Triaging women with a pregnancy of unknown location using a two-step triage protocol including the M6 model: a clinical implementation study

Shabnam Bobdiwala[1], Evangelia Christodoulou[2], Jessica Farren[3], Nicola Mitchell-Jones[4], Christopher Kyriacou[1], Maya Al-Memar[1], Francis Ayim[5], Baljinder Chohan[6], Emma Kirk[7], Osama Abughazza[8], Bramara Guruwadahyarhalli[4], Sharmistha Guha[4], Veluppillai Vathanan[6], Cecilia Bottomley[4], Debbie Gould[3], Catriona Stalder[1], Dirk Timmerman[2,9], Ben Van Calster[2,10], Tom Bourne[1,2,9#]

1. Tommy's National Centre for Miscarriage Research, Queen Charlotte's & Chelsea Hospital, Imperial College, London, United Kingdom.
2. KU Leuven, Department of Development & Regeneration, Leuven, Belgium.
3. St Marys' Hospital, London, United Kingdom.
4. Chelsea and Westminster NHS Trust, London, United Kingdom.
5. Hillingdon Hospital, London, United Kingdom.
6. Wexham Park Hospital, Slough, United Kingdom.
7. Royal Free NHS Foundation Trust, London, United Kingdom.
8. Royal Surrey County Hospital, Guildford, United Kingdom.
9. University Hospital Leuven, Leuven, Belgium.
10. Leiden University Medical Centre, Leiden, Netherlands.

#correspondence to: Professor Tom Bourne (email: t.bourne@ic.ac.uk)

Running title: Safety of using the M6 model to triage PUL

Key Words: ectopic pregnancy, miscarriage, pregnancy of unknown location, decision support techniques, ultrasonography, triage, adverse events

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/uog.20420

CONTRIBUTION

What does this work add to what is already known?

This paper is on a multi-centre interventional trial performed on 3272 prospective cases of PUL which evaluates the diagnostic performance and more importantly the safety, of using a two-step protocol, incorporating the M6 model, to manage women with a PUL in clinical practice.

What are the clinical implications of this work?

The two-step protocol, incorporating the M6 model, is an effective and clinically safe way of rationalizing the management of women with a PUL.

ABSTRACT

Introduction: The M6 risk prediction model has been shown to have good triage performance for stratifying women with a PUL as being at low or high-risk of harboring an ectopic pregnancy. There is evidence that M6 has better overall test performance than the hCG ratio (serum hCG at 48 hours/hCG at presentation) and older models such as the M4 model. M6 was published as part of a two-step protocol using an initial progesterone ≤2nmol/l to identify likely failing pregnancies (step 1), followed by M6 (step 2). This study validated the triage performance of this protocol in clinical practice by evaluating (1) the number of protocol-related adverse events and (2) how patients are effectively triaged.

Methods: This was a prospective multi-centre interventional study of 3272 women with a PUL carried out between January 2015 and January 2017 in four district general hospitals and four university teaching hospitals in the United Kingdom. We defined the final pregnancy outcome as: a failed PUL (FPUL), an intrauterine pregnancy (IUP) or an ectopic pregnancy (EP) (including persistent PUL (PPUL)). FPUL and IUP were grouped as low-risk and EP and PPUL as high-risk PUL. Patients had a serum progesterone and hCG level at 0 hours and repeat hCG at 48 hours. In seven centres, if the initial progesterone was ≤2nmol/l, patients were discharged with a follow-up urine pregnancy test in two weeks to confirm a negative result. If the progesterone was >2nmol/l or had not been taken, a 48 hour hCG level was taken and results entered into the M6 model. Patients were managed according to their predicted outcome: those classified with pregnancies likely to resolve (FPUL) were advised to perform a urine pregnancy test in two weeks and those with a likely IUP were invited for a

scan a week later. When a women with a PUL was classified as high-risk (i.e. those with a risk of EP ≥ 5%) were reviewed clinically within 48 hours. One centre used a progesterone cut-off ≤10nmol/l and was analysed separately. If the recommended management protocol was not adhered to, this was recorded as an 'unscheduled visit (participant reason)', 'unscheduled visit (clinician reason)' or 'differences in timing (blood test/ultrasound)'. The classifications outlined in UK Good Clinical Practice guidelines were used to evaluate adverse events. Data were analysed with descriptive statistics.

Results: Data was available for main analysis in 2625/3272 women with PUL (317 met the exclusion criteria or were lost to follow-up, 330 were evaluated with a progesterone cut-off of ≤10nmol/l). Progesterone results were available for 2392 (91%). 407 (15.5%) patients were classified as low-risk at step 1 (progesterone ≤2nmol/l), of which 7 (1.7%) were ultimately diagnosed with an EP. For 279/2218 remaining women with a PUL, M6 was not used or the outcome was already known usually on the basis of an ultrasound scan before a second hCG level was taken (30 were EP). 1038 women with a PUL were classified as low-risk at step 2, of which 8 (0.8%) were EP. 901 women with a PUL were classified as high-risk at step 2, of which 275 (30.5%) were EP. 275/320 (85.9%) EP were correctly classified as high-risk. Overall, 1445/2625 PUL (55.0%) were classified as low-risk, of which 15 (1.0%) were EP. None of these 15 EP ruptured.

Conclusions: This study has shown that the two-step protocol incorporating the M6 model effectively triaged a majority of women with a PUL as being at low-risk of an EP,

minimizing the follow-up required for these patients after just two visits. There were few misclassified EP and none of these women came to significant clinical harm or suffered a serious adverse clinical event. The two-step protocol, incorporating the M6 model, is an effective and clinically safe way of rationalizing the management of women with a PUL.

Introduction

Between 8-31% women seen in Early Pregnancy Assessment Units (EPAU) will be classified as having a pregnancy of unknown location (PUL)[1-6]. This is when the location of a pregnancy cannot be seen on a transvaginal ultrasound scan in a woman who has presented with a positive urinary pregnancy test. PUL is an intermediary classification whilst the final outcome is determined: failed PUL (FPUL), intrauterine pregnancy (IUP), ectopic pregnancy (EP) or persistent PUL (PPUL).

The modern management of PUL focuses on triaging women into either high or low-risk of harboring an ectopic pregnancy. High-risk PUL (i.e. probable EP/ PPUL) will need more intensive follow-up with regular blood tests and ultrasound scans. Most women with a PUL, however, will be low-risk and only require a urinary pregnancy test to confirm pregnancy failure or repeat ultrasound scan to confirm an IUP. Currently, many women in this situation undergo multiple unnecessary blood tests and scans and there is a need for protocols that triage women effectively and rationalize follow-up.

There is no clear international consensus on how best to identify which women with a PUL require more intensive follow-up. Numerous management protocols exist and these include single serum progesterone levels[7, 8], the hCG ratio (hCG 48 hours/ hCG 0 hours)[9] and risk prediction models combining these variables[10-13]. The care women with a PUL receive is therefore not consistent[14].

One of the most widely evaluated risk prediction models is M4, a logistic regression model based on the initial serum hCG and hCG ratio. A recent systematic review has shown this performs better than other commonly used management strategies[15,16]. An updated version of M4, the M6 model, is based on initial progesterone levels as well as the initial hCG and hCG ratio[13]. We previously published a two-step triage protocol based on the serum progesterone at presentation (step 1) and the M6 model (step 2). This protocol outperforms M4[13] but has never been validated. This study aimed to assess the use of the two-step protocol in everyday clinical practice by evaluating (1) adverse events related to using the protocol and (2) how patients with a PUL are effectively triaged.

Materials and methods.

Design and settings

This prospective multi-centre interventional study was carried out on consecutive women classified as having a PUL at their first visit to a dedicated EPAU in one of eight participating hospitals in the United Kingdom. Four were university teaching hospitals: Queen Charlottes' and Chelsea Hospital (QCCH), St. Marys' Hospital (SMH), Chelsea and Westminster (C&W) and West Middlesex University Hospital (WMUH). The remaining four units were district general hospitals: Hillingdon Hospital (HH), North Middlesex Hospital (NMH), Wexham Park (WXP) and Royal Surrey Hospital (RSH). Details on recruitment by each participating centre are summarized in supplementary table S1.

Consultation with a REC (Research Ethics Committee) and the research and development departments within each participating centre authorized the registration of the study as an audit, after guidance that formal ethical approval and written consent was not required from patients. Inclusion and exclusion criteria set in a previously published study[12] were also used here:

Inclusion criteria - any patient classified as a PUL after their first TVS and suitable for outpatient management.

Exclusion criteria - any participant that did not meet the classification criteria for a PUL (e.g. initial diagnosis of an EP); any patient who was haemodynamically unstable or could not be

safely managed as an outpatient (e.g. moderate to severe pelvic pain, haemoperitoneum on TVS). We initially excluded patients who had a final diagnosis of a molar pregnancy (as serum hCG levels in these patients are often markedly elevated and do not follow a trend that can be accurately interpreted by any management protocol for PUL) but on review, felt it was more clinically accurate to include these patients.

At their initial visit to an EPAU, all women were asked about their current symptoms and past medical history. A transvaginal ultrasound scan (TVS) was performed by a trained sonographer (gynaecologist, nurse specialist or sonographer). Women were classified as having a PUL if they had a positive urine pregnancy test and the location of the pregnancy could not be clearly defined on TVS. . All women with a PUL had blood samples taken for serum hCG and progesterone at the initial visit and a repeat serum hCG 48 hours later if the initial progesterone level was >2nmol/l (>10nmol/l at WMUH).

Triage using the two-step protocol

The two-step protocol was embedded into a protected Microsoft excel document that was used to determine management and allow data collection at each EPAU. After classification as a PUL, women had an initial serum hCG and progesterone level measured. These results were entered onto the excel datasheet. If the progesterone result was ≤2nmol/l, the algorithm predicted a final outcome of FPUL and advised the patient could be discharged and asked to perform a urine pregnancy test to confirm a negative result two weeks later. There was a telephone follow-up service at each unit to contact each patient and provide reassurance and

counselling as well as check the urine pregnancy test result. If the result was positive, women were advised to re-attend the EPAU for review. If the progesterone result was >2nmol/l, women were asked to re-attend 48 hours later for a repeat serum hCG blood test. The two hCG levels were entered onto the excel datasheet which contained an embedded algorithm for the M6 model which provided an instant risk prediction and recommended a management plan to the clinician using it (see figure 1). A version of M6 without progesterone as a predictor was used if the patient was on progesterone supplementation or if progesterone was not measured at all. Units were advised they did not have to measure serum progesterone levels if this was not part of their routine practice and they could move straight onto step 2. When using serum hCG results, units were advised not to use the second hCG result to obtain a risk prediction if it had been taken 8 hours before or after the 48 hour time point.

The M6 model classified a PUL case as 'high-risk' (likely diagnosis of EP/ PPUL) if the predicted risk of EP was ≥5%[11]. If the risk was <5%, the predicted final outcome was either likely 'low-risk, probable FPUL' or 'low-risk, probable IUP' depending on which of these two outcomes had the highest likelihood based on M6. Women with a 'low-risk probable FPUL', were asked to perform a urine pregnancy test 2 weeks later to ensure the result was negative. If this test was positive, the patient was brought back for review with a senior clinician and a repeat serum hCG and/or TVS. 'Low-risk probable IUP' cases were advised to have a repeat TVS to confirm the presence of an IUP after 1 week. Women with a PUL predicted to be 'high-risk, probable EP' were asked to attend the EPAU within 48 hours for senior clinical review and a repeat TVS (figure 1). If the pregnancy was still not seen, a

repeat serum hCG was taken and an individualized plan was made dependent on the clinical situation. At each visit, women were advised to seek medical advice promptly if they had any concerns or worsening of their symptoms. Clinicians were informed that the two-step protocol was a guide offering advice on how to manage a women with a PUL and should not be followed if the clinical situation required a different management strategy. Clinicians were also told that the M6 model should be used for guidance on follow-up and should not be used to determine whether medical/ surgical intervention was necessary – this responsibility always rested with the managing clinician.

Reference standard

The final outcome was defined as either[11, 12]: (1) IUP, where a gestational sac was seen within the endometrial cavity; (2) FPUL (cases where the hCG levels declined spontaneously or those that confirmed a negative urine pregnancy test result at the two week follow-up phone consultation); (3) EP, i.e. a mass outside the endometrial cavity seen on ultrasound. The appearance of the EP was either an extra-uterine gestation sac with a yolk sac +/- a visible embryo, an extra-uterine gestation sac (i.e. a 'bagel' sign) or an inhomogeneous extra-uterine mass (i.e. a 'blob' sign)[17, 18]; (4) persistent PUL, PPUL (cases where the location of the pregnancy was never confirmed on TVS, with at least three serum hCG levels taken at 48 hour intervals remained relatively static (i.e. increased/ decreased by <15% each time).

Deviations from protocol

A protocol deviation was defined as any departure from the management recommended by the two-step protocol. This included three sub-classifications:

(1) Unscheduled visit (clinician reason): these were subdivided into non-technical and technical. A non-technical clinician deviation was when the responsible clinician(s) did not adhere to the management advised by the two-step protocol. This was either due to their individualized clinical assessment of the situation, interpretation of serum hCG/ progesterone results and/or ultrasound images taken. A technical clinician deviation was when there was an error in data entry by the clinician into the excel spreadsheet or there was an error in the M6 formula embedded into the excel providing a risk prediction.

(2) Unscheduled visit (patient reason): where women made an unexpected hospital visit due to symptoms such as vaginal bleeding or abdominal pain

(3) Incorrect timing (blood test/ ultrasound scan): where the timing of a follow-up blood test or TVS did not comply with the management suggested by the two-step protocol.

For patients that had more than one deviation, we applied a hierarchy so a clinician deviation trumped a patient deviation and so on, following the order outlined above.

Data were also collected for interventions that took place (e.g. treatment with methotrexate or laparoscopic salpingectomy). The total number of blood tests and scans required to reach a final diagnosis was also documented.

Adverse events

Adverse events, as defined by national Good Clinical Practice (GCP) guidelines[19] were also recorded. These cover the regulation and guidance related to the organization of clinical trials

in the UK. An 'adverse event' (AE) is "any untoward medical occurrence in a patient or clinical study subject". A 'serious adverse event' (SAE) is "any untoward and unexpected medical occurrence or effect that results in death (or) is life-threatening". We extended this definition to involve any untoward occurrence that may have caused the death of a potentially viable IUP.

Statistical Analysis

We used descriptive statistics to analyze data, counts and percentages for categorical variables and medians and ranges for continuous variables. 95% confidence intervals (CI) were calculated using Wilson's score method. In order to describe what actually happened, we did not impute missing data.

Results

Table 1 shows the descriptive statistics for all eight participating EPAUs, further details are included in supplementary tables S2-4. A total of 3272 women from the eight participating centres were classified as a PUL during the study period (N per centre ranged between 202 and 551, table S1). Six participants met the exclusion criteria and were omitted from the final analysis, while 311 (10%) participants were lost to follow-up (figure 2; table S1). This left 2955/3272 (90%) women with a PUL for the final analyses. In 1683 (57%) the final outcome was a FPUL, there were 936 (32%) IUP and 336 EP/ PPUL (11%). One centre (WMUH, n=340) used a progesterone cut-off of ≤10nmol/l rather than ≤2nmol/l for step 1 and was therefore analysed separately. All data relating to this center is provided in supplementary information (supplementary figure S1, tables S5-8. This left 2625 patients for the main analysis.

233 participants had no initial progesterone taken. One center (HH, n=534), had a lab assay for serum progesterone that had a lower limit of 5nmol/l so all these patients automatically required triage by step 2. Therefore, step 1 of the protocol was applied to 2392 PUL. 407 women with a PUL were classified as low-risk at step 1 (15.5% of all 2625, 95% CI 14.2-16.9). 386 (94.8%) had a final outcome of FPUL, 14 (3.4%) an IUP and 7 (1.7%, 95% CI 0.8-3.5) an EP. 2218 women with a PUL required a 48-hour hCG blood test to apply step 2 (the M6 model): 233 without initial progesterone and 1985 where step 1 recommended M6. For 62 cases, the PUL outcome was known by the time M6 should be implemented whereas for 217 cases, step 2 was not applied, usually secondary to a clinical deviation. Of the

remaining 1038 women with a PUL, 727 were classified as 'low-risk, likely FPUL', of which 685 (94.2%) were FPUL, 40 (5.5%) IUP and 2 (0.3%, 95% CI 0.1-1.0) EP. 311/1038 women with a PUL were classified as 'low-risk, likely IUP', of which 305 (98.1%) IUP and 6 (1.9%, 95% CI 0.9-4.1) EP. 901/1038 women with a PUL were classified as 'high-risk', of which 200 (22.2%) FPUL, 426 (47.3%) IUP and 275 (30.5%, 95% CI 27.6-33.6) EP. The triage results by PUL outcome are summarized in table 2 and triage results by center are given in table S9. For completeness, table S10 provides triage results for cases that were lost to follow-up.

Interventions and safety: EP misclassified by the two-step protocol

Of the 320 EP among the 2625 women with a PUL, 275 were classified as high risk at step 2 (85.9%, 95% CI 81.7-89.3). In addition, 7 (2.2%, 95% CI 1.1-4.4) were classified as low-risk at step 1, and 8 (2.5%, 95% CI 1.3-4.9) as low-risk at step 2. 14 EP were visualized on ultrasonography at 48 hours (i.e. when M6 should have been applied) and for 16 EP M6 was not used. Hence, when M6 was used, 275/283 EP were classified as high risk (97.2%, 95% CI 94.5-98.6). Of the 275 correctly classified EP, 99 (36.0%) were successfully managed expectantly, 71 (25.8%) had successful medical management with methotrexate and 100 (36.4%) were managed surgically via laparoscopy. Five (1.8%) cases underwent other interventions such as a dilatation and curettage.

In 9 of 16 EP where M6 was not used, M6 results were absent due to a technical error. When using the triage protocol in hindsight, all would have been classified as high-risk. In 4/16,

second hCG results were not used by the investigator because they were taken 5-8 days after the first. In 3/16, no 2nd hCG result was available.

Step 1 misclassifications:

Seven EP were misclassified as FPUL in step 1, of which 4 were successfully managed expectantly and 3 were treated with methotrexate. Three of these cases were identified as the patient had a positive UPT two weeks later and the other four cases were brought back due to the clinical picture (1 patient had already previously had suboptimal hCG readings in the IVF unit, in 2 cases there was a suspicion of an EP on the initial TVS and 1 case had ongoing pain and PVB). None of these cases resulted in a ruptured EP or significant clinical harm. See supplementary table S11 for individual case details.

Step 2 misclassifications:

Eight EP were misclassified in step 2 (table S11). Two were misclassified as FPUL and both were successfully managed expectantly. Both of these EP were detected as they had an unscheduled repeat TVS as the clinician, based on the presentation and initial TVS findings, had been suspicious of an EP. Six cases were misclassified as an IUP in step 2, of which 1 was managed expectantly. This case was identified as the managing clinician felt the patient needed more intensive follow-up and the EP was seen on a repeat scan on day 4. Five patients required surgical intervention (laparoscopy) and the EP was diagnosed either because the patient re-presented with symptoms or a repeat ultrasound scan was carried out against protocol as the initial hCG was felt to be high. One of these cases was an interstitial EP that

underwent surgical resection which was misclassified as low-risk as the hCG levels were 60 hours apart. This EP was diagnosed due to an initial clinical suspicion of EP on the first TVS. None of these cases resulted in a ruptured EP or significant clinical harm. See supplementary table S11 for individual case details.

Protocol deviations

1918/2625 (73%) cases had no protocol deviations and clinicians followed the recommended management. 328 (12%) had a protocol deviation for a 'clinician reason' (276 of which were non-technical and the remaining 52 secondary to a technical error), 92 (4%) for a 'patient reason' and 287 (11%) for a 'timing reason'. Deviations for non-technical clinical reasons were most commonly a scan with a senior doctor at 48 hours as the initial hCG was clinically felt to be high. Other reasons included a clinical picture of a miscarriage (e.g. history of heavy PVB and features suggestive of a probable miscarriage on TVS), free fluid seen on an initial TVS which was felt to warrant more intensive follow-up than that recommended by the two-step protocol and less commonly an operator error where the model was not fully understood and a second hCG was performed despite an initially low progesterone. Sometimes this led to unnecessary intervention (e.g. a negative laparoscopy) but the rationale for deviating from the protocol was dictated by either suspicious or reassuring clinical and TVS findings. Of the misclassified EP/PPUL, 4/7 misclassified at step 1 and 3/8 misclassified at step 2 were identified due to a clinician deviating from the protocol. 312 (12%) patients had an initial hCG ≤25IU/L (the level below which a urine pregnancy test would be negative). In most of these cases, this led to clinicians not performing any further hCG or TVS follow-

up. The vast majority of 'timing deviations' were because the second hCG level was not taken +/-8 hours of 48 hours. See table 3.

Clinical Safety Data

In line with the definitions of adverse events in the GCP guidelines[19], 62 women participating in the study had an AE and no women had a SAE (table 4).

Two-step protocol-related AE: 7/320 women with an EP/ PPUL were misclassified as low-risk at step 1 and 8/320 at step 2. Sixteen women with a final outcome of EP were not classified by the M6 model either because the second hCG level was not taken at 48 hours or there was a technical error with the excel spreadsheet in giving a risk prediction.

Clinician-related AE: 10 patients underwent a negative laparoscopy. Nine of these were in women with a final outcome of FPUL/ IUP and one was ultimately a PPUL.

Patient-related AE: 14 patients had an unscheduled attendance/ admission to hospital with abdominal pain and 7 with vaginal bleeding. The majority of these cases had an earlier than scheduled repeat TVS. Thirteen patients underwent a laparoscopy which confirmed a diagnosis of ectopic pregnancy after their unscheduled attendance but none of these was a ruptured EP. A further three patients underwent a laparoscopy which turned out to be negative. Two patients underwent SMM but neither required a blood transfusion or had a prolonged hospital stay.

Ruptured ectopic pregnancies

There were 7 ruptured EP in this study. Two cases re-presented before a 48-hour hCG could be taken so there was no M6 model risk prediction. For the remaining 5 cases, all were stratified as high-risk by the model. All cases were treated promptly and none required a blood transfusion.

Medical and surgical interventions

The main analysis involved 320 EP. Of these, 6 had uterine curettage and none had been misclassified as low-risk. 3 of these cases were caesarean section scar EP, 2 were cervical EP and 1 was a PPUL with suboptimally rising hCG levels and no evidence on TVS of an extra-uterine pregnancy. 109 cases underwent successful expectant management and 5/109 had been misclassified as low-risk. 81 cases were treated medically with methotrexate, of which 3/81 were misclassified and 122 patients underwent a laparoscopy, of which 5/122 were misclassified as low-risk.

There were 2305 IUP or FPUL in the main analysis. 2092/2305 did not require intervention. 52 patients underwent a termination of pregnancy (TOP), 82 patients underwent surgical management of miscarriage (SMM), 12 patients underwent medical management of a miscarriage and 9 underwent a laparoscopy. All 9 laparoscopies were negative. For 4 cases, the laparoscopy was performed as there was suspicion of an adnexal mass on TVS and 4 patients re-presented with pelvic pain (2 had evidence of a haemoperitoneum on TVS but at

laparoscopy this was found to be secondary to a ruptured corpus luteal cyst rather than a ruptured EP. 1 of these cases required a blood transfusion). The final case had a high hCG with no visible pregnancy on TVS and was felt to clinically warrant a diagnostic laparoscopy. 58 patients did not attend (DNA) their follow-up appointment(s), see table 5.

Number of blood tests and scans

The mean number of scans per woman was 1.2 for PUL with a final outcome of FPUL, 2.2 for IUP and 2.5 for EP. The mean number of blood tests was 1.9 for PUL with a final outcome of FPUL, 2.2 for IUP and 3.1 for EP (see table 6).

Discussion

We have shown that a two-step protocol using a serum progesterone cut-off level of ≤2nmol/l and the M6 decision support model performs well in clinical practice. This approach demonstrated good test performance and was associated with few adverse outcomes. Of the 320 women with an EP/PPUL, only 15 were misclassified as low-risk and none came to clinical harm.

A strength of the study is that it is multi-centre, involving a number of district and university hospitals. Furthermore, a large number of women with a PUL were included. A further strength is that we prospectively demonstrated the performance of the protocol in every day clinical practice. The protocol was implemented by a range of clinicians including nurses, trainee doctors and more senior clinicians. Accordingly, we feel the results are likely to be generalizable to other countries able to provide dedicated early pregnancy care with the appropriate expertise. There are two versions of the M6 model designed to be used with or without progesterone results to accommodate women using progesterone supplementation and units where measuring serum progesterone is not a routine part of clinical practice. PUL management protocols have previously either utilized serum progesterone or serum hCG levels, whereas the current study incorporates the utility of both biomarkers.

One of the limitations of the study was that 10% patients were lost to follow-up, although these were largely women with a predicted pregnancy outcome of FPUL whom we could not contact successfully to confirm a negative urine pregnancy test result after two weeks. In 2%

(n=63) of cases there was a technical error with using the excel datasheet to provide a risk prediction by the M6 model (table 3). This occurred as we needed to collate detailed demographic and outcome data on all participants. However the model itself is available for use free via a website (www.earlypregnancycare.co.uk) which allows fewer operator-led errors to occur. Human factor errors with incorrect data entry will remain a limitation in any clinical study.

We found few women with an EP/ PPUL were misclassified and of the 15 that were, none resulted in a ruptured ectopic or the patient coming to significant clinical harm. These cases were usually identified either because the clinician decided to go against the management recommended by the protocol or the patient had an unscheduled visit to hospital with symptoms such as abdominal pain or vaginal bleeding. Clinicians were instructed to deviate from the recommended management if they felt there was a clinical reason to do so and it was encouraging to see that there were sensible protocol deviations when required. Furthermore, they were instructed to counsel patients (including the provision of written information where possible) about the risk of a potential EP and to re-present if they had any concerning symptoms. It was reassuring that patients in the study both appeared to understand and comply with this advice.

There were seven ruptured EP in this study. None were classified as low-risk at step 1. Two cases re-presented before an M6 risk prediction could be calculated (i.e. before a 48 hour hCG test could be performed). The other 5 cases were all correctly classified as high-risk by

the M6 model but underwent subsequent unsuccessful expectant or medical intervention. This remains a risk with non-surgical treatment options and it has been estimated that 7% cases treated with methotrexate will experience a tubal rupture during follow-up[20]. The two-step protocol is solely meant to advise on the scheduling of follow-up and is not meant to determine if conservative management (i.e. methotrexate or expectant management) is appropriate. This was made explicit to participating centres prior to use.

There were 10 negative laparoscopies in the study and for the purposes of the study we classified these as an 'adverse event'. The reasons for proceeding to surgery were, however, clinically justifiable as either there was a suspicion of an adnexal mass/ haemoperitoneum on TVS or the patient re-presented with pelvic pain. It could reasonably be argued that these were not true adverse events as one would expect a proportion of negative laparoscopies in the safe management of women with a PUL.

The management of women with a PUL can be haphazard and lack a clear evidence-base. This two-step protocol is able to guide clinicians in their decision making so they can provide consistent, evidence-based care. It allows the streamlining of follow-up for those that do not need it and therefore the opportunity to allocate resources to those that are at greatest clinical risk of an EP. This large prospective, multi-centre study has now shown it maintains good test performance with few adverse events when used by clinicians in their daily medical practice. It is available for use as a free online application at www.earlypregnancycare.co.uk and as an application that can be downloaded onto smartphones or tablets (search 'early

pregnancy Leuven') for clinicians wishing to use it.

Funding

SB is supported by NIHR CLAHRC NWL (Collaboration for Leadership in Applied Health Research & Care, NorthWest London grant RDIP033). TB is supported by the NIHR Biomedical Research Center based at Imperial College Healthcare NHS Trust and Imperial College 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). EC, BVC, and DT are supported by Research Foundation – Flanders (FWO) grant G0B4716N and Internal Funds KU Leuven grant C24/15/037.

References

1. 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**: 1744-1750.
2. Kirk E, Condous G, Bourne T. Pregnancies of unknown location. *Best Pract Res Clin Obstet Gynaecol* 2009; **23**: 493-499.
3. 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**: 250-261.
4. Hahlin M, Thorburn J, Bryman I. The expectant management of early pregnancies of uncertain site. *Hum Reprod* 1995; **10**: 1223-1227.
5. Banerjee S, Aslam N, Zosmer N, Woelfer B, Jurkovic D. The expectant management of women with early pregnancy of unknown location. *Ultrasound Obstet Gynecol* 1999; **14**: 231-236.
6. van Mello NM, 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**: 603-617.
7. 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**: 693-697.
8. Condous G, Okaro E, Khalid A, Lu C, Van Huffel S, Timmerman D, Bourne T. A prospective evaluation of a single-visit strategy to manage pregnancies of unknown location. *Hum Reprod* 2005; **20**: 1398-1403.
9. Condous G, Lu C, Van Huffel SV, Timmerman D, Bourne T. Human chorionic gonadotrophin and progesterone levels in pregnancies of unknown location. *Int J Gynaecol Obstet* 2004; **86**: 351-357.
10. 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; **29**: 680-687.
11. 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**: 609-616.
12. Bobdiwala S, Guha S, Van Calster B, Ayim F, Mitchell-Jones N, Al-Memar M, Mitchell H, Stalder C, Bottomley C, Kothari A, Timmerman D, Bourne T. The clinical performance of the M4 decision support model to triage women with a pregnancy of unknown location as at low or high risk of complications. *Hum Reprod* 2016; **31**: 1425-1435.
13. Van Calster B, Bobdiwala S, Guha S, Van Hoorde K, Al-Memar M, Harvey R, Farren J, Kirk E, Condous G, Sur S, Stalder C, Timmerman D, Bourne T. Managing pregnancy of unknown location based on initial serum progesterone and serial serum hCG levels:

development and validation of a two-step triage protocol. *Ultrasound Obstet Gynecol* 2016; **48**: 642-649.
14. Bobdiwala S, Al-Memar M, Farren J, Bourne T. Factors to consider in pregnancy of unknown location. *Womens Health (Lond)* 2017; **13**: 27-33.
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**: 938-945.
16. Bobdiwala S, Saso S, Verbakel JY, Al-Memar M, Van Calster B, Timmerman D, Bourne T. Diagnostic protocols for the management of pregnancy of unknown location: a systematic review and meta-analysis. *BJOG* 2018. DOI:10.1111/1471-0528.15442.
17. Condous G, Okaro E, Khalid A, Lu C, Van Huffel S, Timmerman D, Bourne T. The accuracy of transvaginal ultrasonography for the diagnosis of ectopic pregnancy prior to surgery. *Hum Reprod* 2005; **20**: 1404-1409.
18. Kirk E, Daemen A, Papageorghiou AT, Bottomley C, Condous G, De Moor B, Timmerman D, Bourne T. Why are some ectopic pregnancies characterized as pregnancies of unknown location at the initial transvaginal ultrasound examination? *Acta Obstet Gynecol Scand* 2008; **87**: 1150-1154.
19. (MHRA). MaHpRA. The Good Clinical Practice Guide. Medicines and Healthcare products Regulatory Agency (MHRA). In *TSO (The Stationery Office)*.(MHRA) MaHpRA (ed). 2014.
20. Yao M, Tulandi T. Current status of surgical and nonsurgical management of ectopic pregnancy. *Fertil Steril* 1997; **67**: 421-433.

Figure legends

Figure 1: Flowchart of the two-step protocol after inputting the initial (0 hour) serum progesterone (nmol/l) results and 0 and 48 hour hCG (IU/L) results. PUL = pregnancy of unknown location; hCG = human chorionic gonadotrophin; EP = ectopic pregnancy; FPUL = failed pregnancy of unknown location; IUP = intrauterine pregnancy; US = ultrasound; UPT = urine pregnancy test; hr = hour

Figure 2: Flowchart of patient recruitment, final outcome data and correct versus incorrect risk stratification by the M6 model. FPUL = failed pregnancy of unknown location; IUP = intrauterine pregnancy; EP = ectopic pregnancy; hCG = human chorionic gonadotrophin; prog = progesterone; M6NP = M6 model no progesterone (progesterone result not available/ not used in model risk prediction)

Table 1: Pooled descriptive statistics from all participating early pregnancy units except West Middlesex Hospital, n=2625. PUL = pregnancy of unknown location; hCG = human chorionic gonadotrophin; MTX = methotrexate; IQR = interquartile range

Patient characteristic	**Median (IQR, range), or N (%)**	**Missing values, N (%)**
Age	32 (27-36, 14-50)	2 (0.1%)
Initial progesterone (nmol/L)	10 (4-38, 0.3-219)	291 (11.1%)*
Initial serum hCG (IU/L)	407 (92-1257, 0.1-105006)	0 (0%)
2nd serum hCG (IU/L)$	496 (121-1345, 0.5-109658)	500 (19.0%)#
2nd serum hCG if at D2 (IU/L)	515 (126-1361, 0.8-109658)	511 (19.5%)
hCG ratio$	0.88 (0.34-1.91, 0.002-25.0)	500 (19.0%)#
hCG ratio if 2nd hCG at D2	0.96 (0.36-1.95, 0.01-6.31)	511 (19.5%)
Pain score (0-10)	3 (0-4, 0-10)	113 (4.3%)
Vaginal bleeding		9 (0.3%)
No (0)	688 (26%)	
Minimal (1)	728 (28%)	
Moderate (2)	616 (24%)	
Soaked (3)	295 (11%)	
Clots (4)	289 (11%)	
History of EP		7 (0.3%)
None	2446 (93%)	
Any	172 (7%)	
Laparoscopy	124 (5%)	
MTX	26 (1%)	
Expectant	22 (1%)	
Indication for scan		3 (0.1%)
Bleeding and pain	1212 (46%)	
Bleeding only	705 (27%)	
Pain only	485 (18%)	
Maternal reassurance	72 (3%)	
Unsure dates	65 (2%)	
Previous ectopic	42 (2%)	
Previous miscarriage	41 (2%)	
Other	0 (0%)	
Type of PUL		5 (0.2%)
True PUL	1323 (50%)	
Probable miscarriage	831 (32%)	
Probable intra-uterine	338 (13%)	
Probable ectopic	128 (5%)	

$ This describes any 2nd hCG level recorded in the database, irrespective of interval with the initial measurement and irrespective of whether this 2nd level was used to calculate the M6 result.

* This includes known progesterone levels from cases who were on progesterone supplements
# This includes cases where step 1 triage was ‘low-risk’ and therefore a 2nd hCG was not indicated

Table 2: Overview of patient triage, overall and by final PUL outcome.

Final classification	N	Outcome		
		FPUL N (%)	IUP N (%)	EP N (%)
Low risk				
at Step 1 (probable FPUL)	407 (16%)	386 (27%)	14 (2%)	7 (2%)
at Step 2 (all)	1038 (40%)	685 (48%)	345 (40%)	8 (3%)
Probable FPUL	727	685	40	2
Probable IUP	311	0	305	6
High risk (probable EP)	901 (34%)	200 (14%)	426 (49%)	275 (86%)
Outcome known at D2	62 (2%)	1 (<1%)	47 (5%)	14 (4%)
Not classified	217 (8%)	161 (11%)	40 (5%)	16 (5%)
Total	2625	1433	872	320

Table 3: Protocol deviations and other events associated with two-step algorithm use (n=2625). Step 1 = progesterone cut-off of ≤2nmol/l. Step 2 = M6 risk prediction model. hCG = human chorionic gonadotrophin

	n (%)
Protocol deviation	
No protocol deviation	1918 (73.1%)
Clinician reason	328 (12.5%)
- non-technical*	- 276 (10.5%)
- technical*	- 52 (2.0%)
Patient reason	92 (3.5%)
Timing reason	287 (10.9%)

* A technical deviation relates to incorrect data entry or a software malfunction causing an incorrect M6 risk prediction. A non-technical clinician deviation relates to the clinician making a judgment that they would not adhere to the management recommended by the M6 model.

Table 4: Breakdown of adverse events (AE) in the study cohort. EP = ectopic pregnancy; MTX = methotrexate; PVB = per vaginal bleeding

Adverse event	**N**
Model related	
EP classified as at low risk in step 1	7/320
EP classified as at low risk in step 2	8/320
EP not classified (no 2nd bloods, technical error)	16/320
Clinician related	
Negative laparoscopies*	10
Incorrectly given MTX#	0
Patient related	
Unscheduled attendance with pain	14
Unscheduled attendance with bleeding	7

* 9 of these were in patient who had a final outcome of FPUL/ IUP, 1 had a final outcome of PPUL.
# MTX given to women with an intrauterine pregnancy

Table 5: Triage results by outcome and intervention. FPUL = failed pregnancy of unknown location; IUP = intrauterine pregnancy; EP = ectopic pregnancy; PPUL = persistent pregnancy of unknown location; MTX = methotrexate; MMM = medical management of miscarriage; SMM = surgical management of miscarriage; DNA = did not attend; TOP = termination of pregnancy

Outcome	**Intervention**	**N**	**Triage result**			
			Unkown	**Sac seen on D2**	**High risk**	**Low risk**
EP/PPUL	SMM	6		1	5	
	Expectant	99	2	1	99	7
	MTX	81	6	1	71	3
	Laparoscopy	123	8	11	100	4
	Other	1				1
	All	320	16	14	275	15
Failed PUL/IUP	Expectant	2092	187	33	513	1359
	TOP	52	4	8	22	18
	SMM	82	3	3	48	28
	MMM	12		1	9	2
	Laparoscopy	9			6	3
	DNA	58	7	3	28	20
	All	2305	201	48	626	1430

Table 6: Number of scans and blood tests by outcome. FPUL = failed pregnancy of unknown location; IUP = intrauterine pregnancy; EP = ectopic pregnancy; SD = standard deviation

Outcome	Number of scans		Number of blood tests	
	Mean (SD), range	Not documented	Mean (SD), range	Not documented
All (n=2625)	1.7 (0.84), 1-9	676	2.1 (0.95), 1-8	726
FPUL (n=1433)	1.2 (0.62), 1-9	443	1.9 (0.9), 1-7	449
IUP (n=872)	2.2 (0.57), 1-5	164	2.2 (0.7), 1-5	202
EP (n=320)	2.5 (0.85), 1-6	69	3.1 (1.3), 1-8	75