# Managing pregnancy of unknown location based on initial serum progesterone and serial serum hCG: development and validation of a two-step triage protocol

**Van Calster B[1]†, Bobdiwala S[2]†, Guha S[3], Van Hoorde K[4], Al-Memar M[2], Harvey R[5], Farren J[2], Kirk E[6], Condous G[7], Sur S[2], Stalder C[2], Timmerman D[1,8], Bourne T[1,2,8]**

[1]KU Leuven, Department of Development and Regeneration, Leuven, Belgium
2Tommy's National Miscarriage Centre, Queen Charlotte's & Chelsea Hospital, Imperial College, Du Cane Road, London, W12 0HS, UK
[3]West Middlesex Hospital, Twickenham Road, Isleworth, Middlesex TW7 6AF, UK
[4]Open Analytics, Antwerp, Belgium
[5]Charing Cross Oncology Laboratory and Trophoblastic Disease Center, Charing Cross Hospital Campus of Imperial College Healthcare National Health Service Trust, London, UK
[6]North Middlesex Hospital, Sterling Way, London N18 1QX, UK
[7]Acute Gynaecology, Early Pregnancy and Advanced Endosurgery Unit, Nepean Medical School, Nepean Hospital, University of Sydney, Kingswood, NSW, Australia
[8]Department of Obstetrics and Gynecology, University Hospitals Leuven, Leuven, Belgium

† The authors consider that the first two authors should be regarded as joint First Authors.

**Correspondence to:** Professor Tom Bourne, Tommy's National Miscarriage Centre, Queen Charlotte's & Chelsea Hospital, Imperial College, Du Cane Road, London, W12 0HS, UK
+44203 313 1000 (e-mail: tbourne@imperial.ac.uk)

Keywords: pregnancy of unknown location, ectopic pregnancy, progesterone, human chorionic gonadotrophin, decision support techniques

Competing Interests: No competing interests are declared.

**Abstract**

*Objectives*: A uniform rationalized management protocol for pregnancies of unknown location (PUL) is lacking. We developed a two-step triage protocol based on presenting serum progesterone (step 1) and hCG ratio two days later (step 2) to select PUL at high-risk of ectopic pregnancy (EP).

*Methods*: Cohort study of 2753 PUL (301 EP), involving a secondary analysis of prospectively and consecutively collected PUL at two London-based university teaching hospitals. Using a chronological split we used 1449 PUL for development and 1304 for validation. We aimed to select PUL as low-risk with high confidence (high negative predictive value, NPV) while classifying most EP as high-risk (high sensitivity). The first triage step selects low-risk PUL at presentation using a serum progesterone threshold. The remaining PUL are triaged using a novel logistic regression risk model based on hCG ratio and initial serum progesterone (second step), defining low-risk as an estimated EP risk <5%.

*Results*: On validation, initial serum progesterone ≤2nmol/l (step 1) selected 16.1% PUL as low-risk. Second step classification with the risk model '$M6_P$' selected an additional 46.0% of all PUL as low-risk. Overall, the two-step protocol classified 62.1% of PUL as low-risk, with an NPV of 98.6% and a sensitivity of 92.0%. When the risk model was used in isolation (i.e. without the first step), 60.5% of PUL were classified as low-risk with 99.1% NPV and 94.9% sensitivity.

*Conclusions*: The two-step protocol can efficiently classify PUL into being at high or low risk of complications.

## Introduction

Pregnancy of unknown location (PUL) describes a woman who has a positive pregnancy test without evidence of a pregnancy inside or outside the endometrial cavity using transvaginal sonography (TVS). The management of PUL can be prolonged, costly and lacks uniformity. The serum biomarkers human chorionic gonadotrophin (hCG)[1-4] and/or progesterone[5-7] are used to guide management. PUL prediction now focuses on triaging PUL as low-risk or high-risk of complications.[8-13] Low-risk implies the final outcome is an intrauterine pregnancy (IUP) or a failed PUL (FPUL), high-risk that the final outcome is an ectopic pregnancy (EP) or persistent PUL (PPUL). Risk stratification allows resources to be rationalized so women with low-risk PUL avoid unnecessary additional blood tests, visits to hospital and ultrasound scans whilst focusing resources on high-risk PUL that are at greater risk of potentially life-threatening EP. For this streamlining of resources to be safe, the classification tool used to determine risk must be accurate, particularly regarding the classification of EP. Inevitably some EP will be misclassified as low-risk, but many are likely to resolve without harm, especially if they are associated with low initial progesterone or declining hCG values.

In a multi-center study on 1962 PUL, Van Calster et al[14] externally validated the performance of the 'M4' model, a logistic regression model utilizing the initial hCG and hCG ratio. Based on a risk of EP <5%, M4 classified 70% PUL as low-risk, with 97% being an FPUL/IUP. Eighty-eight percent of EP were correctly classified as high-risk. Although triage based on single progesterone has good results[5,6], M4-based triage outperforms hCG ratio or serum progesterone cut-offs.[15] Previous work suggested that clinical symptoms do not meaningfully improve prediction.[16]

To avoid unnecessary visits and blood tests, the interpretation of serum hormone levels used to manage PUL needs standardizing. Our previously published M4 model was developed on a relatively small number of PUL and required a minimum of two visits. The aim of the current study was to develop a novel two-step protocol, combining both the initial serum progesterone level and hCG ratio, that is developed on a much larger cohort of PUL and allows discharge of some PUL after just one visit.

## Methods

### Design, setting and participants

This observational cohort study for diagnostic accuracy involves a secondary analysis of data from the early pregnancy assessment units (EPAU) at two London-based university teaching hospitals: St Georges' (SGH) and Queen Charlottes' & Chelsea (QCCH). Women attend EPAUs up to 12 weeks gestation and are either asymptomatic and undergo ultrasonography for reassurance or present with symptoms such as pain, bleeding and/or hyperemesis. Data on consecutive women with a PUL was prospectively collected from July 2003 to February 2007 at SGH and April 2009 to December 2013 at QCCH using a standardized protocol[8,12]. We excluded women known to be on progesterone supplementation and those with initial hCG ≤25IU/L, because this is the cut-off at which most urine pregnancy tests would be negative and the majority of clinicians in this study did not bring these PUL back for a 48 hour hCG level but a urine pregnancy test in two weeks to confirm a repeat negative result. The data were divided into a development dataset and a temporal validation dataset using a chronologic split within each center. As the normal clinical management of patients was not changed, data collection was registered as an audit at Imperial College Healthcare NHS trust. We followed the recently published TRIPOD (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) guidelines when preparing this report.[17]

### Data collection

Immediately following identification of a PUL, women had blood samples taken for measurement of serum progesterone and hCG as a part of routine clinical management provided at SGH and QCCH. Patients had the serum hCG measurement

repeated approximately 48 hours later. At QCCH, blood samples were separated within one hour and assayed using the Abbott progesterone and hCG assays run onboard the Abbott Architect i2000SR instrument (Abbott Laboratories, Abbott Park, IL). The Architect progesterone assay is a one-step, competitive, single-site type immunoassay that uses chemiluminescent paramagnetic microparticle immunoassay (CMIA 'Chemiflex$^{TM}$ technology'). The analytical sensitivity of the assay is ≤0.1 μg/l, has a precision of <10% (total interassay CV) and <5% cross-reactivity across a range of over forty tested cholesterol-derived molecules. The Architect hCG assay is a two-step, non-competitive, two-site type immunoassay that employs two mouse monoclonal antibodies and CMIA 'Chemiflex$^{TM}$ technology'. The analytical sensitivity of the assay is ≤1.2 IU/L, has a precision of <10% (total interassay CV) and <10% cross-reactivity with FSH, LH and TSH. At SGH, blood samples were separated within one hour and assayed using the Roche (Basel, Switzerland) Elecsys E180 immunoassay analyzer. The Roche progesterone assay (Elecsys 2010 progesterone II) is a one-step, competitive, single-site type immunoassay that uses an electrochemiluminescent paramagnetic microparticle immunoassay (ECL technology). The analytical sensitivity of the assay is 0.48 nmol/l, has a precision of <10% (total interassay CV) and <5% cross-reactivity across a range of >40 tested cholesterol-derived molecules. The Roche HCG+β assay (Elecsys 2010) is a two-step, non-competitive, two-site type immunoassay that employs a mouse monoclonal antibody and streptavidin-coated paramagnetic microparticle immunoassay system. The analytical sensitivity of the assay is < 0.6 mIU/mL, has a precision of <10% (total interassay CV) and <5% cross-reactivity with TSH, LH and FSH. The two units, therefore, had assays with similar performance characteristics.

All samples were processed and tested in accordance with standard medical laboratory practice including daily analysis of QC samples to check assay accuracy, precision and further procedures as specified by Clinical Pathology Accreditation (CPA) quality standards.

**Reference standard: pregnancy outcome definitions**

Pregnancies were followed up until the final clinical outcome was known. This was either an FPUL, IUP, EP or PPUL.[18-19] FPUL was defined as a PUL with declining serum hCG levels successfully managed expectantly but there is no histological confirmation of the location of the pregnancy. IUP was defined as a visible intrauterine gestational sac TVS that did or did not contain a yolk sac or fetal pole. If products of conception were histologically confirmed after surgical evacuation, this was also defined as an IUP. EP was defined as a pregnancy where an IUP could not be visualized using TVS but one of the following was: an inhomogeneous adnexal mass separate from the ovary ('blob sign')[12], an empty extra-uterine gestational sac in the adnexa ('bagel sign')[20] or an extra-uterine gestation sac with a yolk sac, fetal pole, or fetal cardiac activity. If after laparoscopy or laparotomy an EP was confirmed histologically, this was also classified as an EP. Finally, PPUL was defined as a pregnancy that remained a PUL but with serum hCG plateauing and varying by <15% on three consecutive samples 48 hours apart. These are most likely to be EPs not visualized on TVS as the hCG levels behave in the same way biochemically. Therefore, PPUL were analyzed as EP, i.e. high-risk PUL.

**Two-step triage protocol for PUL**

The aim of the triage protocol is to select a group of PUL as low-risk with high confidence, and classify most EP as high-risk. We developed a triage protocol that involves two steps (Figure 1). The first triage step classifies PUL as low-risk at presentation using a simple threshold value for serum progesterone. This is to reduce unnecessary follow-up for PUL that are very likely to have a final outcome of FPUL. Patients not classified as low-risk at presentation proceed to the second step that applies a risk model using the hCG ratio as the key predictor. Other predictors used in the model are the initial hCG and progesterone levels. We have called this risk model $M6_P$. For patients using progesterone supplements, a variant of $M6_P$ can be used in which the initial progesterone is not included in the model as a predictor ($M6_{NP}$). When using $M6_P$ or $M6_{NP}$, we define PUL with an estimated risk of EP below 5% as low-risk.

Using this two-step protocol, some PUL (likely outcome: FPUL) are classified as low-risk after the initial visit, and the remainder as either low or high-risk after the follow-up visit. Note that the risk model can also be used in isolation, i.e. without a first step based on initial progesterone.

**Sample size and statistical analysis**

Overall, data on 2753 women were available, see Results for patient flow. The triage protocol was developed on the first set of patients from each center (N=1449, of which 785 FPUL, 501 IUP, and 163 EP), and validated on the second half (N=1304, of which 665 FPUL, 501 IUP, and 138 EP).

Missing values for initial serum progesterone and hCG levels at approximately 48hrs were handled using multiple imputation,[21] see Supplementary Material for details.

Second hCG levels taken more than three days after the initial hCG were considered missing.

For the first step of the protocol, a cut-off on initial progesterone was determined. Knowing that hCG ratio is in fact more predictive than initial progesterone, we aimed for a cut-off that limits the number of EP classified as low-risk based on only the initial progesterone level to a strict minimum.

The risk model for the second step of the triage protocol is based on multinomial logistic regression. We considered initial hCG, the hCG ratio and initial progesterone as predictors, and used the method of multivariable fractional polynomials[22] to help determine whether these variables have linear effects, and if not to help select the optimal transformation. Further, we considered an interaction term between the hCG ratio and progesterone. More information is provided in Supplementary Material.

***Validation***

The main performance measures were negative predictive value (NPV, percentage of non-EP among patients classified as low-risk) to assess the confidence of the classification as low-risk, and sensitivity (percentage of EP classified as high-risk) to assess how many EP are correctly selected for close follow up. Other important measures are the overall percentage of PUL classified as low-risk, as this is the group that would receive reduced follow up, and the false positive rate (percentage of non-EP classified as high-risk). Performance was then compared with that of other available triage tools for PUL: M4-based triage (again using an estimated risk of EP <5% to define low-risk[14]), and a single visit protocol using a serum progesterone level of ≤10nmol/l to define low-risk.[5]

In addition, for $M6_P$, $M6_{NP}$, and M4 the AUC for EP (i.e. based on the estimated risk of EP) was computed, as well as a multinomial AUC to quantify overall discrimination between FPUL, IUP, and EP.[23] Calibration curves were constructed for the risk of EP to investigate whether the estimated risks are correct.[24]

Finally, we compared the clinical utility of triage based on $M6_P$, $M6_{NP}$, and M4 (with an estimated risk of EP <5% defining low-risk) using the Net Benefit statistic.[25] Net Benefit combines results for sensitivity and specificity, while taking into account that the 5% risk cut-off implies that classifying an EP as low-risk is clearly worse than classifying a non-EP as high-risk. See Supplementary Material for full details.

**Results**

A total of 3152 PUL were recruited during the study period, 2058 from SGH and 1094 from QCCH. 162 PUL were lost to follow up (5.1%). We excluded 271 of 3152 patients upfront: 58 from SGH that were lost to follow up on whom we have no information, 212 with an initial hCG ≤25IU/L, and 1 taking progesterone supplements, leaving 2881 patients (1858 SGH, 1023 QCCH). In the remaining group, there were 164 missing values for serum hCG at 48hrs (5.7%) and 188 for initial progesterone (6.5%). Of the 164 patients with a missing value for serum hCG at 48hrs, in 62 the second hCG measurement was taken more than 3 days after the first and hence not used. For 128 PUL, the final outcome of the pregnancy was unknown (4.4%). These 128 patients were discarded, leaving 2753 patients (1786 from SGH, 967 from QCCH).

Of the 2753 PUL, there were 1450 (52.7%) FPUL, 1002 (36.4%) IUP, and 301 (10.9%) EP. The EP rate was 9.5% at SGH and 13.7% at QCCH. The development dataset included data from the first 921 (52%) patients from SGH and the first 528 (55%) patients from QCCH (1449 in total). The remaining patients formed the validation data (1304 in total, 865 from SGH and 439 from QCCH). Descriptive statistics per center are presented in Table 1, and separately for development and validation data in Table S1.

***Development of the two step triage protocol***

*First step*. Based on the performance of initial progesterone cut-offs ranging from ≤2 to ≤10nmol/l on the development data (Table 2; Figure 2a), we defined low-risk as having a serum progesterone ≤2nmol/l.

*Second step*. The following variables were selected for the $M6_P$ model: log(initial hCG), log(initial progesterone), log(hCG ratio), log(hCG ratio) squared, and the interaction between log(hCG ratio) and log(initial progesterone). The $M6_{NP}$ model does not use log(initial progesterone) and the interaction between log(initial progesterone) and log(hCG ratio), although a joint likelihood ratio test for these two terms suggested that they had a strong effect ($p<0.0001$).

***Evaluating the triage protocol on the temporal validation set***

The selected progesterone cut-off (≤2 nmol/l to define low-risk) classified 210/1304 (16.1%) patients as low-risk at the initial visit (Table 3; Figure 2b). In 206 of these 210 cases the final outcome confirmed they were not an EP (i.e. FPUL or IUP) (NPV 98.1%). Four of the 138 EP were classified as low-risk, yielding a sensitivity of 97.1% (134/138). Finally, when only applying step 1, 960/1166 FPUL and IUP were classified as high-risk (false positive rate 82.3%).

When all 1094 PUL classified as high-risk on the basis of step 1 were triaged using $M6_P$ as the second step of the protocol, an additional 600 (54.8%) were re-classified as low-risk, with an NPV of 98.8% (593/600), a sensitivity of 94.8% (127/134, i.e. 7 EP classified as low-risk), and a false positive rate of 38.2% (367/960).

Taken together, the two-step protocol classified 62.1% PUL as low-risk (810/1304) with an NPV of 98.6% (799/810), a sensitivity of 92.0% (127/138) (i.e. 11 EP were misclassified as low-risk), and a false positive rate of 31.5% (367/1166) (Table 3). When using the $M6_{NP}$ model as the second step, 57.7% PUL were classified as low-risk with an NPV of 98.1%, a sensitivity of 89.6%, and a false positive rate of 36.6%.

***Validation performance when using the $M6_P$ (or $M6_{NP}$) model in isolation as a single step approach***

The $M6_P$ model in isolation (i.e. without the first step) classified 60.5% PUL (789/1304) as low-risk with an NPV of 99.1% (782/789), a sensitivity for EP of 94.9% EP (131/138) (i.e. 7 EP misclassified as low-risk), and a false positive rate of 32.9% (385/1166) (Table 3). The AUC for EP vs. other PUL was 0.903 (95% CI 0.880 to 0.922) (Table S2, Figure S1). The $M6_{NP}$ model classified 54.5% PUL as low-risk with an NPV of 98.6%, a sensitivity of 92.5% (128/138), and a false positive rate of 39.9%. The AUC for EP vs. other PUL was 0.870 (95% CI 0.837 to 0.897). The estimated risks of EP given by the $M6_P$ and $M6_{NP}$ models were accurate as indicated by the calibration plots (Figure 3), even when evaluated for both centers separately (Figure S2).

***Performance of currently available triage tools: M4 model and a single visit progesterone protocol***

The M4 model classified 70.6% PUL as low-risk with an NPV of 97.2%, a sensitivity for EP of 81.4%, and a false positive rate of 23.2% (Table 3). The AUC for EP vs. other PUL was 0.847 (95% CI 0.808 to 0.880) (Table S2, Figure S1), which is 0.056 lower than the $M6_P$ model (95% CI 0.029 to 0.083), and 0.023 lower than the AUC of the $M6_{NP}$ model (95% CI 0.009 to 0.036). The calibration plots indicates that the risk of EP given by M4 was underestimated (Figure 3). The Standardized Net Benefit to assess clinical usefulness of triage at the 5% risk cut-off for EP was 0.71 for M4, 0.80 for the $M6_P$ model, and 0.75 for the $M6_{NP}$ model. This is equivalent to an increase in

sensitivity of 9% for $M6_P$ vs M4 and of 4% for $M6_{NP}$ vs M4 at the same level of specificity.

The existing single visit progesterone protocol (≤10nmol/l indicating a low-risk PUL) classified 43.8% PUL as low-risk with an NPV of 94.7%, a sensitivity for EP of 78.1%, and a FPR of 53.6%.

***Updating the $M6_P$ and $M6_{NP}$ models using all data***

Finally, to make full use of all the available data after successful temporal validation, we updated the coefficients of $M6_P$ and $M6_{NP}$ using all data (n=2753). The final coefficients are given in Table S3, with additional information in Supplementary Material.

**Discussion**

We have proposed a two-step approach to triage PUL using the initial serum progesterone level at the first step and a risk prediction model based on hCG ratio, with or without the initial progesterone level, at the second step. We have shown that this protocol classifies the majority of PUL as low-risk or high-risk with a high level of confidence. When an initial progesterone ≤2nmol/l is used to classify PUL as low-risk at the first step, a significant proportion of women (about 1 in 6) with a PUL can be discharged following an initial consultation with just a urinary pregnancy test at home in two weeks as follow-up and minimal risk of misclassifying EP.

A strength of the study is the large sample size and the inclusion of a temporal validation. As a result, the new $M6_P$ and $M6_{NP}$ models have better predictive ability (discrimination) and provide more accurate risk estimates than M4, which was based on a much smaller number of PUL. A further strength is that the two-center nature of the study increases the likelihood of the protocol and models being applicable in other populations. The protocol is based solely on measurements of progesterone and hCG rather than more subjective information, therefore the approach is more likely to be generalizable. A weakness of the study is that serum progesterone levels were missing in 188 (6.5%) and serum hCG levels were missing in 164 (5.7%) patients. However, we used multiple imputation to deal with these missing values, which is recommended over the use of complete case analysis.[21]

The advantage of incorporating serum progesterone as part of a two-step strategy into the algorithm is that a proportion of patients can be discharged at the initial visit. Using a relatively low cut-off value for progesterone means only 2.9% of EP will be

misclassified as low-risk at this first step. However, we advise these patients to undergo a urinary pregnancy two weeks later with telephone follow-up and to contact the unit in the event of pain or any other concerns they may have. It is an open question whether the cut-off value for progesterone as a first step test can be set higher. This will result in more women being triaged as low-risk, but at the cost of reduced follow up for more EP (Table 2). It can be argued these EP with relatively low progesterone levels are likely to be failing and resolve without complications. The concern is that there are insufficient data currently to accurately determine the safety of a policy of assuming such EP are low-risk and safe to manage expectantly. This issue was addressed by Cordina et al.[5] In this interventional study, an initial serum progesterone level of ≤10 was used to define low-risk PUL. Of the 227 classified as low-risk in whom there was complete follow-up, 14 women returned for unscheduled visits because of pain and/or bleeding. Five were subsequently found to have an EP. Three cases resolved without needing intervention. Two cases underwent laparoscopic salpingectomies because of pain, and in neither case was significant blood found in the pelvis. Clearly proving safety is difficult, however these data suggest that, using this protocol, the number of clinically important EP classified as low-risk is approximately 1%. We used 2nmol/l as a cut-off, because we felt this reached a sensible balance between the size of the group selected for minimal follow-up (1 in 6 in our study) and the number of EP that are placed in this group. Given the lack of safety data, such a low cut-off seems warranted.

The second step of our two-step protocol uses either the $M6_P$ or $M6_{NP}$ model. These models can also be used in isolation. As such, they perform better than previously suggested single-step protocols such as a serum progesterone cut-off value of ≤10[5,6]

and the M4 model[14] (Table 3). In fact, using the $M6_P$ model as a single-step protocol had higher sensitivity for EP than the two-step protocol (94.9% versus 92.0%). The two-step approach, however, benefits from discharging a proportion of women after just one visit.

Using measurements of progesterone to predict early pregnancy outcomes is confounded by women taking progesterone supplementation, as this significantly raises serum levels. In these circumstances a low serum progesterone at the initial visit (step 1) can still be used to indicate a failing pregnancy. A limitation of the study is that our patient population was not suitable to perform a sub-analysis to investigate model performance in women taking progesterone supplementation, as progesterone is not routinely given in the UK. We only identified one patient taking progesterone, and this case was excluded from the final analysis. To deal with this issue we developed the $M6_{NP}$ model that does not use progesterone as a variable. However, as can be seen in table 3, the number of PUL classified as low-risk with this model is lower than when progesterone is included as a variable (52.6% versus 60.8% respectively).

As urinary pregnancy tests become more sensitive and with the ubiquitous access to ultrasonography that exists in many parts of the world, the chances of pregnant women undergoing a scan that fails to identify the location of their pregnancy is significant[26]. This creates an iatrogenic problem that needs to be managed whilst ensuring that EP in women with symptoms are detected. The two-step protocol we have described offers an effective strategy for selecting low-risk cases where follow-up can be reduced to a minimum, whilst directing resources to women at highest risk

of harboring an EP. We believe using the protocol will make the interpretation of serum hormone levels to manage PUL more straightforward, particularly in the hands of less experienced staff, out of hours or at weekends. We hope this will lead to more consistent decision-making and reduce the number of visits and interventions for women. We have made $M6_P$ and $M6_{NP}$ available as a free online application at www.earlypregnancycare.org. We have also incorporated the models into an application that can be downloaded onto smartphones or tablets, allowing clinicians to incorporate its use into everyday clinical practice.

**Acknowledgements**

BVC was partly supported by a postdoctoral fellowship from the Research Foundation Flanders (FWO). KVH was supported by a fellowship from the Flanders' Agency for Innovation by Science and Technology (IWT-Vlaanderen). TB is supported by the NIHR Biomedical Research Center based at Imperial College Healthcare NHS Trust and Imperial College London. SB and development of the application for smartphones were supported by NIHR CLAHRC NWL (Collaboration for Leadership in Applied Health Research & Care, NorthWest London). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. DT is Fundamental Clinical Researcher of FWO (Research Foundation – Flanders).

## References

1. Van Mello N, Mol F, Opmeer BC, Ankum WM, Barnhart K, Coomarasamy A, Mol BW, van der Veen F, Hajenius PJ. Diagnostic value of serum hCG on the outcome of pregnancy of unknown location: a systematic review and meta-analysis. Hum Reprod Update 2012;18:603–17.

2. Condous G, Kirk E, Van Calster B, Van Huffel S, Timmerman D, Bourne T. Failing pregnancies of unknown location: a prospective evaluation of the human chorionic gonadotrophin ratio. BJOG 2006;113:521–27.

3. Silva C, Sammel MD, Zhou L, Gracia C, Hummel AC, Barnhart K. Human chorionic gonadotrophin profile for women with ectopic pregnancy. Obstet Gynecol 2006;107:605–10.

4. Butts SF, Guo W, Cary MS, Chung K, Takacs P, Sammel MD, Barnhart KT. Predicting the decline in human chorionic gonadotropin in a resolving pregnancy of unknown location. Obstet Gynecol 2013;122:337-43.

5. Cordina M, Schramm-Gajraj K, Ross JA, Lautman K, Jurkovic D. Introduction of a single visit protocol in the management of selected patients with pregnancy of unknown location: a prospective study. BJOG 2011;118:693-7.

6. Day A, Sawyer E, Mavrelos D, Tailor A, Helmy S, Jurkovic D. Use of serum progesterone measurements to reduce need for follow-up in women with pregnancies of unknown location. Ultrasound Obstet Gynecol 2009;33:704-10.

7. Mol BW, Lijmer JG, Ankum WM, van der Veen F, Bossuyt PMM. The accuracy of single serum progesterone measurement in the diagnosis of ectopic pregnancy: a meta-analysis. Hum. Reprod 1998;13:3220-27.

8. Kirk E, Condous G, Van Calster B, Van Huffel S, Timmerman D, Bourne T. Rationalizing the follow-up of pregnancies of unknown location. Hum Reprod 2007;22:1744–50.

9. Kirk E, Condous G, Bourne T. Pregnancies of unknown location. Best Pract Res Clin Obstet Gynaecol 2009;23:493-9.

10. Kirk E, Bottomley C, Bourne T. Diagnosing ectopic pregnancy and current concepts in the management of pregnancy of unknown location. Hum Reprod Update 2014;20:250–61.

11. Condous G, Okaro E, Khalid A, Timmerman D, Lu C, Zhou Y, Van Huffel S, Bourne T. The use of a new logistic regression model for predicting the outcome of pregnancies of unknown location. Hum Reprod. 2004 Aug;19(8):1900-10.

12. Condous G, Van Calster B, Kirk E, Haider Z, Timmerman D, Van Huffel S, Bourne T. Prediction of ectopic pregnancy in women with a pregnancy of unknown location. Ultrasound Obstet Gynecol. 2007 Jun;29(6):680-7.

13. Chen ZY, Liu JH, Liang K, Liang WX, Ma SH, Zeng GJ, Xiao SY, He JG. The diagnostic value of a multivariate logistic regression analysis model with transvaginal power Doppler ultrasonography for the prediction of ectopic pregnancy. J Int Med Res. 2012;40(1):184-93.

14. Van Calster B, Abdallah Y, Guha S, Kirk E, Van Hoorde K, Condous G, Preisler J, Hoo W, Stalder C, Bottomley C, Timmerman D, Bourne T. Rationalizing the management of pregnancies of unknown location: temporal and external validation of a risk prediction model on 1962 pregnancies. Hum Reprod 2013;28:609-16.

15. Guha S, Ayim F, Ludlow J, Sayasneh A, Condous G, Kirk E, Stalder C, Timmerman D, Bourne T, Van Calster B. Triaging pregnancies of unknown location: the performance of protocols based on single serum progesterone or repeated serum hCG levels. Hum Reprod 2014;29:938-45.

16. Condous G, Van Calster B, Kirk E, Haider Z, Timmerman D, Van Huffel S, Bourne T. Clinical information does not improve the performance of mathematical models in predicting the outcome of pregnancies of unknown location. Fertil Steril 2007;88:572–80.

17. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015 Jan 7;350:g7594.

18. Barnhart KT, van Mello NM, Bourne T, Kirk E, Van Calster B, Bottomley C, Chung K, Condous G, Goldstein S, Hajenius PJ, Mol BW, Molinaro T, O'Flynn O'Brien KL, Husicka R, Sammel M, Timmerman D. Pregnancy of unknown location: A consensus statement of nomenclature, definitions and outcome. Fertil Steril 2011;95:857–66.

19. Condous G, Timmerman D, Goldstein S, Valentin L, Jurkovic D, Bourne T. Pregnancies of unknown location: consensus statement. Ultrasound Obstet Gynecol 2006;28:121-2.

20. Condous G, Okaro E, Khalid A, Lu C, Van Huffel S, Timmerman D, Bourne T. The accuracy of transvaginal sonography for the diagnosis of ectopic pregnancy prior to surgery. Hum Reprod 2005;20:1404–09.

21. Sterne JAC, 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.

22. Royston P, Sauerbrei W. Multivariable Model - Building: A Pragmatic Approach to Regression Anaylsis based on Fractional Polynomials for Modelling Continuous Variables. Wiley; 2008.

23. Van Calster B, Vergouwe Y, Looman CWN, Van Belle V, Timmerman D, Steyerberg EW. Assessing the discriminative ability of risk models for more than two outcome categories. Eur J Epidemiol 2012;27:761-70.

24. Van Hoorde K, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW, Van Calster B. Assessing calibration of multinomial risk prediction models. Stat Med 2014;33:2585-96.

25. Mallett S, Halligan S, Thompson M, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. BMJ 2012;345:e3999.

26. 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 2009;24:1811-17.

**TABLES:**

**Table 1.** Descriptive statistics presented as median (interquartile range) in the two participating hospitals. SGH = St Georges' Hospital; QCCH = Queen Charlottes' & Chelsea Hospital; PUL = pregnancy of unknown location; IUP = intrauterine pregnancy; hCG = human chorionic gonadotrophin.

| ***Center*** **Variable** | **Failed PUL** | **IUP** | **Ectopic** |
|---|---|---|---|
| *SGH* | N=898 | N=719 | N=169 |
| Age (years) | 31 (26-36) | 30 (24-34) | 31 (28-35) |
| Initial hCG (IU/L) | 257 (87-742) | 539 (280-982) | 419 (198-1039) |
| 48 hour hCG (IU/L) | 112 (37-326) | 1212 (610-2052) | 518 (221-1238) |
| hCG ratio | 0.40 (0.28-0.61) | 2.18 (1.90-2.48) | 1.18 (0.94-1.45) |
| Initial progesterone (nmol/l) | 5 (3-10) | 68 (52-91) | 23 (12-43) |
| | | | |
| *QCCH* | N=552 | N=283 | N=132 |
| Age (years) | 33 (27-37) | 31 (26-35) | 33 (29-37) |
| Initial hCG (IU/L) | 425 (129-1726) | 712 (322-1747) | 480 (188-1139) |
| 48 hour hCG (IU/L) | 180 (58-604) | 1329 (635-2785) | 550 (234-1409) |
| hCG ratio | 0.37 (0.26-0.52) | 2.03 (1.36-2.44) | 1.20 (0.98-1.54) |
| Initial progesterone (nmol/l) | 5 (2-10) | 51 (31-70) | 21 (11-32) |

SGH = St Georges' Hospital; QCCH = Queen Charlottes' & Chelsea Hospital; PUL = pregnancy of unknown location; IUP = intrauterine pregnancy; hCG = human chorionic gonadotrophin.

**Table 2.** Development data performance of different initial progesterone (measured in nmol/l) cut-offs to define low-risk at presentation. EP = ectopic pregnancy; PUL = pregnancy of unknown location; Prog = progesterone

| Progesterone cut-off to define low-risk | PUL classified as low-risk | Negative predictive value | Sensitivity for EP | False positive rate |
|---|---|---|---|---|
| ≤2nmol/l | 200/1449, 13.8% | 197/200, 98.5% | 160/163, 98.2% | 1089/1286, 84.7% |
| ≤3nmol/l | 293/1449, 20.2% | 286/293, 97.6% | 156/163, 95.7% | 1000/1286, 77.8% |
| ≤4nmol/l | 377/1449, 26.0% | 367/377, 97.3% | 153/163, 93.9% | 919/1286, 71.5% |
| ≤5nmol/l | 445/1449, 30.7% | 429/445, 96.4% | 147/163, 90.2% | 857/1286, 66.6% |
| ≤6nmol/l | 505/1449, 34.8% | 486/505, 96.2% | 144/163, 88.3% | 800/1286, 62.2% |
| ≤7nmol/l | 549/1449, 37.9% | 526/549, 95.8% | 140/163, 85.8% | 760/1286, 59.1% |
| ≤8nmol/l | 583/1449, 40.2% | 555/583, 95.3% | 136/163, 83.3% | 731/1286, 56.8% |
| ≤9nmol/l | 609/1449, 42.0% | 579/609, 95.2% | 134/163, 81.9% | 707/1286, 55.0% |
| ≤10nmol/l | 636/1449, 43.9% | 603/636, 94.9% | 130/163, 79.9% | 683/1286, 53.1% |

**Table 3.** Performance of PUL classification approaches. Confidence intervals are given between parentheses. EP = ectopic pregnancy; PUL = pregnancy of unknown location; Prog = progesterone

| *Data*<br>Classification approach | PUL classified as low-risk | Negative predictive value | Sensitivity for EP | False positive rate |
|---|---|---|---|---|
| *Development data* | | | | |
| Step 1 only:<br>Progesterone cut-off | 200/1449, 13.8% | 197/200, 98.5% | 160/163, 98.2% | 1089/1286, 84.7% |
| Two-step protocol:<br>Step 1 + $M6_P$ model | 896/1449, 61.9% | 884/896, 98.7% | 151/163, 92.6% | 402/1286, 31.2% |
| Two-step protocol:<br>Step 1 + $M6_{NP}$ model | 797/1449, 55.0% | 783/797, 98.3% | 149/163, 91.7% | 503/1286, 39.1% |
| $M6_P$ model in isolation | 880/1449, 60.8% | 871/880, 99.0% | 154/163, 94.4% | 415/1286, 32.3% |
| $M6_{NP}$ model in isolation | 762/1449, 52.6% | 751/762, 98.6% | 152/163, 93.4% | 535/1286, 41.6% |
| *Validation data* | | | | |
| Step 1 only:<br>Progesterone cut-off | 210/1304, 16.1%<br>(14.2-18.2) | 206/210, 98.1%<br>(95.0-99.3) | 134/138, 97.1%<br>(92.4-98.9) | 960/1166, 82.3%<br>(80.0-84.5) |
| Two-step protocol:<br>Step 1 + $M6_P$ model | 810/1304, 62.1%<br>(58.8-65.3) | 799/810, 98.6%<br>(97.5-99.3) | 127/138, 92.0%<br>(85.9-95.6) | 367/1166, 31.5%<br>(28.1-35.0) |
| Two-step protocol:<br>Step 1 + $M6_{NP}$ model | 754/1304, 57.7%<br>(53.2-62.1) | 740/754, 98.1%<br>(96.8-98.9) | 124/138, 89.6%<br>(83.0-93.9) | 426/1166, 36.6%<br>(32.0-41.5) |
| $M6_P$ model in isolation | 789/1304, 60.5%<br>(57.1-63.8) | 782/789, 99.1%<br>(98.1-99.6) | 131/138, 94.9%<br>(89.4-97.6) | 384/1166, 32.9%<br>(29.5-36.5) |
| $M6_{NP}$ model in isolation | 716/1304, 54.5%<br>(49.8-59.2) | 706/716, 98.6%<br>(97.3-99.2) | 128/138, 92.5%<br>(86.4-96.1) | 460/1166, 39.9%<br>(35.0-45.1) |
| M4-based triage | 921/1304, 70.6%<br>(68.0-73.1) | 895/921, 97.2%<br>(95.9-98.1) | 112/138, 81.4%<br>(73.9-87.2) | 271/1166, 23.2%<br>(20.8-25.8) |
| Single visit prog ≤10nmol/l | 572/1304, 43.8%<br>(41.1-46.6) | 542/572, 94.7%<br>(92.5-96.3) | 108/138, 78.1%<br>(70.3-84.3) | 625/1166, 53.6%<br>(50.7-56.5) |

**FIGURES:**

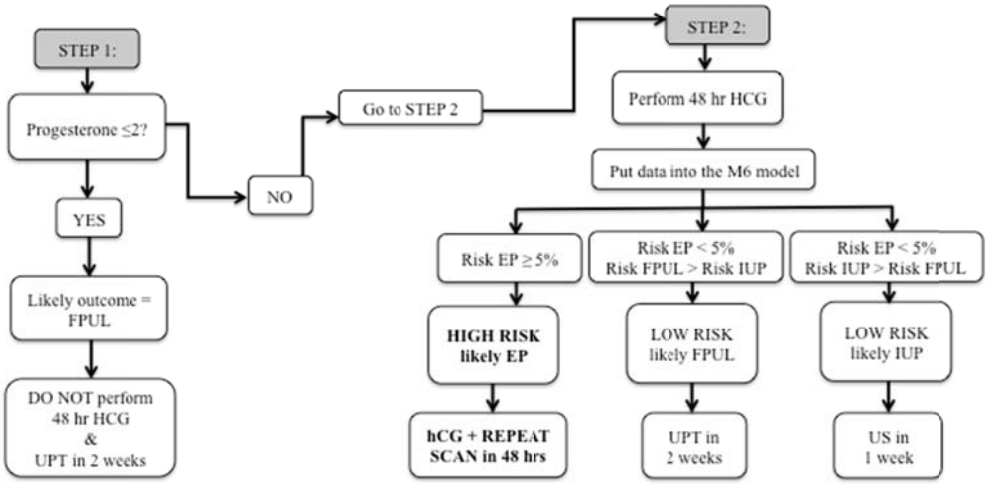
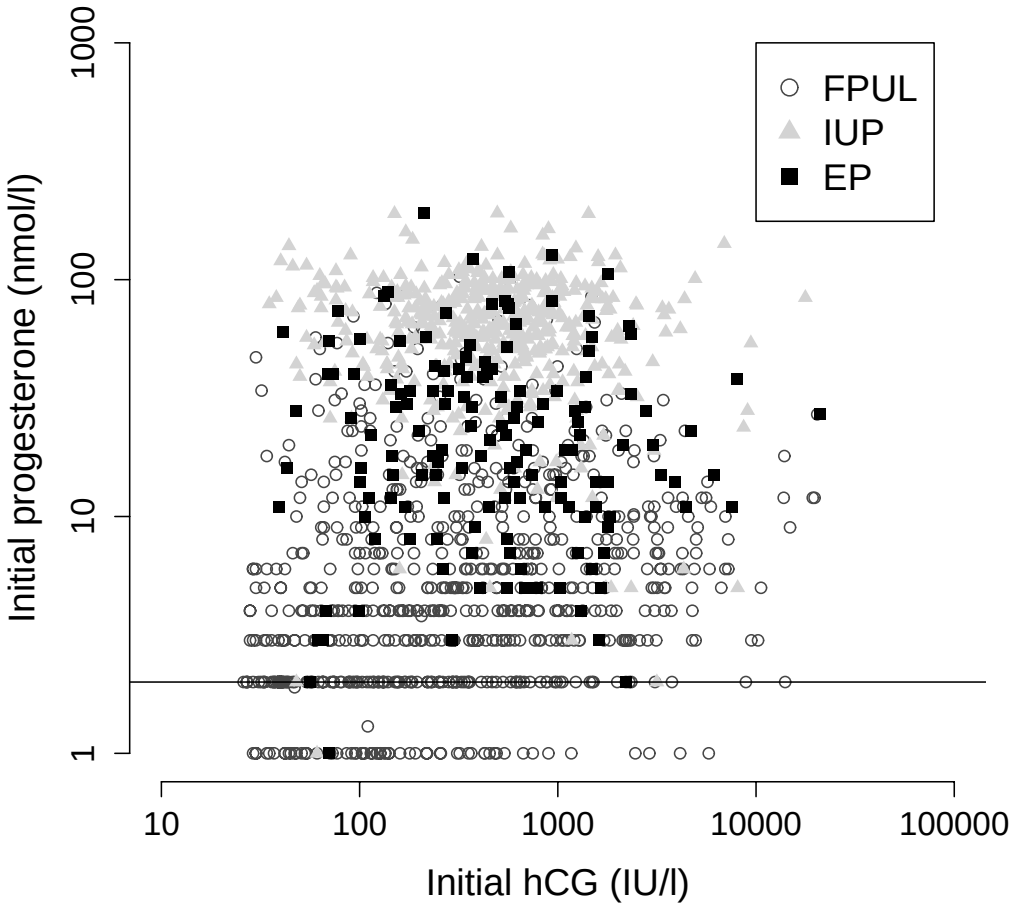



Figure 1: flow diagram of the two-step approach for managing PUL. hCG = human chorionic gonadotrophin; UPT = urine pregnancy test; PUL = pregnancy of unknown location.



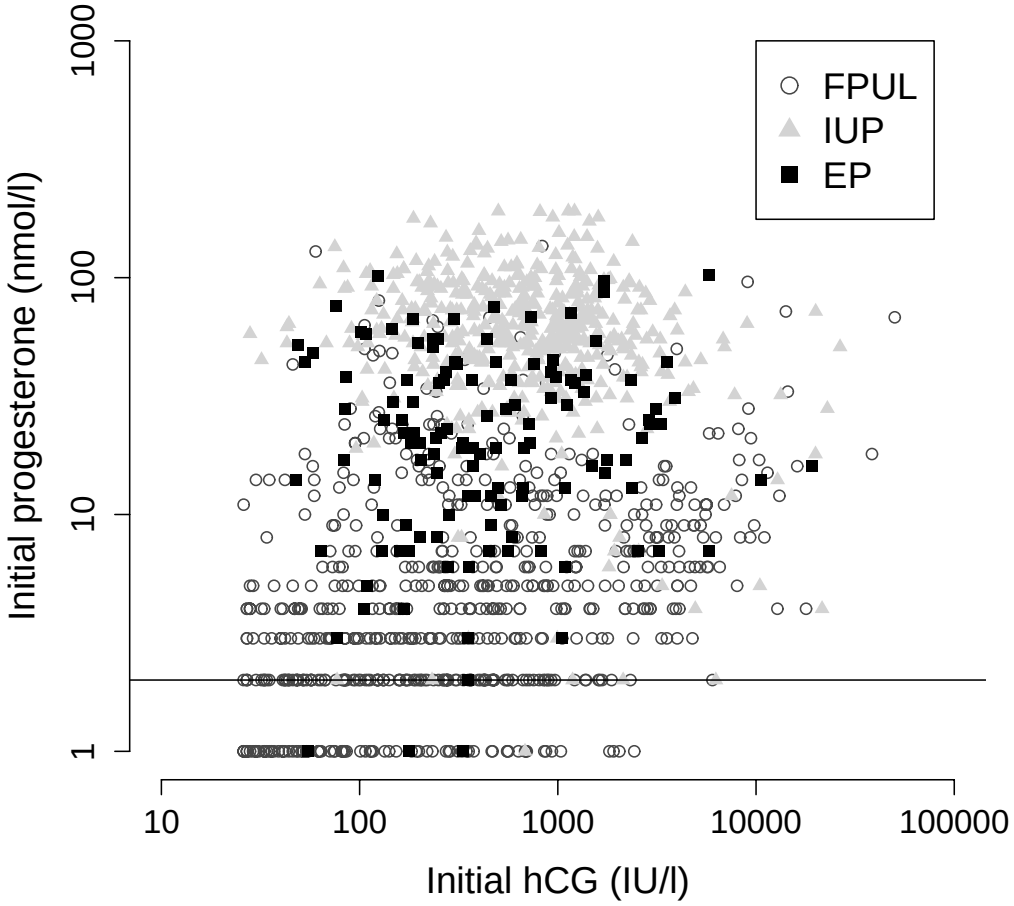
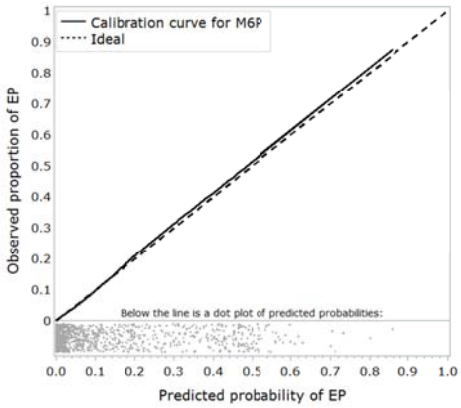
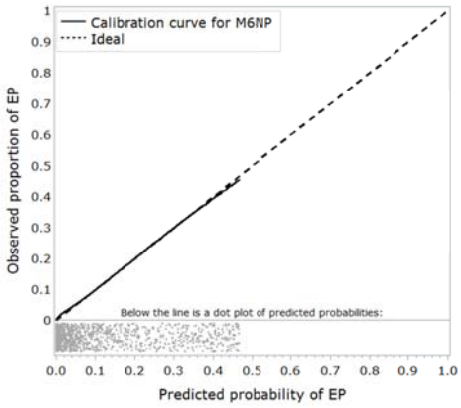



Figure 2: Scatter plot with the cut-off for step 1 of the two-step approach. Panel A is based on the development data, panel B on the validation data. A cut-off of ≤2nmol/l is used for the initial progesterone level to classify PUL as low-risk. Failed = failed PUL; IUP = intrauterine pregnancy; hCG = human chorionic gonadotrophin.



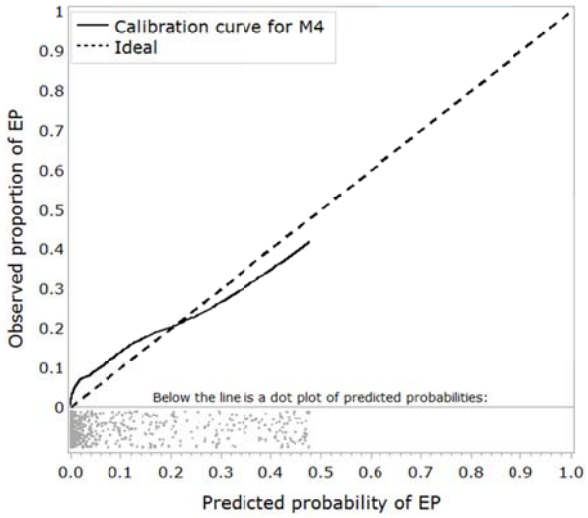



Figure 3: Calibration plots of the risk of EP. Panel A is based on the $M6_P$ model that incorporates initial progesterone as a predictor, Panel B is based on the $M6_{NP}$ model that does not incorporate progesterone, and Panel C is based on the M4 model.