Title page

Title

Role of serum biomarkers in the prediction of outcome in women with threatened miscarriage - A systematic review and diagnostic accuracy meta-analysis

Running title

Biochemical markers and threatened miscarriage

Authors

Rekha N Pillai\textsuperscript{1,2}, Justin C Konje\textsuperscript{2,3}, Douglas G Tincello\textsuperscript{1,2}, *Neelam Potdar\textsuperscript{1,2}

\textsuperscript{1}Women's Health CMG, University Hospitals of Leicester NHS Trust, Leicester, United Kingdom, LE1 5WW

\textsuperscript{2}University of Leicester, Leicester, United Kingdom, LE1 7RH

\textsuperscript{3}Department of Obstetrics and Gynecology, Sidra Medical and Research Center, PO Box 26999, Doha, Qatar

Corresponding author: Neelam Potdar

Email: saphaire_neelam@yahoo.com
Table of Contents

Introduction

Materials and Methods

Study eligibility criteria

Information sources and search strategy

Data extraction and quality assessment

Statistical analysis

Result

Study selection

Study characteristics

Risk of bias assessment

Quantitative data summary and synthesis of results

Serum hCG

Serum progesterone

Serum estradiol

Serum PAPP-A

Serum CA 125

Discussion
Abstract

Background: Threatened miscarriage affects one in five women and is associated with significant emotional distress. The uncertainty around the prognosis of threatened miscarriage makes it equally challenging to the healthcare professionals. Various biochemical markers have been investigated in the past to predict the outcome of threatened miscarriage, however the results have been conflicting. Therefore, we have conducted a systematic review and meta-analysis to determine the diagnostic accuracy of biochemical markers in predicting the outcome in women presenting with threatened miscarriage.

Methods: This is a systematic review and meta-analysis of prospective studies that investigated biochemical markers to determine outcomes for women with threatened miscarriage between 5-23 weeks gestational age. Electronic databases were searched up to June 2015 and study quality assessment was performed using QUADAS-2 (Quality Assessment for Diagnostic Accuracy Studies-2: A Revised Tool) for evaluating the diagnostic accuracy studies. Statistical analysis was performed using the Cochrane systematic review software (Review Manager 5.3) and Stata version 13.0.

Results: A total of 19 studies were included in the qualitative data synthesis of which 15 were eligible for the meta-analysis. The review highlights the role of biochemical markers (serum progesterone, hCG, PAPP-A, estradiol and CA 125) in the prediction of outcome in women with threatened miscarriage. Interestingly, serum CA 125 appears to be the most promising marker (n=648 women in seven studies), whereas serum progesterone and hCG are less useful once fetal viability is established. The summary receiver operator characteristics for CA125 showed sensitivity of 90% (95% CI 83-94%), specificity of 88% (95% CI 79-93%), positive likelihood ratio of 7.86 (95% CI 4.23-14.60) and negative likelihood ratio of 0.10 (95% CI 0.06-0.20). The inverse of negative likelihood ratio was 9.31 (95% CI 5-17.1) indicating that a negative test is likely to identify those who are likely to continue with the pregnancy. Serum estradiol was the next best marker with a sensitivity of 45% (95% CI 6-90%), a specificity of 87% (95% CI 81-
92%), a positive likelihood ratio of 3.72 (95% CI 1.01-13.71) and a negative likelihood ratio of 0.62 (95% CI 0.20-1.84).

**Conclusion:** In women with threatened miscarriage, serum CA125 has high predictive value in identifying pregnancies that are ‘likely to continue’; whereas, the most commonly used biomarkers of serum hCG and progesterone are not useful in predicting outcome of a pregnancy with a viable fetus. Other markers such as inhibin A and a combination of markers need to be investigated to hopefully improve the prediction of outcome in women with threatened miscarriage.

Keywords: miscarriage/biomarkers/threatened/meta-analysis/outcome
Introduction

Miscarriage is the most common early pregnancy complication affecting 20% of recognized pregnancies (Savitz et al., 2002; NICE Guideline CG 154, 2014). Threatened miscarriage is diagnosed when the woman presents in early pregnancy with vaginal bleeding, a closed cervix on clinical examination and subsequent ultrasound scan demonstrates fetal cardiac activity (Saraswat et al., 2010; NICE Guideline CG 154). It is reported to occur in about one-fifth of pregnancies (Everett, 1997) but an estimated 3-16% of these subsequently miscarry (Cashner et al., 1987; Hill et al., 1991; Makrydimas et al., 2003; Siddiqi et al., 1988). Women presenting with threatened miscarriage are often extremely distressed and providing care can be challenging to the health care professionals, more so since it is difficult to provide reasonable information on the potential outcome. These women end up with repeated scans in early pregnancy units to allay their anxieties, which in turn adds to the increase in waiting times and costs. In the presence of reliable predictive biomarkers, the above challenges can be mitigated and potentially new therapeutics can be directed at those identified at an increased risk of miscarriage.

Various biochemical markers have been studied to establish if they are able to predict the outcome of threatened miscarriage (i.e. identify those at risk of subsequent miscarriage), however results have been conflicting. Some of the commonly studied biochemical markers are serum human chorionic gonadotrophin (hCG), progesterone, estradiol, pregnancy associated plasma protein A (PAPP-A), cancer antigen 125 (CA 125), human placental lactogen (HPL), alpha fetoprotein (AFP), inhibin A, follistatin and activin A (Johns et al., 2007; Maged et al., 2013; Ruge et al., 1990; Scarpellini et al., 1995; Vavilis et al., 2001; Westergaard et al., 1985). In view of the conflicting evidence, we performed a systematic review and meta-analysis to determine which biochemical markers have high diagnostic accuracy to predict the outcome of threatened miscarriage either singly or in combination.
Materials and Methods

Study eligibility criteria

The inclusion criteria for the systematic review were all prospective studies with use of biochemical markers to determine outcomes for women with threatened miscarriage and gestational age between 5-23 weeks. Exclusion criteria were retrospective studies, case reports, case series, letters, and reviews; studies that did not include women in the period 5-23 weeks; studies with infertility, recurrent miscarriage or pregnancy of unknown location (PUL) cohorts or where women had ovulation induction medications, exogenous hormones or any form of treatment for prevention of miscarriage. Studies in languages other than English were also excluded where no translated version of the manuscript was available.

Threatened miscarriage was defined as patients presenting with bleeding with or without lower abdominal pain, closed internal os on cervical examination and subsequent ultrasound scan (USS) confirming a viable intra uterine pregnancy (Saraswat et al., 2010; NICE guidance CG154, 2014). Based on this definition of threatened miscarriage, studies that included women with pregnancy viability confirmed on an initial ultrasound scan were selected for the systematic review. The primary outcome of interest was prediction of miscarriage.

Information sources and search strategy

Electronic databases search included Medline (1946 to June 2015), Embase (1980 to June 2015), Cochrane library, ClinicalTrials.gov, World Health Organization international clinical trials registry, LILAC database and OpenGrey (System for Information on grey literature from Europe). The following MESH terms we used to create two subset of citations (1) miscarriage (abortion, pregnancy loss, early pregnancy outcome) (2) biochemical markers (biomarkers, biological markers, hormonal markers, progesterone, β hCG, hCG, human chorionic gonadotrophin, progesterone, follistatin, CA 125, PAPP-A, activin, activin- A, inhibin, inhibin-A, estradiol, estriol, hydroxy progesterone, human placental lactogen, HPL, alpha fetoprotein
(AFP), schwangerschaft protein (SP1), pregnancy specific beta 1 glyco protein, pregnancy zone protein (PZP)). The two subsets were combined using the Boolean term 'AND' to obtain a subset of citations relevant to our research question. Two authors (RP and NP) performed independent literature searches and the reference lists of all recent reviews and primary articles were examined to identify any articles not captured by the search. Any disagreements in selecting the papers and data extraction were resolved by consensus.

**Data extraction and Quality assessment**

Using predetermined forms, data were extracted independently by 2 authors (RP and NP). Data were collected on study design and conduct, country of study, sample size, gestational age, biochemical markers and miscarriage prediction. From each study, outcome data were extracted in $2 \times 2$ tables or using the mean and standard deviation.

Study quality assessment was performed using QUADAS-2 (Quality Assessment for Diagnostic Accuracy Studies-2: A Revised Tool) for evaluating the diagnostic accuracy of studies (Whiting et al., 2011). The tool consists of four key domains covering patient selection, index test(s), reference standard, the flow and timing. Each domain was assessed in terms of risk of bias, and the first 3 domains were also assessed for concerns regarding applicability. Signalling questions were included in the tool to help judge the risk of bias. The index test(s) for the included studies were the biomarkers and reference standard was miscarriage confirmed clinically or by ultrasound scan or by histopathological examination during follow up.

**Statistical analysis**

Statistical analysis was performed using the Cochrane systematic review software (Review Manager 5.3) and the meta-analysis of the eligible studies performed using the diagnostic test accuracy review stream (Cochrane Collaboration 2011). Data from each primary study were summarized in a $2 \times 2$ table of test results and forest plots constructed showing within-study estimates and confidence interval for sensitivity and specificity of each biomarker. For biomarkers with data from four or more studies, further statistically rigorous modeling was
performed using hierarchical summary receiver operator characteristic model (HSROC) and graphs plotted (Harbord \textit{et al.}, 2007; Rutter and Gatsonis 2001). The graph demonstrated summary receiver operating characteristic (SROC) curve and the prediction region, the summary point and the confidence region. The between study heterogeneity was accounted for in the HSROC model. Posterior predictions (empirical Bayes estimates) of the sensitivity and specificity in each study were obtained and plotted since the empirical Bayes estimates give the best estimate of the true sensitivity and specificity in each study. In addition, sensitivity, specificity, positive and negative likelihood ratio for each biomarker were tabulated.

**Results**

**Study selection**

The electronic searches identified 6727 articles and further 93 articles were found from other sources and review of reference lists of individual manuscripts. After reviewing the titles and removing the duplicates 154 manuscripts were identified, of which 119 were excluded after reading the abstract. Full manuscripts of 35 articles were reviewed in detail and of these 16 studies were excluded (patient population was different in 11 studies, four studies were excluded for retrospective study design and one excluded due to data duplication). A total of 19 studies were included in the qualitative data synthesis. Four studies (Azogui \textit{et al.}, 1996, Johns \textit{et al.}, 2007; Vavilis \textit{et al.}, 2001, Jauniaux \textit{et al.}, 2015) were further excluded in the quantitative meta-analysis as the data could not be obtained for the 2 x 2 tables. Overall, 15 studies were eligible for the quantitative meta-analysis and included 1,263 women (Figure 1). Of the included studies, only one study had results from use of combination markers (Scarpellini \textit{et al.}, 1975); all other studies which used combination markers, could not be included in the review because they did not meet the predefined inclusion criteria (Hertz \textit{et al.}, 1983; Kunz and Keller 1976; Osmanagaoglu \textit{et al.}, 2010) (Table II).

**Study characteristics**
All included studies were prospective cohort studies on women with threatened miscarriage. Of the 15 studies, all excepting four (Hanita et al., 2012, Jouppila et al., 1980; Ruge et al., 1990; Westergaard et al., 1985) included women of gestational age less than 14 weeks. The characteristics of the included studies are summarized in Table 1 and excluded studies in Table 2.

Risk of bias assessment

The risk of bias was assessed in 4 main domains using the 'QUADAS-2: A Revised Tool' for patient selection, index test, reference standard and flow and timing (Figure 2). Few of the studies reviewed did not specify their exclusion criteria and therefore scored ‘high risk’ for patient selection. For the index test most studies had not specified a cut off level to differentiate between ongoing pregnancies and miscarriage, and those that had specified a cut off level had not specified it prior to starting the project. This is an area of major bias for the included studies. The reference standard for the review is occurrence of miscarriage which can be best diagnosed using USS or clinical history followed by histopathological examination of the products of conception. Some studies used telephone interviews or review of case notes to determine the outcome, which can contribute to bias. It was not clearly stated in the studies whether the reference standard was interpreted without the knowledge of the index test. However, this is unlikely to affect applicability of the studies since miscarriage is an objective diagnosis and is not prone to subjective interpretation. In the flow and timing section of the QUADAS-2 tool, although it was difficult to predict a specific time interval from the index test to reference standard (occurrence of miscarriage), we used the sampling question to see whether the patients were followed up until at least 23 weeks, so as not to miss any miscarriages (World Health Organization (WHO, 2001) has defined miscarriage as premature loss of a fetus up to 23 weeks of pregnancy and weighing up to 500 g). Therefore, quality concerns exist for the diagnostic accuracy studies included for the prediction of the miscarriage.

Quantitative data summary and synthesis of results
Data were summarized for the biomarkers serum hCG, progesterone, estradiol, PAPP-A and CA 125. Test results were tabulated in a 2 x 2 table and forest plots constructed for the sensitivity and specificity of the biomarker with their confidence intervals. There were other serum biomarkers for which only single studies were available and therefore meta-analysis could not be performed. These were HPL, AFP, Schwangerschafts Protein 1 (SP1) and Pregnancy Zone Protein (PZP) (Westergaard et al., 1985); Plasma Renin Activity (PRA), Plasma Renin Substrate (PRS) and Sex-Hormone Binding Globulin (SHBG) (Siimes et al., 1983); inhibin A, activin A, follistatin (Johns et al., 2007, Phupong and Hanprasertpong, 2011) and estriol (Dessaive et al., 1982).

**Serum human chorionic gonadotrophin**

There were 8 studies with a total of 584 women that investigated either intact hCG (International Federation of Clinical Chemistry denotes intact hCG as ‘HCG’) (Stenman et al., 2006) or beta hCG to predict the outcome in women with threatened miscarriage. Of these, 3 studies used intact hCG (Siimes et al., 1983; Stopelli et al., 1981; Westergaard et al., 1985) and five used β hCG (Dessaive et al., 1982; Jouppila et al., 1980; Leylek et al., 1997; Maged et al., 2013; Scarpellini et al., 1975). The forest plots were plotted separately for studies that used β hCG and intact hCG (Figure 3a and 3b). Further analysis using HSROC (β hCG and intact hCG) showed a sensitivity of 44% (95% CI 17-75%), a specificity of 86% (95% CI 80-91%), a positive likelihood ratio of 3.37 (95% CI 1.98-5.74%) and a negative likelihood ratio of 0.63 (95% CI 0.36-1.11) (Table 3 and Figure 3c).

**Serum Progesterone**

Six studies with 481 women used serum progesterone to predict outcome in threatened miscarriage (Jouppila et al., 1980; Stopelli et al., 1981; Dessaive et al., 1982; Westergaard et al., 1985; Leylek et al., 1997 and Maged et al., 2013) (Supplement Figure 1a). Further analysis using HSROC showed a sensitivity of 30% (95% CI 2-87%), a specificity of 86% (95% CI 78-
91%), a positive likelihood ratio of 2.24 (95% CI 0.32-15.80%) and a negative likelihood ratio of 0.81 (95% CI 0.35-1.86) (Table 3 and Supplement Figure 1b).

**Serum estradiol**

Four studies, with 244 women investigated serum estradiol to predict outcome in women with threatened miscarriage (Stopelli *et al.*, 1981; Dessaive *et al.*, 1982; Siimes *et al.*, 1983 and Westergaard *et al.*, 1985) (Supplement Figure 2a). Further analysis using HSROC showed a sensitivity of 45% (95% CI 6-90%), a specificity of 87% (95% CI 81-92%), a positive likelihood ratio of 3.72 (95% CI 1.01-13.71) and a negative likelihood ratio of 0.62 (95% CI 0.20-1.84) (Table 3 and Supplement Figure 2b).

**Serum PAPP-A**

Three studies with 236 women studied PAPP-A to predict miscarriage (Hanita *et al.*, 2012; Ruge *et al.*, 1990; Westergaard *et al.*, 1985). PAPP-A had a poor and wide sensitivity that ranged from 25-64% but a high specificity ranging from 88-94% (Supplement Figure 3).

**Serum CA 125**

Seven studies with 648 women investigated the accuracy of CA 125 in predicting miscarriage in women with threatened miscarriage (Fiegler *et al.*, 2003; Leylek *et al.*, 1997; Maged *et al.*, 2013; Ocer *et al.*, 1992; Scarpellini *et al.*, 1995; Sherif *et al.*, 2000; Xie *et al.*, 2013) (Figure 4a). Further analysis using HSROC showed a sensitivity of 90% (95% CI 83-94%), a specificity of 88% (95% CI 79-93%), a positive likelihood ratio of 7.85 (95% CI 4.23-14.60) and a negative likelihood ratio of 0.10 (95% CI 0.05-0.20) (Table 3 and Figure 4b). The inverse of the negative likelihood ratio was 9.31 (95% CI 5-17.1) indicating that a negative test is likely to identify those who are likely to continue with the pregnancy. Empirical Bayes estimate gives the best estimate of the true sensitivity and specificity in each study and the estimates are shrunk towards the summary point compared to the study specific estimates (Figure 4b). Figure 4a shows 0 value in false negative group by Ocer *et al.*, therefore sensitivity analysis was performed after adjusting the
values for all cells and similar effect estimates were obtained. The confidence interval for the
estimates of sensitivity and specificity are not symmetric, therefore log odds scale was used and
similar results were obtained.

Further sensitivity analysis was done after excluding the study with a higher miscarriage rate
(Stopelli et al., 1981), however, there were no significant differences noted in the prediction
parameters for the biomarkers of hCG, serum progesterone and estradiol. The shape of the
prediction region on the SROC plots indicates between study heterogeneity, which was
considerable.

**Discussion**

To the best of our knowledge, this is the first systematic review of various serum biochemical
markers for predicting the outcome of threatened miscarriage. This review has highlighted that
biochemical marker of serum progesterone, hCG, PAPP-A, estradiol and CA 125 have been
studied in the prediction of outcome in women with threatened miscarriage. Interestingly,
serum CA 125 is the most reliable marker for predicting the outcome of threatened miscarriage
(sensitivity of 90%, specificity of 88%, positive likelihood ratio of 7.85 and negative likelihood
ratio of 0.10) (Table 3, Figure 4a, 4b).

In this review, the positive likelihood ratio for CA 125 is closer to 10 indicating that this could be
an accurate test. The only negatively reported study for CA125 (Vavilis et al., 2001) was not
included in the meta-analysis as the results were presented using the statistical tool of mean
and standard deviation. Furthermore, this study included only 39 women compared to this
meta-analysis, which has an aggregate sample size of 648. In view of this sample size difference
it is likely that the absence of the study will not alter the results significantly. It has been shown
that the chorio-decidual plate produces large amount of CA 125 in early pregnancy and with the
tropho-decidual detachment at the time of miscarriage, CA 125 is released into the blood stream
(Check et al., 1990; Hornstein et al., 1987; Scarpellini et al., 1995). The caveat is that CA 125 is a
non-specific biochemical marker of cellular activation of mesothelial derived tissues (Scarpellini et al., 1995), therefore its utility as a predictor of miscarriage can become questionable. More so, in the infertility population in the presence of an endometrioma it would not be a reliable marker. In women who have had in vitro fertilization conception, again CA 125 can be raised in presence of ovarian hyperstimulation, therefore interpretation can be difficult in these cases. In clinical practice CA125 is often used as a tumor marker in the presence of ovarian cysts in pregnancy therefore, this should be interpreted with caution because of the additional source of CA125 from the chorio-decidual plate.

Though several pregnancy hormones have been proposed as useful diagnostic markers for early pregnancy, hCG, the earliest detectable marker, is still the mainstay of modern pregnancy diagnosis. hCG can be detected as early as 8-11 days following ovulation (i.e. shortly after implantation) (Carmona et al., 2003). The level of hCG in blood increases rapidly with a maximum level of 50,000-1,000,000 IU/ml attained at about 8-10 weeks of gestation. The consistent nature of this pattern has made quantitative determinations of hCG a valuable tool in the clinical assessment of early pregnancy abnormalities (Duan et al., 2011). hCG is a glycoprotein with a nonspecific α subunit, which is similar to LH and FSH and has a specific β subunit which is unique to it (Stenman et al., 2006). Hence some studies have used β hCG subunit for early pregnancy prognosis (Dessaive et al., 1981; Jouppila et al., 1980; Leylek et al., 1997; Maged et al., 2013; Scarpellini et al., 1983) and others have used intact hCG for early pregnancy prognosis (Siimes et al., 1983; Stopelli et al., 1981; Westergaard et al., 1985).

However, it is proven that the measurement of free β hCG subunits offer no clinical advantage over measurement of intact hCG during the first half of pregnancy (Thomas et al., 1990). In this meta-analysis, therefore, studies on βhCG and intact hCG were combined to create a single SROC curve (Figure 3c).

Lin and Liu (1995) found that the sensitivity of estradiol and hCG in predicting pregnancy outcome at week eight of gestation was better than that of serum progesterone (80 and 85%,
respectively vs. 56%). We found similar results (Supplement Figure 1a and Figure 1a, 1b) but there was significant heterogeneity among the reported studies with regards to the sensitivity of estradiol (Dessaive et al., 1982; Westergaard et al., 1985).

Serum progesterone maintains a crucial role in the maintenance of pregnancy via the inhibition of oxytocin induced myometrial activity and prostaglandin excitation. Johansson et al. (1969) were the first to demonstrate that abnormal early gestation had lower progesterone concentrations than those of viable intrauterine pregnancies. Despite these observations because of the large biological variability of serum progesterone in early pregnancy, choosing a discriminatory value to predict viable and nonviable pregnancy is difficult (Williams et al., 1992). In a systematic review conducted in the pregnancy of unknown location (PUL) population, Verhaegen et al. (2012) determined serum progesterone cut off value of 3.2-6 ng/ml to differentiate between viable and non-viable pregnancies. Most of the studies included in their review had not specified a cut off level for serum progesterone except that of Leylek et al. (1997). In our review also, significant heterogeneity was noted among studies using serum progesterone to predict miscarriage. We noted that older studies (Jouppila et al., 1980; Dessaive et al., 1982) generally had lower sensitivity levels compared to recent studies (Leylek et al., 1997; Maged et al., 2013). Overall, the results of this meta-analysis illustrate that once fetal cardiac activity is demonstrated, serum progesterone and hCG have lower diagnostic accuracy compared to other markers.

Westergaard et al. (1983) were the first to evaluate PAPP-A in the prediction of pregnancy outcome in women presenting with a threatened miscarriage. They concluded that PAPP-A measurement might be useful in differentiating pregnancies that will have normal outcome from those which will not (Westergaard et al., 1983). The abnormal levels were frequently observed weeks before the clinical progression of spontaneous miscarriage while the fetus was still alive (Westergaard et al., 1983). Ruge et al. (1990) observed that serum levels of PAPP-A were significantly lower in women with vaginal bleeding in early pregnancy than normal
pregnant women, however, they failed to differentiate between those who either later miscarried or continued with their pregnancy. PAPP-A as a biochemical marker for the prediction of early pregnancy outcome has certain limitations, which include (1) inability to differentiate between normal and abnormal pregnancies at very early gestation (< 6-7 weeks) (Ruge et al., 1990; Yovich et al., 1986) and (2) ethnic variation of serum concentrations (Leung et al., 2006; Spencer et al., 2000).

There is an extensive list of biomarkers that have been investigated for the prediction of early pregnancy outcome but these were not included in this meta-analysis as the studies did not meet the eligibility criteria. Some of these are activin A (Florio et al., 2007; Kürk et al., 2009; Muttukrishna et al., 2002; Warrick et al., 2012), maternal serum angiogenic factors like placental growth factor, vascular endothelial growth factor and soluble endoglin (Muttukrishna et al., 2011; Ugurlu et al., 2009; Senapati et al., 2013), macrophage inhibitory growth factor (Tong et al., 2012), endocannabinoids (Habayeb et al., 2008; Taylor et al., 2011), cytokine and chemokines (Hannan et al., 2014). Johns et al. (2007) studied inhibin A, activin A, hCG, PAPP-A and follistatin in a threatened miscarriage population. They showed significantly lower concentrations of inhibin A, PAPP-A and hCG in those who had first trimester miscarriage compared to those who had term pregnancies. We could not include this study in our meta-analysis as the results were expressed as mean and standard deviation. Furthermore, although combination of biomarkers may give higher predictive value, there was only one study (Scarpellini et al., 1975) that used combination markers of serum CA125 and hCG with sensitivity of 78.9% and specificity of 96.5%. All other studies that used combination markers did not meet the inclusion criteria for the review (Hertz et al., 1983; Kunz and Keller 1976; Osmanagaoglu et al., 2010) (Table II).

We used demonstrable fetal heartbeat on USS as a strict inclusion criterion for the studies because prediction of miscarriage will benefit this population the most. Similarly, PUL population was excluded, as undiagnosed ectopic pregnancies could skew the outcomes.
There are a few limitations for this meta-analysis. Most of the included studies had not specified a cut off value for the specific biochemical marker in the prediction of the outcome of miscarriage. Because of this drawback, we could only comment on the utility of each biochemical marker in predicting miscarriage and could not determine a useful 'cut off level'. Also it is known that the levels of serum progesterone, hCG, CA 125, estradiol and serum PAPP-A change with each week of gestation. Most of the included studies did not take this into consideration. Ideally, the levels of these biochemical markers should be compared against gestation specific normal values.

Another drawback is the quality and reporting of the included studies. The STARD checklist (Bossuyt et al., 2003) for reporting of the diagnostic accuracy studies was published in 2003. Most of the studies included here were published before 2003 except for four (Hanita et al., 2012; Maged et al., 2013; Phupong and Hanprasertpong, 2011; Xie et al., 2013). The older studies have missing information and have an inadequate reporting format. Nevertheless, it is interesting that even recently published studies have pitfalls in their reporting format. The difference in reporting statistics prevented us from including some of these studies in the meta-analysis (Azogui et al., 1996, Johns et al., 2007; Vavilis et al., 2001).

In conclusion, biochemical markers can be used to predict the outcome of threatened miscarriage particularly serum CA125. Recently, high-sensitivity C-reactive protein (HSCRP) has been studied in threatened miscarriage (Jauniaux et al. 2015) and its role along with CA 125 needs to be further investigated in larger studies. In order to reliably interpret the biochemical markers in early pregnancy, gestational age specific normograms are required and pre specified cut-off values would be important for the study design. Ultrasound markers may have a role in accurately predicting outcome of threatened miscarriage either alone or in combination with the biochemical markers. Moreover, oxidative stress markers in maternal serum or urine may have potential role in predicting miscarriage, but this needs further research. Overall, it is important to consider biomarkers that can reliably predict an ongoing pregnancy rather than
predicting miscarriage since this would allay patient anxiety and be cost-effective. Future, large well-designed prospective cohort studies are needed with rigorous quality control and reporting methodology to accurately predict miscarriage outcome.

**Author's role**

RN Pillai: Contributed to the concept, study design, database search, data extraction and quality analysis, statistical analysis, writing the manuscript and final approval of the manuscript

JC Konje: Contributed to the concept, writing of the manuscript and final approval of the manuscript

DG Tincello: Contributed to the concept, reviewing of the manuscript and final approval of the manuscript

N Potdar: Conceived the idea, study design, database search, data extraction and quality analysis, statistical analysis, writing the manuscript and final approval of the manuscript

**Acknowledgments**

None

**Funding**

This study was conducted without any external funds.

**Conflict of Interest**

No conflict of interest.

**References**


Johansson ED. Plasma levels of progesterone in pregnancy measured by a rapid competitive protein 

Johns J, Muttukrishna S, Lygnos M, Groome N, Jauniaux E. Maternal serum hormone concentrations for 
prediction of adverse outcome in threatened miscarriage. *Reproductive biomedicine online* 2007:15; 413-
421.

Jouppilla P, Ilpo H; Juha T. Early pregnancy failure: study by ultrasonic and hormonal methods. *Obstetrics 
& Gynecology* 1980:55.1; 42-47.

Kirk E, Papageorghiou AT, Van Calster B, Condous G, Cowans N, Van Huffel S, Timmerman D, Spencer K, 
Bourne T. The use of serum inhibin A and activin A levels in predicting the outcome of 'pregnancies of 
unknown location'. *Hum Reprod* 2009:24; 2451-2456.

Kunz J, Keller P. HCG, HPL, oestradiol, progesterone and AFP in serum in patients with threatened 

La Marca A, Morgante G, De Leo V. Human chorionic gonadotropin, thyroid function, and immunological 

Leung T, Spencer K, Leung T, Fung T, Lau T. Higher median levels of free beta-hCG and PAPP-A in the first 
trimester of pregnancy in a Chinese ethnic group. Implication for first trimester combined screening for 

Leylek O, Senocak F, Erselcan T. The prognostic predictive value of maternal serum CA 125 levels in 


Maged A, Walaa A. Biochemical and ultrasonographic predictors of outcome in threatened abortion. 
*Middle East Fertility Society Journal* 2013:18.3; 177-181.

Makrydimas G, Sebire N, Lolis D, Vlassis N, Nicolaides K. Fetal loss following ultrasound diagnosis of a live 


Warrick J, Gronowski A, Moffett C, Zhao Q, Bishop E, Woodworth A. Serum activin A does not predict ectopic pregnancy as a single measurement test, alone or as part of a multi-marker panel including progesterone and hCG. *Clinica Chimica Acta* 2012;413:707-711.


Figure 3a Forest plot of study results for serum $\beta$ hCG in women with threatened miscarriage. FN=false negative; FP=false positive; TN=true negative; TP=true positive.

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dessauve et al., 1982</td>
<td>2</td>
<td>3</td>
<td>16</td>
<td>29</td>
<td>0.11 [0.01, 0.35]</td>
<td>0.90 [0.74, 0.96]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jouppila et al., 1990</td>
<td>7</td>
<td>2</td>
<td>8</td>
<td>9</td>
<td>0.10 [0.00, 0.45]</td>
<td>0.94 [0.66, 0.96]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leyko et al., 1997</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>18</td>
<td>0.80 [0.52, 0.96]</td>
<td>0.72 [0.51, 0.88]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maged and Mostafa, 2013</td>
<td>43</td>
<td>16</td>
<td>22</td>
<td>69</td>
<td>0.66 [0.53, 0.77]</td>
<td>0.81 [0.71, 0.89]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scarpellini et al., 1995</td>
<td>11</td>
<td>4</td>
<td>8</td>
<td>25</td>
<td>0.58 [0.33, 0.80]</td>
<td>0.86 [0.68, 0.95]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3b Forest plot of study results for serum intact hCG in women with threatened miscarriage. FN=false negative; FP=false positive; TN=true negative; TP=true positive.

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slimes et al., 1983</td>
<td>23</td>
<td>6</td>
<td>6</td>
<td>20</td>
<td>0.79 [0.60, 0.92]</td>
<td>0.77 [0.56, 0.91]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slupski et al., 1981</td>
<td>29</td>
<td>3</td>
<td>5</td>
<td>25</td>
<td>0.85 [0.69, 0.95]</td>
<td>0.89 [0.72, 0.96]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Westergaard et al., 1985</td>
<td>9</td>
<td>4</td>
<td>11</td>
<td>62</td>
<td>0.00 [0.00, 0.00]</td>
<td>0.94 [0.85, 0.95]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3c Summary receiver operating curve for hCG (intact and $\beta$ hCG) and empirical Bayes estimate.

Figure 4a Forest plot of study results for serum CA 125 in women with threatened miscarriage. FN=false negative; FP=false positive; TN=true negative; TP=true positive.

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiegl et al., 2003</td>
<td>92</td>
<td>12</td>
<td>5</td>
<td>71</td>
<td>0.93 [0.83, 0.98]</td>
<td>0.86 [0.76, 0.92]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leyko et al., 1997</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td>24</td>
<td>0.87 [0.69, 0.98]</td>
<td>0.96 [0.80, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maged and Mostafa, 2013</td>
<td>52</td>
<td>19</td>
<td>13</td>
<td>66</td>
<td>0.80 [0.68, 0.92]</td>
<td>0.78 [0.67, 0.86]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ozer et al., 1992</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>19</td>
<td>1.00 [1.00, 1.00]</td>
<td>0.96 [0.75, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scarpellini et al., 1995</td>
<td>15</td>
<td>7</td>
<td>4</td>
<td>22</td>
<td>0.79 [0.54, 0.94]</td>
<td>0.76 [0.56, 0.90]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheff et al., 2000</td>
<td>41</td>
<td>1</td>
<td>2</td>
<td>56</td>
<td>0.95 [0.84, 0.99]</td>
<td>0.98 [0.91, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xie et al., 2014</td>
<td>72</td>
<td>9</td>
<td>7</td>
<td>47</td>
<td>0.91 [0.83, 0.96]</td>
<td>0.84 [0.72, 0.93]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4b Summary receiver operating curve and empirical Bayes estimate for serum CA 125
Supplementary Figure 1a Forest plot of study results for serum progesterone in women with threatened miscarriage. FN=false negative; FP=false positive; TN=true negative; TP=true positive

Supplementary Figure 1b Summary receiver operating curve and empirical Bayes estimate for serum progesterone

Supplementary Figure 2a Forest plot of study results for serum estradiol in women with threatened miscarriage FN=false negative; FP=false positive; TN=true negative; TP=true positive

Supplementary Figure 2b Summary receiver operating curve and empirical Bayes estimate for serum estradiol
Supplementary Figure 3  Forest plot of study results for serum PAPP-A in women with threatened miscarriage FN=false negative; FP=false positive; TN=true negative; TP=true positive