Brien B, Newburn-Cook C, Heaman M, Nimrod C. Bed rest and activity restriction for women at risk for preterm birth: a survey of Canadian prenatal care providers. *J Obstet Gynaecol Can* 2008; **30**: 317–326.
7. To MS, Alfirevic Z, Heath VC, Cicero S, Cacho AM, Williamson PR, Nicolaides KH; Fetal Medicine Foundation Second Trimester Screening Group. Cervical cerclage for prevention of preterm delivery in women with short cervix: randomised controlled trial. *Lancet* 2004; **363**: 1849–1853.
8. Berghella V, Tolosa JE, Kuhlman K, Weiner S, Bolognese RJ, Wapner RJ. Cervical ultrasonography compared with manual examination as a predictor of preterm delivery. *Am J Obstet Gynecol* 1997; **177**: 723–730.
9. Heath VC, Southall TR, Souka AP, Elisseou A, Nicolaides KH. Cervical length at 23 weeks of gestation: prediction of spontaneous preterm delivery. *Ultrasound Obstet Gynecol* 1998; **12**: 312–317.
10. Heath VCF, Southall TR, Souka AP, Novakov A, Nicolaides KH. Cervical length at 23 weeks of gestation: relation to demographic characteristics and previous obstetric history. *Ultrasound Obstet Gynecol* 1998; **12**: 304–311.
11. Erasmus I, Nicolaou E, van Gelderen CJ, Nicolaides KH. Cervical length at 23 weeks' gestation – relation to demographic characteristics and previous obstetric history in South African women. *S Afr Med J* 2005; **95**: 691–695.
12. Palma-Dias RS, Fonseca MM, Stein NR, Schmidt AP, Magalhães JA. Relation of cervical length at 22–24 weeks of gestation to demographic characteristics and obstetric history. *Braz J Med Biol Res* 2004; **37**: 737–744.
13. Theron G, Schabort C, Norman K, Thompson M, Geerts L. Centile charts of cervical length between 18 and 32 weeks of gestation. *Int J Gynaecol Obstet* 2008; **103**, 144–148.
14. Bergelin I, Valentin L. Patterns of normal change in cervical length and width during pregnancy in nulliparous women: a prospective, longitudinal ultrasound study. *Ultrasound Obstet Gynecol* 2001; **18**: 217–222.
15. Zorzoli A, Soliani A, Perra M, Caravelli E, Galimberti A, Nicolini U. Cervical changes throughout pregnancy as assessed by transvaginal sonography. *Obstet Gynecol* 1994; **84**: 960–964.
16. van der Ven AJ, van Os MA, Kazemier BM, Kleinrouweler E, Verhoeven CJ, de Miranda E, van Wassenaer-Leemhuis AG, Kuiper PN, Porath M, Willekes C, Woiski MD, Sikkema MJ, Roumen FJ, Bossuyt PM, Haak MC, de Groot CJ, Mol BW, Pajkrt E. The capacity of mid-pregnancy cervical length to predict preterm birth in low-risk women: a national cohort study. *Acta Obstet Gynecol Scand* 2015; **94**: 1223–1234.

17. The Fetal Medicine Foundation. Cervical assessment. https://fetalmedicine.org/cervical-assessment [Accessed 19 May 2015].
18. Fonseca EB, Celik E, Parra M, Singh M, Nicolaides KH. Progesterone and the risk of preterm birth among women with a short cervix. *N Engl J Med* 2007; **357**: 462–469.
19. Hassan SS, Romero R, Vidyadhari D, Fusey S, Baxter JK, Khandelwal M, Vijayaraghavan J, Trivedi Y, Soma-Pillay P, Sambarey P, Dayal A, Potapov V, O'Brien J, Astakhov V, Yuzko O, Kinzler W, Dattel B, Sehdev H, Mazheika L, Manchulenko D, Gervasi MT, Sullivan L, Conde-Agudelo A, Phillips JA, Creasy GW. Vaginal progesterone reduces the rate of preterm birth in women with a sonographic short cervix: a multicenter, randomized, double-blind, placebo controlled trial. *Ultrasound Obstet Gynecol* 2011; **38**: 18–31.
20. van Os MA, van der Ven AJ, Kleinrouweler CE, Schuit E, Kazemier BM, Verhoeven CJ, de Miranda E, van Wassenaer-Leemhuis AG, Sikkema JM, Woiski MD, Bossuyt PM, Pajkrt E, de Groot CJ, Mol BW, Haak MC. Preventing preterm birth with progesterone in a low-risk population: a randomized controlled trial. *Am J Perinatol* 2015; **32**: 993–1000.
21. van Os MA, van der Ven AJ, Kleinrouweler E, Pajkrt E, de Miranda E, van Wassenaer A, Porath M, Bossuyt PM, Bloemenkamp KW, Willekes C, Woiski M, Oudijk MA, Bilardo KM, Sikkema MJ, Duvekot JJ, Veersema D, Laudy J, Kuiper P, de Groot CJ, Mol BW, Haak MC. Preventing preterm birth with progesterone: costs and effects of screening low risk women with a singleton pregnancy for short cervical length, the Triple P study. *BMC Pregnancy Childbirth* 2011; **11**: 77.
22. van Os MA, van der Ven AJ, Bloemendaal PM, Pajkrt E, de Groot CJ, Mol BW, Haak MC. Effect of e-learning on quality of cervical-length measurements. *Ultrasound Obstet Gynecol* 2015; **46**: 327–331.
23. van der Ven AJ, van Os MA, Kleinrouweler CE, de Groot CJ, Haak MC, Mol BW, Pajkrt E, Kazemier BM. Is cervical length associated with maternal characteristics? *Eur J Obstet Gynecol Reprod Biol* 2015; **188**: 12–16.
24. Hróbjartsson A, Skou Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaud P, Brorson S. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 2012; **344**: e1119.
25. Hróbjartsson A, Skou Thomsen AS, Emanuelsson F, Tendal B, Rasmussen JV, Hilden J, Boutron I, Ravaud P, Brorson S. Observer bias in randomized clinical trials with time-to-event outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *Int J Epidemiol* 2014; **43**: 937–948.

