



Research Report
Statistical Research Unit
Göteborg University
Sweden

Detection of intra-uterine growth restriction

Max Petzold
Christian Sonesson
Eva Bergman
Helle Kieler

Research Report 2003:3
ISSN 0349-8034

Mailing address: Statistical Research Unit P.O. Box 660 SE 405 30 Göteborg Sweden	Fax Nat: 031-773 12 74 Int: +46 31 773 12 74	Phone Nat: 031-773 10 00 Int: +46 31 773 10 00	Home Page: http://www.stat.gu.se/stat
--	--	--	---

DETECTION OF INTRA-UTERINE GROWTH RESTRICTION

MAX PETZOLD¹, CHRISTIAN SONESSON¹, EVA BERGMAN² AND HELLE KIELER^{2,3}

¹Department of Statistics, Göteborg University, Göteborg, Sweden

²Department of Women's and Children's Health, Uppsala University Hospital, Uppsala, Sweden

³Department of Medical Epidemiology, Karolinska Institute, Stockholm, Sweden

SUMMARY

A new methodology for on-line detection of intrauterine growth restriction (IUGR) is proposed where traditional methods for statistical surveillance are applied. Here, deficient growth rate is used to detect IUGR instead of the common surrogate measure "small for gestational age" (SGA). Foetal growth is estimated by repeated measurements of symphysis-fundus (SF) height. At each time point the new method, based on the Shiryaev-Roberts method, is used to evaluate the growth in SF height. We use Swedish data to model a normal growth pattern, which is used to evaluate the capability of the new method to detect IUGR in comparison with a method used in practice today. Results from simulations indicate that the new method performs considerably better than the method used today. We also illustrate the effect of some important factors, which influence the detection ability and illuminate the tendency of the method used today to misclassify SGA cases as IUGR.

Keywords: Growth; IUGR; Monitoring; Shiryaev-Roberts method; Surveillance; Symphysis-Fundus.

1 INTRODUCTION

IUGR, considered as a sub optimal growth compared with the genetic growth potential in utero¹, has for decades been known to substantially contribute to perinatal mortality and morbidity². An early detection of IUGR may allow for intensive monitoring during pregnancy and for planning of time and method of delivery. There are evidence that screening for IUGR can reduce e.g. the risk of perinatal death and the number of antenatal admissions and caesarean sections for foetal distress.^{3, 4} There are also evidence of long-term complications for several organ systems as a result of retarded foetal growth.⁵

During a pregnancy foetal growth can be screened by sequential repeated SF heights or ultrasonic measurements of the foetus.⁶ In contrast to ultrasound the technique to collect SF measurements is simple and inexpensive and without any side effects. It is thus widely studied as a screening instrument and used routinely in Scandinavia in antenatal care.⁷ The capability of repeated SF measurements to detect deviant foetal growth will be treated in the sequel.

The common methodology for detecting IUGR is to identify foetuses suspected to be SGA, i.e. have a low weight. By successively comparing the SF measurements from a pregnant woman with a reference curve, SGA is suspected if one or a series of measurements is below some chosen centile. This method originates from the work of Westin⁸ and will be referred to as the 'Gravidogram' method. Several reference curves have been developed⁹⁻¹², but the Gravidogram method will here be applied as a general methodology not using any of these particular curves.

The limitations by using SGA as an indicator of IUGR are well-known. A major problem by using actual weight instead of growth is the risk of misclassification. Among infants classified as SGA there will be both genetically small but healthy infants as well as growth restricted. Some growth restricted and potentially ill infants might be classified as average for gestational age and therefore not detected in time. To reduce the problem of misclassification customised SF reference curves, which include information on background variables, such as the mothers booking weight, parity and sex of the infant, have been suggested.¹³ A major problem with the customised curves is that the sex of the infant is a crucial predictor of the level of the curve and the sex is normally not known. However, customised compared to standard curves are an improvement but the key problem remains as long as actual weight (via actual SF height) instead of growth is used to measure growth restriction.¹⁴

A good indicator of the current growth rate would be the increase in SF height from the previous time point to the current.¹⁵ By using the current growth we can expect less impact from background

variables since each individual acts as its own control. Furthermore, it seems plausible that the prognosis of a pregnancy with SF measurements following the 50:th percentile in the reference curve then drifting down to the 5:th percentile some time points later will differ from a pregnancy with SF heights remaining stable on the 5:th percentile. Therefore, the growth rate may be a better indicator of IUGR than the SF height itself.

The question is now how to use growth data optimally to detect IUGR. In a recent study ultrasonic growth data have been used to detect IUGR by applying the Gravidogram method to conditional scores.¹⁶ However, from the theory of statistical surveillance we know that the use of only the last observation is generally not effective to detect changes in an observed process.¹⁷ Instead all available observations should be used to improve the ability to detect IUGR. It is thus of interest to develop a theoretically well-motivated methodology based on the repeated SF measurements.

The paper is organised in the following way. In Section 2 we estimate the normal growth pattern using data from a previous Swedish study. In Section 3 we discuss the collection of SF data to use for surveillance. A new way using self-measurements is suggested. We discuss the general framework for statistical surveillance in Section 4. Here we also present the Gravidogram method in detail and derive the new SR method. In Section 5 inferential issues of the methods are evaluated using extensive simulations. In Section 6 we discuss different types of possible improvements of the proposed surveillance system, both from a theoretical and a practical point of view and give some concluding remarks.

2 ESTIMATING THE NORMAL GROWTH PATTERN

In this section the aim is to derive a model capturing the essential parts of the growth pattern, which enables an inferential examination of the surveillance methods in the sequel. To model the normal growth pattern a data set with repeated SF measurements on 2255 healthy women living in Sweden, who participated in a multicentre prospective trial on routine ultrasound in the second trimester between 1986 and 1987, was used.⁹ In the data gestational age, expressed as completed gestational weeks, was calculated from measurements of ultrasonic biparietal diameter in the second trimester. Women with miscarriages, induced abortions, multiple pregnancies, pregnancy induced hypertension or delivery before 37 weeks of gestation were excluded.

Having repeated measurements on each subject the response variable has two levels of variability: within and between subjects. Here, genetic differences will cause between-subject variability which

can be expressed in the model as e.g. different intercepts. A genetically small foetus will have a small value of the intercept while a genetically large foetus will have a large value. For an arbitrary i :th individual the relation between $\ln SF$ and gestational age t (measured in completed weeks) can then be modelled as:

$$\ln SF_i(t) = B_{0i} + \beta_1 \cdot t + \beta_2 \sqrt{t} + e_i(t) \quad (1)$$

where B_{0i} and $e_i(t)$ are independent realisations from the $N(\beta_0, \sigma_{B_0}^2)$ and the $N(0, \sigma_e^2)$ distributions, respectively. The repeated observations of an individual are further assumed to be independent over time, i.e. $\text{cov}(e_i(u), e_i(v)) = 0$ for $u \neq v$. The model in (1) was found to be the simplest model among the best fitting models using Akaike's information criterion in a systematic comparison of models. Both more and less flexible linear and non-linear models were compared including stochastic slopes and other transformations of the response variable and the covariate. To emphasise the inferential matters the model in (1) was chosen for its simplicity. The estimates for the model in (1) were obtained as: $\hat{\beta}_0 = -0.8949$, $\hat{\sigma}_{B_0}^2 = 0.001621$, $\hat{\beta}_1 = -0.08256$, $\hat{\beta}_2 = 1.2322$ and $\hat{\sigma}_e^2 = 0.001455$.

The growth between successive time points for the i :th individual can easily be found from (1) as:

$$\Delta_i(t) = \ln SF_i(t) - \ln SF_i(t-1) = \beta_1 + \beta_2(\sqrt{t} - \sqrt{t-1}) + \varepsilon_i(t) \quad (2)$$

where $\varepsilon_i(t) = e_i(t) - e_i(t-1)$. Since each subject now acts as its own control, the individual effect is eliminated. The index i will be suppressed from here. Successive Δ values will not be independent but follow a moving average process of order one. A time sequence of Δ values will have a variance-covariance matrix of Toeplitz type; $\Sigma = (\sigma_w^2)$ where $\sigma_w^2 = 2\sigma_e^2$ if $u = v$, $\sigma_w^2 = -\sigma_e^2$ if $|u - v| = 1$ and zero otherwise. However, this time dependency is taken care of in the analysis.

In general, the SF measures can be suspected to be dependent of sex of the infant and maternal variables such as smoking and parity.¹⁸ However, if these effects are assumed to be additive in (1), i.e. included as a dummy intercept, they will be eliminated in (2). Thus, the background individual effects are avoided when using Δ instead of $\ln SF$ and no customised analysis on maternal variables is then needed.

3 SAMPLING SCHEME OF THE SYMPHYSIS-FUNDUS MEASUREMENTS

The sampling scheme of the SF measurements is of great importance for the detection ability. The risk of not measuring the SF height at every time point is the risk of not observing the growth when the IUGR occurs. In this case we will have a delay in the detection of the IUGR, which is at least as long as the time to the next measurement. Normally the SF measurements are not regularly spaced in time. In general, the time intervals between the measurements are shorter near delivery (every week instead of every third or fourth as in the beginning of the series). In practice, a fixed schedule can be hard to follow and in our data set the number of SF measurements differs between individuals from minimum 2 to maximum 17 spread from the 14:th to the 42:nd week of gestation.

However, in this study we will assume a sampling scheme where we collect measurements of *SF* every week starting from week 20. To shorten the intervals between measurements without increasing the number of antenatal visits would be possible if the SF heights are measured by the pregnant women themselves. A recently published pilot study in England reports positive results of training of women to measure their own SF heights.¹⁹ The study population was found to perform as well as midwives. The study was blinded in the way that the women marked the length on a blank tape. The SF height was then measured from the blank tape using an ordinary measuring-tape. In a coming Swedish study of self-measurements the women are supposed to take repeated SF measurements at each occasion (once a week) facilitating estimation of within- and between individual variance. The measurements will also be compared to the measurements from the routine visits at the health centre.

4 METHODS FOR STATISTICAL SURVEILLANCE

Statistical surveillance (or SPC, statistical process control) was first used in industrial settings by Shewhart²⁰ to control the quality of a manufactured products. The fields of applications have broadened since and statistical surveillance is now used in several areas. Examples include the monitoring of foetal heart rate during labour²¹ and detection of an increased incidence of a disease²².

Statistical surveillance is a sequential decision procedure, where we at each decision time point want to determine whether the process we observe is in- or out of control. The process is said to be out of control if it has undergone an important change, which we are interested in. Here we say that the process is in control if the foetus is growing at the expected rate, while it is out of control if an IUGR has occurred. The change occurs at an unknown time point τ , and our aim is to detect it as quickly as possible. Thus, by collecting data sequentially in time we want to discriminate between the two events;

$D(s) = \{\tau > s\}$ and $C(s) = \{\tau \leq s\}$ where s is the decision time point. To do this we use an alarm system, which consists of two parts; an alarm statistic, which is a function of the observations up to the current time point and an alarm limit. The time of an alarm (the indicator that the process is out of control) is the first time the alarm statistic exceeds the alarm limit.

There are several ways to construct the alarm system depending on the desired features of the system. The features of the methods are generally evaluated both with respect to false alarms as well as motivated ones. How to optimally choose the alarm statistic and the alarm limits depend on the process under study, what is desired to detect and at what time points the highest detection strength is desirable. These are the same type of questions that arises also in a traditional hypothesis testing situation, although the time component is not present. However, the sequential decision situation makes the standard hypothesis testing tools non-applicable.

4.1 *The Gravidogram method*

The simplest way to construct a surveillance method is to use only the last observation made of the process to form the alarm statistic. If this observation deviates enough from what is expected, an alarm is triggered.²⁰ In surveillance of foetus growth, this type of surveillance method originates from the work of Westin⁸ to detect SGA. Today this method is widely used also for the detection of IUGR. The Gravidogram method is based on the absolute size of the foetus. An alarm is triggered if the current SF value is below the 2.5:th population-percentile in a chosen reference curve. Note that the current SF value is an indicator of the total growth until time s and not an indicator of the current growth rate. In this study, we have modelled $\ln SF(s)$ in (1) using a longitudinal approach. As our main focus is on the surveillance methodology (and not in any particular reference curve from the literature), we will evaluate the Gravidogram methodology using the model in (1) as our reference curve. The time of an alarm can then be written as

$$t_A = \min\{s; \ln SF(s) < G(s)\}$$

where $G(s) = E[\ln SF(s)] - 2(\sqrt{\sigma_{B_0}^2 + \sigma_e^2})$.

In models without any random intercept, properties of surveillance methods which use only the last observation can generally be derived analytically. However in this case, the random intercept complicates matters.

4.2 The SR method

In this section, we will derive a new method for the detection of IUGR studied in this paper. It is based on the Shiryaev-Roberts method^{23,24} and will be referred to as the SR-method. Many of the commonly used surveillance methods can be formulated in terms of conditional likelihood-ratios. The optimal method for minimizing the expected delay (from the event to the detection) of an alarm is the full likelihood-ratio method. However this method requires an assumption of the distribution of the change point. Since no reliable prior information about the distribution of the time point at which IUGR occurs is available, we will instead base our method on the Shiryaev-Roberts method. It can be regarded as being the full likelihood-ratio method using a non-informative prior for the time of the change point.

We will base the SR method on $\Delta(t)$. At time point s , let Δ_s be the vector of observations with variance-covariance matrix Σ_s . Let $L(s, t')$ be the (conditional) likelihood-ratio for the case $\tau = t'$. It follows that

$$L(s, t') = f(\Delta_s | \tau = t') / f(\Delta_s | \tau > s) = \exp\{[\Delta_s - \mu_s(t')]'\Sigma_s^{-1}[\Delta_s - \mu_s(t')] - [\Delta_s - \mu_s]'\Sigma_s^{-1}[\Delta_s - \mu_s]\}$$

where $\mu_s(t')$ is the vector of expected values of Δ_s for the case $\tau = t'$ and μ_s is the vector of expected values of Δ_s for the case $\tau > s$. For the SR method, we can specify the out of control state and thus optimize the method to detect a certain type of IUGR. All conditional likelihoods are then weighted equally in the alarm statistic together with a constant alarm limit and an alarm is given at

$$t_A = \min\{s; \sum_{t'=20}^s L(s, t') > K\}.$$

Note that although the alarm statistic implies that the conditional likelihoods are weighted equal, recent observations have more weight than old ones which is intuitively appealing.

A restricted growth (the event to be detected) will correspond to a decrease in the growth rate compared to the expected one. Here we will optimize the SR method for the case of an IUGR with a completely stopped growth, see Figure 1. This is usually the result of placental insufficiency secondary to extrinsic factors such as hypertension and diabetes mellitus.²⁵

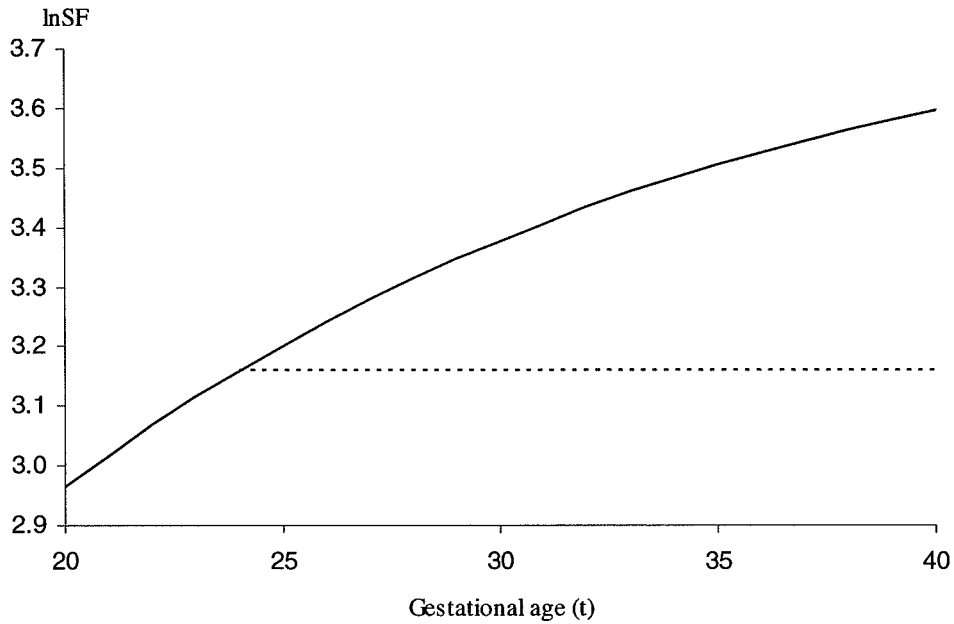


FIGURE 1: The expected growth pattern for a normal growth (solid line) and for the case when IUGR (here a completely stopped growth) occurs in the 25:th week of gestation (dotted line).

5 PROPERTIES OF THE METHODS

There are two types of alarms to consider; false alarms and motivated ones and we have to face a trade off between the rate of false alarms and short delay times for motivated ones. We will here illustrate the distribution of false alarms and the effect of some factors on the ability to detect IUGR (the length of the time period for detection, the time point when the IUGR occurs, the residual variance). These factors influence both the level of detection ability and the relative performance of the methods. For the Gravidogram method the dependency of the genetic size of the foetus will also be illustrated. However, every situation is a combination of these factors, which contribute to a complex pattern of performance. We will therefore illustrate the performance with some examples and discuss the general implications of the various factors on both the level of detection ability and the relative performance of the methods.

The surveillance methods were evaluated using simulations. In this way we can investigate with high accuracy how the methods will perform when applied in clinical practice. The $\ln SF$ values were simulated using the model (1) with the estimated parameter values from Section 2, and the Δ values

were then calculated from the $\ln SF$ values using (2). For each evaluation of the performance of the methods, 1.000.000 replicates have been used.

5.1 *The distributions of the false alarms*

To make the evaluation of the detection ability comparable the probability of a false alarm is fixed within a normal length of a pregnancy (here 40 weeks). The Gravidogram method was found to have a 19.2% probability of a false alarm within this period. In medical terms, the specificity of the method is 19.2% if we regard the whole pregnancy to represent one test period. For comparability, the alarm limit for the SR method was set to achieve the same probability.

Of great importance when comparing the methods are the distributions of the false alarms. In Figure 2 the cumulative false alarm distributions are displayed. For the Gravidogram method the distribution of the false alarms is approximately geometric, with highest probability of a false alarm at the first decision time point and thereafter a geometrically decreasing probability giving the retarding shape of the cumulative distribution. This is not the case for the SR method. For the SR method we need two SF measurements to start the surveillance. Hence, we can not have a false alarm at the first time point. The low frequency of false alarms for early time points is typical for the SR method. For later time points we have a higher rate. Since the false alarm distribution is an indicator of the sensibility of the methods at different time points, the ability to detect IUGR will also depend on it. This means that in general the Gravidogram method will perform relatively better for cases of IUGR occurring at the start of the monitoring period. However, for severe cases of IUGR, the ability to save a foetus is small for early time points.²⁵ Therefore, it could be argued that the false alarm distribution of the SR method is preferable to that of the Gravidogram method from a medical point of view.

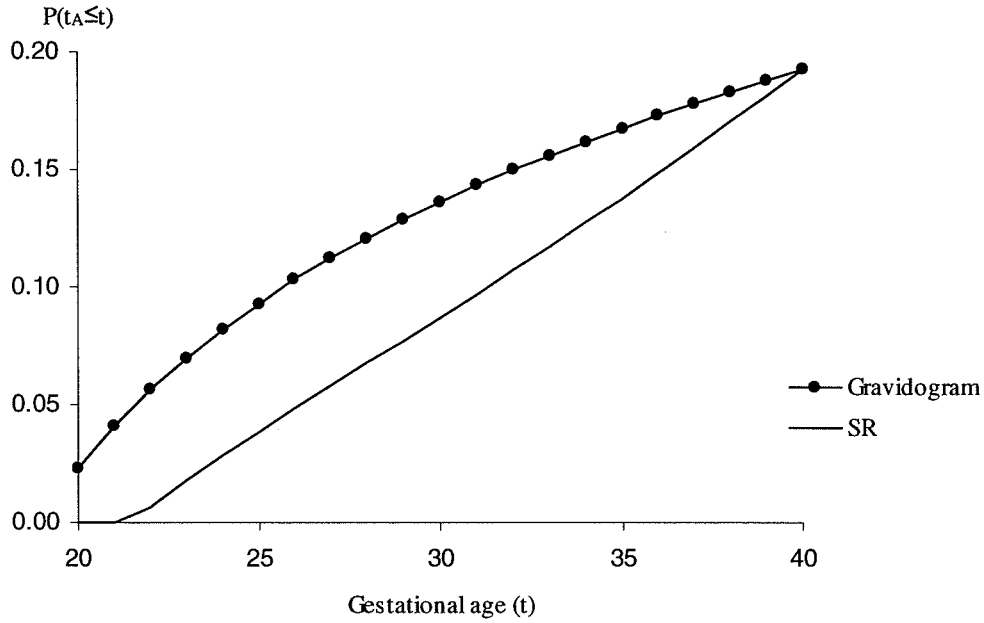


FIGURE 2: The cumulative false alarm distributions over gestational age (t).

A most important issue to deal with is the difference between a genetic small but healthy foetus and a foetus suffering from IUGR. The way the Gravidogram method is constructed, a misclassification of SGA foetuses as IUGR is unavoidable. As previously stated, the genetic differences are expressed as different intercepts for different individuals in the mixed model for the expected growth pattern, see formula (1). The value of the intercept is highly influential on the rate of false alarm. By using a fixed intercept set to a certain percentile from the distribution of B_0 , this was examined for the Gravidogram method. The probability of a false alarm was then found to be 93.3%, 25.8%, 3.8% and 0.3% respectively for the 5:th, 25:th, 50:th and the 75:th percentile of the intercept. Note that the probability of a false alarm is independent of the genetic size of the foetus for the SR method since Δ is independent of the intercept B_0 .

5.2 Ability of successful detection

A common way to evaluate test procedures in medicine is to use the sensitivity. Usually the sensitivity is defined as the proportion of the “positives” which are correctly identified by a test procedure. In this application the sensitivity is the proportion of IUGR foetuses, which are correctly classified. However the sensitivity does not say anything about when the classification is made. The delay time in detection

of IUGR is crucial and therefore we must use measures of evaluation, which includes a time component as well. When evaluating motivated alarms, we find it most appropriate to consider the probability of successful detection,

$$PSD(d, \tau) = P(t_A - \tau \leq d \mid t_A \geq \tau).$$

It stands for the probability that the IUGR is detected with a delay time no longer than d when the IUGR occurred at time point τ . This measure of evaluation has been used in medical applications before²¹, e.g. when monitoring a foetus heart rate during labor. Both the sensitivity and the probability of successful detection thus deal with the proportion of correctly classified “positives”. However for situations when the sensitivity is generally used, the time component is not present. Therefore, the probability of successful detection can be regarded as a measure of sensitivity taking also the delay time of the detection into account. There are also other ways of evaluating surveillance methods. The most commonly used ones however focus on the expected delay time. For the detection of IUGR the important feature is not the expected delay time but rather the proportion of foetuses we can detect within a reasonable time.

5.2.1 The influence of the time available for the detection of IUGR

The detection ability as a function of the delay time available for the detection, d , is displayed in Figure 3. Most important is to notice that the detection ability of the SR method is considerably higher than for the Gravidogram method.

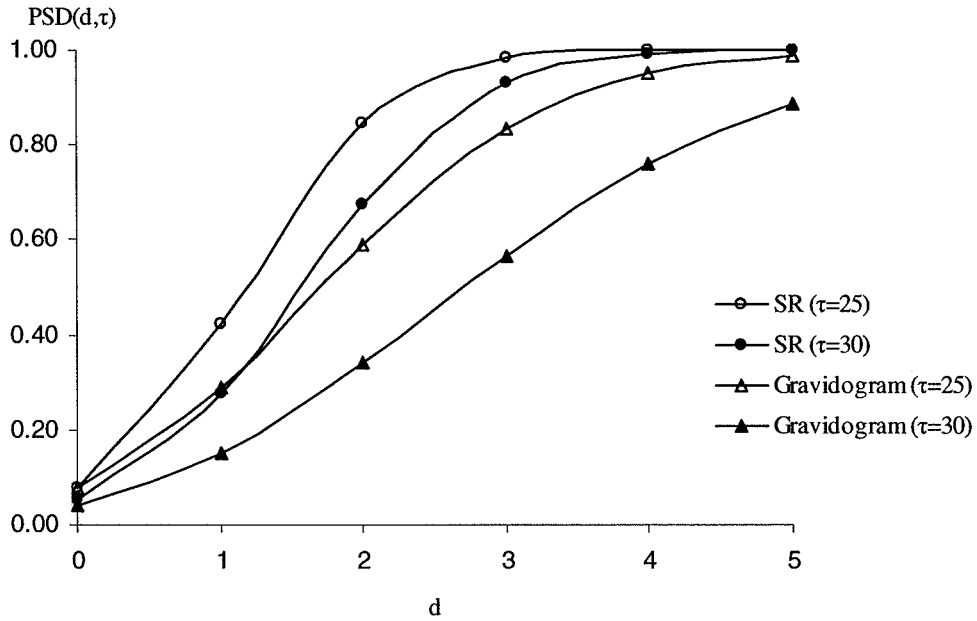


FIGURE 3: The probability of successful detection as a function of the time in weeks (d) available for detection. The week of gestation when the IUGR occurs (τ) is indicated in the labels.

Observe that the level of detection is almost 70% within two weeks for the SR method if the IUGR occurs in the 30:th week of gestation. In this period of the pregnancy, two weeks alone would correspond to a normal interval between the regular visits to a health centre. The corresponding sensitivity for the Gravidogram method is only about 30%.

5.2.2 The influence of the time point of the IUGR

It should be clear that a large decrease in the expected value of $\ln SF(t)$ is easier to detect than a small one. Therefore the level of detection is decreasing as a function of time due to a decreased expected growth rate (Figure 1) as can be seen in Figure 4.

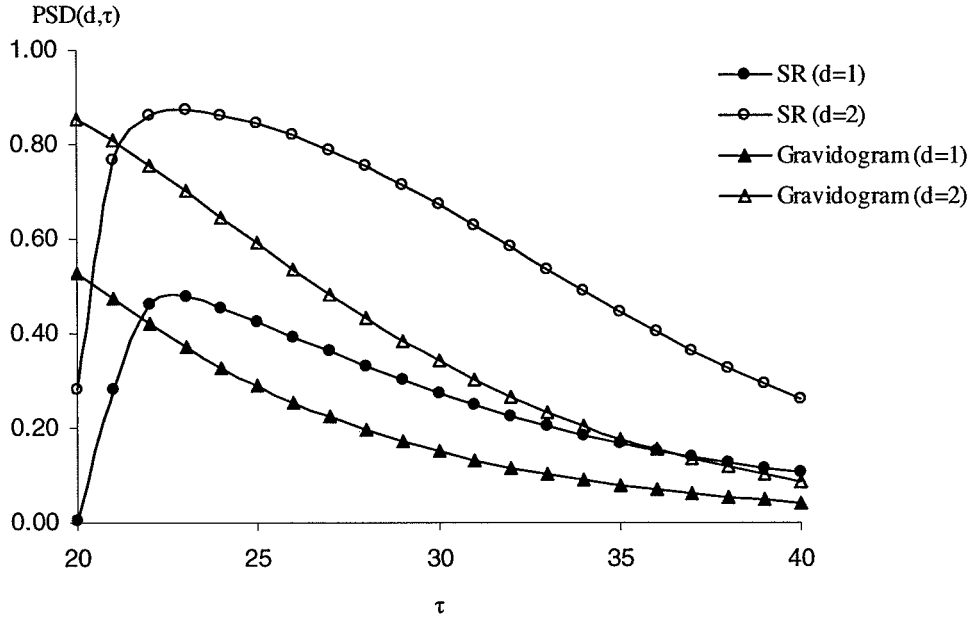


FIGURE 4: The probability of successful detection as a function of the week of gestation the IUGR occurs. The value of d is indicated in the labels.

Although the SR method needs two observations before an alarm can be triggered, it has a higher PSD than the Gravidogram method also for IUGR occurring before 25 weeks of gestation. Normally IUGR with a completely stopped growth is believed to occur in the third trimester²⁶ in which case the detection ability is much higher for the SR method. A large value of the time available for detection (d) favours the SR method even more.

5.2.3 The influence of the residual variance

The technique of measuring SF heights has been criticised for being imprecise. However, standardised measuring procedures and training can substantially decrease the variance in the data.^{27, 28} One crucial issue introducing self-measurements is the quality of the data reported. The residual variance, σ_e^2 , influences both the level of the detection ability as well as the relative performance between the Gravidogram- and the SR method. Here we have examined the effect of a decrease in the residual variance by 50%. This could be the result of several factors such as extensive training in measuring SF height or as a result of measuring the SF height more than once at each occasion and thus reducing the residual variance by considering the mean of the observations taken.

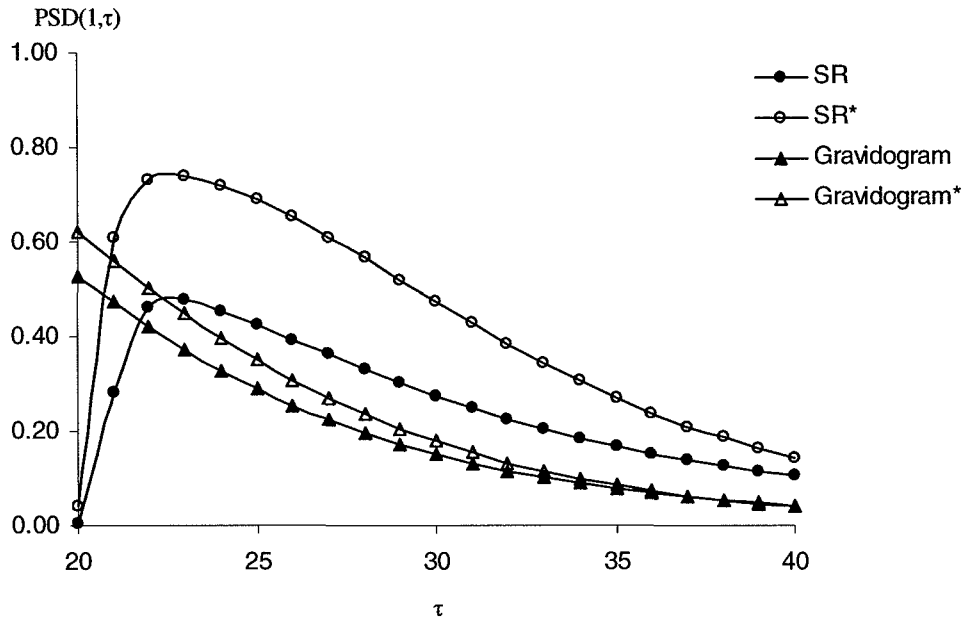


FIGURE 5: The probability of successful detection within one week as a function of the week of gestation the IUGR occurs. The cases with a reduced residual variance by 50% are indicated by * in the labels.

In Figure 5, $PSD(1,\tau)$ values are presented for the two values of the residual variance. Since $V(\Delta(t)) = 2\sigma_e^2$ and $V(\ln SF(t)) = \sigma_{B_0}^2 + \sigma_e^2$ the effect of reducing the residual variance is larger for the SR method than for the Gravidogram method. For the SR method the detection ability is here generally increased by over 50%. Practical clinical work to reduce the residual variance in SF measurements is thus of greatest interest.

5.2.4 The influence of the genetic size of the foetus

As was seen in Section 5.1, the genetic size of the foetus was important for the probability of a false alarm for the Gravidogram method. This is also the case for the detection ability when the IUGR occurs as can be seen in Figure 6. The Gravidogram method is very sensitive to detect IUGR occurring for genetically small foetuses, but not for genetically large ones. Note that the detection ability of the SR method is independent of the value of the intercept. This means that the detection ability is

independent of the genetic size of the foetus, which might be preferred from an ethical point of view when applying the method as a screening instrument in practice.

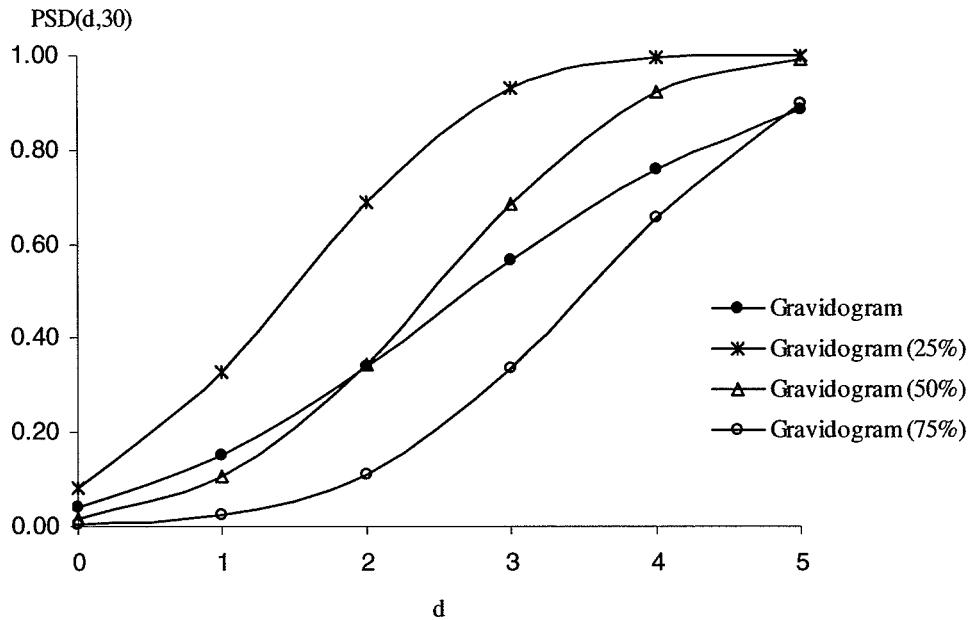


FIGURE 6: The probability of successful detection as a function of the time in weeks (d) available for detection, when the IUGR occurs in the 30:th week of gestation. The percentile of the intercept used is indicated in the labels.

6 DISCUSSION

In this paper we have shown that the methodology used in practice today in Sweden and several other countries for detection of IUGR can be considerably improved still using SF measurements. Three important aspects of this paper are the following. Firstly, we have focused on the growth rate in symphysis-fundus height as being the important characteristic of naturally growing foetus and not the absolute size of the foetus. An important advantage of using the growth rate is that additive effects from background variables (smoking, parity, sex of the child) in the model describing the absolute size do not influence the growth rate. Secondly, we have introduced a theoretically well-motivated surveillance method to detect IUGR. In order to achieve an effective surveillance method, the information in the observed data must be treated correctly as by the SR method. Thirdly, we have used

a measure of evaluation, PSD, which focus not only on whether the IUGR is detected or not, but also on the delay time of the detection.

In our approach we have not used any assumptions on the distribution of the time point when the IUGR occurs. If we had information about the distribution from theory or previous studies we could focus the detection ability to the most probable time points and thus achieve more efficient methods. Knowing the distribution of the time point of IUGR and a value of the incidence of IUGR, we can calculate the predictive value²¹ of the methods which helps us in the choice of actions to take if an alarm is triggered. The predictive value is a function of the time point of an alarm and defined as $PV(s) = P(C(s) | t_A = s)$, which is the probability that the IUGR has occurred given that an alarm is triggered at a specific time point. A constant predictive value is in many situations desirable since it implies that the actions to follow an alarm can be the same whenever it is triggered. The alarms of the Gravidogram method will not be trustworthy due to the high intensity of early false alarms. As a result, the SR method will have a considerably higher predictive value than the Gravidogram method. Also in this respect the SR method will be preferable to the Gravidogram method.

We have chosen to study the behavior of the SR method since it is known to have several desirable properties such as a near optimal expected delay and for some cases an almost constant predictive value. If the requirement of the method is very specific, such as stated here where focus has been solely on the PSD, the SR method is not the optimal method. Another weighting of the conditional likelihoods could be chosen. However that would require a focus on a particular value of time of delay (d).

An advantage of the SR method is that it can be optimised to detect a specific kind of growth. Here, the SR method has been specified for the detection of a completely stopped growth. For the detection of other types of IUGR it would be preferable to specify the SR method according to this type. An example is when the growth rate decreases and remains at a lower rate than expected throughout the pregnancy. Often this pattern occurs with extrinsic conditions such as intrauterine infections or intrinsic embryonic conditions such as congenital malformations.²⁵ To investigate the robustness of the IUGR specification, we have evaluated the detection of a completely stopped growth when specifying the SR method for a decreased growth rate. Notable deterioration in detection ability were only present for early changes (before the 25:th week of gestation), but not for later ones, where the non-optimised SR method performed almost identical to the correctly specified version. This indicates that the SR method is robust to the specification of the type of IUGR.

We believe that self-measurements used together with the SR method have a large potential as a screening instrument in clinical practice. In practice today, cases where we suspect IUGR (e.g. where SGA is suspected) are measured more frequently than those considered to be out of risk. However, IUGR can occur at any time point during the pregnancy. One major advantage using self-measurements is that all women are screened equally often throughout the whole pregnancy and not based on the initial measurement values. We also believe that self-measuring at home will be better and easier for the mothers and that the number of missing data will be small. Self-measuring can also contribute to a more cost-effective antenatal-care program by reducing the number of visits to the health centre. A recent study of standard antenatal-care programs in terms of clinical outcomes, perceived satisfaction and costs concluded that a model with a reduced number of antenatal visits could be introduced without risk to the mother or baby.²⁹ A combined program of self-measuring and visits to the health centre could then be very cost-effective and increase the performance of the antenatal care.

For the clinical practitioner, the SR method is not as easy to grasp as for example the Gravidogram method. This should not however be taken as a motivation for using the non-satisfactory Gravidogram method. A computer program, which performs all necessary calculations, can be used by the practitioner, who only has to register the SF measurements in a database. Also the mothers can be equipped at home with a simple calculator designed for this monitoring. What needs to be done before applying the technique in practice is a large study of self-measurements in order to construct an appropriate expected growth curve. An interesting extension of this work is to apply multivariate surveillance^{30, 31} where the monitoring can be based on both weekly SF measurements and infrequent ultrasonic measurements.

ACKNOWLEDGEMENTS

The authors thank Professor Marianne Friséén for much good advice during the work with this paper. Robert Jonsson PhD and Eva Andersson PhD are acknowledged for valuable comments.

REFERENCES

1. Goldenberg, R. L. and Cliver, S. P. 'Small for gestational age and intrauterine growth restriction: Definitions and standards', *Clinical Obstetrics and Gynecology*, **40**, 704-714 (1997).
2. Resnik, R. 'Intrauterine growth restriction', *Obstetrics and Gynecology*, **99**, 490-496 (2002).
3. Westergaard, H. B., Langhoff-Roos, J., Lingman, G., Marsal, K. and Kreiner, S. 'A critical appraisal of the use of umbilical artery Doppler ultrasound in high-risk pregnancies: use of meta-analyses in evidence-based obstetrics', *Ultrasound in Obstetrics & Gynecology*, **17**, 466-476 (2001).
4. Alfircvic, Z. and Neilson, J. P. 'Doppler Ultrasonography in High-Risk Pregnancies - Systematic Review with Metaanalysis', *American Journal of Obstetrics and Gynecology*, **172**, 1379-1387 (1995).
5. Barker, D. J. P. 'The long-term outcome of retarded fetal growth', *Clinical Obstetrics and Gynecology*, **40**, 853-863 (1997).
6. Mongelli, M. and Gardosi, F. 'Fetal growth', *Current Opinion in Obstetrics & Gynecology*, **12**, 111-115 (2000).
7. Challis, K., Osman, N. B., Nystrom, L., Nordahl, G. and Bergstrom, S. 'Symphysis-fundal height growth chart of an obstetric cohort of 817 Mozambican women with ultrasound-dated singleton pregnancies', *Tropical Medicine & International Health*, **7**, 678-684 (2002).
8. Westin, B. 'Gravidogram and fetal growth', *Acta Obstet Gynecol Scand*, **56**, 273-282 (1977).
9. Kieler, H., Axelsson, O., Hellberg, D., Nilsson, S. and Waldenstrom, U. 'Serial measurements of symphysis-fundus height in women with ultrasonically dated pregnancies', *Journal of Obstetrics and Gynaecology*, **16**, 228-229 (1996).
10. Steingrimsdottir, T., Cnattingius, S. and Lindmark, G. 'Symphysis-fundus height: construction of a new Swedish reference curve, based on ultrasonically dated pregnancies', *Acta Obstet Gynecol Scand*, **74**, 346-351 (1995).
11. Theron, G. B. and Thompson, M. L. 'A centile chart for birth weight for an urban population of the Western Cape', *South African Medical Journal*, **85**, 1289-1292 (1995).
12. Walraven, G. E. L., Mkanje, R. J. B., Vandongen, P. W. J., Vanroosmalen, J. and Dolmans, W. M. V. 'The Development of a Local Symphysis-Fundal Height Chart in a Rural Area of Tanzania', *European Journal of Obstetrics Gynecology and Reproductive Biology*, **60**, 149-152 (1995).
13. Gardosi, J. and Francis, A. 'Controlled trial of fundal height measurement plotted on customised antenatal growth charts', *British Journal of Obstetrics and Gynaecology*, **106**, 309-317 (1999).
14. Altman, D. G. and Hytten, F. E. 'Intrauterine Growth-Retardation - Lets Be Clear About It', *British Journal of Obstetrics and Gynaecology*, **96**, 1127-1128 (1989).
15. Royston, P. 'Calculation of Unconditional and Conditional Reference Intervals for Fetal Size and Growth from Longitudinal Measurements', *Statistics in Medicine*, **14**, 1417-1436 (1995).
16. Hooper, P. M., Mayes, D. C. and Demianczuk, N. N. 'A model for foetal growth and diagnosis of intrauterine growth restriction', *Statistics in Medicine*, **21**, 95-112 (2002).
17. Frisén, M. and de Maré, J. 'Optimal Surveillance', *Biometrika*, **78**, 271-280 (1991).
18. Mongelli, M. and Gardosi, J. 'Symphysis-fundus height and pregnancy characteristics in ultrasound-dated pregnancies', *Obstet Gynecol*, **94**, 591-594 (1999).
19. Boulos, A. N., Griffiths, M., Allott, H. and Holt, E. M. 'Trial of self-administered antenatal care: maternal symphysis fundal height measurements', *Journal of Obstetrics and Gynaecology*, **19**, 623-623 (1999).
20. Shewhart, W. A. *Economic Control of Quality of Manufactured Product*, MacMillan and Co., London, 1931.
21. Frisén, M. 'Evaluations of Methods for Statistical Surveillance', *Statistics in Medicine*, **11**, 1489-1502 (1992).

22. Sonesson, C. and Bock, D. 'A review and discussion of prospective statistical surveillance in public health', *Journal of the Royal Statistical Society A*, **166**, 5-21 (2003).
23. Shiryaev, A. N. 'On optimum methods in quickest detection problems.' *Theory of Probability and its Applications.*, **8**, 22-46 (1963).
24. Roberts, S. W. 'A Comparison of some Control Chart Procedures', *Technometrics*, **8**, 411-430 (1966).
25. Kurjak, A. *Textbook of perinatal medicine : a comprehensive guide to modern clinical perinatology*. Vol. 2 Parthenon, Carnforth, 1998:1194.
26. Owen, P., Maharaj, S., Khan, K. S. and Howie, P. W. 'Interval between fetal measurements in predicting growth restriction', *Obstetrics and Gynecology*, **97**, 499-504 (2001).
27. Rogers, M. S. and Needham, P. G. 'Evaluation of fundal height measurement in antenatal care', *Australian & New Zealand Journal of Obstetrics & Gynaecology*, **25**, 87-90 (1985).
28. Johnsen, T. S., Jacobsen, G. and Knoff, T. 'The Effect of Practical Training in Obstetrics among Medical- Students - Symphysis Fundal Height Measurements', *Medical Education*, **22**, 438-444 (1988).
29. Carroli, G., Villar, J., Piaggio, G., et al. 'WHO systematic review of randomised controlled trials of routine antenatal care', *Lancet*, **357**, 1565-1570 (2001).
30. Wessman, P. 'Some principles for surveillance adopted for multivariate processes with a common change point', *Communications in Statistics-Theory and Methods*, **27**, 1143-1161 (1998).
31. Lundbye-Christensen, S. 'Monitoring pregnancy', *Communications in Statistics-Stochastic Models*, **5**, 383-399 (1989).

Research Report

- | | | |
|--------|--------------------------------------|---|
| 2002:1 | Frisén, M. & Sonesson, C.: | Optimal surveillance. Based on exponentially weighted moving averages. |
| 2002:2 | Frisén, M.: | Statistical surveillance. Optimality and methods. |
| 2002:3 | Jonsson, R. & Persson, A.: | Bayes prediction of binary outcomes based on correlated discrete predictors. |
| 2002:4 | Persson, A.: | Prediction of work resumption among men and women with lower back- and neck pain in a Swedish population. |
| 2002:5 | Persson, A.: | Prediction of work resumption in theory and practice. |
| 2002:6 | Andersson, E., Bock, D. & Frisé, M.: | Detection of turning points in business cycles. |
| 2002:7 | Andersson, E., Bock, D. & Frisé, M.: | Some statistical aspects on methods for detection of turning points in business cycles. |
| 2002:8 | Andersson, E., Bock, D. & Frisé, M.: | Statistical surveillance of cyclical processes with application to turns in business cycles. |
| 2002:9 | Holgersson, T.: | Testing for non-normality in multivariate regression with nonspherical disturbances. |
| 2003:1 | Holgersson, T. & Shukur, G.: | Testing for multivariate heteroscedasticity. |
| 2003:2 | Holgersson, T.: | Testing for multivariate autocorrelation. |