Fundal height screening in antenatal care: an audit of Norwegian performance

MARIT LIED[1], AASE D. PAY[2,3], ELI SAASTAD[3,4], J. FREDERIK FRØEN[3]

[1]Faculty of medicine, University of Oslo, Norway, [2]Department of Obstetrics and Gynecology, Women and Children's Division, Oslo University Hospital, [3]Department of Genes and Environment, Division of Epidemiology Norwegian Institute of Public Health, [4]Oslo and Akershus University College

Correspondence:
J.Frederik Frøen
Division of Epidemiology
Norwegian Institute of Public Health
PB 4404 Nydalen
N-0403 Oslo
Norway
E-mail: Frederik.Froen@fhi.no

Abstract

Objective. Audit screening properties of fundal height measures (FHM) in identifying fetal growth aberrations in low risk pregnancies in Norway. *Design.* A prospective cohort recruited for a controlled trial was retrospectively audited. *Setting.* Routine Norwegian antenatal care. *Population.* 789 women with singleton pregnancies. *Methods.* FHM were registered and dichotomized to positive or negative tests according to common practice. FHM were excluded from analysis in the low risk group from the day a risk factor that requires follow up of fetal growth by ultrasound measures was observed. *Main outcome measures.* SGA, FGR, LGA, sensitivity, specificity and predictive values of the FHM. *Results.* The sensitivity of screening for FGR by FHM in the low risk population was nil, as it was for SGA among overweight women. Among the lean and low risk women, the sensitivity for SGA seemed to be twice that of pure chance, but still capturing a mere 13% of SGA infants in this group of women. Compared with the screening by FHM alone, the clinical judgment only added marginal benefits to the specificity of screening for the low risk population. A negative FHM screening offered no reassurance compared with pure chance. In the screening for LGA in our population, clinician's judgment had much larger impact on the screening properties, mainly by being restrictive in referrals for large FHM, and thus reducing the overall sensitivity by half from 45 to 22%. *Conclusion.* The current reference curves and standards for FHM in Norway have poor screening properties and must be revised.

Key words

Fundal height measurement

Fetal growth restriction

Low risk pregnancies

Antenatal care

Abbreviations

FHM - Fundal height measurement

SGA - Small for gestational age

FGR - Fetal growth restriction

LGA - Large for gestational age

Introduction

Fetal growth is associated with pregnancy outcome. Pregnancies affected by fetal growth restriction (FGR), and the commonly used approximation for FGR, small for gestation (SGA) infants, are at increased risk of preterm delivery, perinatal death and infant morbidity and mortality largely due to placental insufficiencies of a variety of causes (1-6). The large for gestational age (LGA) infants are at risk of stillbirth, perinatal asphyxia, meconium aspiration and birth injuries such as dystocia and fractures (7;8). The early diagnosis and proper assessment and management of these infants can improve their outcome (9).

One of the principle aims of antenatal care is the early detection of aberrations of fetal growth (10;11). Fundal height measures (FHM) are a routine component of antenatal care from 24 weeks until term (12). Low cost, high applicability, acceptable validity, and no potential harm have been arguments in favor of FHM as a screening tool to detect fetal growth disturbances (13), despite the lack of conclusive evidence for its effectiveness to detect fetal growth aberrations (14). Due to the paucity of evidence for benefits of alternative screening methods, such as serial ultrasound measures in third trimester pregnancies (15), screening for fetal growth problems in the Norwegian low risk population remains reliant on FHM (12), in line with international recommendations of NICE (16).

Still, the clinical screening for FGR is the most commonly criticized aspect of antenatal care in retrospective audits of perinatal mortality in our community (17). Observational studies of FHM screening in total populations have been highly inconclusive, indicating anything from very low (17%) to high (86%) sensitivity in detection of SGA, and specificities ranging from 64 to 95% (18). Additionally, the inclusion of unselected populations in these screening evaluations may overestimate the usefulness of FHM screening: High risk pregnancies are not intended to be part of the FHM screening, as there is evidence of benefit of serial Doppler ultrasound examinations (19). When estimating effectiveness, the FHM of individual pregnancies must thus be included only when they belong to the low risk population, and excluded from analysis from the day they leave this cohort due to clinical findings of risk factors. Including their FHM in analyses after they have been defined as risk pregnancies, as is commonly done (20-22), will provide overly optimistic estimates of the screening properties in the low risk population. This is due both to higher prevalence of adverse outcomes and biased FHM towards low measures in pregnancies with e.g. known growth restriction confirmed by repeated ultrasound measures.

Several issues question the Norwegian reference curves and their clinical use. Unfortunately, history has been lost, and it has become impossible to trace the original material or statistic analyses on which the standardized reference curve was devised in the 1970ies (23;24), commonly known as the "Westin-curves". The upper and lower limits of "normality" (figure 1) have unknown origins, and the entire graph is known to have been manually corrected twice without any data on which to base the corrections (24). The apparent limits of normality, as drawn on the original curve, represented both 1 and 2 SD above and below the mean. However, this was not the actual limits used in the screening, which was restricted to use a deviation ± 2 cm from the mean at all gestations (23). On the current Norwegian reference curve, the shaded area is not symmetrical over and under the mean curve, which does neither correspond to the original measures of uncertainty nor the intended limits for screening purposes. Also the mean itself seems to be in error, and compared to the original materials, the mean FHM at term is drawn approximately one centimeter lower than the original curve. The Norwegian reference curve also has poor, if any, accompanying definitions of a positive screening result other than the shaded "limits of normality" drawn on the chart. In the Norwegian guidelines the pregnant woman is to be referred to further investigation in case of "large deviations" (12). This ambiguity and lacking guidance in defining a positive screening test makes the clinical judgment based on discretion and results in a significant inter observer variation. No controlled trials have been performed to confirm any effectiveness of the Norwegian FHM screening.

Fundal height reference curves are used in a setting where care providers make their decisions for care based on the overall clinical information they have, not FHM in isolation. They may discard FHM findings if they do not trust the validity of previous measures, often performed in alternation by midwives and general practitioners, or if they do not find the reference curves "fitting" for the woman in their care, based on her size, clinical history or other clinical findings. Thus, the screening properties of FHM have both theoretical screening properties based solely on the FHM, and clinical properties based on the level of information it provides clinicians in their decision-making. Both sets of properties are clinically important, as the latter reflects a real life "intention to treat" effectiveness of FHM.

This study was therefore conducted to audit both the theoretical and clinical screening properties for SGA and LGA infants by the current fundal height screening in a retrospective population-based cohort, from which pregnancies with established indication for ultrasound screening were excluded from analysis from the day this indication was documented.

Materials and methods

This study is an audit of anonymous copies of case notes with FHM. Data was obtained from a multi center randomized controlled trial in Norway (Eli Saastad & al, resubmitted PLoSOne Oct. 2011, awaiting final reference). Eligible women were Norwegian-speaking with singleton pregnancies. Women were recruited from September 2007 through November 2009 at nine Norwegian hospitals from both urban and rural populations, handling a total of 8,200 births annually. In Norway, the public antenatal care services are free of charge and encompass almost all pregnant women. Demographic and obstetric information was obtained from case notes received from the hospitals after delivery, and an anonymized file was used for this audit. The randomized controlled trial included 1076 women, but as that trial was not designed specifically to evaluate FHM, many of the photocopies of maternity cards left the FHM graph unreadable or missing, completely or partially, and they were excluded from analysis. We included 789 women in our audit – all descriptives in the following are provided with a 95% confidence interval of the mean. The women had a mean age of 30 years (30-31), 18% (15-21) were older than 35 years, 45% (42-49) were primiparous, and 11% (9-14) were obese with a BMI > 30. Among them, 3% (1-4) reported smoking in the third trimester, 64% (60-68) had 12 or more years of education (equivalent of high school graduates), and 93% (91-95) were either married or co-habiting. Among the newborns there were 50% (47-54) boys, 4% (2-5) were born preterm, and the mean birth weight was 3625 grams (3587-3663). The subgroup drawn from the randomized trial did not differ from those not included in the audit in any aspects.

Small for gestational age infants were defined as infants with a birth weight below the 10th centile of the population's distribution of birth weights (25). As a closer approximation to FGR, the randomized controlled trial from which we drew the material defined FGR as a birth weight < 2.5^{th} centile or an antenatal ultrasound measure estimating fetal growth < 2.5^{th} percentile birth weight (i.e. < 21.5% negative deviation), or a negative trend on serial antenatal ultrasounds. This was thus used in this audit as FGR. The large for gestational age (LGA) infants were defined as a birth weight above the 90th centile.

The anonymized maternity cards containing the FHM were reviewed by two of the authors (ML, AP) and a FHM was defined as abnormal according to the common use of the FH reference curves in Norway: An abnormal FHM being 1) no measurable growth for three successive measurements in a time period of minimum three weeks, or 2) a measurement above or below the normality limits of the reference curve.

We based our definition of risk pregnancies on the Norwegian guidelines for antenatal care (12), which is almost identical to the British NICE guidelines (16). Any risk factor that require follow up of fetal growth by ultrasound measures lead to the exclusion from the low risk group and inclusion in the high risk group from the day the risk factor was observed and documented in the medical records. This would include both pre-existing risk factors such as hypertension, chronic renal or coronary disease or diabetes type I or II, as well as complications of pregnancy such as pregnancy induced hypertensive disorders, prolonged preterm rupture of membranes or gestational diabetes. This is described in further detail in the report of the randomized controlled trial.

All data were analyzed with the Statistical Package for the Social Sciences (SPSS, Chicago, IL, USA). The results are expressed as frequencies and mean values.

Ethical approval for the randomized trial was obtained from The Regional Committee for Medical Research Ethics (reference S-07188a) 7 May 2007, and by The Norwegian Data Inspectorate and Directorate for Health (reference 07/2504) 19 July 2007. The study was registered in www.clinicaltrials.gov protocol registration system (number NCT00513942). Additional approval for this audit study was obtained from The Regional Committee for Medical Research Ethics on 23 February 2010.

Results

In screening for SGA/FGR in our low risk population, we found no reassurance in a negative FHM screening compared with pure chance: The probability of being SGA/FGR despite a negative screening (1-NPV) was identical to the pre-test incidence of SGA and FGR. Even in the most optimistic end of the 95% confidence interval, a negative FHM screening using Norway's current reference charts offers little, if any, clinically relevant reassurance (table 1). The sensitivity of screening for FGR by FHM in the low risk population was nil, as it was for SGA among overweight women. Among the lean and low risk women, the sensitivity for SGA seemed to be twice that of pure chance, but still capturing a mere 13% of SGA infants in this group of women. Yet, the numbers of SGA and FGR cases in our population was small, and in the most optimistic end of the 95% confidence interval, a 25% sensitivity for SGA among lean low risk women cannot be excluded. Compared with the screening by FHM alone, the clinical judgment only added marginal benefits to the specificity of screening for the low risk population, but was otherwise identical (data not shown).

In screening for LGA in our population, clinician's judgment had much larger impact on the screening properties, mainly by being restrictive in referrals for large FHM, and thus

reducing the overall sensitivity by half from 45 to 22% (table 2). Among women at highest risk (overweight or obese) FHM has no certain benefit over chance once clinical judgment is taken into account, neither in capture of cases through a positive test nor in reassurance by a negative test. Among lean women, the combined properties of FHM screening and clinical judgment had acceptable specificity of a positive test, and the sensitivity remained twice the incidence at 20%, but a negative screening did not provide substantial reassurance over pure chance.

Discussion

We found that FHM offers little help to the clinician in identifying fetuses with abnormal growth. As a screening tool the main problem in screening for SGA was a low sensitivity, and in screening for LGA a low specificity. Thus the screening may have adverse effects "in both ends" by offering false reassurance to the majority of pregnancies affected by growth restriction, and by causing unneeded anxiety and interventions among women screened "positive" for risk of LGA.

A contributing factor to the poor test properties of FHM is that the fundal height growth curve might not be optimal for the population. The "Westin-curve" was constructed in the 70's and was based on 100 highly selected healthy Caucasian Swedish women with weight for height between the 10^{th} and 90^{th} centile with "completely uneventful singleton pregnancies" leading to term "uneventful intrapartum periods" and the delivery of a non-SGA non-LGA healthy and normally formed baby (23). In contrast, the Norwegian population is increasingly diverse in both ethnicity and body mass index. The prevalence of maternal overweight and obesity has increased significantly (26). Both are examples that may reduce the utility of the existing fundal height curve in antenatal care, and combined with the fact that the screening tool was not derived from a total population, this may be part of the explanation of why we cannot reproduce the good screening properties originally described (11;23;27).

The only study we have found that evaluated the Norwegian curve as it exists today concluded that the FHM seemed to be of only limited value in determining whether a fetus is lighter or heavier than average (28). With sensitivity on 41% for detecting SGA (defined as birth weight below the 10th centile for gestational age) they still were far more optimistic than our 20% in the total group of women. However, women in that study, as in all other evaluations of FHM, were not excluded from screening once they entered the group of high risk of fetal growth aberrations which should be referred to ultrasound measures, and not followed by FHM. This may not only influence results due to increasing prevalence of the

outcome being screened for, as seen in our material (table 1), but it may also introduce bias among screeners when they know that the woman being screened is at high risk of growth aberrations. In contrast to our study design, the original evaluations of the "Westin-curves" either included high risk pregnancies, leading to favorable evaluations of the screening properties with a sensitivity of 59% (11), or specifically used high risk pregnancies as the population to study (27). The latter compared a group of known SGA children and a control group and used an indirect analysis with extrapolation to a total population to suggest a sensitivity and specificity for SGA of 62 and 88%, respectively. Studies that are not conducted in a low risk setting provide little information about the screening properties in a low risk population.

In our audit, we examined a "real life" prospective FHM collection. Poor standardization of FHM methodology may have contributed to the small predictive values of FHM in this study through large inter- and intra-examiner variations (29;30). Yet, we believe this is the most realistic evaluation of actual performance, similar to an "intention to treat" analysis. This reflects not only the limitations of non-standardized measurements in itself, but also how screening fares in a setting where clinical staff is well aware of such limitations when they make clinical decisions regarding referrals for suspicion of SGA or LGA. The only clinically significant effect we found of such clinical deliberations was a reduction of pregnancies referred for further evaluations due to large FHM.

A weakness is that we have a small material of 789 women, and thus wide confidence intervals. Yet, even in the most optimistic end of our confidence intervals, there is insufficient evidence to recommend the use of FHM screening with the current reference curves. In addition, recruitment rate in the RCT was only 14%, and this may hamper generalization of the results. Only 8% of women in this group had babies below the 10^{th} centile of the Norwegian population, and thus had somewhat better health outcomes than average. Importantly, this group of women was Norwegian-speaking, and thus less ethnically diverse than our total population - on the other hand, they should thus be a better fit for the current FHM reference curve based on an equally homogenous population.

Conclusion

In conclusion, the results of this study confirm earlier anxieties about poor screening properties of the current FHM in identifying aberrations of fetal growth. FHM as practiced today does not provide reliable information about fetal growth. In a public health context, an improvement of reference curves or a good alternative to FHM has to be found. A robust

evaluation of the screening properties and FHM in a low risk population is necessary, and a randomized controlled trial should be conducted to show benefit in health outcomes.

References

(1) Clausson B, Cnattingius S, Axelsson O. Outcomes of post-term births: the role of fetal growth restriction and malformations. Obstet Gynecol 1999 Nov;94(5 Pt 1):758-62.

(2) Froen JF, Gardosi JO, Thurmann A, Francis A, Stray-Pedersen B. Restricted fetal growth in sudden intrauterine unexplained death. Acta Obstet Gynecol Scand 2004 Sep;83(9):801-7.

(3) Goldenberg RL, Hoffman HJ, Cliver SP. Neurodevelopmental outcome of small-for-gestational-age infants. Eur J Clin Nutr 1998 Jan;52 Suppl 1:S54-S58.

(4) Markestad T, Vik T, Ahlsten G, Gebre-Medhin M, Skjaerven R, Jacobsen G, et al. Small-for-gestational-age (SGA) infants born at term: growth and development during the first year of life. Acta Obstet Gynecol Scand Suppl 1997;165:93-101.

(5) Selling KE, Carstensen J, Finnstrom O, Sydsjo G. Intergenerational effects of preterm birth and reduced intrauterine growth: a population-based study of Swedish mother-offspring pairs. BJOG 2006 Apr;113(4):430-40.

(6) Simchen MJ, Beiner ME, Strauss-Liviathan N, Dulitzky M, Kuint J, Mashiach S, et al. Neonatal outcome in growth-restricted versus appropriately grown preterm infants. Am J Perinatol 2000;17(4):187-92.

(7) Berard J, Dufour P, Vinatier D, Subtil D, Vanderstichele S, Monnier JC, et al. Fetal macrosomia: risk factors and outcome. A study of the outcome concerning 100 cases >4500 g. Eur J Obstet Gynecol Reprod Biol 1998 Mar;77(1):51-9.

(8) Wojtasinska M, Belfrage P, Gjessing L. [Large fetus--a retrospective study]. Tidsskr Nor Laegeforen 2000 Jun 20;120(16):1848-50.

(9) Lindqvist PG, Molin J. Does antenatal identification of small-for-gestational age fetuses significantly improve their outcome? Ultrasound Obstet Gynecol 2005 Mar;25(3):258-64.

(10) Bergman E, Kieler H, Petzold MG, Sonesson C, Axelsson O. Symphysis-fundus measurements for detection of small for gestational age pregnancies. Acta Obstet Gynecol Scand 2006;85(4):407-12.

(11) Cnattingius S. Antenatal screening for small-for-gestational-age, using risk factors and measurements of the symphysis-fundus distance--6 years of experience. Early Hum Dev 1988 Dec;18(2-3):191-7.

(12) Klovning A et al. Retningslinjer for svangerskapsomsorgen. 2005. Oslo, Sosial- og Helsedirektoratet.
Ref Type: Generic

(13) Quaranta P, Currell R, Redman CW, Robinson JS. Prediction of small-for-dates infants by measurement of symphysial-fundal-height. Br J Obstet Gynaecol 1981 Feb;88(2):115-9.

(14) Neilson JP. Symphysis-fundal height measurement in pregnancy. Cochrane Database Syst Rev 2000;(2):CD000944.

(15) Bricker L, Neilson JP, Dowswell T. Routine ultrasound in late pregnancy (after 24 weeks' gestation). Cochrane Database Syst Rev 2008;(4):CD001451.

(16) Antenatal care: routine care for the healthy pregnant women. 2003. London, National Institute for Clinical Excellence.
Ref Type: Generic

(17) Saastad E, Vangen S, Froen JF. Suboptimal care in stillbirths - a retrospective audit study. Acta Obstet Gynecol Scand 2007;86(4):444-50.

(18) Jacobsen G. Prediction of fetal growth deviations by use of symphysis-fundus height measurements. Int J Technol Assess Health Care 1992;8 Suppl 1:152-9.

(19) Alfirevic Z, Stampalija T, Gyte GM. Fetal and umbilical Doppler ultrasound in high-risk pregnancies. Cochrane Database Syst Rev 2010;(1):CD007529.

(20) Hargreaves K, Cameron M, Edwards H, Gray R, Deane K. Is the use of symphysis-fundal height measurement and ultrasound examination effective in detecting small or large fetuses? J Obstet Gynaecol 2011 Jul;31(5):380-3.

(21) Lindhard A, Nielsen PV, Mouritsen LA, Zachariassen A, Sorensen HU, Roseno H. The implications of introducing the symphyseal-fundal height-measurement. A prospective randomized controlled trial. Br J Obstet Gynaecol 1990 Aug;97(8):675-80.

(22) Indraccolo U, Chiocci L, Rosenberg P, Nappi L, Greco P. Usefulness of symphysis-fundal height in predicting fetal weight in healthy term pregnant women. Clin Exp Obstet Gynecol 2008;35(3):205-7.

(23) Westin B. Gravidogram and fetal growth. Comparison with biochemical supervision. Acta Obstet Gynecol Scand 1977;56(4):273-82.

(24) Geirsson RT, Steingrimsdottir T. [Symphysis-fundus measurements in Scandinavia]. Nord Med 1990;105(11):304-6.

(25) Skjaerven R, Gjessing HK, Bakketeig LS. Birthweight by gestational age in Norway. Acta Obstet Gynecol Scand 2000 Jun;79(6):440-9.

(26) Droyvold WB, Nilsen TI, Kruger O, Holmen TL, Krokstad S, Midthjell K, et al. Change in height, weight and body mass index: Longitudinal data from the HUNT Study in Norway. Int J Obes (Lond) 2006 Jun;30(6):935-9.

(27) Wallin A, Gyllensward A, Westin B. Symphysis-fundus measurement in prediction of fetal growth disturbances. Acta Obstet Gynecol Scand 1981;60(3):317-23.

(28) Jensen OH, Larsen S. Evaluation of symphysis-fundus measurements and weighing during pregnancy. Acta Obstet Gynecol Scand 1991;70(1):13-6.

(29) Crosby ME, Engstrom JL. Inter-examiner reliability in fundal height measurement. Midwives Chron 1989 Aug;102(1219):254-6.

(30) Engstrom JL, McFarlin BL, Sittler CP. Fundal height measurement. Part 2--Intra- and interexaminer reliability of three measurement techniques. J Nurse Midwifery 1993 Jan;38(1):17-22.

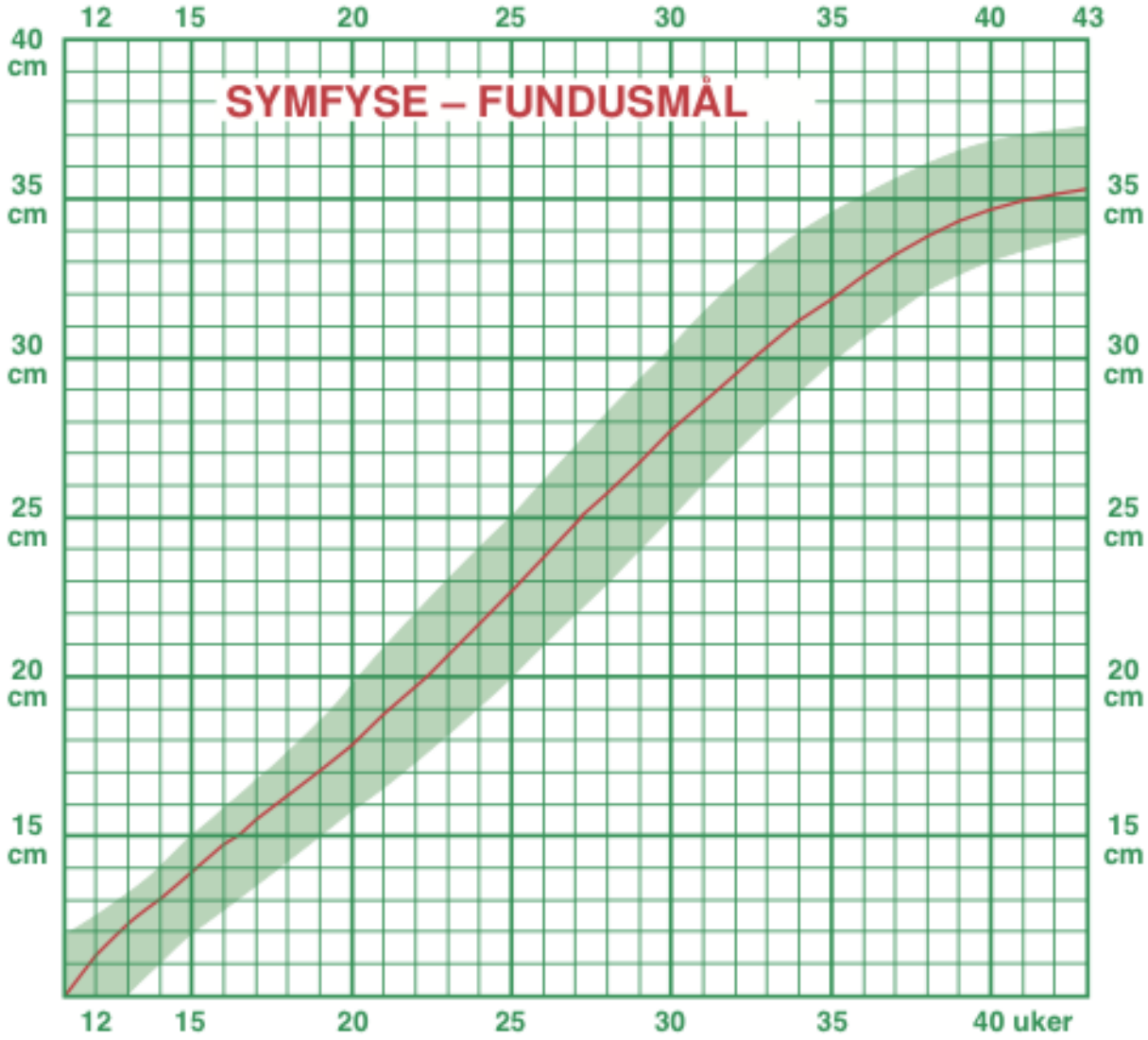

Figure 1. The Norwegian fundal height curves in use today.

Table 1. Screening properties for FHM concerning SGA and FGR. FHM: fundal height measurement, SGA: small for gestational age, FGR: fetal growth restriction, PPV: positive predictive value, NPV: negative predictive value, SE: sensitivity, SPE: specificity

SGA		SGA	non-SGA	SGA	PPV	NPV	SE	SPE
		n	n	%	%(95%CI)	%(95%CI)	%(95%CI)	%(95%CI)
Total		64	725	8	28(15-43)	93(91-95)	20(10-30)	95(94-97)
Risk	Total	28	203	12	47(23-72)	91(87-95)	32(14-51)	95(92-98)
	BMI≥25	7	86	8	67(0-100)	94(90-99)	29(0-73)	99(97-100)
	BMI<25	21	117	15	44(16-71)	89(83-94)	33(11-55)	92(87-97)
Low risk	Total	40	618	6	15(0-29)	94(92-96)	10(0-20)	96(95-98)
	BMI≥25	8	218	4	0(0-)	96(94-99)	0(0-)	98(96-100)
	BMI<25	32	399	7	18(1-37)	93(91-96)	13(0-25)	95(93-98)

FGR		FGR	non-FGR	FGR	PPV	NPV	SE	SPE
		n	n	%	%(95%CI)	%(95%CI)	%(95%CI)	%(95%CI)
Total		31	758	4	22(9-34)	97(96-98)	32(15-50)	95(94-97)
Risk	Total	26	205	11	53(28-77)	92(89-96)	38(18-59)	96(93-98)
	BMI≥25	5	88	5	67(0-100)	97(93-100)	40(0-100)	99(97-100)
	BMI<25	21	117	15	50(22-78)	89(84-95)	38(15-61)	93(89-98)
Low risk	Total	7	651	1	0(0-)	99(98-100)	0(0-)	96(94-97)
	BMI≥25	3	223	1	0(0-)	99(97-100)	0(0-)	98(96-100)
	BMI<25	4	427	1	0(0-)	99(98-100)	0(0-)	95(93-97)

Table 2. Screening properties for FHM concerning LGA . FHM: fundal height measurement, LGA: large for gestational age, PPV: positive predictive value, NPV: negative predictive value, SE: sensitivity, SPE: specificity

LGA		LGA	non-LGA	LGA								
					FHM alone				Clinical screening			
					PPV	NPV	SE	SPE	PPV	NPV	SE	SPE
		n	n	%	%(95%CI)	%(95%CI)	%(95%CI)	%(95%CI)	%(95%CI)	%(95%CI)	%(95%CI)	%(95%CI)
Total		104	685	13	29 (22-36)	91 (89-93)	45 (35-55)	83 (80-86)	36 (24-48)	89 (87-91)	22 (14-30)	94 (92-96)
	BMI≥30	22	66	25	37 (21-53)	84 (73-95)	64 (42-85)	64 (52-76)	37 (13-61)	78 (68-88)	32 (11-53)	82 (72-91)
	BMI≥25	54	222	20	32 (23-42)	87 (82-92)	56 (42-69)	72 (66-78)	35 (19-51)	83 (78-88)	24 (12-36)	89 (85-93)
	BMI<25	50	462	10	24 (14-34)	93 (90-95)	34 (20-48)	88 (85-91)	37 (18-57)	92 (89-94)	20 (9-31)	96 (95-98)