# Symphysis-fundus measurement and prediction of SGA neonates

Aase Serine Devold Pay



### Institute of Clinical Medicine, Faculty of Medicine

### UNIVERSITY OF OSLO

2016



#### © Aase Serine Devold Pay, 2016

Series of dissertations submitted to the Faculty of Medicine, University of Oslo

ISBN 978-82-8333-119-6

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard Printed in Norway: 07 Media AS – www.07.no

## Acknowledgements

The work presented in this thesis was carried out while I held a part-time (50%) research, part-time (50%) clinical midwife position at the Department of Obstetrics and Department of Gynecology, Oslo University Hospital, Ullevål, from January 2010 until July 2015.

My work has been supported by many people to whom I am very grateful. My main supervisor, Håkon K. Gjessing, who has shown a long-term commitment to fetal growth research and statistical modeling, introduced me to this exciting field of research. Thank you, Håkon, for your ability to convey complex statistics in a competent and manageable manner and for supporting me throughout the dissertation process.

I also want to thank my co-supervisors, Frederik Frøen, Bo Jacobsson, and Annetine Staff. Thank you, Frederik, for initially planning and designing this study, for guiding my scientific work and for providing the facilities and support necessary to accomplish the project. Thank you, Bo, for giving me access to the data, and for engaging in interesting discussions and providing skilled assistance whenever I needed. You included me in your own research projects, and encouraged me to take responsibility and leadership. Thank you, Annetine, for supervising and supporting me during this period, and for giving me access to your vast amounts of knowledge in all aspects of research; you have been enormously inspirational and valuable.

Grethe Gaarder Næsje, Hanne Knutsen, Kristi Hjelle, Bjørn Busund, and Thomas Åbyholm – your efforts to implement a research position for a midwife in the Department of Obstetrics and Department of Gynecology were impressive and very stimulating. Thank you for giving me the opportunity to do research in combination with clinical work in difficult times of cost cutting and hospital mergers. To all contributing co-authors and colleagues at Oslo University Hospital, Norwegian Institute of Public Health, and Sahlgrenska University Hospital, thank you for never ending support and encouragement.

Brita Winje Askeland, Solveig Myking, and Ingvild Fjeldheim, who have been my officemates during different periods of this work, thank you for fruitful scientific discussions, for sharing the ups and downs of everyday work life, and for all of the humorous and warm discussions concerning important – and not so important – topics of daily life.

Finally, I want to thank my family; my dear husband and best friend Jens and my sons Peter, Stian and Emil. Thank you, Jens for your support, patience, and encouragement. Dear Peter, Stian, and Emil, I am so grateful to have you in my life.

This project has been financially supported by the Norwegian Extra Foundation for Health and Rehabilitation through EXTRA funds, the Unexpected Child Death Society of Norway, and Oslo University Hospital, which is gratefully acknowledged.

## Summary

Symphysis-fundus (SF) measurement, i.e., the systematic recording of the distance from the pubic symphysis to the uterine fundus, is used widely as a simple method to screen for fetal growth restriction (FGR). Pregnancies affected by FGR are at increased risk of adverse outcomes in the short and long term, and across generations. They constitute 40–60% of stillbirths, in which the failure to identify FGR is the most common finding of substandard care. Antenatal detection may improve perinatal outcome. Although it is used widely, the accuracy of SF measurement for FGR screening is controversial. Observational studies using SF measurements have reported findings ranging from very low to high sensitivity in the detection of small neonates. A previously published Cochrane review concluded that insufficient evidence was available to evaluate the effectiveness of the test in preventing adverse outcomes. As SF measurement is the sole tool for FGR screening for the vast majority of pregnancies globally, improved methodology and evidence are of great public interest.

In the research performed for this thesis, we conducted a systematic review to examine the use of SF height for the prediction of small-for-gestational-age (SGA) status in unselected and low-risk pregnancies. This review revealed a lack of relevant studies. Most studies were small and conducted in high-risk populations. The findings of this systematic review prompted the second study, in which we analyzed SF measurements from a large population-based cohort. We developed a reference standard for SF height and determined the effects of maternal and fetal factors on SF height. In the third study, we assessed the relationship between SF deviations and SGA at birth, and presented risk curves that delineate SF values corresponding to specific degrees of elevation in SGA risk relative to pregnancies with normal patterns of SF values.

We successfully characterized covariate values to individualize the population standard to suit individual pregnancies. The new reference standard shows a pattern that differs from that of currently used Scandinavian standards of older origin, reflecting changes in the pregnant population, as well as other factors. Our results refute the common belief that fundal growth flattens toward term. SF height continued to increase throughout pregnancy. We have shown that the quality of SGA risk prediction is only moderate in early pregnancy, but that quality and predictive value increase closer to term. The SGA detection rate was not improved by consideration of the pattern of change in serial SF measurements. This finding suggests that the current common practice of using falling or static SF values as criteria of abnormality is not necessarily useful. Given its consistent but still modest predictive value, SF measurement should preferably be used in combination with other available information to provide the best first-line screening. Further work is needed to establish the relationship between SF values and perinatal morbidity and mortality.

# List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals. Reprints were made with permission from the respective publishers.

- I. Pay ASD, Wiik J, Backe B, Jacobsson B, Strandell A, Klovning A. Symphysis-fundus height measurement to predict small-for-gestational-age status at birth: a systematic review. *BMC Pregnancy Childbirth* 2015;15(1):461.
- II. Pay ASD, Frøen JF, Staff AC, Jacobsson B, Gjessing HK: A new population-based reference curve for symphysis–fundus height. *Acta Obstet Gynecol Scand* 2013;92(8):925–933.
- III. Pay ASD, Frøen JF, Staff AC, Jacobsson B, Gjessing HK. Prediction of small-forgestational-age status at birth by symphysis-fundus height: a registry-based population cohort study. *BJOG* 2015, DOI: 10.1111/1471-0528.13727 [Epub ahead of print].

# Abbreviations

- AC, abdominal circumference
- ANC, antenatal care
- AUC, area under the curve
- BMI, body mass index
- BPD, biparietal diameter
- BW, birth weight
- CTG, cardiotocography
- CI, confidence interval
- dSF, deviation in symphysis-fundus height
- DOR, diagnostic odds ratio
- EFW, estimated fetal weight
- FGR, fetal growth restriction
- FL, femur length
- FN, false negative
- FP, false positive
- GAMLSS, generalized additive models for location, scale, and shape
- HC, head circumference
- LMP, last menstrual period
- MBRN, medical birth registry of Norway
- NLR, negative likelihood ratio
- NPV, negative predictive value
- pdBW, percent deviation in birth weight
- PLR, positive likelihood ratio

- PPV, positive predictive value
  ROC, receiver operating characteristic
  RR, relative risk
  SD, standard deviation
  SF, symphysis-fundus
  SGA, small-for-gestational-age
  SROC, summary receiver operating characteristic
  SU, Sahlgrenska University
- TN, true negative
- TP, true positive

# **Table of contents**

1	Gei	neral	introduction	1
	1.1	Nor	mal and abnormal fetal growth	2
	1.1	.1	Normal growth	2
	1.1	.2	Commonly used ultrasound standards	3
	1.1	.3	Small-for-gestational-age status	6
	1.1	.4	Fetal growth restriction	9
	1.2	Sma	all-for-gestational age screening using symphysis-fundus measurement	. 11
	1.2	.1	Historical background	. 11
	1.2	.2	Clinical performance of SF measurement	. 13
	1.2	.3	Longitudinal approaches	. 17
	1.2	.4	Concepts in screening	. 21
2	Stu	dy ra	ationale and aims	. 25
3	Ma	teria	l and methods	. 26
	3.1	Des	ign and data sources	. 26
	3.1	.1	Paper I	. 26
	3.1	.2	Papers II and III	. 27
	3.2	Stat	istical methods	. 28
	3.2	.1	Paper I	. 28
	3.2	.2	Paper II	. 29
	3.2	.3	Paper III	. 30
	3.3	Eth	ical approvals	. 31
4	Syr	nopsi	s of study results	. 32
	4.1	Pap	er I	. 32
	4.2	Pap	er II	. 33
	4.3	Pap	er III	. 34
5	Dis	cuss	ion	. 37
	5.1	Met	hodological considerations	. 37
	5.2	Inte	rpretation of results and comparison with previous findings	. 39
6	Co	ncluc	ling remarks and future perspectives	. 46
7	Ret	feren	ces	. 50
8	Co	rrecti	ons	. 56

# **1** General introduction

A major focus of antenatal care (ANC) is the identification of small-for-gestational-age (SGA) fetuses, in particular those with fetal growth restriction (FGR). Pregnancies affected by FGR are at increased risk of a wide range of adverse outcomes in the short and long term, as well as across generations (1-4). SGA is defined as a weight below a specific (distribution-based) threshold for gestational age, usually the 10th percentile, whereas FGR indicates the presence of a pathophysiological process occurring in utero that inhibits fetal growth. Antenatal identification of SGA aims to facilitate the identification of at-risk pregnancies requiring further investigation due to potential FGR. The appropriate identification and management of FGR improves perinatal outcome (5).

In Norway, screening for SGA includes the routine use of serial symphysis-fundus (SF) measurement, complemented by ultrasound measurement of fetal size in women with pregnancy complications or relevant histories or clinical evidence of FGR. SF measurement is a technique involving measurement of the maternal abdomen from the symphysis pubis to the uterine fundus with a tape measure. If the measurement is lower than expected according to a reference, further investigations for fetal growth and wellbeing are to be performed. Although used routinely in pregnancy to screen for SGA fetuses, evidence for the efficiency of this method remains unclear. Observational studies of SF measurement have been highly inconclusive, indicating anything from very low (27 %) to high (76 %) sensitivity in detection of SGA, and specificities ranging from 60 to 94 % (6). Additionally, studies have applied different SF curves, cut-off values for calling an SF result abnormal, and SGA definitions. As an important route of entry to ultrasound for most pregnancies in our population, improved methodology and evidence for SF screening is of great public health interest.

In this thesis, we investigate the use of SF measurement for the prediction of SGA status in unselected and low-risk pregnancies. We present high-quality replacement curves for the Norwegian population and explore new methodological approaches to the analysis and interpretation of SF deviation. To improve the predictive value of SF height, we look closely at the effects of maternal and fetal factors, as well as longitudinal measurements. Because population data sets of sufficient size and quality are not currently available in Norway, we use ANC data from Western Götaland County dating from 2005–2010, which we believe to cover a population reasonably similar to the Norwegian target population.

#### 1.1 Normal and abnormal fetal growth

The development of screening tools for the identification of deviation in fetal growth begins with the definition of normal and abnormal fetal growth, which includes the consideration of accurate dating and ultrasound standards (reference curves).

#### 1.1.1 Normal growth

Normal fetal growth pattern can be divided into three phases which involves hyperplasia and hypertrophy on a cellular level. During the first phase, which takes place from the beginning of development through the early part of the fourth month, rapid cell division and multiplication (hyperplasia) occurs as the embryo grows into a fetus. In the second phase, cell division declines and the cells increase in size (hypertrophy). In the third phase (after 32 weeks), there is rapid increase in cell size, rapid accumulation of fat, muscle, and connective tissue. Disturbance of fetal growth dynamics in these phases can lead to a reduced cell number, cell size, or both, resulting in restricted growth (7). In typically developing singleton fetuses, growth is less variable in early gestation, a period that is mainly genetically influenced. Considerably more variance occurs as gestational age advances and genetic and environmental factors influence fetal size and growth (8).

In Norway and Sweden, routine ultrasound examination to determine fetal age and term date is recommended in the second trimester, during weeks 17–19. Ultrasound dating assigns the same fetal age and birth term to all fetuses of the same size. Ultrasound-based fetal dating partially eliminates the effects of uncertain reporting of the day of the last menstrual period (LMP) and individual biological variability inherent in LMP-based fetal aging. In particular, the effect of individual variability in follicular phase length is avoided when dating is based on ultrasound measurement of fetal size (9). This approach achieves better agreement between the ultrasound-based term date and the actual date of birth than is possible with LMP-based term dating (10-13).

Around gestational week 18, the variability in normal fetal growth is moderate, and the assumption that all fetuses of the same size are of the same age is – for the majority of the population – reasonable. However, this assumption has some counterintuitive and potentially misleading consequences. Variables that influence fetal size at week 18, such as fetal sex, also influence fetal dating. As male fetuses are larger than female fetuses at this time, they will be

assigned an age that is slightly older than their "true" age; the opposite is the case for female fetuses (14). This difference can lead to erroneous conclusions about the effects of exposures during pregnancy in epidemiological studies, as well as systematic differences in, for instance, the day on which a given pregnancy is considered to be overdue. A potentially serious consequence is that fetuses with first- and early second-trimester FGR, which are small at week 18, may be assigned an ultrasound age that is younger than the "true" age. Consequently, they will be treated as young, normal fetuses, rather than older fetuses with FGR. In particular, the term date computed from the LMP is typically shifted to a later date when ultrasound is used, thereby also shifting the date on which an overdue pregnancy should be induced (12, 13, 15, 16). A common recommendation in Norwegian pregnancy care is that a fetus should be followed as a potential case of FGR when the LMP date is relatively certain and ultrasound dating extends the term date by 14 or more days (17).

Whereas week 18 dating is still being used in Sweden and Norway, international recommendations focus increasingly on dating in the first trimester, e.g., around week 12. Earlier dating has two potential advantages: first, dating closer to the LMP reduces individual size variability, leading to a more precise age estimate (18, 19); and second, repeated measurement of fetal size at weeks 12 and 18 improves the ability to detect possible early FGR, as it enables comparison of size at weeks 12 and 18, rather than size at week 18 and LMP-based age. One possible objection to early dating is that decisions concerning the management of extremely premature births depend to a considerable extent on estimated fetal age, particularly around weeks 22–24. Dating based on ultrasound-measured fetal size is an important predictor of survival after birth. Although numerous studies have shown improved precision of age estimates by early ultrasound relative to late ultrasound (20-22), very few studies have examined the quality of term prediction from the same perspective; some studies have suggested that early ultrasound achieves little improvement in term prediction (23, 24).

#### 1.1.2 Commonly used ultrasound standards

Ultrasound-based fetal size monitoring is usually based on two main components. The first component is the measurement and monitoring of a set of ultrasound parameters – e.g., head circumference (HC), biparietal diameter (BPD), femur length (FL), and abdominal circumference (AC) – in one or more instances during pregnancy. The second component is

the translation of ultrasound parameters into a fetal size (weight) estimate at each examination. The individual parameter estimates and weight estimates are then compared with standard fetal growth or size curves to detect signs of growth restriction.

Standard curves for individual parameter size and growth are typically developed from population data or data from smaller clinical studies involving repeated longitudinal measurement of individual fetuses. A population-based standard has the advantage of following the population as it is; the sample is large, as a large number of mothers undergo at least one follow-up examination. In addition, a large number of ultrasound operators have contributed to the data, which will average out specific biases that may be present in clinical studies performed by only a few - or a single - ultrasound operator(s). On the other hand, longitudinal measurement of individual fetuses is often the recommended approach; population-based approaches are argued to produce primarily cross-sectional data for any given day of pregnancy, whereas longitudinal data follow the growth of individual fetuses. Accordingly, longitudinally developed reference standards are argued to represent fetal growth, whereas cross-sectionally developed curves represent fetal size. However, the end product is always a reference chart of size by age percentiles. Percentiles as such are population concepts, and percentiles of size at a given age are by nature cross-sectional. If cross-sectional data are derived from a sufficiently large sample at each fetal age, longitudinal data are not actually needed to produce size-by-age curves. Although longitudinal data consist of more measurements per individual, only measurements obtained close to a given fetal age provide information about size percentiles at that age. Thus, there is little reason to expect the two approaches to produce substantially different charts (25).

In clinical practice, the monitoring of fetal size using size-by-age charts is usually based on clinicians' judgment. A fetus exhibiting normal growth remains largely in the same percentile throughout pregnancy; a fetus showing a clear fall to lower percentiles is likely to exhibit pathological growth. The exact rate of decline that is considered to be pathological is a matter of individual judgment. To improve the assessment of growth, some researchers suggest the use of conditional percentiles, which are developed from longitudinal data. By conditioning on fetal size at the first measurement, size percentiles for subsequent measurements can be developed for individual fetuses (26). A drawback of this method is, however, that conditional percentiles are strongly model dependent; very few fetuses fall into extreme percentiles at the first measurement, and the development of reliable conditional curves would require the

undertaking of very large studies. Currently, such models entail the assumption that all fetuses, whether they fall on the median or in the 1-percentile, will have similar growth patterns in the future. A second drawback is that a fetus in the 1-percentile at first measurement is potentially already pathological; the establishment of a new, conditional standard for this fetus essentially normalizes its small size. As conditional standards are also harder to understand and use in clinical practice, whether the added complexity results in improved detection is not clear.

The second component of fetal size monitoring is the estimation of fetal weight. Ultrasound determination of HC, BPD, FL, and AC can be used to derive an estimated fetal weight (EFW) using a formula, such as that of Combs et al. (27) or Hadlock et al. (28). Reference curves, i.e., fetal weight percentile curves, are developed from a study population by computing EFWs for all fetuses examined at a given age, and then constructing curves directly from these values (29). An advantage of this approach is related to the likely greater intrauterine size of fetuses born at term, relative to those born very or extremely preterm (Figure 1); the application of an intrauterine standard for ultrasound monitoring procedures is thus natural. A weakness of this approach is that formulas for weight estimation, such as Combs et al.'s (27) formula, are developed almost exclusively using data from term births. The application of such formulas to fetuses at 20–30 weeks of gestation involves gross extrapolation, with the assumption that fetal proportions in the early weeks of gestation are the same as those around term. Thus, this approach entails the risk of obtaining incorrect weight estimates in late second- and early third-trimester examinations.

An alternative approach to the Combs and Hadlock formulas is weight estimation based on population data. In the Norwegian eSnurra model (30), ultrasound parameters are used to predict the percentage of deviation in birth weight (pdBW), i.e., actual fetal weight as a percentage of the median weight of all births at the same age. The pdBW is related directly to fetal size and is used extensively for fetal monitoring in Norwegian and Swedish clinical practice. In the population-based approach, the number of births from week 24 onward is sufficient for the estimation of percentile curves based on observed BWs. Using ultrasound measurements for fetuses that remain in utero until term and those that are born preterm, the population-based system enables the computation of reference curves for actual BWs and for intrauterine fetal weight, without the need for the extrapolation of formulas.

Figure 1. Comparison of the ultrasound-based EFW reference chart constructed by Marsal et al. (29) (blue lines), with the eSnurra BW reference chart (31) (black lines). The 10th, 50<sup>th</sup>, and 90th percentiles are shown from 24 to 42 weeks of gestation. The mean ultrasound-based EFW is noticeably larger than the actual mean BW. At 28–32 weeks of gestation, the 50th percentile for BW is approximately comparable to the 10th percentile for EFW



The use of intrauterine standards is natural in the assessment of growth restriction and the identification of SGA fetuses; however, when using actual BW as the endpoint, SGA is most conveniently defined using standards from observed weights.

#### 1.1.3 Small-for-gestational-age status

SGA is defined as a weight below a certain limit compared with a reference population. SGA is important because the fetal/neonatal percentile is related inversely to adverse perinatal outcomes, with the greatest risk at weights below the 3rd percentile for gestational age (32). In Norway, SGA is defined as BW < 10th percentile for gestational age and sex. Several SGA standards are applicable to the Norwegian population. Skjærven et al. (33) developed a standard based on data from the MBRN using LMP dating. A more recent standard was developed on BWs in the Trondheim region over the years 1987-2005 in the eSnurra dating and weight prediction system (31), using ultrasound dating. The present thesis uses the eSnurra reference curves.

#### Secular trends in BW and SGA incidence in Norway

To assess secular trends in SGA incidence in the Norwegian birth population, application of the same standard to data from all years is important. Despite the existence of noticeable differences between LMP-based and ultrasound-based standards, the pattern of change over years is not influenced greatly by the standard used, as long as the standard is kept fixed.

Figure 2. Secular trend in median BW in Norway, 1967–2008 (data from the MBRN, n = 2,451,242). Håkon Gjessing, personal communication, November 16, 2015



Figure 2 shows that median BW in Norway increased moderately between 1967 and 1990. In subsequent years, a clear and strong increase of about 60 g is visible, with a peak of 3580 g in 2000. BW then again dropped rapidly, reaching the 1990 level in 2008.

To assess whether the marked peak around 2000 was due to a change in clinical practice, one must look specifically at how the BW distribution of spontaneous births changed over the years 1980–2008. Figure 3 shows that the lower (5th percentile) and upper (95th percentile) parts of the distribution changed at almost exactly the same rate as the middle part of the distribution (the median). Thus, the increase in BW up to 2000 clearly represents a complete shift in the BW distribution at all levels.

Figure 3. Change in BW in Norway, 1980–2008, for spontaneous births only (data from the MBRN, n = 1,354,227). Black line: change in median BW since 1980. Solid gray line: change in the 95th percentile since 1980. Dotted gray line: change in the 5th percentile since 1980. Håkon Gjessing, personal communication, November 16, 2015



Figure 4. Secular trend in the incidence of SGA in Norway, 1998–2008 (data from the MBRN, n = 429,026). Black curve: all births. Gray curve: spontaneous births. Ultrasound-based dating was available only from 1998 onward. Håkon Gjessing, personal communication, November 16, 2015



Figure 4 shows that the drop in SGA for spontaneous births around 2000 corresponds to the increase in all parts of the BW distribution, as shown in Figure 3. For all births, the pattern of

change in SGA remains the same, but the incidence increases. This pattern strongly suggests that the drop in SGA incidence around 2000 is due to a general shift (increase) of BW, rather than a change in clinical practice.

When the SGA standard is kept constant over all years of evaluation, the choice of standard does not influence secular trends to a noticeable degree. This finding strengthens the hypothesis that the trends are due to the underlying shift in BW.

#### 1.1.4 Fetal growth restriction

FGR describes a decrease in the fetal growth rate that prevents a fetus to achieve its genetically determined growth potential (34). This condition is common, affecting 3–10% of all pregnancies, depending on the definition used (35), and it is linked strongly to antepartum death (36, 37). Growth-restricted fetuses are at increased risk of several neonatal morbidities, prematurity, and decreased fetal reserve during labor (1, 3, 4, 38). Evidence also suggests that FGR has long-term complications extending well beyond childhood, such as hypertension, hyperlipidemia, coronary heart disease, and diabetes mellitus in adults (39).

FGR and SGA are commonly used interchangeably in the literature. They are not synonymous, however, and the distinction between growth-restricted fetuses and those that are constitutionally normal but SGA is important. SGA is a descriptive term and means that the fetal/neonatal weight is less than expected, regardless of the cause. In other words, it refers to the size, irrespective of the individual growth velocity. An estimated 50–70% of fetuses with BWs < 10th percentile for gestational age are constitutionally small, with fetal growth appropriate for parental size and ethnicity; they are not at high risk of perinatal mortality or morbidity (40). SGA fetuses with BWs < 2nd percentile for gestational age have a greater likelihood of being growth restricted (41).

#### Pathophysiology of FGR

Placental insufficiency is the major cause of growth restriction, which prevents fetuses from reaching their genetically determined growth potential (42). A fetus exposed to insufficient supplies of oxygen and nutrients due to a malfunctioning placenta responds to this unfavorable environment by making adjustments that maximize the chance of survival (43). These adjustments include preferential preservation of fetal growth over placental growth,

changes in fetal movement pattern, and the eventual deceleration of the fetal growth rate (Figure 5). With continuing deprivation, compensation gives way to decompensation. For example, a decreased fetal growth rate is an adaptive and protective reflex in the early stages of deprivation; in later stages, however, it can represent final decompensation associated with chronic hypoxia and acidosis. A critical component of the fetal homeostatic response involves flow redistribution, which favors perfusion of the vital organs (the brain, heart, and adrenals) at the expense of muscle, viscera, skin, and other less critical tissues and organs. These adaptive phenomena provide the basis for clinical surveillance.

Figure 5. Proposed fetal response to stress, which results in fetal death. Reprinted from Warrander and Heazell (44), with permission from Elsevier.

#### Stress

↓ Chronic respiratory and nutritive insufficiency

**Primary adaptive response**  $\psi$  Decreased fetal growth rate

#### Secondary adaptive response

Fetal energy conservation

 Decreased fetal movement
 Decreased fetal heart rate reactivity

 Circulatory redistribution

 Falling cerebral flow impedance
 Rising umbilical and aortic impedance
 Fetal growth preferred over placental growth
 Increased efficiency of placental exchange

 ✓ Polycythemia

Greater O2 carrying capacity

#### Progressive decompensation

Hypoxia → respirtory acidosis → metabolic acidosis High impedance in fetoplacental and systemic circulation results in absent end diastolic flow in umbilical arteries Declinig amniotic fluid volume → oligohydramnios Loss of fetal movement Loss of fetal heart rate reactivity and variability

 $\downarrow$  Persistent late decelerations

Agonal pattern

#### Death

In Norway, clinical suspicion of FGR based upon SF measurement or high-risk evaluation is followed by detailed ultrasound assessment of fetal weight and growth, amniotic fluid evaluation, Doppler examination, and cardiotocography (CTG) (17). Initially, decreased fetal growth can be seen as reduced EFW. The impaired efficiency of the placenta may be reflected as abnormal blood flow in the umbilical artery and vein, as well as in the uterine artery. As

fetal circulation is redistributed to maintain the blood supply to the vital organs, decreased amniotic fluid volume reflects reduced blood flow to the kidneys and decreased urinary production. The blood flow to the fetal brain, examined in the medial cerebral artery, primarily increases in these situations, but returns to normal if the fetus is severely compromised. One of the final effects in the fetus may be reflected as change in blood flow pattern in the ductus venosus. Later, pathological CTG findings may appear before intrauterine fetal death occurs (45).

As no effective treatment to reverse FGR is available, prenatal management is aimed primarily at determining the ideal timing and mode of delivery. This assessment must be individualized, depending on several variables: gestational age, maternal health, severity of growth restriction, and fetal well-being (46). Operative deliveries are common in SGA/FGR pregnancies (47). When a fetus experiences relative hypoxia before the onset of labor, even normal uterine contractions can lead to fetal distress. Active decision making about delivery mode and high awareness of the need for intensive surveillance during labor are probably of great importance for outcome in individual FGR fetuses.

# 1.2 Small-for-gestational age screening using symphysis-fundus measurement

#### 1.2.1 Historical background

Serial SF measurement gained attention as a screening tool for the detection of growth restriction with the publication of promising results from early studies by Westin in the 70s (48). In a large uncontrolled study of low-risk, uncomplicated pregnancies in Sweden, Westin showed that SF measurement was superior to maternal weight gain, maternal girth measurement, and biochemical analyses (uE3, HPL) for the detection of SGA status. The routine introduction of a reference SF chart (gravidogram) in the case notes of all patients was associated with a significantly steeper fall in the local perinatal mortality rate compared with the overall Swedish statistics. This chart was constructed based on data from 100 healthy, completely uneventful singleton term pregnancies, with no known factor affecting fetal growth, normal pre-pregnancy weights within the 10–90th percentiles, known dates of LMP,

uneventful intrapartum periods, and birth to healthy neonates with weights within 1 standard deviation (SD) of the mean according to Swedish growth charts.

Westin (48) provided the following instructions for SF measurement: "the patient should be supine. Legs should be straight otherwise the pubic symphysis moves upwards. The uterus should be relaxed and the bladder empty. The measurement should be performed along the longitudinal axis of the uterus defined both in the frontal and lateral projections. During the third trimester, measurements should be performed along the longitudinal axis of the fetus whereby fetal crown-rump length will be reflected." SF values  $\geq$  3 cm above and below the mean were regarded as signs of growth acceleration and growth retardation, respectively. Static or falling SF values, measured weekly on more than two occasions, were regarded as signs of retarded growth, irrespective of whether previous observations were close to, above, or below the mean. The method presented by Westin became implemented as a screening tool for the detection of SGA/FGR in the general population in Scandinavian countries.

In 1996, Steingrimsdottir et al. (49) introduced an updated SF chart that replaced the Westin chart (48) for use in routine clinical practice in Sweden. This chart was based on measurements from 1650 Swedish women with ultrasound-dated pregnancies. Criteria for inclusion in the study were similar to those used by Westin (48), except that Steingrimsdottir et al. (49) made no exclusion on the basis of maternal height or weight. This curve is situated about 1 cm higher until 37 weeks. After 37 weeks the curves almost coincide with each other. Two other Swedish curves for SF measurement were published in the same time period (50, 51), but have never been in clinical use. The largest study, by Kieler et al. (51), produced a curve based on data from 2225 women who participated in a multicenter randomized trial; the second curve (50) was based on data from 403 women attending routine ANC. Both curves closely resemble the curve by Steingrimsdottir et al. (49) until the end of pregnancy; at pregnancy week 40, the curve developed by Kieler et al. (51) shows stronger SF growth than does Steingrimsdottir et al.'s (49) curve.

In Norway, a modified version of Westin's curve has been used routinely in ANC to screen for growth restriction in the low-risk population (52). As the original material and statistical analyses on which the curve's development in the 1970s was based are not available, neither the definition nor the significance of the limits of normality is known. Recently, the modified Westin curve was replaced by the SF reference curve presented in paper II (53), which was based on data from 42,018 ultrasound-dated pregnancies. Women with medical conditions or complications of pregnancy were not excluded from that study because the aim was to construct population-based references, rather than those of a "healthy" selected population. The shape of the curve differs substantially from the modified references of approximately 50 years ago presented by Westin (48), with a considerable increase in SF height at all gestational ages. The new curves are more similar to those of Steingrimsdottir et al. (49) for pregnancy weeks < 34, but with an approximately 0.5 cm higher mean. From pregnancy week 34, the new curve (50th percentile) shows greater SF growth than does the curve of Steingrimsdottir et al. (49); this pattern is more similar to the curve developed by Kieler et al. (51). The observed differences may be due to differences in analytical strategy, study design, and increased maternal body size in recent years.

#### 1.2.2 Clinical performance of SF measurement

Early reports of the performance of SF measurement as a screening test for growth restriction were encouraging, with a sensitivity of 70% (48, 54). Later reports were not quite so enthusiastic (55, 56). Different definitions of abnormal SF measurement results have been presented in the literature. The populations tested also differ, and some have been high risk, which has artificially increased the detection rate. In most studies, the designation of abnormality requires at least two or three consecutive measurements below the 10th percentile. Theoretically, increasing the number of abnormal measurements required for the diagnosis of SGA should reduce the false-positive (FP) rate (1 – specificity).

Table 1 shows the performance of SF measurement in the detection of SGA in 19 SF studies (48, 54, 55, 57-71) conducted in northwestern Europe. SGA is defined as BW < 10th percentile, > 1 SD below the mean, and  $\geq$  2 SDs below the mean. Also shown is the prevalence of SGA in the study populations, the number of women in each study, and definitions of positive test results. The ranges of sensitivity and specificity are wide, from 27% to 86% and 47% to 97%, respectively. This variation is partly the result of the use of different SGA definitions, cut-off criteria used to define abnormal results, and populations. As expected, positive predictive values (PPVs) varied, given that the prevalence of SGA differed in the populations studied.

gative predictive value in 19 SF studies
alue, ne
, positive predictive v
specificity,
. Sensitivity,
Table 1.
14

First author	SGA definition	Population	Definition of + test	Sensitivity	Specificity	PPV	NPV
Westin, 1977	BW > 1SD below mean	Sweden,	A. ≥3 cm below mean	A. 68 %	A. 89 %	A. 44 %	A. 96 %
	Prevalence 11 %	n=428	B. Static/falling values	B. 52 %	B. 95 %	B. 54 %	B. 95 %
Quaranta, 1981	BW < 10th percentile Prevalence 30 %	England, n=138	Two consecutive/three isolated values < 10th percentile	73 %	79 %	60 %	88 %
Wallin, 1981	BW ≥ 2 SD below mean Drevalance not stated	Sweden, n=812	Two values > 3cm below mean	62 %	88 %	42 %	94 %
Rosenberg, 1982	BW < 10th percentile Prevalence 6.6 %	Scotland, n=761	<ul> <li>A. Two consecutive/three isolated values</li> <li>&lt; 10th percentile</li> </ul>	A. 56 %	A. 85 %	A. 21 %	A. 96 %
			B. 20 %, or	B.61 %	B. 79 %	B. 17 %	B. 97 %
			C. 30 %, or	C. 52 %	C. 92 %	C. 31 %	C. 96 %
			D. 40 % values < 10th percentile	D. 48 %	D. 92 %	D. 30 %	D. 96 %
Calvert, 1982	BW < 5th percentile	England,	Applied:	< 5th	< 5th	< 5th	< 5th
	Prevalence 6.6 %	n=381	A. Westin criteria	A. 72 %	A. 58 %	A. 11 %	A. 97 %
	BW < 10th percentile		B. Belizian criteria	B. 60 %	B. 76 %	B. 15 %	B. 96 %
	Prevalence 12 %		C. Quaranta criteria	C. 36 %	C. 92 %	C. 24 %	C. 94 %
				< 10th	< 10th	< 10th	< 10th
				A. 76 %	A. 60 %	A. 20 %	A. 95 %
				B. 64 %	B. 79 %	B. 29 %	B.4%
				C. 36 %	C. 94 %	C. 43 %	C. 92 %
Cox, 1983	BW < 5th percentile	Ireland,	Two values < 10th percentile	< 5th	< 5th	< 5th	< 5th
	Prevalence 30 %	n=123		70 %	87 %	70 %	87 %
	BW < 10th percentile			< 10th	< 10th	< 10th	< 10th
	Prevalence 48 %			58 %	95 %	92 %	71 %
Cnattingius, 1984	l BW ≥ 2 SD below mean	Sweden,	Low (L): Last value $\geq$ 3 cm below the mean	L 50 %	r 98 %	L 39 %	7 66 %
	Prevalence 2.7 %	n= 527	Static (S): Last three values same, but no value	S 7 %	S 87 %	S 1 %	S 97 %
			> 2 cm below mean	C 29 %	C 94 %	C 12 %	C 98 %
			Catch up (C): One value > 3 cm below mean,	L+C 79 %	L+C 92 %	L+C 21 %	L+C 99 %
			but last < 3 cm below mean	L+C+S 86 %	L+C+S 79 %	L+C+S 10 %	L+C+S 99.5 %
Rogers, 1985	BW < 10th percentile	England,	Westin criteria	73 %	92 %	51%	97 %
	Prevalence 10.4 %	n=250					
Persson, 1986	BW < 10th percentile	Sweden,	> 2 SD below mean	27 %	88%	18 %	92 %
	Prevalence 9 %	n=2941					

Okonofua, 1986	BW < 10th percentile	England,	Two consecutive values < 10th percentile	71%	85 %	31 %	98 %
	Prevalence 7 %	n=100					
Pearce, 1987	BW < 10th percentile Drevelence 14.3 %	England, n=200	< 10th percentile	76%	% 62	36 %	95 %
Cnattingius, 1988	3 BW 2 2 SD below mean	Sweden,	≥ 3 cm below mean or falling or static values	59 %	97 %	15 %	% 66
	Prevalence 0.8 %	n=3038					
Stuart, 1989	BW < 10th percentile	England,	< 10th percentile	A. 51 %	A. 88 %	A. 26 %	A. 96 %
	Prevalence:	A. n=1139		B. 65 %	B. 88 %	B. 21 %	B. 97 %
	A. General population, 7.6 %	B. n=319					
	B. Women with ≥ four SF height						
	values, 7.2 %						
Lindhard, 1990	BW< 10th percentile	Denmark,	Two values with a fall of > 20 %/two values	A. 28 %	A. 97 %	A. 41 %	A. 94 %
	Prevalence:	n=1639	< 10th percentile/three values with no	B. 48 %	B. 97 %	B. 47 %	B. 97 %
	Group A (open), 8 %	(RCT)	increase				
	Group B (masked), 6 %						
Jensen, 1991	BW < 10th percentile	Norway,	≥3 cm below mean	41 %	87 %	33 %	93 %
	Prevalence 11 %	n=831					
Gardosi, 1999	BW < 10th percentile	England, n=1272	<pre>! &lt; 10th percentile/two values indicating a</pre>	A. 48 %	not stated	not stated	not stated
	Prevalence:		curve flatter than the 10th percentile	B. 29 %	not stated	not stated	not stated
	A. Study group, 10,6 %	Customized					
	B. Control group, 11.9 %						
Wright, 2006	Prevalence:	England,	< 10th percentile/two values indicating a	A. 36.2 %	not stated	not stated	not stated
	A. Study group, 19 %	n=598	curve flatter than the 10th percentile	B. 16.1 %	not stated	not stated	not stated
	D. Comuni group, 20.0 %	Customized					
Bergman, 2006	BW > 2 SD below mean	Sweden,	Applied:	A. 51 %	A. 83 %	63 %	75 %
	Prevalence:	Case, n=169	A. Value > 2 SD below mean/ K-curve	B. 32 %	B. 90 %	64 %	70 %
		Control, n=296	B. Value > 2 SD below mean/ S-curve	C. 85 %	C. 47 %	48 %	85 %
			C. Value > 1 SD below mean/ K-curve	D. 72 %	D. 71 %	59 %	82 %
Hargreaves, 2011	1 5000g <term bw<2500g<="" td=""><td>England, n=3200</td><td>Criteria not stated</td><td>20 %</td><td>94 %</td><td>7 %</td><td>% 66</td></term>	England, n=3200	Criteria not stated	20 %	94 %	7 %	% 66

Definition of		SGA (BW<10-pe	rcentile)	
positive test	Sensitivity (%)	Speificity (%)	PPV (%)	NPV (%)
2,5	23	93	29	90
5	32	88	26	91
10	47	79	22	92
15	58	93		
2,5	30	92	11	98
5	38	87	9	98
10	54	77	8	98
15	64	68	7	98

Table2. Sensitivity, specificity, positive predictive value, and negative predictive value of the SF reference curve constructed by Pay et al. (53)

Table 2 shows the clinical usefulness of the population-based reference presented in paper II (53). The sensitivity and specificity for the detection of SGA defined as BW < 10th percentile, with the 10th percentile SF value serving as the cut-off value, were 47% and 79%, respectively. For SGA defined as BW < 2.5th percentile, the corresponding values were 54% and 77%. With stricter SF cut-off values, sensitivity decreases and specificity increases. For example, use of the 2.5th percentile cut-off value yields a sensitivity of 23% and a specificity of 93% for the detection of SGA (BW < 10th percentile).

Observational studies suggest that customized SF curves involving adjustment for maternal and fetal factors influencing SF height improve the detection of SGA status. Gardosi et al. (61) found that the use of customized charts improved the sensitivity of neonatal SGA detection relative to abdominal palpation (48% versus 29%). The use of customized charts was also associated with fewer referrals for investigation and fewer admissions. In an audit, Wright et al. (71) showed that the use of customized SF charts resulted in the detection of 36% of SGA neonates, whereas only 16% were detected when customized charts were not used (this group received traditional ANC). Serial measurement of SF height and plotting on customized growth charts are now recommended by The National Institute for Health and Care Excellence and the Royal College of Obstetricians and Gynecologists in the UK.

#### Lack of improved outcomes with SF measurement screening

No good evidence shows that the introduction of routine SF measurement leads to reductions in perinatal morbidity and mortality. The positive results reported by Westin (48) may have been due to chance, as the study was not a controlled study. Lindhard et al. (63) conducted a

randomized controlled clinical trial investigating the clinical performance of SF measurement versus clinical palpation in a population of 1639 Danish women. Pregnant women were allocated randomly to the intervention (SF measurement) and control groups at around 14 weeks of gestation. SF height was measured routinely after 28 weeks, and the results were plotted on a locally derived chart. Women in the control group underwent observations made with a fabric tape with no marking. The tape was cut and was not measured until after delivery. No significant difference was found between methods in the SGA detection rate, numbers of interventions and additional diagnostic procedures, or the condition of newborns. This study contributed decisively to a Cochrane review (72), which showed that evidence was insufficient to recommend the use of SF measurement in ANC; however, the authors opined that "it would be unwise to abandon the use of symphysis-fundus value measurement unless a much larger trial likewise suggested it is unhelpful."

The inability of randomized controlled studies to document an improvement in outcome following SF measurement screening may to some extent reflect the limited accuracy of SF measurement. Many studies have also determined that the reliability of SF measurement is unacceptably low (57, 73). These studies have highlighted what is probably the major shortcoming of SF measurement: the unacceptably large measurement error due to interobserver variation, even between experienced practitioners. An unreliable test has lower accuracy in terms of sensitivity and specificity, which in turns reduces its value as a screening tool and demonstrates that it will not result in improved health indicators. In addition, the value of a screening method depends on the general clinical circumstances into which it is introduced. The beneficial effect of screening in terms of improved outcome depends not only on the predictive values of the method, but also on the value of additional diagnostic procedures, the implementation of proper measures in reaction to pathological values, and the possibility of correcting or preventing adverse outcomes.

#### 1.2.3 Longitudinal approaches

The standard approach to the detection of growth restriction by SF measurement is natural from a clinical point of view; a pregnancy is followed along percentile charts and an "alarm" is triggered when the SF measure drops below a set percentile limit. This method ensures that a specific percentage of the population triggers the alarm on any given day of measurement. For a single measurement on a given day, this method is nearly optimal; the smallest

measurements, possibly adjusted for maternal and fetal covariates as in the customized setting, trigger the alarm. Clearly, this approach is related more closely to SGA detection than to FGR detection. However, SF measurements are taken serially; the specification of a reasonable procedure for improved FGR detection based on serial measurement is considerably more complicated. In practice, percentile charts are often employed in a repetitive fashion, with at least one measurement below the cut-off triggering the alarm. The drawback of this approach is that pregnancies with large numbers of measurements have a greater likelihood of triggering alarms, due simply to the greater probability that at least one of many measurements falls below the cut-off value by chance. As a consequence, the percentage of pregnancies with positive test results increases substantially. Although this trend typically increases sensitivity, it also reduces specificity. There is also no particular reason to believe that it will improve detection of FGR. Rather, screening of the same type is simply repeated.

To bring procedures closer to FGR detection, informal recommendations, such as observation to detect a drop in percentile values over the course of a pregnancy, are frequently given. The required magnitude of this decline and number of measurements showing a consistent drop are often left to clinical judgment. As for ultrasound procedures, the potential decline in future SF measurements could be evaluated in a conditional manner, with percentiles for future change computed depending on the most recent measurement (26, 74). Although this approach is closer to the FGR paradigm, it has other challenges, as discussed for ultrasound detection in Section 1.1.2.

Bergman et al. (75) proposed and evaluated a comprehensive method that targets FGR detection and controls the number of alarms triggered during a pregnancy. To focus on FGR detection, the authors recommend the assessment of relative, rather than absolute, growth. Relative growth can be computed from week to week, and the next week's change can be assessed relative to the current SF measurement. The model was developed in terms of logarithms of SF measurements; thus, relative change is assessed as log[SF(t)] - log[SF(t - 1)], where SF(t) is the measurement taken at week t. By employing relative change, the actual change is assessed in relation to its importance. Presumably, a moderate change in an already large SF measure would be of less clinical importance than the same change in a smaller SF measure and/or earlier in pregnancy. Evidence also indicates that the use of relative measures reduces or eliminates the influence of maternal and fetal covariates, such as maternal body

mass index (BMI) and fetal sex. To improve control of the number and timing of alarms triggered during a pregnancy, a method for stochastic control of detection in a serial system is employed.

All of the approaches described above have one issue in common; the curves and testing standards are developed from SF measurements alone. Normal and deviating growth are defined intrinsically, with reference only to the standards themselves. The actual clinical utility of such standards must thus be evaluated in a different setting, in which the growth standards and detection limits are related to outcomes of direct clinical importance, such as stillbirth and neonatal morbidity. As a consequence, the clinically important endpoints do not directly influence the development of the growth standards, but appear only in after-the-fact evaluation of the results. In the study reported in paper III (76), our approach was to link curve development to endpoints of clinical relevance. Although targets such as fetal death and morbidity are clearly of great clinical relevance, they are difficult to incorporate in the practical development of growth standards. This difficulty is due partly to the low incidence of the most important outcomes, as well as the number of possible outcomes, which raises the question of which outcomes are consequences of FGR and could potentially be averted if detected through FGR at an early stage. As a middle ground, in the study reported on in paper III (76), BW and SGA status at birth served as targets. BW and SGA status at birth have frequently been related to perinatal mortality and morbidity (77). At the same time, as BW of almost all neonates is measured, by definition, about 10% of the population is SGA at birth, which provides a large and stable endpoint for curve development. BW and SGA status thus function as accessible proxies for endpoints of clinical interest. The focus of curve development thus shifted from intrinsic description of the population to the construction of a predictive model with BW and SGA status as targets. The development of curves for the relative risk (RR) of SGA, rather than percentile curves or relative growth curves, is thus natural. As expected, fetuses following the median SF value through pregnancy have been shown to have an approximately 10% risk of SGA status at birth. Deviations below the median SF value at a given age translate into an increased risk of SGA status at birth and a corresponding increase in the RR. RR curves can be drawn over the span of a pregnancy, providing an alternative measure of deviation from expected size. A result derived directly from the RR prediction curve is that a specific deviation from the median SF measure at a given age is of less importance in the earlier weeks than in the later weeks of pregnancy. This finding is in line with most prediction problems; the endpoint is harder to predict at an early

19

stage, and deviation from normal is thus less indicative of an adverse outcome. This obvious fact is not taken into account in intrinsic studies in which growth is defined without reference to an endpoint.

This SF method is new in that it connects the endpoint directly to curve development, and does not require separate follow-up evaluations of SGA prediction for the assessment of prediction quality, although such evaluations always have independent value. Importantly, prediction models developed using data from small clinical studies must be evaluated separately for at least two reasons. First, a small dataset results in a model that is over fitted to the data; the prediction quality is exaggerated and is not representative of future applications. Second, the data may be skewed and not completely representative of the total population for which the screening tool is designed. In contrast, with population-based model development, the study population achieves nearly 100% coverage of the region it represents. The selection problem is absent, and statistical estimates are stable. In principle, the models could be developed using, for instance, a random sample of half of the population data and then evaluated with tools such as receiver operating characteristic (ROC) curves using the second half of the data. However, due to the size of such datasets, this approach has no practical difference from the use of the entire dataset for development and quality assessment.

Despite these advantages, this approach has drawbacks. Serial SF measurements are assessed using lines of constant RR, rather than percentiles, which places all measurements obtained during a pregnancy on equal footing in terms of risk prediction. However, the approach does not entirely account for the overall test characteristics, such as total sensitivity and specificity, which are the targets of the relative growth SF approach. It also, by definition, targets the prediction of SGA status at birth, rather than the assessment of FGR specifically. However, it enables evaluation of the effect of change in growth pattern directly in terms of clinical outcome. We have shown that only the last SF measurement has predictive value; change in growth prior to this value adds little to the prediction (76). This finding is a clear indication that actual FGR detection using SF measurement may be difficult, in line with results showing that the added quality of the detection model used in the relative growth SF approach has a limited payoff in terms of outcome detection (75).

#### 1.2.4 Concepts in screening

A diagnostic test is any kind of medical test performed to aid in the diagnosis or detection of a suspected disease or condition. This type of test differs from a screening test, which is used in low risk settings, when no disease or condition is suspected a priori. Screening tests are offered to asymptomatic people who may or may not have early disease or disease precursors, and test results are used to guide whether a diagnostic test should be offered. Diagnostic tests are offered to people who have specific indications of possible illness (history, symptom, sign or positive screening test result) to determine whether they have the disease in question.

Screening has benefits, costs, and harms, and clinicians have an ethical obligation to maximize benefits and minimize harm; the overall benefits should outweigh any harm that results from screening (78). In addition, the benefits of screening should justify its expense. Screening tests will always need to balance sensitivity against specificity; setting a liberal cutoff in the screening test will obviously increase sensitivity, but at the same time lead to reduced specificity and thus increased cost in terms of expensive follow-up, unnecessary treatment etc.

SF measurement is a simple, non-invasive, and harmless test with no negative side effect except the potential risk of increased maternal anxiety in the case of deviant results. The accuracy of SF measurement is, however, low. EFW in the third trimester has been suggested as a way to detect SGA fetuses. A recently published study by Sovio et al (79), reported a sensitivity of 57 %, and a specificity of 90 % for detection of SGA fetuses with universal ultrasound in the third trimester. EFW by ultrasound is a costly method demanding trained personnel and expensive equipment. Additionally, from a global perspective EFWs by ultrasound is not available for most pregnant women. Accordingly, after refinement of the method, SF measurement might be applicable as a screening test and EFW by ultrasound as a diagnostic method in pregnancies of suspected FGR.

#### Sensitivity and specificity

Test results have four possible interpretations: true positive (TP), true negative (TN), FP, and false negative (FN; Figure 6).

Figure 6. Four possible interpretations of test results

	Disease present	Disease absent
Test positive	True positive (TP)	False positive (FP)
Test negative	False negative (FN)	True negative (TN)

Sensitivity (also called the TP rate) is defined as the proportion of patients with the disease who have positive test results. It is calculated by the formula: sensitivity (TP) rate = TP / (TP + FN). Specificity (also called the TN rate) is defined as the proportion of patients without the disease who have negative test results. It is calculated by the formula: specificity (TN) rate = TN / (FP + TN). Most tests are characterized by a trade-off between sensitivity and specificity. A change in the cut-off value to achieve higher sensitivity results in lower specificity, with more FP results. Two other proportions, the FP and FN rates, are sometimes reported and may be calculated (Figure 7).

Figure 7. Summary of terms used to describe the accuracy of a diagnostic test. TP, true positive; FP, false positive; TN, true negative; FN, false negative

True positive rate	=	TP	=	sensitivity
		TP + FN		
True negative rate	=	TN FP + TN	=	specificity
False positive rate	=	FP	=	1 - specificity
		FP + TN		i specificity
False negative rate	=	FN	=	1 - sensitivity
		TP + FN		
Positive predictive value	=			
		1F + FP		
Negative predictive value	=	TN		
-0		TN + FN		

#### **Predictive values**

Sensitivity and specificity convey information about whether a test is useful in making a diagnosis. Once the test has been performed, sensitivity and specificity do not indicate whether a positive result truly indicates the presence of disease. That information is given by predictive values. The PPV is defined as the proportion of patients with positive test results who have the disease. It is calculated by the formula: PPV = TP / (TP + FP). The negative

predictive value (NPV) is defined as the proportion of patients with negative test results who do not have the disease. It is calculated by the formula: NPV = TN / (TN + FN). Importantly, predictive values vary with the prevalence of a disease, defined as the proportion of the population that has the disease at a given time.

#### Receiver operating characteristic curves

ROC curves are used to illustrate and evaluate the accuracy of one or more tests over a range of cut-offs. The ROC curve on a graph for each cut-off plots the sensitivity (TP rate) on the *y* axis, and 1 – specificity (FP rate) on the *x* axis. For an ideal test, the plotted curve ascends along the *y* axis; for a test of no diagnostic value, the plot ascends along the y = x line.

Summary ROC (sROC) curve analysis is a recently developed statistical technique that can be applied to meta-analysis of diagnostic tests. This technique can overcome some of the limitations associated with the pooling of sensitivities and specificities from published studies. The sROC curve is initially constructed by plotting the sensitivity (true positivity) and false positivity (1 – specificity) of each study. After mathematical manipulation of the true and false positivities, linear regression is performed to calculate the slope and y intercept. These coefficients are then entered into the sROC equation to generate the sROC curve (80).

#### Likelihood ratios and diagnostic odds ratios

Likelihood ratios compare the proportions of patients with and without the disease who have positive (or negative) test results. They are calculated by the formulas: positive likelihood ratio (PLR) = sensitivity / (1 - specificity) and negative likelihood ratio (NLR) = (1 - sensitivity) / specificity.

A likelihood ratio > 1 indicates that the test result is associated with the presence of the disease, whereas a likelihood ratio < 1 indicates that the test result is associated with the absence of disease. Evidence for the presence or absence of disease strengthens with the distance of a likelihood ratio from 1. Likelihood ratios > 10 and < 0.1 are considered to provide strong evidence to rule diagnoses in and out, respectively, in most circumstances (81).

The diagnostic odds ratio (DOR) is used commonly as an overall indicator of diagnostic performance. It is calculated as the odds of a positive test result among those with the target condition, divided by the odds of a positive test result among those without the condition

(DOR = PLR / NLR). As a general rule, DORs > 100 indicate high accuracy, values of 25–100 indicate moderate accuracy, and those <25 indicate that the test is not useful (81).

#### **Relative risk**

The result of a test (positive or negative) can be considered to be an exposure. In the setting of paper III (76), the event is an SGA outcome, and the exposure is an SF value falling below a certain percentile. The RR thus corresponds to: RR = [TP / (TP + FP)] / [FN / (FN + TN)] = PPV / (1 - NPV).

The RR is thus equivalent to the PLR, but reversed to reflect the probability of an outcome conditional on the test result. The numerical values of the RR and PLR are close for rare outcomes; when SGA is defined as BW < 10th percentile, they usually remain close.

When defining SF percentile curves based on exposure, the RR measures the increased SGA risk for a fetus in a low SF percentile relative to one at the median value.

# 2 Study rationale and aims

In The Lancet's series on stillbirth prevention, an expert panel ranked improved methodology and evidence for SF height measurement among the top 10 research priorities in efforts to reduce stillbirth (82). This perspective is in line with the Cochrane review's call for research to establish the sensitivity and specificity of SF height in detecting SGA, and to determine its effectiveness in reducing perinatal morbidity and mortality (72).

The use of SF measurement is widespread, despite its uncertain effects. It has significant impacts in clinical practice, leading to interventions that increase the rates of hospital admission and ultrasound examination. Thus, substantial benefits to patients and health care providers would likely accrue from an increased understanding of the clinical usefulness of this method. Studies conducted to date have focused mainly on the definition of percentile cut-offs and evaluation of their ability to identify the risk of SGA/growth restriction. This is not evident; growth trajectories differ in health and disease, and a given distribution-based percentile may indicate different risks according to gestation (77).

The objective of this study was threefold: to systematically identify, appraise, and synthesize published evidence for the accuracy of SF height measurement for the prediction of SGA status at birth; to present a new population-based SF reference curve; and to develop and evaluate a new SGA risk prediction model based on SF measurement.

The aims of the individual studies were:

Paper I: to conduct a systematic literature review of the sensitivity and specificity of SF height for the prediction of SGA status at birth in unselected and low-risk pregnancies, and to perform meta-analyses of test accuracy.

Paper II: to present a new reference curve for SF height, and to determine the effects of maternal and fetal factors on SF height.

Paper III: to develop and evaluate a new method to assess the risk of SGA status at birth by SF measurement.

# 3 Material and methods

#### 3.1 Design and data sources

#### 3.1.1 Paper I

Paper I reports on a systematic review (6). We systematically searched PubMed, Medline, Embase, CINAHL, the Cochrane Library, and SweMEd for studies that assessed the accuracy of SF measurement for the identification of SGA neonates. The search strategy involved combinations of SF-related terms appearing in subject headings and as keywords. We searched the databases from inception through September 2014. Inclusion criteria were: cohort study of test accuracy performed in a routine ANC setting; measurement of SF height in all participants; classification of SGA, defined as BW < 10th, 5th, or 3rd percentile, or  $\geq 1$ or 2 SDs below the mean; study conducted in Northern, Western, or Central Europe, USA, Canada, Australia, or New Zealand; and reporting of sufficient data for 2 × 2 table construction. Our search returned 722 citations, of which 8 studies fulfilled the inclusion criteria. Quality of the included studies was assessed in duplicate using the Quality Assessment of Diagnostic Accuracy Studies tool (83, 84).

#### Study characteristics

The included studies were published between 1982 and 1991. Five studies (57, 65, 67-69) were conducted in Great Britain, and three (58, 62, 66) were conducted in Scandinavia. Participant numbers ranged from 381 to 3038, with a mean of 1252. The total population was 10,018. Most studies used locally derived SF charts. Different cut-off criteria were used to identify abnormal results, including one value < 10th percentile (57, 65), one value < 10th percentile or static or falling values (69), two consecutive or three isolated values < 10th percentile (57, 68), one value  $\geq$  3 cm below the mean (62), one value  $\geq$  3 cm below the mean or static or falling values (57, 58, 62, 67), and one value  $\geq$  2 SDs below the mean (66). Definitions of SGA included BW < 10th percentile (57, 62, 65-69), BW < 5th percentile (57), and BW  $\geq$  2 SDs below the mean (58), according to local standards.

#### 3.1.2 Papers II and III

Papers II and III present the results of population registry–based cohort studies. All women who delivered singleton neonates at Sahlgrenska University (SU) Hospital, Sweden, between January 2005 and September 2010 and attended ANC between gestational weeks 24 and 42 (n= 44,056) were included after identification using the hospital's obstetric database. The SU Hospital has three delivery wards in two locations in Gothenburg and is a tertiary referral unit for a wider region. Data from pregnancies with no registered dating scan or measurement of SF height (n = 2038) were excluded. Thus, the study sample consisted of data from a total of 42,018 pregnancies with complete ultrasound-confirmed information about gestational age and a total of 282,713 measurements of SF height, with an average of 6.8 measurements per pregnancy.

To investigate the effect of the use of SF measurements from a selected cohort versus those from the entire cohort, we created two subgroups of selected pregnancies by excluding (i) pregnancies with severe congenital anomalies/chromosomal abnormalities or resulting in stillbirth (n = 41,205); and (ii) pregnancies in women with registered severe chronic diseases, pregnancy complications, BMI > 30 kg/m<sup>2</sup>, operative deliveries, preterm deliveries (<259 gestational days), post-term deliveries (>294 gestational days), delivery of neonates with low Apgar scores (<7 at 5 min), SGA status (BW < 10th percentile), and large-for-gestational-age status (BW > 90th percentile; n = 23,832).

#### Maternal and neonatal characteristics

The women in the study sample (n = 42,018) had a mean age at childbirth of 30 (SD = 5) years, and 33.7% were giving birth to their first child. Mean maternal pre-pregnancy weight and height were 67 (SD = 12.6) kg and 166 (SD = 6.7) cm, respectively. Mean BMI was 22.2 (SD = 4.5) kg/m<sup>2</sup>. About 6% of women were cigarette smokers at the time of registry.

The neonates in the study sample (n = 42,018) had a mean BW and gestational age of 3498 (SD = 545.8) g and 283 (SD = 12) days, respectively. Among the neonates included in this cohort (51.3% were male), 88.2% were born at term (259–294 days), 7.4% were born postterm (>294 days), and 4.3% were born preterm (<259 days). The sample included 0.3% antepartum stillbirths.

We compared the excluded 2038 pregnancies without registered dating scan or measurement of SF height to the final study population on key variables. We found no relevant differences in maternal age, pre-pregnancy BMI or nulliparity. The excluded pregnancies had a slightly lower mean BW, 3332 grams versus 3498 grams in the study population.

#### **Description of variables**

**SF measurement.** The SF technique is standardized in Sweden according to Westin's (48) method. SF height is measured from the upper border of the pubic symphysis along the longitudinal axis of the uterus to the highest point of the uterus, within or outside the midline. The woman should be in a supine position with her legs extended and bladder empty.

**Gestational age determination.** All pregnancies were ultrasound dated using the BPD in the second trimester, according to the Swedish ANC program (85). Gestational ages (in weeks and days) were derived from these values using a pregnancy length of 283 days (13, 30, 86).

**SGA.** SGA was defined as BW < 10th percentile for gestational age and sex relative to the population standard (30, 31).

#### 3.2 Statistical methods

#### 3.2.1 Paper I

Statistical analyses were performed using the MIDAS module (87) for STATA (version 12; StataCorp, College Station, TX, USA) and Review Manager 5.3 software. Data on sensitivity, specificity, and TP, FP, TN, and FN results were extracted directly from the source papers or, if necessary, calculated from the data provided. PLRs, NLRs, DORs, and 95% confidence intervals (CIs) were calculated.

Hierarchical bivariate random-effects meta-analysis (88) was used. This model takes into account potential threshold effects and the correlation between sensitivity and specificity. If there appeared to be no or minimal threshold differences between the studies clinically or on the ROC plot, then use of a summary statistic in the form of sensitivity and specificity was planned. If clinical and visual appearance of a threshold effect was present, then the use of the summary ROC curve as the most appropriate summary measure. As our data was a mixture of different positivity thresholds, which also was reflected in the ROC plot, we decided to

estimate a summary ROC curve. The summary estimation of sensitivity and specificity by pooling studies with mixed thresholds would produce an estimate that related to a notional, unspecified average of the thresholds in the included studies, which is clinically unhelpful and is not recommended (89).

The heterogeneity of the results between studies was assessed statistically using the quantity *I2*, which describes the percentage of total variation across studies that is attributable to heterogeneity rather than chance (90). Subgroup analyses was intended for investigation of heterogeneity, however, there was an insufficient number of studies (and thus low statistical power).

#### 3.2.2 Paper II

Statistical analyses were performed using the R Project Statistical Computing environment (version 2.14.1; http://www.r-project.org) and STATA Statistical Software version 11 (StataCorp). Gestational age was measured in days, and SF height was measured in centimeters.

Nonlinear regression of SF height on day of pregnancy was used to construct a reference chart for the median and other percentiles of SF height. SF curves were estimated by the generalized additive models for location scale and shape (GAMLSS) method (91) using the GAMLSS R library. For all gestational ages, the SF distribution was very close to normal. However, at the extreme ends of the distribution, such as the 1st and 99th percentiles, small deviations from normal could cause noticeable biases in estimates. To account for minor deviations from normal distribution, the GAMLSS model was fitted using the Johnson SU distribution, which extends the normal distribution by allowing the estimation of not only means and SDs, but also skewness and kurtosis, the latter being a measure of the "pointedness" of the distribution.

To assess the effects of fetal and maternal characteristics on SF height, we computed day-byday residuals of SF height by subtracting the estimated median SF value from the observed values. Gestational age was then categorized into three groups (24–30, 31–36, and 37–42 completed weeks of gestation), and the effects were assessed using the residuals as dependent variables in three separate multiple regression analyses, one for each gestational age group. As independent variables, the fetal characteristic was sex (female, male); the maternal characteristics were parity group (0, 1, 2+), age group (14–20, 20–35, 35–52 years), smoking at first antenatal visit (yes, no), height (<162, 162–165, 166–168, 169–172, >172 cm), and pre-pregnancy weight group (<57.4, 57.4–62.5, 62.6–67.7, 67.8–75.2, >75.2 kg).

#### 3.2.3 Paper III

In this study, SGA was defined as pdBW  $\leq -13.8\%$  (30), corresponding to BW < 10th percentile for gestational age. We computed pdBW as (BW – BW0) / BW0 × 100%, where BW0 is the population median BW for a given gestational age at birth. Similarly, we computed the (absolute) symphysis-fundus deviation (dSF) as SF – SF0, where SF0 is the population median SF for the given day of measurement (53). A linear regression for each week of pregnancy with pdBW serving as the dependent variable and dSF and gestational age serving as independent parameters was performed. The ability of dSF to predict pdBW in each week was determined using the  $R^2$  value. To assess the added predictive value of maternal factors, regression was also performed using dSF in combination with fetal sex (male, female) and the following maternal characteristics: parity (0, 1, 2+), age (14–20, 21–35, 36–52 years), smoking at first antenatal visit (yes, no), height (<162, 162–165, 166–168, 169–172, >172 cm), and pre-pregnancy weight (<57.4, 57.4–62.5, 62.6–67.7, 67.8–75.2, >75.2 kg).

To predict the risk of SGA status at birth, we similarly performed a binary regression analysis for each pregnancy week, with dSF serving as a linear predictor and SGA as a dichotomous outcome. The results allowed us to define dSF values predicting specific RRs (e.g., 1, 2) for SGA. To assess the predictive value of the logistic regression models, we computed an ROC curve and the area under the curve (AUC) for each week, and overall. The overall ROC curve was computed by assuming a positive test result for SGA, when at least one measurement exceeded the specified RR cut-off value during a pregnancy.

To assess whether longitudinal SF measurements could increase the predictive value of the model, the linear regression model described above was used. This model allowed examination of the relationship between dSF at any given gestational age and BW outcome. We computed the rate of change in SF measurements prior to that age, and included it in the regression model. The rate of SF change in the final model was computed using the difference between the *Z* score of the SF measure at the given gestational age and that of the previous measurement, divided by the number of days between measurements. Several other options

for computing the rate of change, such as the slope along all previous SF measurements for that pregnancy, were investigated.

## 3.3 Ethical approvals

No ethical approval was required for the study reported on in Paper I, as it was a systematic review based on published data. The studies reported on in Papers II and III received ethical approval from the local Institutional Board and the ethics committee of Gothenburg (Regionala etikutprövningsnämden i Göteborg, ref. no. 305-10).

# 4 Synopsis of study results

### 4.1 Paper I

# Symphysis-fundus height measurements to predict small-for-gestational-age status at birth: a systematic review

In this study, we conducted a systematic review to assess the accuracy of SF height for the prediction of SGA status in unselected and low-risk pregnancies. Eight studies (57, 58, 62, 65-69) were included in the final dataset and seven (57, 62, 65-69) were included in summary analyses.

The sensitivity of SF height for prediction of SGA (BW < 10th percentile) ranged from 0.27 to 0.76 (range of 0.49) and specificity ranged from 0.79 to 0.92 (range of 0.13). All studies reported DORs > 1 and CIs that did not include 1, implying that the positive association of SF height with SGA is not accounted for by chance alone. One study (66) had a narrower CI than the other studies, which may be attributed to its large size. PLRs were >1 in all studies, indicating that an abnormal SF height value is associated with SGA status at birth. However, all PLRs were <10, the threshold generally accepted for a useful test. The same seven studies obtained NLRs < 1, indicating that a normal SF height value is correctly associated with the absence of SGA. However, no study met the accepted level of <0.1 in this group of women. The sROC curve fell to the left of the diagonal, signifying that the SF measurement test has value. Accuracy measured by the AUC was 0.84 (95% CI = 0.81–0.87). The  $I^2$  value was high (98 %).

One study (57) assessed the accuracy of SF height for the prediction of SGA defined as BW < 5th percentile. This study used several cut-off points, with stricter criteria yielding lower sensitivity and higher specificity values. NLRs and PLRs did not meet the accepted criteria for classification of SF height measurement as a useful test. One study (58) assessed the outcome of SGA defined as  $BW \ge 2$  SDs below the mean. For a less strict SF cut-off point (one value > 2 cm below mean or falling or static values), the authors reported low sensitivity (59%) and high specificity (97%). The PLR exceeded 10, but the NLR did not meet the required criterion of <0.1.

#### 4.2 Paper II

#### A new population-based reference curve for symphysis-fundus height

In paper II, we presented a new reference chart for SF height based on SF measurements from a large population-based maternity cohort, and determined the effects of maternal and fetal factors on SF height. The new SF reference curve was compared with other relevant Scandinavian SF charts.

The new reference curve for SF height showed linear growth until term. Maternal weight and height were found to influence SF height moderately, whereas parity, maternal age, smoking, and fetal sex had no clinically relevant effect. We found that the median value of our new reference curve was considerably greater than those of the older Westin curve (48), and the modified curve used previously in Norway (52) (Figure 8). Compared with the Steingrimsdottir et al. (49) curve, the median values were slightly higher (0.5 cm) at gestational ages < 34 weeks. From pregnancy week 35, the differences increased to approximately 1 cm.

Figure 8. Comparison of our new curve with the modified curve previously used in Norway



To investigate the effect of the use of SF measurements from selected cohorts versus those from the entire cohort, SF curves were drawn for two sub-populations. The exclusion of stillbirths and fetuses with severe congenital abnormalities caused no change in the SF percentiles. Exclusions based on maternal and fetal pathologies shifted the 90th and 95th percentiles downward, but all remaining percentiles remained nearly unchanged from those for the total population (Figure 9).

Figure 9. Comparison of SF percentile curves from the entire population (black lines) with SF percentile curves from the "hyper-normal" population (green lines)



#### 4.3 Paper III

Prediction of small-for-gestational-age status at birth by symphysis-fundus height: a registry-based population cohort study

In this study, we presented a prediction model for SGA during pregnancy by predicting fetal size at birth using SF measurements. An additional objective was to develop risk curves that delineate SF values corresponding to specific degrees of elevation in SGA risk relative to pregnancies with normal patterns of SF height. We also compared the SGA risk curves with

the SF percentile curves presented in paper II (53) to determine similarities and differences across the gestational age range.

We showed that the overall level of variance in BW explained by dSF increased gradually with gestational age, from 3% in pregnancy week 24 to 20% in week 42. The addition of maternal and fetal factors to the model explained an additional 10 percentage points of variance in BW. The explained variation in BW was not improved by including the SF *Z* score (SF change) as the predictor, indicating that only the most recent SF measurement is useful in the detection of SGA.

ROC curve analysis confirmed that SF measurements obtained in late pregnancy were stronger predictors of SGA than were those obtained in early pregnancy. The overall sensitivity and specificity for SGA detection using an RR cut-off limit of  $\geq$ 2-fold were 50% and 80%, respectively. Tuning the cutoff correspondingly changed the sensitivity and specificity. A stricter cutoff resulted in a higher specificity, but lower sensitivity. Vice versa, a less stringent cutoff gave higher sensitivity but lower specificity.





Comparison of the percentiles of the SF population distribution extracted from paper II (53) showed that the 50th percentile corresponded closely to 1.0 RR line, indicating that women with SF values close to the population median give birth to children with no increased SGA risk. At the 2.5th percentile of the SF population, the RR of SGA status increased gradually from approximately 2 at 24 weeks of gestation to 3.5 at 42 weeks of gestation. Figure 10 shows the relationship between SF percentiles and the corresponding RR of SGA.

# **5** Discussion

In this chapter, an overview of methodological strengths and weaknesses is presented, and whether limitations may have influenced results is assessed. The main results are discussed and compared with results from other studies.

## 5.1 Methodological considerations

#### Paper I

This study was carried out according to a predefined protocol that set out the scope of the review and details of the methodology used throughout the review process. The review was conducted according to the Meta-Analysis of Observational Studies in Epidemiology and Preferred Reporting Items for Systematic Reviews and Meta-Analyses statements (92-94), and registered in the PROSPERO international prospective register of systematic reviews (no. CRD42014008928).

#### Limitations

A large number of abstracts were reviewed. With regard to study design, suitable publications may have been omitted due to the sole use of electronic searches, reviewer error, or the use of limited search terms. Further detail could be added to the searches, including the use of limited text terms. Publication bias may occur, consensus is lacking on its importance or how to assess its impact on systematic reviews of diagnostic test accuracy (81). As the current use of SF height in clinical practice is debated (95), whether publication bias would result in the exclusion of papers showing low or high test accuracy is not clear.

The lack of large cohort studies conducted in routine ANC settings that were suitable for inclusion in our analysis was the main limitation of this review. Given the small number of studies included in the review, statistical subgroup analysis and covariate hierarchical modeling for the investigation of heterogeneity were not performed, resulting in an incomplete assessment of controlling bias.

#### Generalizability

The populations of studies included in the review were limited to unselected and low-risk pregnancies. Thus, the results cannot be applied to high-risk populations. In such settings, the prevalence of SGA would be higher, and SF measurement would likely show greater accuracy (96). In addition, the studies included in the review were conducted in countries with health care systems comparable to those in Scandinavia. Thus, the results cannot be applied to different and less well-resourced national health systems. Overall, the results can be applied to unselected and low-risk pregnancies in women attending health care systems comparable to those in Scandinavia.

#### Papers II and III

The studies reported on in papers II and III utilized a population-based registry containing approximately 300,000 longitudinal measurements. The large amount of population data minimizes random errors. The data were collected prospectively, eliminating recall bias. The material mirrors clinical practice, with gestational age computed by ultrasound and measurements performed by numerous health care workers.

#### Limitations

These studies have several limitations. First, our findings may have been influenced by knowledge of fetal and maternal complications affecting fetal growth among those who measured SF height, which would bias the SF values. We had no information regarding the timing of obstetric diagnosis, and hence could not determine whether the examiners had any knowledge of pathology prior to obtaining the measurements. Such bias could have resulted in under- or overestimation of the true SF values.

Furthermore, the number of SF measurements differed among the women in the sample, with an average of 6.8 measurements per pregnancy. SF measurements may have been recorded more frequently when SGA status was suspected. However, we found that the number of SF measurements was greater (0.2) among women with low-risk pregnancies than in the rest of the population; it thus seems unlikely that this potential source of bias influenced the results.

Finally, inter-observer differences could arguably bias the results. Several health care workers contributed to the SF measurements in this dataset. However, this weakness did not likely

systematically influenced our results because the importance of bias in the general interpretation of the effects of observer variation declines with the increase in the number of observers involved in a study, as the biases cancel out.

#### Generalizability

The data used for these studies came from a community-wide, compulsory, population-based reporting system and represent a real-life sample from ordinary ANC. Maternal and neonatal characteristics were largely in agreement with those of births occurring during the study period in Sweden and reported to the Swedish Medical Birth Registry (97), except for the parity distribution. The nulliparity was lower compared with the Swedish population during the study period (33.7 % vs. 45 %). In paper II we assessed the effect of parity on SF measurements and found no clinical effect of parity on SF height, suggesting that moderate differences in parity distributions by themselves should have little impact on the result. We cannot rule out, however, that there may be other population differences.

# 5.2 Interpretation of results and comparison with previous findings

# Systematic review: SF height can be the first parameter raising suspicion of SGA (BW < 10th percentile) in low-risk and unselected pregnancies

SF measurement seems to have some significance for the prediction of SGA defined as BW < 10th percentile. All studies included in this review reported DORs > 1. The sROC curve fell to the left of the diagonal (AUC=0.84, 95% CI= 0.81-0.87), indicating that the SF test has value. Adequate levels of sensitivity appear to be achieved at the expense of lower specificity, with large number of FP results. The results can be applied to low-risk and unselected pregnancies in routine ANC settings. The diagnostic accuracy of SF measurement in other populations of pregnant women has recently been reviewed. In a meta-analysis, including 46 studies conducted mainly in developing countries, Goto (98) reported pooled sensitivity and specificity of 58% and 87%, respectively. The AUC was 0.82 (95% CI, 0.78–0.85). The review of Goto included studies involving a wide range of ethnic groups, clinical settings, and disease spectrums. In our review we applied more strict inclusion criteria, focusing mainly on the analysis of data from a more homogenous and relevant population. The included studies could not be case-control studies nor could they be studies that were conducted in high risk

settings. Case-control studies have been demonstrated to overestimate diagnostic accuracy compared with cohort studies conducted using an appropriate spectrum (96, 99). Similarly, studies conducted in high risk settings provide overly optimistic estimates compared with the screening in in a low risk population (100). Despite the differences in study inclusion criteria between the two reviews, sROC curves and AUC estimates are quite similar.

The observed similarity in sROC curves may seem surprising in light of the many differences between study populations, choice of included studies, and methodology used in the selected studies. There may still be good reasons for the similarities. It is reasonable to assume that the basic underlying correlation between BW and the SF measures at a given gestational age will be relatively constant across populations and unaffected by population differences such as mean BW, maternal BMI etc. This again leads to general similarities between tests based on dichotomizing SF and BW. Specific population-specific cutoff values will clearly produce different sensitivities and specificities, but when summarizing the test properties in a ROC curve, many of the dissimilarities will disappear. The sROC curve is the endpoint – the "average" ROC curve - of all included studies. The individual study differences will thus again average out, leading to an end result that is likely to be relatively unaffected by the differences in study populations.

#### New population-based reference curve: no decrease or flattening of the curve toward term

We showed that the pattern of the new reference standard differed from that of currently used Scandinavian standards of older origin. Our results refute the common belief that fundal height flattens toward term. SF height continued to increase throughout pregnancy. The median of our new curve fell about 1 cm above that of Westin's (48) curve developed in 1977, for the whole gestational range. It was 1 cm above the median of the modified curve used previously in Norway (52) for pregnancy weeks 24–36, and gradually approached 2 cm above that median after 37 gestational weeks . The median of our curve fell 0.5 cm above that of Steingrimsdottir et al.'s (49) curve, developed in 1995, for pregnancy weeks 24–34, and then gradually approached 1 cm above that median after 34 gestational weeks.

The lower SF values in some previous studies may be explained by differences in the population samples, analytical strategy, and study design. With respect to the observed difference at the end of pregnancy in the Steingrimsdottir et al. (49) curve, this curve fits cubic polynomials for the mean and SD of the SF measurement, which may explain the

notably lower SF values in advanced pregnancy weeks. The disadvantage of using standard polynomial regression is that fits at the endpoints may be less satisfactory when an optimal fit is obtained in the central region. In particular, the use of polynomials often results in sharp up- or downturns at the ends. This issue may have caused the bending/drop at the ends of some previously published curves. However, without access to the data, determination of how much of the downturn can be explained in this manner is difficult.

#### New population-based reference: moderate effects of maternal covariates

We investigated the effects of maternal height, pre-pregnancy weight, age, parity, smoking habit at first antenatal visit, and fetal sex on SF measurement. The adjusted results were presented in Table 3 in Paper II. Only maternal weight and height had clinically relevant effects, with differences in late pregnancy of about 2 cm between the lightest and heaviest women, and about 0.7 cm between the shortest and tallest women. Figure 11 shows in detail how the covariate effects change during pregnancy. Note that values in the figures are unadjusted and thus differ somewhat from the numbers in Table 3 in Paper II. The most noticeable effect of adjusting the covariates for one another is seen in maternal weight and height. In the unadjusted analyses, there is a strong effect of maternal weight, whereas maternal height is almost without effect. After adjustment, there is an effect also of maternal height; higher women have shorter SF measurements, with a span of approximately 0.7 cm around term. The lack of effect of height before adjustment is likely due to the correlation between weight and height; smaller women are both shorter and lighter, and the effects cancel.

In the literature, basic pregnancy characteristics, alone and in combination, have been reported to influence SF height measurement. Mongelli and Gardosi (101) reported on the influences of maternal booking weight, parity, ethnic group, and fetal sex in 325 pregnancies. Booking weight was the most influential maternal characteristic, with every 10 kg above average increasing expected SF height by 0.7 cm. Multiparas and male fetal sex were each associated with SF heights of 1 cm above average. Ethnic group did not appear to affect SF height significantly.



Figure 11. Effects of maternal and fetal covariates on SF measurement

Steingrimsdottir et al. (49) examined maternal characteristics using univariate statistics. They also found that maternal booking weight was a significant determinant of SF height, with a difference of approximately 2 cm between the lightest and heaviest women. The only other factors included were parity and smoking, and these were not significant predictors.

Challis et al. (102) reported on maternal weight at booking and parity in a Mozambique population (n = 770). Average SF values were 0.5–1 cm higher for parous than for nulliparous women. Values were approximately 1 cm lower and 1 cm higher among women with BMIs < 19 kg/m<sup>2</sup> and those with BMIs > 27 kg/m<sup>2</sup>, respectively, compared with women with normal BMIs. The authors also compared SF curves from various studies conducted in different populations and showed that American (103) and Indian curves (104) were highest and lowest, respectively. Unfortunately, we were unable to study the effects of ethnicity.

# SGA prediction: low, but consistent and increasing, predictive ability for the detection of SGA neonates

We showed that the quality of BW and SGA risk predictions is low in early pregnancy, but that quality and predictive value increase closer to term. The addition of maternal and fetal factors to the model explained an additional 10 percentage points of variance in BW, but they did not confound the SF-BW relationship to a substantial degree. Thus, accounting for them is not essential when assessing the predictive strength of SF measurements.

The predictive SF model resulted in a much smaller degree of explained variation than obtained with the use of standard ultrasound fetal biometric algorithms; such model estimates, including measurement of BPD, AC, and FL, were previously reported to explain 18% of the variation in BW in early pregnancy, and close to 65% of explained variation at the end of pregnancy (31). Even without AC, greater precision was reported, with these factors explaining about 35% of variation in BW at the end of pregnancy.

Despite the low degree of explained variation in BW compared with ultrasound, SF measurement may have a place in the antenatal detection of SGA. SF measurement is used mainly as a screening test, not as a diagnostic test. The main purpose of screening tests is to identify a population at risk that requires follow up with diagnostic testing. Screening tests are generally accepted to be less accurate than diagnostic tests. In addition, SF measurement does not exist in total absence of other ANC; it is only a part of antenatal screening for the

detection of fetuses at risk. Furthermore, it is low cost, easy to use, and requires very little training, with no direct risk to the fetus. The main risk is the possibility of FP and FN results, which may produce excess maternal anxiety.

#### SGA prediction: no effect of longitudinal measurements on BW prediction

We found no effect of longitudinal SF measurement on BW prediction. Various attempts to analyze the current and preceding measurements, the entire trend, the slope of the SF curve, the declining growth rate, and the accelerated growth rate (Figure 12) did not improve the explained variation in BW. Such SF changes are commonly regarded as signs of increased risk of aberration in fetal growth. That SF measurement obtained from the same pregnancy prior to a given measurement time-point provided no useful additional information for the purpose of BW prediction is probably due to the low proportion of BW variance explained by SF measurements.





Theoretically, additional information on a falling or static trend may be clinically useful; in practice, however, correlation with BW may not be sufficiently strong to be of independent

value. The current clinical practice to use falling or static SF values as criteria of abnormality is thus not necessary. The use of only the most recent SF measurement for evaluation may facilitate the provision of clear guidelines, thereby improving the clinical usefulness of this method. It should, however, be kept in mind that our results have direct relevance only to BW as an outcome. They do not rule out the possibility that static or falling SF values are indicative of perinatal morbidity and mortality in general.

# 6 Concluding remarks and future perspectives

The goals of antenatal fetal surveillance in cases of FGR are to avoid imminent fetal jeopardy, and to identify pregnancies at increased risk of adverse pregnancy outcomes. ANC should provide appropriate care in order to reduce this risk while avoiding unnecessary interventions. The results of this study can be summarized as follows:

- SF height has high FN and FP rates for SGA, and should preferably be used in combination with other clinical findings, information about medical conditions, and previous obstetric history (Paper I).
- The new reference curve indicates that currently used screening tools for SGA may not represent the current Scandinavian population (Paper II).
- The clear increase in predictive value from week 24 until term shows that SF deviations should be taken more seriously in late pregnancy, and underscores the difference between BW prediction curves and traditional SF percentile curves (Paper III).
- Individualized prediction using the current SF measurement and additional covariates for maternal and fetal characteristics increases the predictive value of the most recent SF measurement (Paper III).
- Patterns of static or decreasing SF values do not add to the predictive value of the most recent SF measurement (Paper III).

As antenatal detection of SGA and FGR may improve neonatal outcome, and many affected pregnancies are undetected, continuous efforts to construct better screening and diagnostic methods are requisite. The presented curves should be further examined and tested in clinical settings, and the effects of their use must be evaluated systematically. Some key questions are: how will health care providers relate to the new reference curves? How many pregnant women will be referred to specialists for further investigation of fetal growth and well-being based on the chosen cut-off value? Will the introduction of the new curves actually improve pregnancy outcomes for mothers and children?

The Norwegian Directorate of Health recently revised the personal health record (Helsekort for gravide), and included the percentile curves for SF height from paper II (53). In Paper III, we demonstrate that curves delineating relative risk of SGA are an alternative to the percentile curves; relative risk of SGA is a concept that relates to clinical practice and may be more directly interpretable than population percentiles. To compare the overall predictive value of the distribution-based references (53) (paper II) with that of the RR curves (76) (paper III), we computed ROC curves for each method, defining a positive test as at least one SF measurement below the appropriate cutoff (either population percentile or RR value). The results showed almost no difference in predictive quality (Figure 13). This is somewhat unexpected since the RR-based curves are made explicitly to focus on SGA prediction. However, as is seen from Figure 2 in Paper III, the RR curves are relatively parallel with the SF percentile curves from around week 33 and up. Thus, the largest difference between the RR and percentile curves is in the early part, where predictive value is low, regardless of model. The choice of RR charts versus percentile charts is thus not very important for the overall test characteristics. Rather, it is more a matter of choice of presentation and clinical relevance.

Related to the introduction of the SF curves in the new Helsekort for gravide, we also assessed the referral rates for pregnancies with at least one measurement below specific percentile limits (Figure 14). The cut-off percentile value should be chosen not only to improve the sensitivity of the method, but also with consideration of the practical possibilities for further supervision of fetal growth and well-being. Our analyses showed that the previously used SF chart results in a referral rate of 7% for pregnancies with one measurement below normal limits. With the new curve from paper II (53), this referral rate corresponds roughly to using the 2.5th percentile cutoff limit.

In our studies we have seen that the predictive value of SF measurements is very low in the early weeks; later in pregnancy there is a moderate but consistent correlation with BW and SGA risk as endpoints. The utility of the SF measurements seem to be fundamentally restricted by the inherent uncertainty and measurement errors. The numerous charts and detection methods suggested in the literature include the gravidogram approach, methods for direct detection of FGR (for instance, the relative growth method), and our endpoint-guided SGA prediction model.

However, they all reveal the fundamental limitation of SF measurements, in that they all remain at modest levels of accuracy. This underscores the need to combine SF measurements with other clinical criteria that together may reach useful levels of predictive quality. Due to its ease of use and applicability across populations of varying levels of health care standards, it will likely remain a helpful supplement to standard pregnancy care.

Figure 13. Performance of distribution-based references (gray line) and RR curves (solid line) for prediction of SGA ( $BW \le 10$ th percentile) expressed by ROC curves



Specificity





# 7 References

1. Frøen 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.

2. Goldenberg RL, Hoffman HJ, Cliver SP. Neurodevelopmental outcome of small-forgestational-age infants. Eur J Clin Nutr. 1998 Jan;52 Suppl 1:S54-8.

3. 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.

4. Jacobsson B, Ahlin K, Francis A, Hagberg G, Hagberg H, Gardosi J. Cerebral palsy and restricted growth status at birth: population-based case-control study. BJOG. 2008 Sep;115(10):1250-5.

5. 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.

6. Pay AS, Wiik J, Backe B, Jacobsson B, Strandell A, Klovning A. Symphysis-fundus height measurement to predict small-for-gestational-age status at birth: a systematic review. BMC Pregnancy Childbirth. 2015 Dec;15(1):461.

7. Winick M. Cellular growth in intrauterine malnutrition. Pediatr Clin North Am. 1970 Feb;17(1):69-78.

8. Gluckman PD, Pinal CS. Regulation of fetal growth by the somatotrophic axis. J Nutr. 2003 May;133(5 Suppl 2):1741s-6s.

9. Baird DD, Wilcox AJ, Weinberg CR, Kamel F, McConnaughey DR, Musey PI, et al. Preimplantation hormonal differences between the conception and non-conception menstrual cycles of 32 normal women. Hum Reprod. 1997 Dec;12(12):2607-13.

10. Gardosi J, Geirsson RT. Routine ultrasound is the method of choice for dating pregnancy. Br J Obstet Gynaecol. 1998 Sep;105(9):933-6.

11. Kieler H, Axelsson O, Nilsson S, Waldenstrom U. The length of human pregnancy as calculated by ultrasonographic measurement of the fetal biparietal diameter. Ultrasound Obstet Gynecol. 1995 Nov;6(5):353-7.

12. Mongelli M, Wilcox M, Gardosi J. Estimating the date of confinement: ultrasonographic biometry versus certain menstrual dates. Am J Obstet Gynecol. 1996 Jan;174(1 Pt 1):278-81.

13. Tunon K, Eik-Nes SH, Grøttum P. A comparison between ultrasound and a reliable last menstrual period as predictors of the day of delivery in 15,000 examinations. Ultrasound Obstet Gynecol. 1996 Sep;8(3):178-85.

14. Källen K. Mid-trimester ultrasound prediction of gestational age: advantages and systematic errors. Ultrasound Obstet Gynecol. 2002 Dec;20(6):558-63.

15. Gardosi J, Vanner T, Francis A. Gestational age and induction of labour for prolonged pregnancy. Br J Obstet Gynaecol. 1997 Jul;104(7):792-7.

16. Geirsson RT. Ultrasound instead of last menstrual period as the basis of gestational age assignment. Ultrasound Obstet Gynecol. 1991 May 1;1(3):212-9.

17. Ebbing C, Acharya G, Sitras V, Helbig A, Husby H. Intrauterin veksthemming. 2014. In: Veileder i fødselshjelp 2014 [Internet]. Oslo: Norsk gynekologisk forening. Available from: http://legeforeningen.no/Fagmed/Norsk-gynekologisk-forening/Veiledere/Veileder-ifodselshjelp-2014/Intrauterin-veksthemming/. 18. Drooger JC, Troe JW, Borsboom GJ, Hofman A, Mackenbach JP, Moll HA, et al. Ethnic differences in prenatal growth and the association with maternal and fetal characteristics. Ultrasound Obstet Gynecol. 2005 Aug;26(2):115-22.

19. Pedersen NG, Wojdemann KR, Scheike T, Tabor A. Fetal growth between the first and second trimesters and the risk of adverse pregnancy outcome. Ultrasound Obstet Gynecol. 2008 Aug;32(2):147-54.

20. Bennett KA, Crane JM, O'Shea P, Lacelle J, Hutchens D, Copel JA. First trimester ultrasound screening is effective in reducing postterm labor induction rates: a randomized controlled trial. Am J Obstet Gynecol. 2004 Apr;190(4):1077-81.

21. Caughey AB, Nicholson JM, Washington AE. First- vs second-trimester ultrasound: the effect on pregnancy dating and perinatal outcomes. Am J Obstet Gynecol. 2008 Jun;198(6):703.e1-5; discussion .e5-6.

22. Saltvedt S, Almstrom H, Kublickas M, Reilly M, Valentin L, Grunewald C. Ultrasound dating at 12-14 or 15-20 weeks of gestation? A prospective cross-validation of established dating formulae in a population of in-vitro fertilized pregnancies randomized to early or late dating scan. Ultrasound Obstet Gynecol. 2004 Jul;24(1):42-50.

23. Gjessing HK, Grøttum P, Økland I, Eik-Nes SH. Comment and reply on: prediction of the date of delivery based on first trimester ultrasound measurements: an independent method from estimated date of conception. J Matern Fetal Neonatal Med. 2010 Aug;23(8):944-7.

24. Taipale P, Hiilesmaa V. Predicting delivery date by ultrasound and last menstrual period in early gestation. Obstet Gynecol. 2001 Feb;97(2):189-94.

25. Wade AM, Ades AE. Incorporating correlations between measurements into the estimation of age-related reference ranges. Stat Med. 1998 Sep 15;17(17):1989-2002.

26. Royston P, Altman DG. Design and analysis of longitudinal studies of fetal size. Ultrasound Obstet Gynecol. 1995 Nov;6(5):307-12.

27. Combs CA, Jaekle RK, Rosenn B, Pope M, Miodovnik M, Siddiqi TA. Sonographic estimation of fetal weight based on a model of fetal volume. Obstet Gynecol. 1993 Sep;82(3):365-70.

28. Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements--a prospective study. Am J Obstet Gynecol. 1985 Feb 1;151(3):333-7.

 Marsal K, Persson PH, Larsen T, Lilja H, Selbing A, Sultan B. Intrauterine growth curves based on ultrasonically estimated foetal weights. Acta Paediatr. 1996 Jul;85(7):843-8.
 Gjessing HK, Grøttum P, Eik-Nes SH. A direct method for ultrasound prediction of day of delivery: a new, population-based approach. Ultrasound Obstet Gynecol. 2007 Jul;30(1):19-27.

31. Gjessing HK, Grøttum P, Økland I, Eik-Nes SH. Fetal size monitoring and birth weight prediction using ultrasound: A new, population-based approach [Submitted]. 2016.

32. Pilliod RA, Cheng YW, Snowden JM, Doss AE, Caughey AB. The risk of intrauterine fetal death in the small-for-gestational-age fetus. Am J Obstet Gynecol. 2012 Oct;207(4):318.e1-6.

33. Skjærven R, Gjessing HK, Bakketeig LS. New standards for birth weight by gestational age using family data. Am J Obstet Gynecol. 2000 Sep;183(3):689-96.

34. Bakketeig LS. Current growth standards, definitions, diagnosis and classification of fetal growth retardation. Eur J Clin Nutr. 1998 Jan;52 Suppl 1:S1-4.

35. Gardosi J. Intrauterine growth restriction: new standards for assessing adverse outcome. Best Pract Res Clin Obstet Gynaecol. 2009 Dec;23(6):741-9.

36. Frøen JF, Arnestad M, Frey K, Vege A, Saugstad OD, Stray-Pedersen B. Risk factors for sudden intrauterine unexplained death: epidemiologic characteristics of singleton cases in Oslo, Norway, 1986-1995. Am J Obstet Gynecol. 2001 Mar;184(4):694-702.

37. Mongelli M. Factors influencing fetal growth. Ann Acad Med Singapore. 2003 May;32(3):283-8.

38. Pallotto EK, Kilbride HW. Perinatal outcome and later implications of intrauterine growth restriction. Clin Obstet Gynecol. 2006 Jun;49(2):257-69.

39. Barker DJ, Gluckman PD, Godfrey KM, Harding JE, Owens JA, Robinson JS. Fetal nutrition and cardiovascular disease in adult life. Lancet. 1993 Apr 10;341(8850):938-41.

40. Illanes S, Soothill P. Management of fetal growth restriction. Semin Fetal Neonatal Med. 2004 Oct;9(5):395-401.

41. de Jong CL, Gardosi J, Dekker GA, Colenbrander GJ, van Geijn HP. Application of a customised birthweight standard in the assessment of perinatal outcome in a high risk population. Br J Obstet Gynaecol. 1998 May;105(5):531-5.

42. Sankaran S, Kyle PM. Aetiology and pathogenesis of IUGR. Best Pract Res Clin Obstet Gynaecol. 2009 Dec;23(6):765-77.

43. Maulik D, Figueroa R. Doppler Velocimetry for Fetal Surveillance: Adverse Perinatal Outcome and Fetal Hypoxia. In: Maulik D, editor. Doppler Ultrasound in Obstetrics and Gynecology: Springer Berlin Heidelberg; 2005. p. 363-74.

44. Warrander LK, Heazell AE. Identifying placental dysfunction in women with reduced fetal movements can be used to predict patients at increased risk of pregnancy complications. Med Hypotheses. 2011 Jan;76(1):17-20.

45. Marsal K. Obstetric management of intrauterine growth restriction. Best Pract Res Clin Obstet Gynaecol. 2009 Dec;23(6):857-70.

46. Bernstein IM, Horbar JD, Badger GJ, Ohlsson A, Golan A. Morbidity and mortality among very-low-birth-weight neonates with intrauterine growth restriction. The Vermont Oxford Network. Am J Obstet Gynecol. 2000 Jan;182(1 Pt 1):198-206.

47. A randomised trial of timed delivery for the compromised preterm fetus: short term outcomes and Bayesian interpretation. BJOG. 2003 Jan;110(1):27-32.

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

49. Steingrimsdottir T, Cnattingius S, Lindmark G. Symphysis-fundus height: Construction of a new Swedish reference curve, based on ultrasonically dated pregnancies. Acta Obstet Gynecol Scand. 1995;74(5):346-51.

50. Hakansson A, Aberg A, Nyberg P, Schersten B. A new symphysis-fundus height growth chart based on a well defined female population with ultrasound-dated singleton pregnancies. Acta Obstet Gynecol Scand. 1995;74(9):682-6.

51. Kieler H, Axelsson O, Hellberg D, Nilsson S, Waldenstrom U. Serial measurements of symphysis-fundus height in women with ultrasonically dated pregnancies. J Obstet Gynaecol. 1996;16(4):228-9.

52. Sosial- og helsedirektoratet. Retningslinjer for svangerskapsomsorgen [Guidelines for antenatal care] ( in Norwegian) Oslo: Sosial- og helsedirektoratet; 2005. Available from: https://helsedirektoratet.no/retningslinjer/nasjonal-faglig-retningslinje-for-svangerskapsomsorgen.

53. Pay ASD, Frøen JF, Staff AC, Jacobsson B, Gjessing HK. A new population-based reference curve for symphysis–fundus height. Acta Obstet Gynecol Scand. 2013;92(8):925-33.

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

55. 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.

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

57. Calvert JP, Crean EE, Newcombe RG, Pearson JF. Antenatal screening by measurement of symphysis-fundus height. Br Med J. 1982;285(6345):846-9.

58. 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;18(2-3):191-7.

59. Cnattingius S, Axelsson O, Lindmark G. Symphysis-fundus measurements and intrauterine growth retardation. Acta Obstet Gynecol Scand. 1984;63(4):335-40.

60. Cox G, Walsh P, Stack J, Murphy H. The value of fundal height measurement in prediction of fetal growth retardation. Ir Med J. 1983;76(2):95-6.

61. Gardosi J, Francis A. Controlled trial of fundal height measurement plotted on customised antenatal growth charts. Br J Obstet Gynaecol. 1999;106(4):309-17.

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

63. 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;97(8):675-80.

64. Okonofua FE, Ayangade SO, Chan RCW, O'Brien PMS. A prospective comparison of clinical and ultrasonic methods of predicting normal and abnormal fetal growth. International Journal of Gynecology and Obstetrics. 1986;24(6):447-51.

 Pearce JM, Campbell S. A comparison of symphysis-fundal height and ultrasound as screening tests for light-for-gestational age infants. Br J Obstet Gynaecol. 1987;94(2):100-4.
 Persson B, Stangenberg M, Lunell NO. Prediction of size of infants at birth by measurement of symphysis fundus height. Br J Obstet Gynaecol. 1986;93(3):206-11.

67. Rogers MS, Needham PG. Evaluation of fundal height measurement in antenatal care. Aust N Z J Obstet Gynaecol. 1985;25(2):87-90.

68. Rosenberg K, Grant JM, Tweedie I. Measurement of fundal height as a screening test for fetal growth retardation. Br J Obstet Gynaecol. 1982;89(6):447-50.

69. Stuart JM, Healy TJ, Sutton M, Swingler GR. Symphysis-fundus measurements in screening for small-for-dates infants: a community based study in Gloucestershire. The Journal of the Royal College of General Practitioners. 1989 Feb;39(319):45-8.

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

71. Wright J, Morse K, Kady S, Francis A. Audit of fundal height measurement plotted on customised growth charts. MIDIRS Midwifery Digest. 2006;16(3):341-5.

72. Robert Peter J, Ho JJ, Valliapan J, Sivasangari S. Symphysial fundal height (SFH) measurement in pregnancy for detecting abnormal fetal growth. Cochrane Database of Systematic Reviews. 2012 (7).

73. Bagger PV, Eriksen PS, Secher NJ. The precision and accuracy of symphysis - Fundus distance measurements during pregnancy. Acta Obstet Gynecol Scand. 1985;64(5):371-4.
74. Thompson ML, Fatti LP. Construction of multivariate centile charts for longitudinal

measurements. Stat Med. 1997 Feb 28;16(4):333-45.

75. Bergman E, Axelsson O, Petzold M, Sonesson C, Kieler H. Self-administered symphysis-fundus measurements analyzed with a novel statistical method for detection of intrauterine growth restriction: a clinical evaluation. Acta Obstet Gynecol Scand. 2011 Aug;90(8):890-6.

76. Pay A, Frøen JF, Staff AC, Jacobsson B, Gjessing HK. Prediction of small-forgestational-age status by symphysis-fundus height: a registry-based population cohort study. BJOG. 2015 Dec 8.

77. Boulet SL, Alexander GR, Salihu HM, Kirby RS, Carlo WA. Fetal growth risk curves: defining levels of fetal growth restriction by neonatal death risk. Am J Obstet Gynecol. 2006 Dec;195(6):1571-7.

78. Wilson JMG JG. Principles and practice of screening for disease. WHO Chron. 1968;Public Health Paper(34):22(11).

79. Sovio U, White IR, Dacey A, Pasupathy D, Smith GC. Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study. Lancet. 2015 Nov 21;386(10008):2089-97.

80. Rosman AS, Korsten MA. Application of summary receiver operating characteristics (sROC) analysis to diagnostic clinical testing. Adv Med Sci. 2007;52:76-82.

81. Deeks JJ. Systematic Reviews of Evaluations of Diagnostic and Screening Tests. Systematic Reviews in Health Care: BMJ Publishing Group; 2008. p. 248-82.

82. Bhutta ZA, Yakoob MY, Lawn JE, Rizvi A, Friberg IK, Weissman E, et al. Stillbirths: what difference can we make and at what cost? Lancet. 2011 Apr 30;377(9776):1523-38.

83. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36.

84. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol. 2006;6:9.

85. Svensk förening för obstetrik och gynekologi. Mödrahälsovård, sexuell och reproduktiv hälsa [Antenatal care, sexual and reproductive health] (in Swedish). Stockholm: SFOG; 2008. Available from: https://www.sfog.se/ARG+intressegrupper/arg-forord/Info%2059.pdf.

86. Økland I, Nakling J, Gjessing HK, Grøttum P, Eik-Nes SH. Advantages of the population-based approach to pregnancy dating: results from 23,020 ultrasound examinations. Ultrasound Obstet Gynecol. 2012 May;39(5):563-8.

87. Dwamena BA. Midas: A program for Meta-analytical Integration of Diagnostic Accuracy Studies in Stata. Division of Nuclear Medicine, Department of Radiology, University of Michigan Medical School, Ann Arbor, Michigan. 2007.

88. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001 Oct 15;20(19):2865-84.

89. Macaskill P GC, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. 2010. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1,0 [Internet]. The Cochrane Collaboration. Available from: http://srdta.cochrane.org/.

90. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in metaanalyses. BMJ. 2003 Sep 6;327(7414):557-60.

91. Stasinopoulo D, Rigby R. Generalized additive models for location scale and shape (GAMLSS) R J Stat Softw. 2007;23 (27)(27):1-46.

92. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009 Oct;62(10):e1-34.

93. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Medicine : A Peerreviewed, Independent, Open-access Journal. 2009;3(3):e123-30.

94. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Metaanalysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000 Apr 19;283(15):2008-12.

95. Smith GCS. A bonfire of the tape measures. The Lancet. 2011;377(9774):1307.

96. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med. 2004 Feb 3;140(3):189-202.

97. Socialstyrelsen. Pregnancies, Deliveries and Newborn Infants. The Swedish Medical Birth Register 1973–2012. Assisted Reproduction, treatment 1991–2011. 2013.

98. Goto E. Prediction of low birthweight and small for gestational age from symphysisfundal height mainly in developing countries: a meta-analysis. J Epidemiol Community Health. 2013 Dec 1;67(12):999-1005.

99. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999 Sep 15;282(11):1061-6.

100. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. J Clin Epidemiol. 2009 Jan;62(1):5-12.

101. Mongelli M, Gardosi J. Symphysis-fundus height and pregnancy characteristics in ultrasound- dated pregnancies. Obstet Gynecol. 1999 October;94(4):591-4.

102. Challis K, Osman NB, Nystrom L, Gunnar N, Bergstrm S. Symphysis-fundal height growth chart of an obstetric cohort of 817 Mozambican women with ultrasound-dated singleton pregnancies. Trop Med Int Health. 2002;7(8):678-84.

103. Azziz R, Smith S, Fabro S. The development and use of a standard symphysial-fundal height growth curve in the prediction of small for gestational age neonates. International Journal of Gynecology and Obstetrics. 1988;26(1):81-7.

104. Mathai M, Jairaj P, Muthurathnam S. Screening for light-for-gestational age infants: A comparison of three simple measurements. Br J Obstet Gynaecol. 1987;94(3):217-21.

# 8 Corrections

#### Paper II

Pay ASD, Frøen JF, Staff AC, Jacobsson B, Gjessing HK: A new population-based reference curve for symphysis–fundus height. Acta Obstet Gynecol Scand 2013;92(8):925–933.

In Table 1, the number of "Gestational age < 259 days", "Gestational age 259-294 days", and "Gestational age >294 days" was reported incorrectly. A corrected version of Table 1 is attached.

Maternal characteristics (n=42 018)	Mean (SD)
Maternal age (years)	30 (5.0)
Maternal height (cm)	166 (6.7)
Pre-pregnancy weight (kg)	67 (12.6)
Body mass index (kg/m <sup>2</sup> )	22.2 (4.5)
	n (%)
Nullipara	14 147 (33.7)
Multipara	27 871 (66.3)
Smokers	2 629 (6.3)
Perinatal data (n=42 018)	Mean (SD)
Perinatal data (n=42 018) Birth weight (g)	Mean (SD) 3 498 (545.8)
Perinatal data (n=42 018) Birth weight (g) Gestational age at delivery (days)	Mean (SD) 3 498 (545.8) 283 (12.0)
Perinatal data (n=42 018) Birth weight (g) Gestational age at delivery (days)	Mean (SD) 3 498 (545.8) 283 (12.0) n (%)
Perinatal data (n=42 018) Birth weight (g) Gestational age at delivery (days) Gestational age < 259 days	Mean (SD) 3 498 (545.8) 283 (12.0) n (%) 1 812 (4.3)
Perinatal data (n=42 018) Birth weight (g) Gestational age at delivery (days) Gestational age < 259 days Gestational age 259-294 days	Mean (SD) 3 498 (545.8) 283 (12.0) n (%) 1 812 (4.3) 37 077 (88.2)
Perinatal data (n=42 018) Birth weight (g) Gestational age at delivery (days) Gestational age < 259 days Gestational age 259-294 days Gestational age > 294 days	Mean (SD) 3 498 (545.8) 283 (12.0) n (%) 1 812 (4.3) 37 077 (88.2) 3 129 (7.4)
Perinatal data (n=42 018) Birth weight (g) Gestational age at delivery (days) Gestational age < 259 days Gestational age 259-294 days Gestational age > 294 days Boys	Mean (SD) 3 498 (545.8) 283 (12.0) n (%) 1 812 (4.3) 37 077 (88.2) 3 129 (7.4) 21 547 (51.3)
Perinatal data (n=42 018) Birth weight (g) Gestational age at delivery (days) Gestational age < 259 days Gestational age 259-294 days Gestational age > 294 days Boys Girls	Mean (SD) 3 498 (545.8) 283 (12.0) n (%) 1 812 (4.3) 37 077 (88.2) 3 129 (7.4) 21 547 (51.3) 20 470 (48.7)

Table 1. Characteristics of women and their newborn infants

Appendix Helsekort for gravide

Mor		Føds	selsnr. (11	siffer)			Far/m	nedmor	. Føds	elsnr.	(11 siffe	r)				וך	HI	Ξl	_S	E	K	0	R	ΤI	F(	)F	{	ЗR	A		DE
Navn							Navn										Se lavn	veile fastl	e <mark>der</mark> ege	for	utfyl	ling	av I	helse	ekor Nav	t for /n jor	grav rdmor	ide IS r	3-219	99)	
Adresse					Telefon		Adresse									s	ted								Ste	d					
Postor		Poetetor					Stilling/v	rko				Telefo	0			-	alafo	n							Tel	afon					
POSUII.		USISIEU					Suming/y	i ke	0.00			Teleioi													Ten						
Gift	tus	Ut (he	danning oyest fullfø	rte)	Yrkesak Ja	tiv sist	e 6 mn Nei	d.	Stillin	js-%							lor								Spi	ак					
Sam	boer		Grunns	kole	Stilling/yr	ke	-					Land	lbak	grur	nn	F	ar/m	edm	or						Spr	Beł ak	nov f	or tol	k		
Ugift	/enslig		Videreg	lående	Type bedr	ift						-				N	lor								Far	/med	Imor				
Anne	ət		Høyere	utd.								Tro o	og liv	/ssy	'n										1 cu						
Tidliger	e svang	erskap /	) Chont	ob	Merkna	der tidl	igere s	V.SK. (Ars	stall, fød	ested,	flerfødsl	er, fødse	elsvek	t, sv.:	sk.va	arigh	et, ko	mpli	kasj.	, pre	eklar	npsi	ope	rative	e forlø	snin	ger, n	nisdar	inelse	r, amı	meerf.)
Lev født	itali sv.sr	`	Fx II	a																											
2011.000	Døc	lfødt ≥	500 g/22	u.																											
Tidliger	e/nåvæi	ende s	sykdomr	ner	·			A	rvelige	syk	d.	Me	rkna	der/	/ann	net											Leg	gemi	dler	_	
Intet Hiert	spesielt esykdom		Diabetes Allergi	/sv.sk.dia	a	Autoimr Gvn. svi	nun syk kd /opr	dom	Inge	n kjen e mer	ite rkn																Mec	Dagl dikam	ig ent	A	v og til
Нуре	ertensjon		Epilepsi			Psykisk	helse		Fore	dre i	slekt																				
Nyre	/urinv.		Trombose	e/behand	lling	Annet, s	se merk	n.	Hofte	eledds	sdyspla	si																			
Røyking	ler	Nei	Av og ti	il Dagl.	Sluttet i sv	.sk.	Siga	Ant. aretter Si	daglig nus A	lkohol																					
Snus						Vec	troll																								
Alkoholf	orbr.					Vec 36.	l ca. uke																				A.5	Medi	kame	entall	ergi
Andre n	et om lev	evaner	(se veiled	ler IS-2199	9)																						An	Folat	t	Fø	r sv.sk.
Aktuelt	svangei	skap	l	Jitralyd	termin		F	osterdiagr Ja	nostikk p Nei	å indik	asjon		_	2	5				3	0			_	3	35			_	40		
Siste							A	ssistert be	efruktnin	g		40	⊢							$\vdash$					⊢		$\left  \right $		$\pm$		97.5
Termin	. I	. 1		Vår	Г. Г.		. 1	Ja				cm															$\square$		4		90%
Rutinep	røver		P	conigen		Pro	ove på	Flerling	jer on				⊢		S	YM	FY	SE	-	FU	ND	ับร	SM.	ÅL	$\vdash$		4		4		50%
Dato		i I		R	ubella antis ikke påv	stoff MR	SA neg.	pos.																					<b>1</b>	1	
ABO	RhD	nos	ne		påvist	Нер	peg	nos				35 cm	⊢							⊢					H			4	╈		10%
Kun RhD	-neg. gra	vide	24)	<u>9.</u> 5	ikke påv	ist Hep	patitt B	p00.				1																$\square$	7	_	2.5
pos.		neg.	ikk	e tatt	påvist	Tok	soplasmo	ose					⊢											$\vdash$				H	-		-
RhD-profy	/lakse gitt ke	(anbef.	gitt i u. 28)		neg.	Kla	mydia					30										/							4		30
Ja Dersom fo	oster-RhD	-typing i	kke er utfø	ort: A	pos. sympt.	Fer	ritin						⊢							$\checkmark$					╞		+	$\vdash$	+		
Kontroll ar stoffscreer	nti- ning	32. u.	36	i. u.	akteriuri (A	BU) Gr.F	3 strepto	kokker (Gl	BS)									/									$\square$	$\square$	$\mp$		
Blodtypea	intistoff	Ja, se i	merkn.	-	pos.		5 ou opto		50)			25					/												+		25
Merkna	der											cm			/														$\mp$		cm
															4					$\vdash$					⊢			$\vdash$	+		-
Høyde	Før sva Vekt	ngersk	вмі	Gluk	osebelastr	ling	1	Sv.sk.u.	Faste	nde	2 timer		2					- F	Pay,	l Frø	en,	Sta	ff, J	acol	l	on, G	ajes:	sing	(201:	3) —	
				Dat	.0	Ødom	*Leie/		Kjen-	Mod	Liobh	20 cm		2	5				3	0				3	35				40	uke	r cm
Dato	Uke	Vekt	BT	Hb	Urin	0/1/2/3	Be- veg.	min.	ner liv	+/ -	%									N	otat	er									Sign.
				_																											
				_																											
				_																											
																						_									
Seleksj	jon fødes	sted	Stee	d og grunr	nlag																						Amr	neve	iledni	ing	
			Nav	m, adr.															E	Beho	ov fo	r fø	getj	enes	ste		Fød	selsfo	orb. k	urs	
Helses	tasjon et	ter fød:	sei '				ille	anfal-i				1	10.7													14	Fars	kaps	erklæ	ering	utfylt
*Leie: H Beveg.:	= hodel B = beve	eie, S egelig,	= setele F = fest	ie, T=t et	verrleie		Syke	proiging hus		Spes.	polikl.	H	/OF													KO	n <b>tinu</b> Nei	Jasjo	J <b>nsar</b> J	a a	
							1,1,10			1.0.01																_		_			

IS-2253 Helsedirektoratet 11-2015