Management of pregnancies of unknown location – a randomized controlled trial of two hCG- based decision support models.

Principal investigator: Johan Fistouris, Göteborg johan.fistouris@vgregion.se
Co-workers: Annika Strandell (main supervisor), Göteborg, Christina Bergh (co-supervisor), Katja Bergman, Södra Älvsborg Hospital (SÄS), Helen Garbergs and Klara Törneke, Skaraborg Hospital Skövde.

Revised version 1.1, November 09, 2020.
NCT 03461835 (www.clinicaltrials.gov)

The following changes have been made to the study protocol:
Co-workers
Section 2.2.3
Section 3.2
Section 5.1
Section 8.2

1. Background

Pregnancy of unknown location (PUL) is a term used to describe a situation where there is no evidence of a gestation either intra-or extrauterine on transvaginal sonography (TVS) in women seeking medical attention for various reasons in early pregnancy (1). The principal tool in PUL management besides TVS is the interpretation of the change in serial serum human chorionic gonadotropin (hCG) levels. In many cases (50-70%) the location will never be known; hCG levels resolve without intervention and a pregnancy is never seen on TVS (2). However, a smaller part (6-20%) constitutes of ectopic pregnancies (EP) with an inherent risk of heavy bleeding if diagnostically and therapeutically neglected before rupture or (30-47%) will progress to a normal intrauterine pregnancy (IUP), initially too small to visualize on TVS (2). After just two hCG measurements contemporary practice wish to select PULs that is of high risk of being an EP regardless of rising or declining hCG levels, for an early repeat TVS, while minimizing the burden for the patient and the clinic by correctly selecting PULs with low risk of being an EP to a limited follow-up.

Depending on TVS skills among other things, the prevalence of PUL may vary between units and thus the workload, since many PULs require repeat clinical visits before the final location can be established and an EP needing intervention ruled out, why an efficient and safe triage system is needed.

A recent meta-analysis (3) concluded that there is not enough evidence to recommend a specific strategy using hCG when managing PULs. In a NICE guideline (4) however women with a PUL acceptable for outpatient surveillance, where the rate of change in rising hCG levels surpassing 63 % in 48 hours (h) highly likely represents an IUP and the patient is triaged to a TVS after one week. Those with a decline in hCG levels of more than 50 % are predicted to be failed PULs and a urinary pregnancy test (UPT) after two weeks is appropriate. If the rate of change in hCG levels fail to meet these cut-offs a re-visit within 24 h is warranted in order to possibly diagnose an EP and suitable treatment selected. The efficiency by implementing these cut-offs and management protocol a reduction in hCG sampling has been reported (5) with implied cost- benefits. Also, in another study the patient safety with no ruptured EPs or blood transfusions among 835 PULs was reported (6). In the latter prospective interventional study 85% of all patients were followed-up according to the PUL management protocol but an hCG based logistic regression model, M4, was used instead of the cut-offs published by NICE for classification of the PULs as high or low risk of being an EP.

In a previous study (7); M4 classified 70 % of all PULs as low risk of being an EP, hence qualifying for a reduced follow-up while correctly identifying 88 % of EPs. A study from Sahlgrenska University hospital (SU) evaluated the M4 model with similar results (8), where also the cut-offs by NICE similarly identified 86 % of EPs but to a lesser extent correctly classified low risk PULs. Another finding in the study was that EPs identified by M4 mainly had rising hCG levels opposed to the NICE model, identifying mainly EPs with declining hCG levels which could have clinical consequences since EPs with rising hCG levels more often

seem to necessitate laparoscopic surgery (9) and maybe a postponed diagnosis could be deleterious in these cases. Hence, the M4 model potentially would be better than the NICE cut-offs for PUL management. None of these findings however have been evaluated in a randomized controlled trial (RCT).

2. Objectives

In a randomized multicenter trial, we want to compare the ability of two hCG-based models in correctly classifying EPs as high risk among PULs and correctly classifying IUPs and failed PULs as low risk after two hCG measurements.

After the classification of PULs into high or low risk of EP, the clinical management will be the same within each risk group (high and low) regardless of randomization group (which model classified the patients). The clinical management will be according to the management protocol published in the NICE guideline.

2.1 Primary objectives (Diagnostic performance of the models)

1. To examine if the sensitivity (to identify EP) is non-inferior for M4 compared with the NICE model.
2. To examine if the specificity (to identify IUP or failed PUL) is superior for the M4 model compared with the NICE model.

2.2 Secondary objectives (Clinical outcomes)

2.2.1 To examine if the M4 model compared with the NICE model during the entire follow-up until a final diagnosis has been reached results in a difference in:

-Number of hCG measurements and TVSs within the final outcomes of IUP and failed PULs
-Number of serious adverse events (SAE) and adverse advents (AE) (10)
-Number of management protocol deviations (unscheduled visits initiated by the patient or the physician) during the initial follow-up stated by the algorithm
-Time to final diagnosis (days) by the means of TVS, histology or perioperative
-Number of diagnostic and therapeutic surgical and medical procedures
-Number of persistent PULs (PPULs)
-Sick leave (days)
-Duration of hospital stay (days)
-Overall economic cost to the clinic and the society due to the number of doctors' visits, blood tests, diagnostic and therapeutic interventions and sick leave.
-Number of days until resolution of trophoblast (calculated from randomization) by either;
    -Surgery, measured by hCG (< 5.0 IU/L) if tubotomy
    -Medical treatment (Methotrexate (MTX)) measured by hCG (< 5.0 IU/L)
    -Expectant management measured by hCG (< 5.0 IU/L)

2.2.2 To examine within each model if EPs correctly classified as high risk or failed PULs and IUPs correctly classified as low risk benefit in regard of clinical efficacy and safety in comparison to those being misclassified. That is, if it will result in a difference in:

-Number of hCG measurements and TVSs within the final outcomes of IUP and failed PULs

-Number of SA and AE (10)

-Number of ruptured EPs

-Number of emergency surgeries due to a bleeding EP during follow-up stated by the protocol and during the entire follow-up

-Perioperative blood loss and number of blood transfusions

- Number of accidental disruptions of a possibly viable IUP

-Number of protocol deviations (unscheduled visits (initiated by the patient or the physician)) during the initial management

-Time to final diagnosis (days) by the means of TVS, histology or perioperative

-Time to resolution of an EP

-Number of diagnostic and therapeutic surgical and medical procedures

-Number of PPULs

2.2.3 To evaluate the presence and severity of psychological distress and quality of life (QoL) among women with a PUL by using two self-reporting questionnaires; Hospital anxiety and depression scale (HADS) reported at one and four weeks after the first visit and the 36-Item Short Form Health Survey (SF-36) reported four weeks after the first visit. The HADS comprises 14-items divided in two seven-item subscales that measure anxiety and depression symptoms. For each of the two subscales, a total score is obtained by adding the scores of the seven items in each subscale. The total scores of each subscale range from 0 to 21, a higher score representing more symptoms. A cut-off score ≥ 8 is used for both subscales to identify depression and anxiety. Control subjects constituting women with a normal early pregnancy also fill out the HADS one week after their first visit to the maternity care, enabling comparisons of both dichotomous and continuous results.

SF-36 consists of 36 items covering eight subscales of health, both physical and mental. The eight subscales can be aggregated into two summary scores: the Physical Component Summary and the Mental Component Summary. The summary scores are transformed to a 0–100 scale, with higher scores indicating better quality of life. The interpretation of SF-36 is based on norm-based scores for the Swedish population with a mean of 50 and a standard deviation of 10.

2.3 Tertiary objective (based on a cohort of the first 600 patients)

2.3.1. To compare the diagnostic performance of alternative cut-offs of decreasing hCG levels to identify patients with a spontaneously resolving pregnancy (failed PUL).

3. Endpoints

3.1 Primary endpoints

Diagnostic and predictive performance measured as sensitivity, specificity, and positive predictive value (PPV) and negative predictive value (NPV) and positive and negative likely hood ratio (LR+, LR-). The sensitivity represents EPs correctly classified as high risk of being an EP and the specificity represents IUPs and failed PULs correctly classified as low risk of being an EP. PPV represents EPs among PULs classified as high risk and NPV represents IUPs and failed PULs among PULs classified as low risk.

3.2 Secondary endpoints

I. Clinical efficacy and safety.

-*Number of hCG and TVS* performed overall and within each outcome (EP, IUP, and failed PUL) during the entire follow-up

-Number of *SA* and *AE*

A serious adverse event (experience) or reaction is any untoward medical occurrence that:

• Results in death,

• is life-threatening (10)

- Deviation from PUL management protocol- number of unplanned visits, initiated by the physician or the patient

-Length of follow-up from randomization until final diagnosis (days)

-TVS diagnosis at the first clinical examination after the initial two hCG measurements for PULs classified as either high risk or predicted to be IUPs (low risk)

-Number of surgical and medical procedures

-Duration of hospital stay (days)

-Duration of sick leave (days)

-Number of persistent PULs (PPUL)

-Number of days from randomization until diagnosis of an EP and resolution of hCG after randomization and diagnosis (surgical removal, MTX or expectant management)

II. Cost- analysis based on the number of visits (scheduled and unscheduled clinical examinations and blood tests), hospital stay and diagnostic and therapeutic interventions (medical or surgical)

III. Time to subsequent pregnancy in months and cumulative pregnancy rate (and time to pregnancy resulting in a live birth) calculated from the date of resolution of an EP, PPUL, failed PUL or a non-viable IUP to the first day of the last menstrual period (LMP) before the conception among those with a wish for pregnancy.
IV. The percentage points difference after one week in the incidence of a HADS anxiety score ≥ 8 between IUP starting as PUL compared to women with normal early pregnancies (control subjects). The mean score on each subscale is also calculated for group comparison. Health-related quality of life obtained from the SF-36 where all questions are scored on a scale from 0 to 100, with 100 representing the highest level of functioning possible.
V. Investigate the performance of the models when clinical factors such as risk factors of EP, clinical findings or symptomatology are part of the PUL prediction.

4. Study design

4.1 Generally

The study is conducted at gynecological emergency departments in academic hospitals or teaching hospitals on women with a PUL while managed as outpatients by using serial hCG measurements.

4.2 Study design

The study is a national multicenter RCT. The study has a superiority design regarding the primary endpoint of specificity for M4, given that the primary endpoint of sensitivity is not inferior (non-inferiority design) in comparison to the NICE model's cut-offs.

4.3 Randomization procedure

Two subsequent hCG values of each study patient will be entered into a web-based randomization program. Timing of randomization will be as soon as the second hCG value has been registered. A randomization code concealed to both investigators and patients will be generated and allocation to either of two models (M4 (model 1) or NICE (model 2)) after stratification by center. Randomization will be performed in a ratio 1:1, i.e., half of the patients are allocated to model M4 and half to the NICE model. The randomization program will control that an even distribution between randomization groups within each center, occurs over time, either by applying variable block sizes or optimal allocation.

After randomization a classification (high or low risk for the PUL being an EP), is made by the allocated model and a predicted outcome, i.e., IUP failed PUL or EP is also presented.

4.3.1 Model classification

Model 1 (M4): High risk classification if the chance of the PUL being an EP ≥5 %; a calculation based on the average value of the two hCG and the ratio of these two hCG (0

h/48 h). Otherwise, a low-risk classification is made and a predicted outcome of either an IUP or failed PUL is presented.

Model 2 (NICE): High risk classification if the change in rising hCG levels ≤ 63 % or the change in declining hCG ≤ 50 %. If these cut-offs are exceeded the PUL is classified as low risk and predicted to be either an IUP or failed PUL depending on rising or declining hCG levels.

4.4 Blinding

The study will be double-blind. At the time of randomization, the allocated model will not be known to the participant. The allocation will be concealed to the investigator until recruitment has been completed, at the onset of data analysis. However, it cannot be ruled out that the investigator assumes about which model that has been used, by subjective interpretation of hCG development and having knowledge about the models.

4.5 Intervention

- Randomization after two hCG measurement.

- Initial management of the patients is not based on the randomization groups, but on classification to high or low risk for EP. The initial clinical management follows the PUL management protocol published by NICE.

1. High risk PULs are examined again within 24 h with TVS, and a repeat hCG measurement if the pregnancy location remains unknown.
2. Low risk PULs with
    a. a decline in hCG levels is asked to take a home UPT after two weeks and will be contacted by a nurse- a clinical follow-up is arranged if the UPT is positive.
    b. a rise in hCG levels is followed-up after one week with TVS, and a repeat hCG measurement if the pregnancy location remains unknown.

- After the second hCG measurement study patients receive two self-reporting questionnaires regarding psychological distress to be answered at home during the follow-up.

- A two-year questionnaire will be sent to all patients with a non-viable pregnancy, to obtain subsequent pregnancy outcome.
- All hCG results as well as the clinical information given by the electronical medical record will be reviewed by a physician as per current practice and if he or she is not content by the management given by the randomization program, it can be overruled and the patient managed differently. After the initial management further follow-up

and interventions of suspected miscarriages and EPs will be according to local practices of each participating center.

5. Patients

5.1 Selection of patients

All women seeking medical attention due to early pregnancy concern at any of the participating centers are screened for study eligibility. A patient can only be screened once during the same pregnancy. They receive oral and written study information upon arrival at the unit by the consulted midwife. After the first examination eligible patients that are willing to participate in the study sign an informed consent together with the enrolling physician. At the time for the second hCG sampling eligible patients will be contacted directly by assistant investigators for further study information and those not already included will be asked to participate in the study.
Patients not eligible or not participating for other reasons in the study are suggested to be classified according to the NICE model and managed according to the NICE management protocol or a local protocol if available.

Women with a normal early pregnancy are recruited opportunistically from the local maternity care and fill out HADS forms and a questionnaire. Only women with a normal pregnancy without complications during early pregnancy are included as controls. Women who do not understand Swedish are excluded from participation.

5.2 Number of patients. Sample size calculation.

Approximately 712 patients with a PUL were included in a previous study (8) during a three-year period at the SU, of which all would be eligible for the present study. Thus, approximately 250 patients per year could be eligible at the SU clinic and approximately the same number (or less) in the three regional hospitals. Thus, 400-500 patients would be eligible for randomization per year within Region Västra Götaland (VGR).
Based on data in Fistouris et al. (2016), model 1 (M4) and model 2 (NICE) classified 89% and 86% of EPs as high risk (sensitivity) respectively (p=0.52). The specificity (IUPs and failed PULs classified as low risk) were 78% and 70% for model 1 and model 2 respectively (p=0.01). Five percent of patients were lost to follow-up, while another study reported a 7% lost to follow-up (6). We have estimated the maximum rate of loss to follow-up to be 10%.

We calculated that we would need 121 women in each group to reach a power of 80 % to detect an absolute difference of 15 percentage points after one week, with a 15 % incidence of HADS anxiety score ≥ 8 among control subjects and 30 % among IUP starting as a PUL at a two-sided alpha level of 0.05 (20). For protection against that 10 % of patients will be lost to follow-up a total of 130 women are included in each group. In a second calculation based on

using HADS as a continuous scale, 94 women would need to be included in each group to detect an absolute two points difference on the anxiety subscale of the HADS assuming a mean score of 6.0 and a standard deviation of 3.4 in control subjects, at a significance level 0.05 and power 0.80. This minimally important difference is based on mean scores on the HADS anxiety subscale reported for women in early pregnancy in prior studies (20-22). We plan to include a total of 210 women with a 1:1 ratio to account for an expected 10 % loss to follow-up.

5.2.1 Diagnostic performance; sensitivity (non-inferiority)
The reference sensitivity is 86% (model 2) and we would allow a one percentage point lower sensitivity for model 1 and that the lower limit of the two-sided 95 % confidence interval (CI) for the difference between the two models should not exceed seven percentage points with a probability of 80% (β=20%), to be a non-inferior difference. Given a significance level (α) of 5% and a two-sided test, 541 patients per randomization group are needed. For protection against that 10 % of patients will be lost to follow-up a total of approximately 1200 patients are needed for randomization.
5.2.2 Diagnostic performance; specificity (superiority)
The reference specificity is 70% (model 2). Considering eight percentage points to be a clinically important difference in specificity, given β=20%, α=5% and a two-sided test, 495 patients are needed per group to demonstrate superiority. For protection against that 10 % of patients will not be lost to follow-up a total of approximately 1100 patients are needed for randomization.
Thus, 1200 patients will be recruited and randomized.

5.3 Inclusion criteria

1. Patients with a PUL
2. Patients with mild or no symptoms
3. Patients willing to be randomized
4. First hCG value <10,000 IU/L
5. Patients with an interval between two hCG measurements of 44-56 h

5.4 Exclusion criteria

1. Patients with an IUP
2. Hemodynamically unstable patients
3. Hemoperitoneum
3. Patients not managed as outpatients during the initial two hCG measurements
4. non-understanding of the oral or written study information

6. Variables for analysis

Demographic variables
The following variables are recorded: age, ethnicity, reason for seeking medical attention (i.e. bleeding, pain, dating of pregnancy, previous complications of early pregnancy), gestational age according to LMP, conception desired or not, a history of infertility, fertility treatment, previous pregnancies and locations of previous pregnancies, parity, previous

tubal surgery, known tubal pathology, history of pelvic inflammatory disease or sexually transmitted infection, smoking habits, presence of an intrauterine device. Previous and current psychological morbidity including medication. General health issues. Highest level of education. Perceived spousal support.

Variables related to PUL management.

The following variables are recorded: TVS competence by the examining physician (certified ultrasound course, years of training, specialist), unscheduled follow-up visits outside the triage protocol framework (protocol deviation initiated by patient or physician), reasons for altered follow-up plan initiated by patient or physician, number and values of hCG samples and corresponding interval (h) between two hCG measurements, number and timing of TVS performed. Findings on TVS at each time performed. Days from randomization until a final diagnosis are established. Number of ruptured EPs. Number of emergency surgeries due to a bleeding EP during follow-up stated by the algorithm and during the entire follow-up. Perioperative blood loss and number of blood transfusions. Number of accidental disruptions of possibly viable IUPs. Drug related adverse events.

Biochemical analysis: At SU, Electrochemiluminescence immunoassay (ECLIA, Cobas®) will be used for the in vitro quantitative determination of hCG (intact hCG and free β –hCG) in serum and plasma (short turnaround time). The assay is standardized against the 4th International Standard for Chorionic Gonadotropin from the National Institute for Biological Standards and Control (NIBSC) 75/589. Results are expressed as milli-International Units per milliliter (mIU/mL). No detectable cross-reactivity for thyroid-stimulating hormone, luteinizing hormone 0.12% and follicle-stimulating hormone <0.1%. The functional sensitivity of the assay is <0.6 mIU/mL and it has a precision of <10% total interassay coefficient of variation (CV).

Categorization of TVS criteria modified from Barnhart et al. (1):

1. **Definitive EP**: extrauterine (i.e., tubal, cervical, caesarean scar, interstitial, corneal, ovarian, abdominal and heterotopic) gestational sac with yolk sac and/or embryo (with or without cardiac activity)
2. **Probable EP**: inhomogeneous adnexal mass or extrauterine (i.e., tubal, cervical, caesarean scar, interstitial, corneal, ovarian, abdominal and heterotopic) saclike structure.
3. **PUL**: no signs of either EP or IUP
4. **Probable IUP**: intrauterine gestational sac-like structure
5. **Definitive IUP**: intrauterine gestational sac with yolk sac and/or embryo (with or without cardiac activity).

The terminology is customized to current published literature:

- Intrauterine is used in the same sense as eutopic, i.e., any pregnancy implanted inside the endometrial cavity
-Extrauterine is used in the same sense as ectopic, i.e., any pregnancy implanted outside of the endometrial cavity.

Final outcomes of a PUL (1, 6):

I. A *failed PUL* is defined by spontaneously declining hCG levels to < 5.0 IU/L or with a negative UPT during follow-up.
II. An *IUP*, regardless of viability, is diagnosed by either.
    a. TVS findings outlined above.
    b. Histological findings of chorionic villi after surgical evacuation.
    c. Completion of a suspected miscarriage after medical (prostaglandins) or surgical treatment verified by the removal of products of conception by either; TVS, an eventful decline of hCG to <5.0 IU/L or negative UPT.
III. An *EP* is diagnosed at laparoscopy (hysteroscopy) with or without confirmed histology or by TVS findings outlined above.
IV. A PPUL is defined by either.
    a. Plateauing hCG levels where at least three consecutive hCG measurements with a change in hCG levels <15% during a one-week follow-up (6) or a non-visualised pregnancy where MTX is given.
    b. A non-confirmed location of the pregnancy neither by laparoscopy, nor without evidence of chorionic villi after uterine evacuation where hCG levels persists or rise.

Persistent PULs are incorporated among the EPs in the analysis as in previous studies (7, 8)).

Operative variables

Treatment of EPs: surgical (laparoscopic salpingotomy/salpingectomy), medical (MTX) or expectant management. Number of ruptured EPs. The number of blood- transfusions. Number of EPs classified as low risk needing emergency surgery during the entire follow-up. Indication for laparoscopic surgery: suspicion of intra-abdominal bleeding visualized by TVS, pain, diagnostic intervention. Additional treatment due to persistent trophoblast after the initial treatment. Treatment of miscarriages: surgical, medical or expectant management. Histological findings after salpingectomy, salpingotomy and dilatation and curettage.

Post- operative variables

Length of hospital stay and planned sick leave.

Two self-reporting questionnaires measuring psychological distress.

Hospital anxiety and depression scale (HADS)

Posttraumatic stress disorder checklist (PCL-5)

Subsequent two-year pregnancy outcome questionnaire

Time to pregnancy after a non- viable pregnancy (i.e., EP, failed PUL or a non-viable IUP), cumulative pregnancy rate (biochemical, clinical or ongoing pregnancy (documented by TVS or live birth)) and method of conception among patients with a wish for pregnancy.

Alternatively national register data will be used for information regarding the two-year follow-up on subsequent pregnancy outcome, using the Swedish Medical Birth Register (MBR) and the National Patient Register (formerly the Hospital Discharge Register). Data from MBR will include maternal characteristics (i.e., age, parity, body mass index [BMI], smoking habits, and years of involuntary infertility) and data on delivery. By the use of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) the following diagnosis will be included: bleeding in early pregnancy (*ICD-10* code O20.9), threatened miscarriage (ICD-10 code O20.0), incomplete miscarriage (ICD-10 code O03.0-4), complete miscarriage (ICD-10 code O03.5-9), anembryonic pregnancy (ICD-10 code O02.0), missed abortion (ICD-10 code O02.1, recurrent pregnancy loss (ICD-10 code O26.2 or N96.9), gestational trophoblast disease (ICD-10 code O01.0-1; O01.9; D39.2A,B,C,X;C58.9), positive pregnancy test in early pregnancy with uncertain localization/viability (ICD-10 code Z32.1), infertility (ICD-10 code N97.0-4; N97.8B,C,D,W; N97.9), surveillance of pregnancy (Z35.0; Z35.9; O26.8A).

7. Sub study of progesterone as a biomarker in combination with hCG for PUL management

7.1 Background

Serum progesterone is a biomarker known for its capability to rule out a viable pregnancy and inefficiency to distinguish an IUP and failed PUL from an EP (11, 12). In PUL management progesterone has been used to identify spontaneously resolving pregnancy regardless of location with a low risk for complications or need for intervention after a single visit. In an interventional study a progesterone level ≤ 10 nmol/L singled out 37 % (n=227) of PULs from further follow-up where five EPs later were identified of which two needed laparoscopic surgery (13). In a study by Guha et al. (2014) (14), repeat hCG (cut-offs of 66% rising hCG levels and 13 % declining hCG levels) and M4, a progesterone level <10 nmol/L classified 73 %, 84 % and 52 % of PULs as low risk respectively. The misclassification rate of EPs as low risk was 16 % for M4, 37 % for hCG cut-offs and 24 % for progesterone. These data suggest that progesterone in isolation is not an optimal triage tool. A combination of progesterone and serial hCG measurements to predict spontaneous resolution of a pregnancy or a normal IUP have been studied (15). In a more recent study this approach was revived where the first hCG was accompanied with progesterone sampling and when levels were ≤ 2 nmol/L in a

first step 16 % of PULs were classified as low risk, suitable for a reduced follow-up after just one visit and another 46 % by the means of a second hCG in a second step (16). In the second step a novel hCG model was used ($M6_p$) which also contained progesterone as a variable. The two-step model classified 92 % of EPs as high risk in comparison to 82 % by M4 in the same study, which on the other hand classified 70 % of PULs as low risk. In a sub study we want to externally evaluate these findings.

7.2 Methods

The study will be conducted on the same patients recruited in the main study. The former hCG based classification results will be used together with progesterone analysis and we will examine if the combination of each hCG based model and progesterone can achieve a better diagnostic and predictive performance than the hCG based models alone. We will also compare the results of the two randomization groups with each other when progesterone is added. The same primary end points and the clinical efficacy will be applied as in the main study.

Each PUL will be classified as high risk or low risk by combining progesterone and serial hCG in a two-step procedure. Pregnancies of unknown location with a progesterone level ≤2nmol/L will be classified as low risk in a first step and theoretically they could be discharged after a single visit. The remaining PULs will be classified according to the results from the main study within each hCG-based model. Thus, the sequential use of progesterone and the hCG-based models will generate a summarized high and low risk classification of all PULs. Additionally, a novel model, $M6_p$, will be used in a third combination in the same manner as the two other models. $M6_p$ is based on log (initial hCG), log (initial progesterone), log (hCG ratio), log (hCG ratio)^2 and the interaction between log (hCG ratio) and log(initial progesterone) (Van Calster et al., 2016).

Blood samples collected at the first visit will be stored in a biobank for impending analysis of progesterone after randomization and follow-up has been completed in the main study. At SU, the Abbott Architect i2000SR instrument (Abbott Laboratories, Abbott Park, IL, USA) is used for the progesterone assay. The assay is a one-step immunoassay to determine the presence of progesterone in human serum and plasma using Chemiluminescent Microparticle Immunoassay (CMIA) technology with flexible assay protocols, referred to as 'Chemiflex™ technology'; California Medical Instrumentation Association, Sacramento, CA, USA). It has an analytical sensitivity of ≤ 0.1 ng/mL and a total interassay CV <10%. Cross reactivity is undetectable or <5% for 44 tested compounds. Progesterone concentrations will be presented in nmol/L.

8. Statistical methods and handling of data

8.1 Review of variables and time for registration

Recruited patients fill out a questionnaire regarding demographic variables and the examining physician will fill out another questionnaire about clinical findings etc. The data will subsequently be transferred to the eCRF by appointed personnel. Variables regarding PUL management will be registered continuously in the eCRF of each recruited patient by the examining physicians during follow-up until a final diagnosis is established. Variables will also be available and manually retrieved from each patient's electronical medical record including hCG variables, by appointed personnel and the principal investigators (PIs) and registered in the eCRF.

The randomization rate will be followed continuously, and the study duration governed accordingly. A separate two-year questionnaire will be sent to all patients completing the study, for questions regarding fertility during this time of period unless we choose national registers for this data instead. These datasets will be cross-linked with other national registers.

All data in the eCRF and the questionnaires will be analysed after completion of the study.

The progesterone levels will be analysed separately in the sub study and will not be a part of the randomization and subsequent PUL management. The progesterone sample will be saved in a biobank. We will use the unique personal identification number given to all citizens in Sweden to cross link the eCRF with national registers for subsequent pregnancy data retrieval.

8.2 Statistical methods

Descriptive statistics will be presented by mean, standard deviation (SD), median and interquartile range (IQR) (continuous variables) and number and percentage (categorical variables). The diagnostic performance of the two models (sensitivity and specificity) and the predictive performance (NPV, PPV, LR+ (sensitivity/ (1–specificity)) and LR- ((1–sensitivity)/specificity) will be calculated by means of 2 x 2-tables and presented with percentage where applicable. An exact 95 % confidence interval (CI) will be calculated. We will also construct receiver operating characteristic curves for each cut-off to compare the secondary outcome and test if other cut-offs would be superior in our data set. The sign test will be used when comparing the diagnostic performance of paired samples among the cohort of 600 PULs. Crude and adjusted odds ratios (AOR) with 95% confidence interval will be calculated for comparison of outcomes on HADS. The mean score on both subscales of HADS will also be calculated. A t-test will be used for comparison of continuous variables

between the groups. For comparison of proportions between groups a chi-squared test is used. A p-value of <0.05 will be considered statistically significant.

Post hoc analysis

An estimated 48 h hCG value will be calculated when an exact 48 h hCG value is lacking since the second hCG value will be sampled between 40-56 h after the first hCG sample. As in previous studies (17, 18), we will assume a linear function of log transformed values between the baseline (0 h) and the second measurement for each patient. A second value beyond 48 h will be interpolated according to this line, while a second value within 48 h will be extrapolated according to the line in order to obtain an estimated 48 h hCG value. The diagnostic performance of the models using adjusted hCG values will be compared to the main results where unadjusted hCG values are used.

Two-year pregnancy outcome

Differences in pregnancy and live birth rate (LBR) rate during 24-36 months of follow-up will be expressed as fecundity rate ratio with 95 % CI and calculated through Cox proportional hazards analysis. Kaplan-Meier curves will be constructed to estimate the cumulative probability of a pregnancy and LBR over time and the log-rank test will be used to test the differences between the curves for significance.

8.3 Handling of data

Clinical data are handled within ordinary electronical medical records, patient questionnaires and the eCRF of each patient. All data from the eCRFs and cross-linked register data will be transferred to a statistical program for analysis. Regular security rules for data will be applied and all final datasets will be without any identification of the patients.

9. Ethical considerations

Application for ethical approval will be sent to the regional ethical committee in Göteborg. Written and oral information will be given to the patients at the first consultation by a physician or at the time for the second hCG blood test. Informed consent is signed by the patient and the physician on paper before randomization and documented in the electronical medical record.

10. Follow up- control of the study

10.1 Quality assessment

At each clinic there will be a contact person responsible for the conduct of the study. The PIs will be responsible for checking the consistency of data transfer to the eCRFs. The study will be conducted and reported in compliance with good clinical practice (10)

10.2 Source data and CRF

Documentation of the management of the individual patient regarding follow-up and interventions is performed in ordinary electronical medical records, patient and physician questionnaire as well as in the eCRF. Two self-reporting questionnaires. The MBR has been found to have high validity and covers almost all deliveries in Sweden since 1973 (19).

11. Administration

11.1 Coordination group

For coordination of the study the PIs is responsible. The PIs and co-workers will have repetitive contact with contact persons at each clinic.

11.2 Handling of data and statistics

Responsible for the data in the eCRF is the PI. Statistical consultation will be provided from Statistiska konsultgruppen in Göteborg.

12. Handling of patients after finishing the study

The two-year questionnaire is the last study contact for randomized patients in both arms if there will be a follow-up by the means of a two-year questionnaire. Otherwise the last study contact will be after the self-reporting questionnaire has been filled out by the patient.

13. Reporting policy and publication of data

Several publications are planned. Authorship according to the Vancouver convention.

14. Time schedule

Study preparation will start during 2017. Recruitment period is estimated to be completed within three years based on the participating clinics in VGR. The study may be expanded to Skåne University Hospital in Malmö and other regions in Sweden, which will result in a shorter recruitment time.

15. Amendments

16. Participating clinics

Recruitment of patients will start in four clinics; SU, SÄS, Skaraborg Hospital Skövde and NÄL. Additional clinics may join the study after ethical approval has been obtained.

References

1. Barnhart K, 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**: 857-866.
2. Kirk E, Bottomley C and Bourne T. Diagnosing ectopic pregnancy and current concepts in the management of pregnancy of unknown location. *Human Reprod update* 2014; **20**: 250-2613.
3. 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
4. Ectopic pregnancy and miscarriage: Diagnosis and initial management in early pregnancy of ectopic pregnancy and miscarriage. *NICE Guideline CG154, December 2012*.
5. Frost L. Reducing the overuse of βhCG measurements in the emergency gynaecology clinic BMJ Qual Improv Rep. 2016 Feb 8; 5 (1).
6. Bobdiwala S, Guha S, VanCalster 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 Jul; 31(7):1425-35
7. 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. *Human Reprod* 2013; **28**: 609-616
8. Fistouris J, Bergh C, Strandell A. Classification of pregnancies of unknown location according to four different hCG based protocols. *Human Reprod* 2016 Oct; 31 (10):2203-11
9. Mavrelos D, Nicks H, Jamil H, Hoo W, Jauniaux E, Jurcovic D. Efficacy and safety of a clinical protocol of selected women diagnosed with a tubal ectopic pregnancy. Ultr *Obstet Gynecol* 2013; **42**: 102-107
10. European medicines agency (EMA) 2015
11. Hahlin M, Wallin A, Sjöblom P, Lindblom B. Single progesterone assay for early recognition of abnormal pregnancy. Human Reprod 1989; 5, 62: 622-626
12. Verhaegen J, Gallos ID, van Mello NM, Abdel-Aziz M, Takwoingi Y, Harb H, Deeks JJ, Mol BW, Coomarasamy A. Accuracy of single progesterone test to predict early pregnancy outcome in women with pain or bleeding: meta-analysis of cohort studies. BMJ 2012 Sep 27; 345
13. 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.
14. Guha S, Ayim F, Ludlow J, Sayasneh, 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 progesterone or repeated serum hCG levels. *Human Reprod* 2014; **29:** 938-945
15. Hahlin M, Sjöblom P, Lindblom B. Combined use of progesterone and human chorionic gonadotropin determinations for differential diagnosis of very early pregnancy. Fertil Steril 1991; 56: 492-496

16. 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 Nov; 48 (5):642-649.
17. Barnhart KT, Sammel MD, Appleby D, Rausch M, Molinaro T, Van Calster B, Kirk E, Condous G, Van Huffel S, Timmerman D et al. Does a prediction model for pregnancy of unknown location developed in the UK validate on a US population? *Hum Reprod* 2010; 25: 2434–40. Bignardi T, Condous G, Alhamdan D, Kirk E, Van Calster
18. Zee J, Sammel M.D, Chung K, Takacs P, Bourne T and Barnhart K.T. Ectopic prediction in women with a pregnancy of unknown location: data beyond 48 h are necessary. *Human Reprod* 2013; **29**: 441-447
19. Swedish National Board of Health and Welfare. The Swedish Medical Birth Register: a summary of content and quality. The Swedish Center of Epidemiology. Available at:http://www.socialstyrelsen.se/Lists/Artikelkata log/Attachments/10655/2003-112-3_200311 23.pdf. Accessed Nov. 15, 2015.

20. van de Loo et al. J Mood Disord Ther 2017, 1(1):1-16

21. Farren J, Jalmbrant M, Falconieri N et al. Posttraumatic stress, anxiety and depression following miscarriage and ectopic pregnancy: a multicenter, prospective, cohort study. Am J Obstet Gynecol 2020; 222:367.

22. NICOLINE C. DO Acta Obstetricia et Gynecologica Scandinavica 96 (2017) 190–197