Citation/Reference	**Cecilia Bottomley, Vanya Van Belle, Emma Kirk, Sabine Van Huffel, Dirk Timmerman, Tom Bourne**, (2013), Accurate prediction of pregnancy viability by means of a simple scoring system Human Reproduction, 28 (1), 68-76.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://humrep.oxfordjournals.org/content/28/1/68.long
Journal homepage	http://humrep.oxfordjournals.org/
Author contact	vanya.vanbelle@esat.kuleuven.be *+ 32 (0)16 32.10.65*
IR	https://lirias.kuleuven.be/handle/123456789/355623

Accurate prediction of pregnancy viability by means of a simple scoring system

Cecilia Bottomley1, Vanya Van Belle2, Emma Kirk3, Sabine Van Huffel2, Dirk Timmerman4, Tom Bourne3,4

1. Chelsea and Westminster Hospital, London, UK
2. ESAT, SCD-SISTA, Kasteelpark Arenberg 10, box 2446, 3001, Leuven; IBBT Future Health Department
3. Imperial College London, Hammersmith Campus, Du Cane Road, London, UK
4. Department of Obstetrics and Gynaecology, University Hospitals KU Leuven, Leuven, Belgium

Correspondence:
Dr Cecilia Bottomley MRCOG
Department of Obstetrics and Gynaecology
Chelsea and Westminster Hospital
369 Fulham Road
London SW10 9NH
Tel: +44 7985 937833
Fax: +44 20
E-mail: ceciliabottomley@gmail.com

Key words:
Miscarriage, first trimester, pregnancy, transvaginal ultrasound, viability

Running title: Early pregnancy viability score

Extended abstract

Study question:
What is the performance of a simple scoring system to predict whether women will have an ongoing viable intrauterine pregnancy beyond the first trimester.

Summary answer:
A simple scoring system using demographic and initial ultrasound variables accurately predicts pregnancy viability beyond the first trimester with an area under the curve (AUC) in a receiver operating characteristic (ROC) curve of 0.924 [confidence interval (CI) 0.900-0.947] on an independent test set.

What is known already?
Individual demographic and ultrasound factors, such as maternal age, vaginal bleeding and gestational sac size, are strong predictors of miscarriage. Previous mathematical models have combined individual risk factors with reasonable performance. A simple scoring system, derived from a mathematical model, which can be easily implemented in clinical practice, has not previously been described to predict ongoing viability.

Study design, size, duration:
This was a prospective observational study in a single early pregnancy assessment centre during a nine month period.

Participants/materials, setting, methods:
A cohort of 1881 consecutive women undergoing transvaginal ultrasound scan at a gestational age less than 84 days were included. Women were excluded if the first trimester outcome was not known. Demographic features, symptoms and ultrasound variables were tested for their influence on ongoing viability. Logistic regression was used to determine the influence on first trimester viability from demographics and symptoms alone, ultrasound findings alone and then from combining all variables. Each model was developed on a training data set, and a simple scoring system was derived from this. This scoring system was tested on an independent test dataset.

Main results and the role of chance:
The final outcome based on a total of 1435 participants was an ongoing viable pregnancy in 885 (61.7%) and early pregnancy loss in 550 (38.3%) women. The scoring system using significant demographic variables alone (maternal age and amount of bleeding) to predict ongoing viability gave an AUC of 0.724 (95% CI=0.692-0.756) in the training set and 0.729 (95% CI=0.684-0.774) in the test set. The scoring system using significant ultrasound variables alone (mean gestation sac diameter, mean yolk sac diameter and presence of fetal heart beat) gave an AUC of 0.873 (95% CI=0.850-0.897) and 0.900 (95% CI=0.871-0.928) in the training and test sets, respectively. The final scoring system using demographic and ultrasound variables together gave an AUC of 0.901 (95%

CI=0.881-0.920) and 0.924 (CI 0.900-0.947) in the training and test sets, respectively. After defining the cut-off for which the sensitivity is 0.90 on the training set, this model performed with a sensitivity of 0.92, specificity of 0.73, positive predictive value of 84.7% and negative predictive value of 85.4% in the test set.

Limitations, reasons for caution:
BMI and smoking variables were a potential omission in the data collection and might further improve the model performance if included. A further limitation is the absence of information on either bleeding or pain in 18% of women. Caution should be exercised before implementation of this scoring system prior to further external validation studies

Wider implications of the findings:
This simple scoring system incorporates readily available data which are routinely collected in clinical practice and does not rely on complex data entry. As such it could, unlike most mathematical models, be easily incorporated into normal early pregnancy care, where women may appreciate an individualized calculation of the likelihood of ongoing pregnancy viability.

Study funding/competing interest(s): Research by VvB supported by Research Council KUL: GOA MaNet, PFV/10/002 (OPTEC), several PhD/postdoc & fellow grants; IWT: TBM070706-IOTA3, PhD Grants; IBBT; Belgian Federal Science Policy Office: IUAP P7/ (DYSCO, `Dynamical systems, control and optimization', 2012-2017).

Trial registration number: Not applicable

Introduction

Miscarriage (defined as the loss of pregnancy before 24 completed weeks' gestation (RCOG Press 2006)) is the most common serious pregnancy complication. It is not generally associated with serious physical morbidity or mortality (Trinder et al. 2006) but very frequently has a significant social and psychological impact on the mother (Lok and Neugebauer 2007). The physician has an important role in making an accurate and certain diagnosis of pregnancy failure and relaying this information in a sensitive and professional manner. Evidence-based practice should always be employed to ensure that an accurate diagnosis is made in order to avoid inadvertent termination of a potentially viable pregnancy; but the physician must also avoid giving false assurance regarding a pregnancy that is in fact likely to fail.

The exact incidence of miscarriage is difficult to assess. A commonly quoted figure for the rate of clinical miscarriage is around one in five recognised pregnancies (Savitz et al. 2002). Prospective observational studies suggest a rate of 31-38% when the high number of 'biochemical' pregnancies is included (Zinaman et al. 1996, Ellish et al. 1996). It is likely that the rate of miscarriage of clinically recognised pregnancies is increasing as a result of earlier recognition of pregnancy with commercially available, highly sensitive, urine-based, home pregnancy tests and the easy availability of transvaginal ultrasonography (TVS). The number of women who suffer a miscarriage at home without any medical involvement is unknown but the miscarriage rate amongst women attending an early pregnancy unit is 17–46% (Bigrigg and Read 1991, Shillito and Walker 1997, Schauberger et al., 2005).

Individual maternal and ultrasound factors are known to be associated with a higher likelihood of miscarriage. Maternal age is the most important demographic variable in the prediction of miscarriage with a steep increase in the risk after 35 years of age (De la Rochebrochard and Thonneau 2002) predominantly owing to the increase in chromosomal abnormalities with maternal age. Approximately 20% of pregnancies in 35-year-old women result in fetal loss compared to 55% at 42 years (Andersen et al., 2000).

Pain and bleeding are the most common reasons for women to be seen in an early-pregnancy assessment unit (Bigrigg and Read 1991). However, the influence of these is poorly described in relation to predicting the diagnosis of miscarriage. Vaginal bleeding occurred in at least 20% of women in a prospective community-based early pregnancy study and over half (55%) of the women who presented with bleeding ultimately went on to miscarry (Everett 1997). Women who have moderate or heavy bleeding have more than twice the rate of miscarriage compared with those with light bleeding (Poulose et al., 2006).

Where the diagnosis of miscarriage is not made at first ultrasound assessment, various ultrasound factors have been described as possible predictors of

subsequent outcome (Jauniaux et al., 2005). The rate of miscarriage after initial visualisation of cardiac activity is 2–16% (Cashner et al., 1981, Hill et al., 1991, Siddiqui et al., 1988, Makrydimas et al., 2003). Measurements of fetal size below expected for gestational age or decreased embryonic growth is associated with subsequent miscarriage (Reljic 2001, Mantoni and Pedersen 1982, Choong et al., 2003, Mukri et al., 2008, Bottomley et al., 2009a). Single measurements of small gestation sac size and a small difference between mean sac diameter (MSD) and crown-rump length (CRL) have also been associated with an increased likelihood of miscarriage after initial confirmation of viability (Bromley et al., 1991, Dickey et al., 1992). A yolk sac diameter of more than two standard deviations either above or below the mean for gestation is associated with a high specificity for miscarriage (95.3% below mean, 97.4% above mean). Evidence of a very large sac ('mega yolk sac') is considered pathological and is usually associated with pregnancy loss (Jauniaux et al. 2005).

The role of sub-chorionic haematoma as a predictor of miscarriage remains controversial and a recent prospective study showing that the presence of an intrauterine haematoma in the first trimester is not an independent risk factor for miscarriage after vaginal bleeding has been taken into account (Johns et al., 2003).

Some previous attempts have been made to determine how these variables can be combined to predict miscarriage in women after initial ultrasound. For example in 322 women, a multivariate model including MSD, CRL, fetal heart rate, maternal age and gestational age had an area under the curve (AUC) in the receiver operating characteristic (ROC) curve of 0.87 [95% confidence interval (CI) 0.80–0.95] (Choong et al., 2003). This was in an IVF population of women with an average age of 34 years who may have been at high risk of miscarriage. Another study of 270 women with vaginal bleeding and an initially viable fetus found that independent risk factors for subsequent miscarriage were fetal bradycardia, a small discrepancy between the MSD and CRL and a discrepancy between menstrual and ultrasound age. Where none of the three risk factors was present, the chance of miscarriage was 6% compared to 84% when all were present (Falco et al., 1996).

A further recent retrospective assessment of 768 women who miscarried (compared with 4698 with a normal first trimester outcome) showed that the addition of ultrasound variables to maternal demographics and the presence of bleeding, in a logistic regression model, was associated with a prediction rate for miscarriage of 85.7% (at a false positive rate of 30%) (Papaioannou et al., 2011).

The aim of our study was to prospectively collect demographic and symptom variables and ultrasound data on all women attending our early pregnancy unit for early pregnancy ultrasound assessment in order to develop and evaluate a simple and practical model to predict likelihood of a woman having a viable intrauterine pregnancy at the end of the first trimester.

Methods

The national ethics committee confirmed that this prospective observational cohort study did not require ethical approval (January 2006). The study group was composed of all women attending an early pregnancy assessment centre for ultrasound assessment during a nine-month period (January to October 2006). The unit is located within an inner city university hospital with an open access early pregnancy service. Women either self refer or are referred by the general practitioner or emergency department. Prior to a TVS the referral indication was recorded and all women completed a detailed written questionnaire of their demographic details (including ethnicity and age), obstetric history and current pregnancy symptoms (pain or bleeding) immediately prior to consultation. A pictorial blood loss chart was used to quantify the amount of reported bleeding, allocating a score of zero to four, where zero represented no bleeding and four represented clots or 'flooding'. This is described hereafter as the 'bleeding score'. A visual analogue pain score was used, using a 100mm horizontal line, with zero representing 'no pain at all' and ten representing 'the worst pain I can imagine'.

TVS was then performed using a 5 MHz transducer for B mode imaging (Aloka SSD 900, 2000, 4000 or GE Voluson 730). All operators were experienced sonographers (specialist nurse, midwife or gynaecologist). Ultrasound findings were contemporaneously entered onto an electronic database (Viewpoint PIA database, LB Systems, Vienna, Austria). Whether or not a yolk sac was seen, whether or not an embryo was seen and whether or not a heart beat was seen, were recorded. Measurements of MSD and mean yolk sac diameter were taken in three orthogonal planes and the CRL was recorded.

The criteria for inclusion in the study analysis were first attendance in the index pregnancy, a history of a positive pregnancy test and a gestational age, defined as the number of days from the first day of the last menstrual period (LMP), of no more than 84 days (in women with known LMP only). Women were excluded if they presented postpartum or with complications from termination of pregnancy or previously diagnosed early pregnancy loss. Multiple pregnancies, patients who were not in fact pregnant and those in whom first trimester outcome was unknown were excluded from the study.

The TVS outcomes were pre-defined using the recommended criteria at the time, described previously (Bottomley et al., 2009b) and shown in Table 1. Depending on the outcome of the assessment, women were managed according to standard departmental protocols, consistent with the contemporaneous Royal College of Obstericians and Gynaecologists (RCOG) guideline (RCOG 2006). Women with inconclusive TVS findings (pregnancy of unknown location or intrauterine pregnancy of uncertain viability) were managed expectantly.

First trimester outcome was ascertained from the early pregnancy unit database if the woman had been diagnosed with a subsequent miscarriage or from the

fetal medicine unit database or by direct telephone contact with the woman if she had not had further care within our department. Outcomes were defined as an ongoing intrauterine pregnancy, where a live fetus had been seen at the routine 11-14 week scan, or early pregnancy loss where the pregnancy had resulted in miscarriage, ectopic pregnancy or failed pregnancy of unknown location. Women who could not be contacted for follow up or who underwent termination of pregnancy were excluded from the analysis, as the first trimester outcome could not be assumed.

The terms 'embryo' and 'fetus' are used interchangeably in this paper as the study spanned the embryonic-fetal boundary period of 10 weeks gestation.

Statistical analysis

The data were randomly split into a training set (two thirds) and a test set (one third). Missing data were imputed using multiple imputation (Siddique and Harel 2009). Twenty imputed datasets were used to train logistic regression (LR) models.

A first model solely included demographic and symptom variables. A second model considered only variables measured at first scan. A third model combined both previous models. All variables were first tested for their significance in a univariable logistic regression model (LR). If necessary, variables were transformed in order to deal with non-linear variable effects. Variables with a p-value less than 0.1 were retained for multivariable logistic regression. Relations between variables were checked using Pearson and Spearman correlation coefficients. Variables resulting in unstable coefficient estimates (high variance inflation factor) were removed from the model. When making a selection between highly correlated variables, those variables which are more reliable, easy to measure and clinically more relevant, were selected. Binary interactions were investigated. After training the LR models, scoring systems were derived to improve clinical applicability and interpretability (Sullivan et al., 2004).

A ROC curve was constructed to describe the relationship between the sensitivity and the false-positive rate (1– specificity) for the use of the models in the prediction of ongoing viability. The AUC was calculated combining all imputed datasets. For the final model, a cut-off was defined at 90% of sensitivity on the training set.

The data collected has also been used for further separate analyses regarding growth rates and optimal timing of ultrasound scans (Bottomley et al., 2009a, 2009b, 2009c, 2011).

Results

During the study period 1881 women attended for early pregnancy ultrasound assessment of whom 1435 fitted the criteria for inclusion in the study analysis. The median age was 32 years (range: 15-30), the median gravidity was 2 (range: 1-12) and 866 women were white, 231 black, 149 Asian and 189 of other or unknown ethnicity. The referral indication for assessment was bleeding in 657 women and pain in 522. Of these, 299 had both bleeding and pain. 406 women were asymptomatic and presented for dating or reassurance scans. In 224 women (15.6%) no information on pain was available and in 58 women (4%) no information on bleeding was available.

The median gestational age at presentation was 50 days (interquartile range 42-62 days). At first TVS 713 women had pregnancies with an embryo demonstrating cardiac activity, 376 had intrauterine pregnancies of uncertain viability (gestation sac (GS) only: 51, GS and yolk sac (YS) (no embryo): 290, GS and embryo (no cardiac activity) <6mm: 35). 213 had a pregnancy of unknown location and the remainder had diagnosed nonviable pregnancies at first scan (111 miscarriage and 22 ectopic pregnancy)

The final outcome was an ongoing viable pregnancy in 885 and early pregnancy loss in 550 women. Of early pregnancy losses 408 were miscarriages, 34 were ectopic pregnancies and 108 were failed pregnancies of unknown location.

Prediction of miscarriage using demographics and symptoms alone

Table 2 shows the demographic and symptom variables tested for significance for the prediction of viability and the results of the multivariable analysis. The multivariable model had an AUC of 0.739 (95% CI=0.706-0.772) on the training and 0.731 (95% CI=0.685-0.778) on the test set. The scoring system derived from this model is presented in Table 3 (Panel A). For both variables in the table, different categories are defined. Depending on the category in which the variables of a patient fall, the corresponding points need to be added to obtain a score. The estimated chance of an ongoing viable pregnancy beyond the first trimester which is linked with the obtained score is then found using Table 3 (Panel B). The scoring system obtained an AUC of 0.724 (95% CI=0.692-0.756) on the training and 0.729 (95% CI=0.684-0.774) on the test set.

The following variables were highly correlated: referral indication of bleeding, current bleeding and bleeding score. Only the bleeding score was retained for further analysis since this variable contained the most information. Whether or not the patient was currently reporting pain was highly correlated with the pain being the referral indication for presentation to the unit, therefore only the variable ‘currently reporting pain’ was retained in the multivariable analysis. The referral indication of ‘uncertain dates’ and whether or not the cycles were regular

were also highly correlated, and only the latter was retained for analysis. As regular cycles and the number of previous late miscarriages did not fit with the existing literature or have obvious clinical plausibility, they were not retained in the analysis (Steyerberg 2009).

Prediction of miscarriage using ultrasound variables alone

CRL, MSD and mean yolk sac diameter, whether or not a yolk sac was seen, whether or not an embryo was seen and whether or not a heart beat were seen were all significant in a univariable model, as shown in Table 4. In a multivariable model, MSD, yolk sac diameter and whether or not a heart beat was seen, remained independent predicting factors.

An interaction between the MSD and the presence of a heart beat was introduced, since the effect of the mean gestation sac size was different depending on whether a heart beat was seen or not. CRL and MSD were highly correlated. CRL was therefore no longer considered since MSD was more often measured (CRL only being measurable in those pregnancies with a visible embryo). Table 4 presents the results of the multivariable model. The AUC for the LR model was 0.875 (95% CI=0.851-0.899) and 0.902 (95% CI=0.874-0.931) on the training and test set, respectively.

The scoring system derived from this second LR model is published in Supplementary Table 1. The scoring system resulted in an AUC of 0.873 (95% CI=0.850-0.897) and 0.900 (0.871-0.928) on the training and test set, respectively.

Predicting miscarriage using a combination of all available information

When all possible covariates (demographic information, symptoms and ultrasound findings) were combined to predict viability at the end of the first trimester, the same variables remained significant. The resulting scoring system is summarized in Table 5 (Panel A). Panel B is used to obtain the estimated chance of an ongoing viable pregnancy beyond the first trimester.

The LR model had an AUC of 0.903 (95%CI=0.883-0.923) and 0.927 (95% CI=0.904-0.950) on the training and test sets, respectively. The AUC of the scoring system derived from the model was 0.901 (95% CI=0.881-0.920) and 0.924 (95% CI=0.900-0.947) on the training and test set, respectively. Calibration of the model was excellent on training and test data.

After defining the cut-off for which the sensitivity is 0.900 on the training set, this model performed with a sensitivity of 0.922, specificity of 0.733, positive predictive value of 84.7% and negative predictive value of 85.4% in the test set.

In view of the number of women with missing variables, the model was retrained using only women where all relevant variables were available. The conclusions remained the same as in the presented analysis. The scoring system had an AUC of 0.922 (CI 0.891-0.953) in the test data set. Similarly, in order to assess whether the scoring system was likely to perform as well in an asymptomatic population, the model was retrained using only the women without pain or bleeding. The conclusions remained the same as in the presented analysis. The AUC of the final score system in this asymptomatic population was 0.954 (CI 0.9154-0.993) in the test data set.

Discussion

Our results have demonstrated that routine (but carefully defined) clinical information about the mother and her symptoms as well as some simple ultrasound derived parameters can be combined to very accurately determine an individual's likelihood of maintaining a viable pregnancy at the end of the first trimester. The combination of such factors (age, bleeding score, gestational age, mean gestation and yolk sac sizes and the presence of fetal cardiac activity) provides a more accurate prediction of viability than any of the individual factors alone.

This model is easily transferable to the clinical setting by way of the presented scoring system, where both maternal and ultrasound variables are easily available to the physician and can be used to make an accurate assessment of likelihood of viability, as shown in Table 5. For example, a woman of 36 years with a bleeding score of 1, and mean gestation sac diameter of 25mm, yolk sac diameter of 5mm and visualized fetal cardiac activity has a total score of 2, equivalent to a 91% chance of viability (or 9% chance of early pregnancy loss before 11 to 14 weeks).

The study used only variables easily available or measurable in the early pregnancy setting. The variation between reported LMP and gestation derived from ultrasound is well reported (Papaioannou et al., 2010). However we have shown in a previous study that the ability to make a diagnosis is not influenced by the certainty of dates (Bottomley et al., 2009b) and the performance of our model is good using the information provided by women at presentation. BMI and smoking variables were a potential omission in the data collection in this study and might further improve the model performance if included. A further limitation in this study is the absence of information on either bleeding or pain in 18% of women. We imputed all missing data in order to take all available information into account, as excluding cases with missing data is deemed potentially misleading (Sterne et al., 2009).

The study agrees with our current knowledge that miscarriage is more likely with increasing maternal age and in the presence of vaginal bleeding, and that it is less likely after the visualization of embryonic cardiac activity. The advantage of our model is that it incorporates the more complex interaction of variables that

experienced clinicians are aware of. For example, larger gestational sac size at presentation does not increase the likelihood of a pregnancy being viable, as shown by a more negative score allocated for increasing gestational sac size. However this is compensated for in the scoring system if there is fetal cardiac activity present and the yolk sac is appropriately sized. This point also explains the apparent anomaly that the presence of cardiac activity alone gives a negative score as it is balanced by an increasingly positive score dependent on the size of the gestation sac.

Unsurprisingly using the scoring system to predict outcome in women prior to an ultrasound scan (employing the 'pre-test' demographic and symptoms data only), performs much less well than combining both the pre-test and post-test (ultrasound) variables.

The study was undertaken when the RCOG guideline on early pregnancy loss management recommended that the diagnosis of a missed miscarriage was made when there was an embryo of greater than 6mm without visible heart beat or empty sac of greater than 20mm without embryonic structures. This has subsequently been revised and the recommended criteria have been extended to 7mm and 25mm, respectively (RCOG 2011, Abdallah et al., 2011). The number of cases near the decision boundaries are small and therefore we do not anticipate that this would have made a significant impact on our findings.

One criticism of this study might be that, as miscarriage cannot be prevented, the outcome is not changed by this prediction model. We believe that medical care for women in early pregnancy should be orientated not just to physical medical wellbeing and detection of pathological pregnancy but to the psychological care of hopeful parents to be. For many couples miscarriage is associated with significant psychological morbidity. As many as 50% of women suffer some form of psychological effect from miscarriage which may last from 6 to 12 months after the event, with a major depressive disorder reported in 10–50% (Lok and Neugebauer 2007). Elevated anxiety and depression have been observed in 45% and 15% of the 204 women diagnosed with a delayed miscarriage, respectively, with the mean score of grief being similar to that observed at the time of the loss of a close relative (Nikcevic et al. 1998). Whether early warning of probable miscarriage can modify the psychological morbidity associated with pregnancy loss needs to be assessed. In the meantime, data to support our clinical judgement of the viability of a pregnancy can only improve our patients' care.

To our knowledge this is the largest prospective study to address the prediction of miscarriage in the early pregnancy population. We have collected both maternal and ultrasound data prospectively, using predefined questions and measurements, which increases the likelihood of its accuracy. The analysis has been developed to allow its easy and simple incorporation into everyday clinical practice. As the work relates to relatively high risk women presenting largely with

pain or bleeding to an early pregnancy assessment unit, caution should be exercised before the model is employed for other populations. The results are now being prospectively tested in different early pregnancy populations, including those where women present for dating and viability rather than with predominant symptoms of bleeding or pain.

Acknowledgements
Aris Papageorghiou was instrumental in facilitating access to first trimester outcomes and review of the manuscript. Research supported by Research Council KUL: GOA MaNet, PFV/10/002 (OPTEC), several PhD/postdoc & fellow grants; IWT: TBM070706-IOTA3, PhD Grants; IBBT; Belgian Federal Science Policy Office: IUAP P7/ (DYSCO, `Dynamical systems, control and optimization', 2012-2017).
Funding
(AUTHOR: please would you recheck, and complete this Funding section (and that in the Abstract), as the Acknowledgements above suggest that funding was received? Or state 'None' if appropriate.)
Conflict of Interest
None
Authors Roles
CB collected the majority of data and drafted the manuscript. VvB performed the statistical analysis and contributed to the final manuscript, supervised by SvH. EK was involved in the initial study design, data collection and revision of the manuscript. TB and DT were involved in the initial study idea, study design, review of analysis and manuscript revisions.

References

Abdallah Y, Daemen A, Kirk E, Pexsters A, Naji O, Stalder C, Gould D, Ahmed S, Guha S, Syed S, et al. Limitations of current definitions of miscarriage using mean gestational sac diameter and crown-rump length measurements: a multicenter observational study. *Ultrasound Obstet Gynecol* 2011; 38(5):497-502.

Nybo Andersen AM, Wohlfahrt J, Christens P, Olsen J, Melbye M. Maternal age and fetal loss: population based register linkage study. *Br Med J* 2000; 320: 1708–1712.

Bigrigg MA, Read MD, Management of women referred to early pregnancy assessment unit: care and cost effectiveness. *BMJ* 1991; 302(6776): 577–579.

Bottomley C, Daemen A, Mukri F, Papageorghiou AT, Kirk E, Pexsters A, De Moor B, Timmerman D, Bourne T, Functional linear discriminant analysis: a new longitudinal approach to the assessment of embryonic growth. *Hum Reprod* 2009a; 24(2):278-83.

Bottomley C, Van Belle V, Mukri F, Kirk E, Van Huffel S, Timmerman D, Bourne T. The optimal timing of an ultrasound scan to assess the location and viability of an early pregnancy. *Hum Reprod* 2009b; 24(8):1811-7.

Bottomley C, Daemen A, Mukri F, Papageorghiou AT, Kirk E, Pexsters A, De Moor B, Timmerman D, Bourne T. Assessing first trimester growth: the influence of ethnic background and maternal age. *Hum Reprod* 2009c; 24(2):284-90.

Bottomley C, Van Belle V, Pexsters A, Papageorghiou AT, Mukri F, Kirk E, Van Huffel S, Timmerman D, Bourne T. A model and scoring system to predict outcome of intrauterine pregnancies of uncertain viability. *Ultrasound Obstet Gynecol* 2011; 37(5):588-95.

Bromley B, Harlow BL, Laboda LA, Benacerraf BR. Small sac size in the first trimester: a predictor of poor fetal outcome. *Radiology* 1991; 178(2): 375–377.

Cashner KA, Christopher CR, Dysert GA. Spontaneous fetal loss after demonstration of a live fetus in the first trimester. *Obstet Gynecol* 1987; 70: 827–830.

Choong S, Rombauts L, Ugoni A, Meagher S. Ultrasound prediction of risk of spontaneous miscarriage in live embryos from assisted conceptions. *Ultrasound Obstet Gynecol* 2003; 22: 571–577.

De la Rochebrochard E & Thonneau P, Paternal age and maternal age are risk factors for miscarriage; results of a multi- centre European study. *Hum Reprod* 2002; 17: 1649–1656.

Dickey RP, Olar TT, Taylor SN, Curole DN, Matulich EM. Relationship of small gestational sac-crown-rump length differences to abortion and abortus karyotypes. *Obstet Gynecol* 1992; 79(4): 554–557.

Ellish NJ, Saboda K, O'Connor J, Nasca PC, Stanek EJ, Boyle C. A prospective study of early pregnancy loss. *Hum Reprod* 1996; 11: 406–412.

Everett C. Incidence and outcome of bleeding before the 20th week of pregnancy: prospective study from general practice. *BMJ* 1997; 315(7099):32-4.

Falco P, Milano V, Pilu G, David C, Grisolia G, Rizzo N, Bovicelli L. Sonography of pregnancies with first-trimester bleeding and a viable embryo: a study of prognostic indicators by logistic regression analysis. *Ultrasound Obstet Gynecol* 1996; 7(3): 165–169.

Hill LM, Guzick D, Fries J, Hixson J. Fetal loss rate after ultrasonically documented cardiac activity between 6 and 14 weeks menstrual age. *J Clin Ultrasound* 1991; 19: 221–223.

Jauniaux E, Johns J, Burton GJ, The role of ultrasound imaging in diagnosing and investigating early pregnancy failure. *Ultrasound Obstet Gynecol* 2005; 25(6): 613–624.

Johns J, Hyett J, Jauniaux E, Obstetric outcome after threatened miscarriage with and without a hematoma on ultrasound. *Obstet Gynecol* 2003; 102(3): 483–487.

Lok IH & Neugebauer R. Psychological morbidity following miscarriage. *Best Pract Res Clin Obstet Gynaecol* 2007; 21(2): 229–247.

Makrydimas G, Sebire NJ, Lolis D, Vlassis N, Nicolaides KH. Fetal loss following ultrasound diagnosis of a live fetus at 6–10 weeks of gestation. *Ultrasound Obstet Gynecol* 2003; 22(4): 368–372.

Mantoni M & Pedersen JF. Fetal growth delay in threatened abortion: an ultrasound study. *Br J Obstet Gynaecol* 1982; 89: 525–527.

Mukri F, Bourne T, Bottomley C, Schoeb C, Kirk E, Papageorghiou AT. Evidence of early first-trimester growth restriction in pregnancies that subsequently end in miscarriage. *BJOG* 2008; 115(10): 1273–1278.

Nikcevic AV, Tunkel SA, Nicolaides KH. Psychological outcomes following missed abortions and provision of follow-up care. *Ultrasound Obstet Gynecol* 1998; 11(2): 123–128.

Papaioannou GI, Syngelaki A, Poon LC, Ross JA, Nicolaides KH. Normal ranges of embryonic length, embryonic heart rate, gestational sac diameter and yolk sac diameter at 6-10 weeks. *Fetal Diagn Ther* 2010; 28(4):207-19.

Papaioannou GI, Syngelaki A, Maiz N, Ross JA, Nicolaides KH. Ultrasonographic prediction of early miscarriage. *Hum Reprod* 2011; 26(7):1685-92.

Poulose T, Richardson R, Ewings P, Fox R. Probability of early pregnancy loss in women with vaginal bleeding and a singleton live fetus at ultrasound scan. *J Obstet Gynaecol* 2006; 26(8): 782–784.

Reljic M. The significance of crown-rump length measurement for predicting adverse pregnancy outcome of threatened abortion. *Ultrasound Obstet Gynecol* 2001; 17: 510–512.

Royal College of Obstetricians and Gynaecologists, Management of early pregnancy loss. Guideline number 25, RCOG Press, 2006.

Royal College of Obstericians and Gynaecologists Addendum to Clinical Guideline 25. http://www.rcog.org.uk/files/rcog-corp/Addendum%20to%20GTG%20No%2025.pdf (25 January 2012).

Savitz DA, Hertz-Picciotto I, Poole C, Olshan AF. Epidemiologic measures of the course and outcome of pregnancy. *Epidemiol Rev* 2002; 24: 91–101.

Schauberger CW, Mathiason MA, Rooney BL. Ultrasound assessment of first-trimester bleeding. *Obstet Gynecol* 2005; 105(2): 333–338.

Shillito J & Walker JJ. Early pregnancy assessment units. *Br J Hosp Med* 1997; 58(10): 505–509.

Siddiqi TA, Caligaris JT, Miodovnik M, Holroyde JC, Mimouni F. Rate of spontaneous abortion after first trimester sonographic demonstration of fetal cardiac activity. *Am J Perinatol* 1988; 5(1): 1–4.

Siddique, J. & Harel, O. MIDAS: A SAS Macro for Multiple Imputation using Distance-Aided Selection of Donors. *Journal of Statistical Software* 2009; 29:1-18.

Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; 338:b2393

Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation and Updating. Statistics for Biology and Health. Springer, New York (2009).

Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med* 2004; 23(10):1631-60.

Trinder J, Brocklehurst P, Porter R, Read M, Vyas S, Smith L. Management of miscarriage: expectant, medical, or surgical? Results of randomised controlled trial (miscarriage treatment (MIST) trial). *BMJ* 2006; 332(7552): 1235–1240.

Zinaman MJ, Clegg DE, Brown CC, O'Connor J, Selevan SG. Estimates of human fertility and pregnancy loss. *Fertil Steril* 1996; 65: 503–509.

Tables

Abbreviations:
CRL: crown-rump length

Table 1: Criteria for the classification of findings at transvaginal ultrasound assessment

Viable intrauterine pregnancy	An intrauterine gestational sac containing an embryonic/ fetal pole with visible cardiac activity
Ectopic pregnancy	An empty endometrial cavity with: (i) an inhomogeneous adnexal mass or (ii) an extrauterine gestational sac with or without a yolk sac and/or fetal pole *(Condous et al,, 2005)*
Miscarriage	(i)An intrauterine gestational sac containing a embryonic/fetal pole with a CRL of greater than 6 mm with no cardiac activity Or (ii) An empty intrauterine gestational sac with a mean diameter greater than 20 mm Or (iii)The presence of heterogeneous, hyperechoic irregular tissues within the uterine cavity *(Farquharson et al,, 2005),*
Intrauterine pregnancy of uncertain viability	(i) An empty intrauterine gestational sac measuring less than 20 mm in diameter Or (ii) an intrauterine gestational sac containing an embryonic pole with a CRL less than 6 mm with no cardiac activity *(RCR/RCOG guideline 1995)*
Pregnancy of unknown location	An empty endometrial cavity, with no evidence of an intrauterine gestational sac or retained products of conception and no visualised extra-uterine pregnancy *(Condous et al,, 2005, RCOG 2006)*

Table 2: Pre-test (demographic and symptom) variables. Results of univariable and multivariable logistic regression on the training set. The grey areas indicate the variables which were considered for multivariable logistic regression. Positive coefficient ratios and odds ratios > 1 signify a higher chance of pregnancy viability.

	Univariable analysis				Multivariable analysis			
Variable	coefficient Estimate	Odds ratio	CI1	CI2	coefficient Estimate	Odds ratio	CI1	CI2
Maternal age[1]	-0.212	0.809	0.75	0.872	-0.204	0.815	0.752	0.884
Gestational age by last menstrual period date	0.002	1.002	0.992	1.011				
Gravida	-0.012	0.988	0.918	1.063				
Regular cycles and certain dates[2]	-0.503	0.605	0.463	0.79				
Previous ectopic pregnancy[2]	-0.212	0.809	0.416	1.571				
Previous delivery after 24 weeks[2]	0.014	1.014	0.781	1.317				
Previous early miscarriages[2]	-0.132	0.876	0.664	1.156				
Previous late miscarriages[2]	1.03	2.801	1.051	7.464				
Previous pregnancy of unknown location[2]	-0.477	0.621	0.039	9.955				
Previous termination of pregnancy[2]	0.211	1.235	0.907	1.68				
Reported indication for scan: Anxiety[2]	0.87	2.386	1.172	4.859				
Reported indication for scan: Bleeding[2]	-1.554	0.211	0.16	0.28				
Reported indication for scan: Uncertain dates[2]	1.257	3.516	2.477	4.992				
Reported indication for scan: Previous ectopic, pregnancy of unknown location[2]	-0.195	0.822	0.416	1.628				
Reported indication for scan: Other[2]	-0.017	0.983	0.472	2.05				
Reported indication for scan: Pain[2]	-0.285	0.752	0.579	0.977				
Reported indication for scan: Previous miscarriage[2]	0.466	1.593	1.032	2.459				
Presence of any bleeding[2]	-1.534	0.216	0.161	0.288				
Bleeding score	-1.034	0.356	0.294	0.431	-1.031	0.357	0.294	0.433
Presence of any pain[2]	-0.264	0.768	0.591	0.998				
Visual analogue pain score	-0.043	0.958	0.906	1.014				
Type of pain: Backpain[2]	-0.072	0.931	0.261	3.321				
Type of pain: Left iliac fossa[2]	0.506	1.659	0.901	3.054				
Type of pain: Right iliac fossa	0.611	1.842	0.887	3.826				
Type of pain: Lower abdomen[2]	-0.178	0.837	0.583	1.2				
Type of pain: Upper abdomen [2]	-0.161	0.852	0.339	2.137				
Type of pain: Period-type pain[2]	-0.566	0.568	0.41	0.786				

[1] Since no difference in effect was seen for ages up to 35 years, all ages less than 35 years were made equal to 35.
[2] Binary variable
CI1: lower 95% confidence interval, CI2: upper 95% confidence interval

Table 3 (Panel A): Pre-test (demographic and symptom) variables. Scoring system derived from multivariable logistic regression model.

Variable	Categories	Points
Maternal age (years)		
	<35	0
	35-39	-1
	>=40	-2
Bleeding score		
	0	0
	1	-1
	2	-2
	3	-3
	4	-4

Table 3 (Panel B): Pre-test (demographic and symptom) variables. Link between the points of the scoring system and the estimated chance of ongoing pregnancy viability beyond the first trimester

Sum of Points	Estimated chance of a viable pregnancy beyond the first trimester
0	0.875
-1	0.717
-2	0.477
-3	0.247
-4	0.106
-5	0.041
-6	0.015

Table 4: Post-test (ultrasound) variables. Results of multivariable logistic regression on the training set. The grey areas indicate the variables which were considered for multivariable logistic regression. Positive coefficient ratios and odds ratios > 1 signify a higher chance of pregnancy viability

	Univariable analysis				Multivariable analysis			
Variable	**coefficient Estimate**	**Odds ratio**	**CI1**	**CI2**	**coefficient Estimate**	**Odds ratio**	**CI1**	**CI2**
Crown rump length	0.139	1.149	1.119	1.180				
Fetal heart beat visualised	2.490	12.066	8.668	16.797	-1.93	0.145	0.058	0.366
Presence of a yolk sac	1.695	5.445	3.233	9.170				
Presence of a fetus	1.734	5.664	4.254	7.541				
Mean yolk sac diameter[1]	-0.355	0.701	0.620	0.793	-0.401	0.67	0.509	0.88
Mean gestational sac diameter	0.083	1.086	1.071	1.102	0.196	1.217	1.084	1.366
Mean gestational sac diameter[2]					-0.016	0.984	0.978	0.99
Mean gestational sac diameter[2]* presence of fetal heart beat [3]					0.016	1.016	1.012	1.021

[1] Due to a strong non-linearity of the yolk sac diameter, the difference between the mean yolk sac diameter and 5 was used.
[2] The mean gestational sac diameter to the power of two is considered here
[3] Interaction effect between gestational sac size and presence of fetal heart beat

Table 5 (Panel A): Final scoring system using pre- and post-test (demographic, symptom and ultrasound) variables combined, derived from the multivariable logistic regression model.

Variable	Category	Points
Maternal age (years)		
	<35	**0**

	35-39	**-2**
	>=40	**-4**
Bleeding score		
	0	**0**
	1	**-3**
	2	**-5**
	3	**-8**
	4	**-10**
Mean gestational sac size		
	0	**0**
	0.01-4.99	**1**
	5.00-9.99	**3**
	10.00-14.99	**5**
	15.00-19.99	**7**
	20.00-24.99	**9**
	25.00-29.99	**11**
	30.00-34.99	**13**
	35.00-39.99	**15**
	40.00-44.99	**17**
	45.00-49.99	**19**
	>=50.00	**23**
Presence of a fetal heart beat		
	No	**0**
	Yes	**-4**
Mean yolk sac diameter		
	<1	**-3**
	1.00-1.99	**-2**
	2.00-2.99	**-1**
	3.00-4.99	**0**
	5.00-5.99	**-1**
	6.00-6.99	**-2**
	>=7.00	**-4**
In the absence of a fetal heart beat (only allocate a score here for patients in whom a fetal heart beat is NOT visualised) :		
Mean gestational sac size		
	<5	**0**
	5.00-9.99	**-2**
	10.00-14.99	**-5**
	15.00-19.99	**-10**
	20.00-24.99	**-17**
	25.00-29.99	**-25**
	30.00-34.99	**-35**
	35.00-39.99	**-47**
	40.00-44.99	**-61**

	45.00-49.99	**-76**
	>=50.00	**-109**
In the presence of a fetal heart rate (only allocate a score here for patients in whom a fetal heart beat is visualised):		
Mean gestational sac size		
	<20	**0**
	20.00-29.99	**1**
	30.00-39.99	**2**
	40.00-49.99	**3**
	>=50.00	**5**

Table 5 (Panel B): Final scoring system using pre- and post-test (demographic, symptom and ultrasound) variables combined. Link between the points of the scoring system and the estimated chance of ongoing pregnancy viability beyond the first trimester.

Sum of points	Estimated chance of a viable pregnancy beyond the first trimester
-12	0.042
-11	0.060
-10	0.087
-9	0.123
-8	0.172
-7	0.234
-6	0.311
-5	0.400
-4	0.497
-3	0.593
-2	0.683
-1	0.761
0	0.825
1	0.874
2	0.911
3	0.938
4	0.957
5	0.971
6	0.980
7	0.986
8	0.991
9	0.994
10	0.996
11	0.997
12	0.998

Supplementary material

Supplementary Table 1 (Panel A): Post-test (ultrasound) variables. Scoring system derived from multivariable logistic regression model.

Variable	Categories	Points
Mean gestational sac diameter		
	0	0
	0.01-4.99	1
	5.00-9.99	3
	10.00-14.99	5
	15.00-19.99	7
	20.00-24.99	9
	25.00-29.99	11
	30.00-34.99	13
	35.00-39.99	15
	40.00-44.99	17
	45.00-49.99	19
	>=50.00	23
Fetal heart beat present		
	no	0
	yes	-4
Mean yolk sac diameter		
	0	-3
	0.01-0.99	-2
	1.00-1.99	-2
	2.00-2.99	-1
	3.00-3.99	0
	4.00-4.99	0
	5.00-5.99	-1
	6.00-6.99	-2
	>=7.00	-3
In the absence of a fetal heart beat (no score allocated in this category if a fetal heart beat is visualised):		
Mean gestational sac size		
	<5	0
	5.00-9.99	-2
	10.00-14.99	-5
	15.00-19.99	-10

	20.00-24.99	-17
	25.00-29.99	-25
	30.00-34.99	-35
	35.00-39.99	-47
	40.00-44.99	-60
	45.00-49.99	-75
	>=50.00	-107

Supplementary Table 1 (Panel B): Post-test (ultrasound) variables. Link between score obtained when using the scoring system and the estimated chance of ongoing pregnancy viability beyond the first trimester

Sum of points	Estimated chance of a viable pregnancy beyond the first trimester
5	0.943
4	0.911
3	0.862
2	0.793
1	0.701
0	0.590
-1	0.468
-2	0.351
-3	0.249
-4	0.168
-5	0.110