Management of pregnancy of unknown location

Diagnostic protocols, clinical outcomes, and psychosocial aspects

Johan Fistouris

Department of Obstetrics and Gynecology
Institute of Clinical Sciences
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden

UNIVERSITY OF GOTHENBURG

Gothenburg 2025

Management of pregnancy of unknown location
Diagnostic protocols, clinical outcomes, and psychosocial aspects

Johan.fistouris@vgregion.se

ISBN 978-91-8115-026-1 (PRINT)
ISBN 978-91-8115-027-8 (PDF)

Printed in Borås, Sweden 2025
Printed by Stema Specialtryck AB

SVANENMÄRKET
Trycksak
3041 0234

“Knowing yourself is the beginning of all wisdom”
— Aristotle

Abstract

Management of pregnancy of unknown location
Diagnostic protocols, clinical outcomes, and psychosocial aspects

Johan Fistouris
Department of Obstetrics and Gynecology, Institute of Clinical Science
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden

Background: Ectopic pregnancy (EP) is the leading cause of pregnancy-related first trimester mortality. Many EP remain undiagnosed beyond an initial transvaginal ultrasound assessment. Noninvasive diagnosis of EP in women with a pregnancy of unknown location (PUL) involves serial serum human chorionic gonadotropin (hCG) levels. Published hCG-based protocols, including risk-prediction models, can be of great interest to Swedish patients and physicians as they may improve PUL management. The performance of protocols is commonly assessed by their diagnostic and discriminatory abilities, but comparisons of clinical outcomes are limited. The psychological effect of experiencing a PUL has not been investigated.

Aim: To determine a well-functioning strategy for PUL management in Swedish gynecological emergency settings and to investigate the psychological outcome of having a PUL.

Methods: A comparative study of four hCG-based protocols and a validation study of risk-prediction models (M4 and M6NP) were conducted at the gynecological emergency unit at the Sahlgrenska University Hospital. M4 and the NICE (National Institute for Health and Care Excellence in the UK) algorithm were selected for a multicenter randomized controlled trial (RCT). Psychological distress was measured with the Hospital Anxiety and Depression Scale (HADS) in women with a PUL and in control participants in antenatal care. Health-related quality of life was assessed with the 36-item short form (SF-36).

Results: In the first study M4 and the NICE algorithm had higher sensitivity for EP but lower specificity for non-EP than the two other protocols. In the validation study M6NP made more accurate predictions of EP, had higher discriminatory ability for EP vs non-EP, and higher sensitivity for EP, but lower specificity for non-EP than M4. In the RCT the NICE algorithm had higher discriminatory ability for EP vs non-EP and specificity for non-EP than M4 when tested on cross-sectional data. There was no significant difference in diagnostic performance for M4 vs NICE when analyzed according to randomized groups, and no significant between-group differences in number of hCG or ultrasounds undertaken, time to diagnosis of EP, length of follow-up, first-line treatment success of EP, or adverse events. Women with a PUL had more frequently an anxiety score ≥8 (58.6% vs 29.2%, p=.0002) than women in antenatal care. No significant differences in anxiety or depression were observed for women with a viable vs a nonviable pregnancy within the PUL group after 1 or 4 weeks. For SF-36, the mental component summary was below the norm (50) for women with a viable (38.5) or a

nonviable pregnancy (33.5), but did not differ between groups, *p*=.20. The physical component summary was lower in women with a viable pregnancy (47.7 vs 52.2, p=.020) and below the norm (50).

Conclusion: The diagnostic and discriminatory abilities of the NICE algorithm were mostly superior to M4, but no significant differences of clinical outcomes were observed. The RCT was not powered to detect differences in these secondary outcomes, which should be considered when interpreting the results. The M6NP seems to be preferable for managing PUL given its high sensitivity for EP and discriminatory ability, although it was not tested in clinical practice. Symptoms of anxiety and depression were substantial among women with a PUL, and health-related mental quality of life was low. This aspect of PUL is novel and requires more research.

Keywords: pregnancy of unknown location, human chorionic gonadotropin, ectopic pregnancy, miscarriage, anxiety, depression, health-related quality of life

List of Papers

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Fistouris J, Bergh C, Strandell A. Classification of pregnancies of unknown location according to four different hCG-based protocols. *Hum Reprod* 2016; **31**(10):2203–2211.

II. Fistouris J, Bergh C, Strandell A. Pregnancy of unknown location: external validation of the hCG-based M6NP and M4 prediction models in an emergency gynaecology unit. *BMJ Open* 2022;**12**(11).

III. Fistouris J, Bergh C, Strandell A. Managing pregnancies of unknown location with the M4 prediction model or the NICE algorithm: a randomized controlled trial. *Under revision.*

IV. Fistouris J, Järvholm S, Bergh C, Strandell A. Psychological aspects, and health-related quality of life in women with a pregnancy of unknown location: a prospective multicenter cohort study. *Submitted.*

Sammanfattning på svenska

Bakgrund: Extrauterin graviditet (EP), eller utomkvedshavandeskap är den vanligaste graviditetsrelaterade orsaken till sjukdom och död i första trimestern. På en gynekologisk akutmottagning är kvinnor med misstänkt graviditetskomplikation såsom EP en vanlig patientgrupp. I vissa fall kan vaginalt ultraljud inte lokalisera någon graviditet vilket benämns ”graviditet med oklar lokalisation” (PUL). De flesta patienter med PUL har en normal tidig graviditet som initialt var för liten att se eller ett missfall, men upp till 20% kan ha en EP. Handläggningen av patienter med PUL är ofta utdragen med upprepade blodprover för analys av graviditetshormon (hCG) och ultraljudsundersökningar. Det saknas gemensam strategi i Sverige för hur kvinnor med PUL ska handläggas. Flera studier har jämfört den diagnostiska förmågan hos olika hCG baserade protokoll som kan användas av läkaren för att identifiera patienter med hög eller låg risk för EP. Få studier har jämfört kliniska utfallsmått för olika protokoll och det saknas studier som undersökt hur handläggningen påverkar patienterna psykiskt.

Syfte med avhandlingen: Att undersöka den diagnostiska förmågan hos fyra hCG baserade protokoll samt att även jämföra kliniska utfall hos två av protokollen, för att avgöra vilket protokoll som är mest användbart på en svensk gynekologisk akutmottagning. I en studie undersöks den psykiska påverkan av PUL.

Resultat: I **delarbete I** visade en retrospektiv jämförelse av fyra protokoll att en prediktionsmodell (M4) och en algoritm publicerad av NICE (The National Institute for Health and Care Excellence) i Storbritannien hade högre sensitivitet för EP än de två andra protokollen men lägre specificitet för andra graviditetsutfall hos patienter med PUL. I **delarbete II** jämfördes M4 med en vidareutvecklad prediktionsmodell (M6NP) som var bättre på att predicera EP bland PUL men båda modellerna visade på en övergripande underestimering. M6NP visade signifikant högre diskrimineringsförmåga mellan patienter med hög risk för EP och patienter med låg risk för EP. M6NP hade högre sensitivitet för EP och negativt prediktivt värde än M4 men signifikant lägre specificitet för andra graviditetsutfall. **Delarbete III** var en randomiserad studie (RCT) där hälften av deltagarna fördelades till att handläggas med hjälp av M4 och hälften till NICE algoritm. NICE algoritm hade mestadels bättre diagnostisk förmåga och diskriminering än M4 men det var ingen signifikant skillnad i tid till diagnos, antal blodprover/undersökningar, behandlingsresultat av EP, ruptur av EP eller andra negativa händelser mellan grupperna. I **delarbete IV** mättes patienternas nivåer av ångest och depression med hjälp av självskattningsskalan HADS (the Hospital Anxiety and Depression Scale) en och fyra veckor efter de randomiserats i RCTn. Efter fyra veckor mättes också hälsorelaterad livskvalitet med SF-36 (36-item short form). Tidigt gravida i mödravården utgjorde en kontrollgrupp och besvarade HADS vid ett tillfälle. Resultatet visade att kvinnor med PUL hade en signifikant ökad nivå av ångest och depression jämfört med kontrollgruppen och låg mental livskvalitet. Ingen signifikant skillnad av ångest och depression uppmättes mellan patienter med

viabel eller icke viabel graviditet inom PUL-gruppen. Efter fyra veckor var nivåerna av ångest och depression signifikant lägre hos kvinnor med PUL men fortsatt förhöjda.

Slutsats: Resultaten i avhandlingen tyder på att M6NP har högst diagnostisk och prediktiv förmåga för EP hos kvinnor med PUL. Övriga protokoll hade acceptabel diagnostisk förmåga, men i RCTn hade NICE algoritm bättre resultat än M4. Ingen signifikant skillnad noterades i kliniska utfall mellan M4 och NICE algoritm men på grund av liten studiepopulation och metodologiska begränsningar skall resultaten tolkas med viss försiktighet. M6NP är möjligen bättre lämpad att använda på en gynekologisk akutmottagning än NICE algoritm och studier med fler deltagare skulle eventuellt kunna påvisa kliniska skillnader mellan dessa protokoll. Prevalensen av ångest-och depressionssymptom är hög bland kvinnor med PUL och jämförbar med det som rapporterats för kvinnor med missfall. Detta är ny kunskap som behöver studeras ytterligare och belyser ett behov av strukturerat psykologiskt stöd för kvinnor som genomgår PUL handläggning.

Abbreviations

aOR	Adjusted odds ratio
AUC	Area under the curve
CI	Confidence interval
CRF	Case report form
eCRF	Electronic case report form
EP	Ectopic pregnancy
EPAU	Early pregnancy assessment unit
ESHRE	European Society of Human Reproduction and Embryology
FAS	Full analysis set
HADS	Hospital Anxiety and Depression Scale
hCG	human chorionic gonadotropin
IUP	Intrauterine pregnancy
M4	Model 4 (risk-prediction model)
M6P	Model 6 progesterone (risk-prediction model including progesterone)
M6NP	Model 6 no progesterone (excluding progesterone)
MCS	Mental component summary of the SF-36
NICE	National Institute for Health and Care Excellence
NPV	Negative predictive value
OR	Odds ratio
PCS	Physical component summary of the SF-36
PPV	Positive predictive value
PUL	Pregnancy of unknown location
Q1	First quartile
Q3	Third quartile
RCT	Randomized controlled trial
ROC	Receiver-operating characteristic
SD	Standard deviation
SF-36	36-item Short Form
TVS	Transvaginal ultrasound

Contents

Introduction 13

Terminology and definitions 14

Pregnancy of unknown location 14

Rationale for the terminology and definitions 16

Incidence of PUL and influencing factors 17

Clinical outcomes 21

Pregnancy outcomes 21

Interventions and adverse outcomes 22

Diagnostic biomarkers and protocols 24

Human chorionic gonadotropin 24

Progesterone 26

Novel biomarkers 28

Application 28

Psychological distress and health-related quality of life 30

Aims of the thesis 31

Patients and methods 33

Setting and study design 33

Inclusion and exclusion criteria 34

Diagnostic protocols 36

Reference standard 38

Data sources 39

Procedures and interventions 40

Outcomes 42

Sample size 43

Statistical methods 44

Methodological considerations 48
Ethical permissions and considerations 50
Results 53
Paper I 53
Paper II 56
Paper III 60
Paper IV 76
Discussion 81
Summary, conclusions, and future perspectives 91
Acknowledgements 95
References 97

Introduction

Vaginal bleeding is experienced by 9% to 30% of pregnant women before 12 weeks´ of gestation.[1, 2] Heavy bleeding accompanied by pain is associated with an increased risk of pregnancy loss while spotting or light episodes are not.[3] Symptoms of threatened miscarriage prompt women to seek medical attention. In the UK early pregnancy complications are the most frequent emergency in gynecology departments.[4] Every year around 1500 women are examined for potential early pregnancy complications at the gynecological emergency unit at the Sahlgrenska University Hospital. The aim of transvaginal ultrasound (TVS) assessment of these women is to determine the location and health of the pregnancy. A pregnancy of unknown location (PUL) is considered when TVS fails to visualize a pregnancy in a clinically stable woman with a positive pregnancy test, of which a majority eventually progress to either a normal intrauterine pregnancy (IUP), a miscarriage or a spontaneously resolving PUL.[5]

Ectopic pregnancy (EP) represents 1% to 2% of all pregnancies and is the most serious pregnancy outcome of PUL, as it can be life-threatening and compromise reproductive health.[6-8] In 2019, there were 6.7 million incident cases of EP globally with 6542 deaths, and despite diagnostic and surgical advances, almost 10% of pregnancy-related deaths in the United States during 2010–2013 were from EPs.[9, 10] In the UK, the mortality rate from EP has been relatively unchanged for the past 20 years, occurring in 0.82 per 100,000 pregnancies in 2021.[11]

Up to 40% of EPs have a delayed diagnosis after the first assessment, and outpatient monitoring can continue for several days or weeks with great uncertainty and risk of morbidity.[12] In a clinical context, reducing follow-up visits and unnecessary medical interventions is desirable, as is shortened time to diagnosis of EP, to avoid potential emergency situations. The unintended interruption of a live pregnancy or fatal intraabdominal bleeding from a ruptured EP has been reported among women with PUL.[13, 14] Protocols of PUL management are designed to prevent such events, stemming from hasty or delayed clinical interventions.[15] Procedures surrounding PUL are associated with a significant workload and need to be efficient as well as aimed at alleviating patient and societal burden.[16]

Terminology and definitions

Pregnancy of unknown location

There is no consensus on ultrasound definitions and terminology of an early pregnancy and much of the existing PUL literature uses nomenclature originating from three publications (**Table 1).** Three findings of the first TVS in women with a positive pregnancy test were accepted as PUL for all studies in this thesis. Given parenthetically below is the abbreviated terminology used for the specific PUL to differentiate among them:

- No signs of a pregnancy inside or outside the uterus (true PUL)
- A sac-like structure or suspicion of retained products of conception (heterogeneous or echogenic tissue) within the uterine cavity (probable IUP or probable miscarriage)
- Ectopic inhomogeneous (solid) structure or sac-like structure (probable EP)

The following ultrasonographic terminology and definitions were used to distinguish between pregnancy outcomes after the initial TVS classifying a PUL:

- Pregnancy of unknown location
    - No signs of a pregnancy inside or outside the uterus
- Intrauterine pregnancy
    - Nonviable
        - Fetal pole ≥7 mm and absence of heartbeat
        - Empty sac-like structure ≥25 mm
        - Heterogeneous or echogenic tissue (complete miscarriage if the anterio-posterior diameter was <15 mm)
    - Viable
        - Fetal pole with heartbeat
    - Uncertain viability
        - Fetal pole <7 mm and absence of heartbeat
        - Empty sac-like structure <25 mm
- Ectopic pregnancy
    - Sac-like (empty) structure, either adnexal (partial or complete interstitial; isthmic; ampullary or ovarian), cervical, intramural or at a caesarean scar.
    - Solid structure, either adnexal (partial or complete interstitial; isthmic; ampullary or ovarian), or partial or complete cervical, intramural or at a caesarean scar.
    - Gestational sac with yolk sac, either adnexal (partial or complete interstitial; isthmic; ampullary or ovarian), cervical, intramural or at a caesarean scar.
    - Live, either adnexal (partial or complete interstitial; isthmic; ampullary or ovarian) or partial or complete cervical, intramural or at a caesarean scar.
        - Gestational sac containing fetal pole with heartbeat

Table 1. Terminology and definitions of diagnostic and nondiagnostic transvaginal ultrasound findings in early pregnancy.

Author, journal, publication year	Diagnostic ultrasound		Nondiagnostic ultrasound		
	Inside the uterine cavity	**Outside the uterine cavity**	**Inside the uterine cavity**	**Outside the uterine cavity**	**Not identified**
Barnhart et al.[17] *Fertil Steril* 2011	Terminology Definite intrauterine pregnancy Definition *Intrauterine gestational sac with yolk sac and/or embryo (with or without cardiac activity)*	Terminology Definite ectopic pregnancy Definition *Extrauterine gestational sac with yolk sac and/or embryo (with or without cardiac activity)*	Terminology Probable intrauterine pregnancy Definition *Intrauterine echogenic sac-like structure*	Terminology Probable ectopic pregnancy Definition *Inhomogeneous adnexal mass or extrauterine sac-like structure*	Terminology Pregnancy of unknown location Definition *No signs of either ectopic or intrauterine pregnancy*
Kirk et al.[24] *Hum Reprod Open* 2020	Terminology Normally sited pregnancy (eutopic) Definition *A gestational sac located within the uterine cavity*	Terminology Ectopic pregnancy -Extrauterine Tubal (interstitial[a] ampullary, isthmic) Ovarian Abdominal -Uterine Cervical[a] Cesarean scar[a] Intramural[a] Definition *A gestational sac or solid structure*	Terminology Early pregnancy Definition *Gestational sac without visible embryo*	Terminology - Definition -	Terminology Pregnancy of unknown location Definition *No pregnancy is visualized*
Doubilet et al.[25] *NEJM* 2013	Terminology Pregnancy failure Definition *Mean sac diameter ≥25 mm and no embryo* *Crown–rump length ≥7 mm and no heart pulsation* *Absence of embryo with hearth activity ≥2 weeks after ultrasound that showed a gestational sac without a yolk sac* *Absence of embryo with heart pulsation ≥11 days after a scan that showed a gestational sac with a yolk sac*	Terminology - Definition -	Terminology Suspicion of pregnancy failure Definition Crown–rump length <7 mm and no heart pulsation Mean sac diameter of 16 to 24 mm and no embryo Absence of embryo with heart activity 7 to 13 days after a scan that showed a gestational sac without a yolk sac Absence of embryo with heart activity 7 to 10 days after ultrasound that showed a gestational sac with a yolk sac Absence of embryo ≥6 weeks after last menstrual period Empty amnion (amnion seen adjacent to yolk sac, with no visible embryo) Small gestational sac in relation to the size of the embryo (<5 mm) difference between mean sac diameter and crown–rump length)	Terminology - Definition -	Terminology - Definition -

[a]Partial or complete.

Rationale for the terminology and definitions

Clear diagnostic criteria are crucial to avoid conflating an EP with an IUP and vice versa, diagnostic errors can lead to delayed or inappropriate treatment causing significant morbidity. Stringent ultrasound definitions and terminology for an early pregnancy also enable comparisons and generalizability of study results.

The appearance of a round or oval hypoechoic area within the intrauterine cavity is not universally accepted as confirmation of an IUP. In an international consensus statement and according to the American College of Radiology, a yolk sac or embryo needs to be visible within a gestational sac to verify the presence of an IUP, and in cases involving no detectable embryonic structure or yolk sac, the finding should instead be a probable IUP. [17, 18] In contrast, in the UK, no proof of embryonic content is needed for IUP confirmation in specialist early pregnancy assessment units (EPAUs).[19] The reason for a more conservative definition of IUP is that a sac-like structure or fluid collection within the uterine cavity could represent a pseudogestational sac (**Figure 1**), which is reported in 10% to 16% of women with an EP. [20, 21] In two recent studies the rate of EP was < 2.2% when a intrauterine fluid collection was seen without concomitant adnexal mass, and its clinical relevance has been questioned.[22, 23]

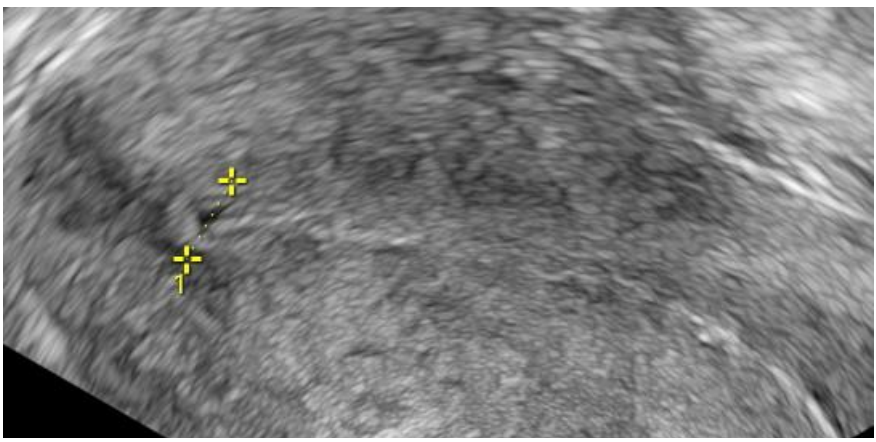


Figure 1. Transvaginal ultrasound image illustrating a pseudogestational sac.
Fluid collection centrally located within the intrauterine cavity, without surrounding decidual changes in a woman with a laparoscopically confirmed ectopic pregnancy.

A pregnancy can be located ectopically within the margins of the uterus and thus not be extrauterine but still extend partially or completely beyond the endometrial-myometrial junction of the uterine (endometrial) cavity. Ectopic pregnancies confined to the uterus are becoming more prevalent, and the European Society of Human Reproduction and Embryology (ESHRE) working group on Ectopic Pregnancy has suggested the term "normally sited (eutopic) pregnancy" instead of "intrauterine pregnancy" to avoid confusion with a uterine EP.[24] The ESHRE group further recommends the term "live" for an early pregnancy when heart pulsation is visible and not using the term "viable" which in obstetric practice signifies a fetus that can survive if delivered. If a gestational sac is seen within the uterine cavity without a visible embryo, the term "early normally sited (eutopic) pregnancy" is advised.

In addition to the possibility of an inconclusive location, the viability of a pregnancy can be unclear. Safety criteria have been developed that virtually eliminate the risk of interrupting a desired pregnancy and are important to recognize during PUL management.[25] In one prospective validation study, the cutoffs for mean gestational diameter, crown–rump length, and presence or absence of heart pulsation were 100% specific for a nonviable pregnancy.[26] In this thesis, the terminology and definitions for the location and viability of a pregnancy are based on the articles listed in **Table 1**. Hereafter, the term "ectopic pregnancy" or EP is used as synonymous with extrauterine pregnancy, as approximately 98% of EPs occur in the fallopian tube.[27] A "viable IUP" is used as synonymous with a live, normally sited (eutopic) pregnancy.

Incidence of PUL and influencing factors

Ultrasound diagnosis in early pregnancy is largely dependent on the operator, and without adequate competence, the chances of a PUL determination likely increase.[28, 29] Technological advancement with high-resolution ultrasound imaging has led to lower PUL rates today compared to a couple of decades ago, when rates of 25% to 43% were reported in emergency gynecology units and emergency departments.[30-32] In specialist clinics, PUL frequencies from 8% to 11% are common, which is below the 15% benchmark recommended for units that evaluate women in early pregnancy.[33-36] The rate can vary widely, however, and in a recent study in 44 EPAUs in the UK, the incidence of PUL ranged from 1% to 27%.[37] Rates of PUL from selected studies are presented in **Table 2**.

Table 2. Studies presenting the incidence of pregnancy of unknown location in clinically stable women managed as outpatients.

Author, journal, publication year	Study design, data period	Country	Setting	Definition of nondiagnostic location	No. of women scanned	PUL incidence n (%)
Hahlin et al.[30] *Hum Reprod* 1995	Prospective 12 months	Sweden	Emergency gynecology unit Single center	Failure to identify the location of the gestational sac without further definitions	261	80 (30.7%)
Mol et al.[31] *Fertil Steril* 1998	Prospective 1993 to 1996	The Netherlands	Gynecology departments Two centers	Adnexal mass or extrauterine sac-like structure and no signs of intrauterine pregnancy	824	354 (43.0%)
Barnhart et al.[32] *Obstet Gynecol* 2011	Retrospective 1999-2007	US	Emergency department Single center	Intrauterine sac-like structure, inhomogeneous adnexal mass or extrauterine sac-like structure	1880	416 (22.1%)
Condous et al.[34] *UOG* 2005	Prospective 2001 to 2003	UK	EPAU Single center	No signs of either ectopic or intrauterine pregnancy	5544	569 (10.3%)
Jin et al.[49] *UOG* 2024	Retrospective 2007 to 2021	Australia	EPAU Single center	No signs of either ectopic or intrauterine pregnancy	10 000+	848 (8.5%)
Goldberg et al. *Obstet Gynecol*[a] 2022	Retrospective 2014 to 2019	US	Abortion clinic Single center	No signs of either ectopic or intrauterine pregnancy	5619	452 (8.0%)

Abbreviations: EPAU, early pregnancy assessment unit.
[a]Obstet Gynecol. 2022;139(5):771-80.

Several factors can contribute to a nondiagnostic TVS, which cannot completely be avoided in some instances because of the timing of the examination or adherence to local routines and national guidelines. Blastocyst implantation occurs from 6 to 12 days after conception, and with implantation, hCG becomes detectable in serum and urine.[38] At this early stage, a TVS will inevitably be inconclusive. A PUL is the most common TVS finding in women examined before 35 days of gestation, and postponement of examinations in asymptomatic pregnant women until 49 days of gestation is advisable.[39] Additionally, many women present after a complete miscarriage when no trophoblastic tissue remains in the uterine cavity. If an IUP was not confirmed earlier during such pregnancies, an EP cannot be ruled out. In one study, 5.9% of women with this clinical scenario had an EP.[40] A current UK guideline recommends further assessment of women in such cases until the pregnancy location can be established.[41]

A gestational sac can be visualized with ultrasound 4 weeks and 3 days after the start of the last menstrual period in women who have a regular 28-day cycle.[42] At this stage, no embryonic contents are visible, and a TVS diagnosis of an IUP is based on decidual changes surrounding the gestational sac, such as the double

decidual sac sign or intradecidual sign.[43] The potential diagnoses that could be made under these circumstances are an early pregnancy of uncertain viability, a pseudo-gestational sac, or a probable IUP, depending on the sonographer and the routines in the specific setting.

One systematic review and meta-analysis showed that the pooled sensitivity for IUP was 42% with the presence of a yolk sac and 81.8% with the double decidual sac sign, with post-test probabilities of 100% and 99.6%, respectively.[44] In a study from the United States, the absence of a yolk sac led to 34.7% of IUPs being defined as probable IUPs and 36.6% of EPs defined as probable EPs.[32] Thus, a more conservative definition of a definitive pregnancy location can substantially affect the number of women who will have additional clinic visits before a diagnosis can be confirmed.

Furthermore, 58% to 71% of EPs visualized with TVS are morphologically characterized by an empty gestational sac or an inhomogeneous mass (**Figure 2**), which in the international consensus statement are nondiagnostic and should be referred to as a probable EP. [45-47] In one study, almost three-quarters of EPs were regarded as nondiagnostic if the definitive diagnosis depended on the presence of a fetal pole or yolk sac.[47] These morphological types of EP have positive predictive values (PPVs) that are similar to those associated with finding embryonic content (96.7% vs 97.7%) and should therefore be reclassified as definitive and not probable EPs, according to some authors.[46]

In summary, a low PUL rate is believed to reduce the number of follow-up visits and in turn the individual and societal burdens. A high PUL rate, on the other hand, potentially leads to more emergency surgery because of undetected EPs and to unnecessary laparoscopy of women with an actual IUP.[33]

Fig. 2a)

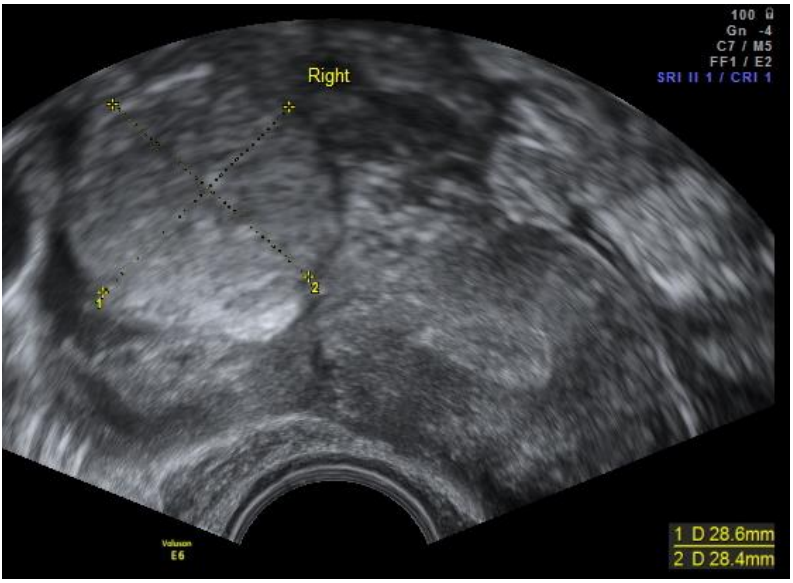


Fig. 2b)

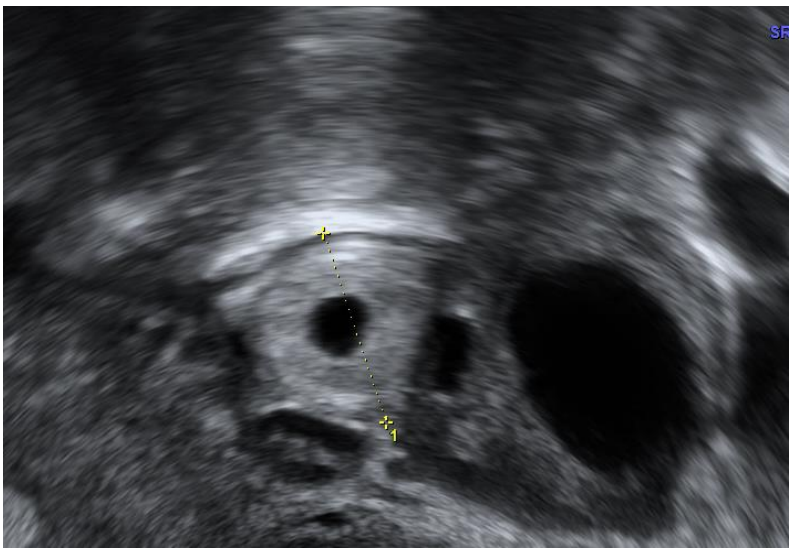


Figure 2. Transvaginal ultrasound images showing different morphological types of ectopic pregnancy in the fallopian tube. (a) Solid inhomogeneous mass (blob sign). (b) Empty gestational sac with a hyperechoic ring around (bagel sign).

Clinical outcomes

Pregnancy outcomes

A PUL is a transient state that eventually progresses to a viable IUP or pregnancy failure, which entails three possible outcomes (**Figure 3**). Up to one-third of PULs are early viable IUPs that are too small to identify when first examined.[15, 48, 49] During surveillance of PULs, 46% to 70% resolve spontaneously without having a verified location.[36, 50-52] A spontaneously resolving PUL signifies self-constraining processes occurring within the intrauterine cavity or ectopically, whereas manifested pregnancy failure consists of either a confirmed miscarriage or an EP.[16, 53-55]

Studies report that 74%-85% of clinically relevant EPs can be diagnosed at the initial TVS, with a cumulative detection rate of 98% prior to surgery.[45-47] Among women with a PUL, 6%-20% will have an EP confirmed either by TVS or at laparoscopy.[30, 32, 54, 56] Ectopic pregnancies starting as PULs have shorter gestational age and lower serum human chorionic gonadotrophin (hCG) levels at presentation than EPs visualized at the first TVS, and some evidence suggests that they also may be less harmful.[57, 58]

A rare clinical scenario that can arise involves a nonidentified pregnancy with sustained hCG levels for a prolonged period, and the term “persistent PUL” is commonly used in such cases.[17, 59, 60] Two definitions currently used in clinical practice are either a daily change of <15% in hCG concentrations or this magnitude of change in 2 days on three occasions.[60, 61] A persistent PUL most often represents either a slow-growing nonviable IUP, a nonvisualized EP, or undiscernible retained products of conception.[59, 62] Subcategories of persistent PUL exist and depend on the management and outcome (**Figure 3**).

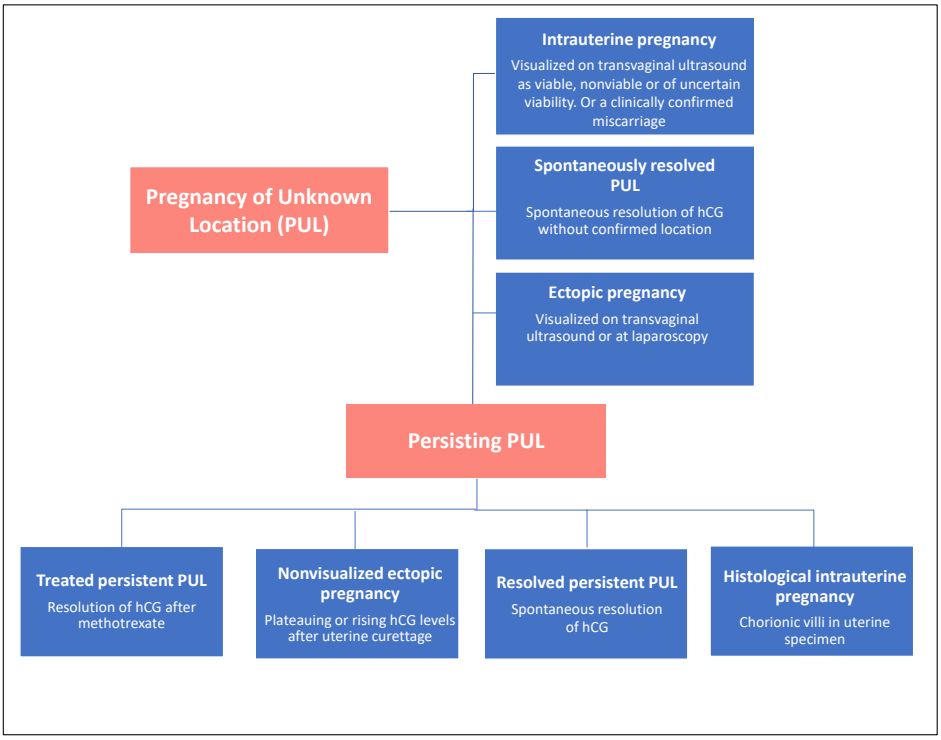


Figure 3. Possible pregnancy outcome of a pregnancy of unknown location (PUL) identified on the initial transvaginal ultrasound. Illustration modified from Barnhart et al. Fertil Steril 2011.

Interventions and adverse outcomes

The clinical workup for women with PUL has undergone a gradual shift from an interventional to an expectant approach. Repeated ultrasound scanning and serial measurement of hCG levels provide highly accurate noninvasive diagnosis of an EP and minimize the risk of initiating a medical intervention without a clear diagnosis. Generally, adverse outcomes are related to hasty or delayed medical interventions, the most severe of which are rupture of an EP with potential life-threatening intra-abdominal bleeding and disruption of a wanted IUP. In the past, laparoscopy was frequently used for diagnostic purposes out of fear of missing an EP.[63] This strategy was eventually abandoned in hemodynamically stable women with minor symptoms because research had shown that most cases with inconclusive TVS were not an EP, and a substantial number of both pathological IUPs and EPs resolved spontaneously.[64-66]

The occurrence of rupturing EPs has been reported to be 2% in women with persistent PUL and from 0% to 0.9% in the PUL population altogether [13, 35, 61, 67] The diagnosis of EP can be challenging, and in cases of ambiguity in symptomatic but clinically stable women, laparoscopy may be initiated prematurely. Rates of emergency laparoscopy or uterine evacuation are reported to be from 0.2% to 1.8% in general PUL populations and 10.5% in women with persistent PUL.[13, 35, 60, 61] A recent study found that 18% of laparoscopies for suspected EP were negative, whereas older studies reported that up to two-thirds were negative.[37, 65, 68, 69] Negative laparoscopies could potentially harm desired viable pregnancies and result in opportunistic salpingectomy of a unaffected fallopian tube because of the fear of missing a small EP.[70]

Consensus is lacking around an optimal strategy for women with a persistent PUL, and both active and expectant management are used. Women with these conditions seem to be at high risk for EP-related complications, and some clinics advocate uterine curettage or methotrexate to reach a diagnosis more quickly or treat an undiagnosed EP. Many persistent PULs resolve spontaneously, and potentially harmful methotrexate could be redundant.[71] In a randomized controlled trial (RCT) from the Netherlands, success rates did not differ between single-dose methotrexate and expectant management in women with persistent PUL (75.6% vs 59.4%), and the resolution time of hCG was the same in both groups (34 vs 38 days).[71] An RCT from the United States showed that a two-dose methotrexate protocol or uterine curettage was more effective than expectant management (51.5% vs 36.0%) and resulted in fewer unplanned surgical interventions.[60] In that analysis, the voluntary crossover of patients was substantial, mostly from active to expectant management, and the results favored active management when analyzed as treated (94.5% vs 56.1%). It was argued that uterine curettage or two-dose methotrexate was a more effective active management than used in previous studies.

According to a US guideline, diagnostic uterine curettage is optional if no IUP is visible and hCG levels exceed 3500 IU/L, but its routine use is not advised in the UK.[72,73] One reason for limiting the use of diagnostic uterine curettage or methotrexate is the risk of terminating a viable pregnancy. Studies have shown that biochemically excluding pregnancy viability is difficult and could be unsafe. In one evaluation of using four different cutoffs of increasing hCG levels, up to 12.3% of ongoing pregnancies would risk uterine curettage, and an estimated >50% of women would be exposed to an unnecessary intervention.[74] A recent study reported that a ratio of 1.02 for hCG change in two days and a progesterone level of 5 nmol/L

(16 ng/mL) were the lowest cutoffs associated with a viable pregnancy among PULs, and that 11% of viable pregnancies involved hCG >3000 IU/L.[75]

Unintended disruption of an ongoing pregnancy is extremely rare and scarcely reported in the literature. In two studies with more than 3000 PULs in total, there was one case of methotrexate given to a woman with a viable IUP, with pregnancy termination because of the teratogenic effect.[13, 61, 76] In the previously referenced US RCT, one woman allocated to expectant management was later found to have a viable IUP and delivered at term without complications.[60] The patient's hCG levels increased from 86 to 92 (7%) in 2 days and to 107 (24%) over 4 days, making the case eligible for the study as a persistent PUL. This case emphasizes the fallacy of solely basing pregnancy viability on hCG progression and that even if conservative criteria are used for considering intervention, the outcome could be devastating. Unless required for a life-threatening maternal condition, hasty intervention without a clear diagnosis should be avoided in women with PUL.

In summary, some differences for diagnosing the pregnancy outcome of PUL exist but with the shared goal to minimize adverse events and unnecessary interventions. The management of PUL is hazardous in many ways for both the health of the woman and pregnancy. A proportion of negative laparoscopy is unavoidable under these circumstances but are not recommended for diagnostic purposes. Although an expectant approach is often feasible, an exception might be persistent PUL that according to one study had better chance of resolution from active management, with uterine curettage or methotrexate.[60]

Diagnostic biomarkers and protocols

Human chorionic gonadotropin

Serum hCG is the most studied biomarker in pregnancy, and guidelines worldwide recommend its use in PUL management, even though no shared strategy exist.[41, 77, 78] In the past, discriminatory hCG levels were used to screen for EP.[79] The prerequisite was that a viable IUP should be visible at a certain hCG level; otherwise, an EP was suspected. This strategy is no longer acceptable because the false-negative rate is too high.[34] Assessing serial hCG trends has proved more useful for identifying abnormal pregnancies that deviate from an expected hCG pattern. In one study, the slowest increase of hCG in a viable pregnancy was 66%.[80] In a more recent study, the minimum rise in hCG was 24% after 24 hours and 53% after 48 hours.[81]

Similarly, the decline in hCG levels for spontaneously resolving PUL has been defined, but an EP does not show a single pattern as it can have either rising and declining hCG levels.[82, 83] An hCG trend that deviates from what would be expected with a viable IUP or spontaneously resolving PUL can indicate a high risk for EP.

Several other cutoffs in the percentage change in hCG have been suggested for use in PUL management. In a systematic review and meta-analysis of 23 studies, the cutoff in hCG change between two measurements solely or as part of a logistic regression model had better diagnostic value for EP than did a single hCG measure.[84] The study did not identify an optimal cutoff, and other pregnancy outcomes could not be evaluated because of substantial heterogeneity in clinical definitions. The authors further concluded that there was a lack of studies comparing the diagnostic value of different protocols. The advantage of serial hCG over progesterone is that the risk of EP, IUP, and spontaneously resolving PUL can simultaneously be ascertained, and a larger proportion of PULs can be managed as low risk compared to proportions when using a single progesterone value.[85]

In a more recent meta-analysis of 43 studies that included hCG and progesterone protocols, a risk-prediction model (M4) proved to be the best method for predicting an EP,[86] with an AUC of 0.87 and sensitivity and specificity of 82% and 80%, respectively (**Table 3**). The M4 included both the initial hCG level and the change in hCG, and in a UK EPAU, both predictor variables clearly added information predicting the outcome of a PUL during development.[87] When M4 was validated in five UK EPAUs, it classified 69.6% of PULs on average as low risk for EP, with a negative predictive value (NPV) of 97.5%.[54] This finding implies that the model theoretically is safe and effective for PUL management. M4 is the most evaluated prediction model, and a high diagnostic accuracy is reported from similar clinical settings in the UK and Australia (**Table 3**).[54, 88] When M4 was tested in a US emergency department and in an assisted reproduction unit in Spain, the sensitivity for EP was too low (**Table 3**) for recommending it in clinical practice.[89, 90] One potential reason for the poor results was that the mean hCG levels for EPs in both studies were lower than for EPs in the development study and previous validation studies. The estimated probability for EP by M4 partly depends on the initial hCG level, and an EP with low hCG can generate a low probability (<5%) and thus be misclassified as a viable IUP or spontaneously resolving PUL. The research team behind M4 has developed a successor, the M6NP, with promising diagnostic performance (**Table 3**) and that can be used with the incorporation of a single progesterone level, M6P.[91] The M6 model was recently updated, and its performance

may be improved with the inclusion of clinical variables such as previous EP, vaginal bleeding and pain score.[92]

Progesterone

Progesterone has been widely studied as a predictor of EP and early pregnancy failure.[93] Sufficient progesterone synthesis from the corpus luteum is crucial for maintaining an early pregnancy in the first 7 weeks of gestation and is relatively constant until the luteoplacental shift.[94] Bleeding can occur during the transfer of progesterone production from the corpus luteum to the placenta because of a transient deficiency in progesterone. The function of the corpus luteum depends primarily on the rate of hCG change, which often is diminished in nonviable IUPs and EPs, resulting in vaginal bleeding correlated with inadequate progesterone levels.[95] The result is a considerable overlap in progesterone levels between EPs and nonviable IUPs, along with slow-growing viable IUPs.[95] In a meta-analysis of 26 studies, progesterone <6 ng/mL had a 99.2% PPV for a nonviable pregnancy but did not satisfactorily distinguish between EP and IUP.[96]

Among women with PUL, progesterone is useful for identifying spontaneously resolving pregnancies with a low risk of complications regardless of location. Progesterone measurement may be safer than serial hCG levels, as the risk for an EP is assessed on the same day the woman presents to the clinic. In one interventional study using a progesterone level of 10 nmol/L, 37% (n=227) of PULs were deemed to carry a low risk of complications and warranted no further clinical follow-up.[35] Five of the women in these cases had an EP that was confirmed later, and two needed laparoscopic surgery. The 1.8% risk of surgery in this group was thought to be acceptable. The downside with progesterone is that many viable IUPs may be misclassified, leading to unnecessary close monitoring.[97] The National Institute for Health and Care Excellence (NICE) in the UK do not endorse progesterone as an adjunct to serial hCG measurement for women with a PUL.[41]

Table 3. Performance of hCG and progesterone protocols for diagnosing pregnancy outcome in women with PUL.

Author, journal, publication year	Study design	No. women	Country, setting	Protocol	Pregnancy outcome (sensitivity/specificity/AUC)[a] IUP	EP	Spontaneously resolving PUL
Bobdiwala et al.[86] *BJOG* 2018	Meta-analysis 43 studies	NA	UK NA	Progesterone 10 nmol Single hCG, 1000 IU/L hCG ratio 0.87 hCG ratio 1.66 M4	91%/59%/0.89 91%/33%/0.71 NA 86%/97%/0.97 59%/71%/0.86	66%/55%/0.69 42%/52%/0.42 55%/67%/0.69 52%/69%/0.69 82%/80%/0.87	82%/79%/0.87 51%/59%/0.59 86%/97%/0.98 NA 71%/57%/0.84
Van Calster et al.[53] *Hum Reprod* 2013	Retrospective temporal validation and prospective external validation cohorts	1962	UK EPAU 5 centers	M4 Temporal cohort External cohort	 NA NA	 84.7%/79.3%/- 88.0%/79.6%/-	 NA NA
Barnhart et al.[89] *Hum Reprod* 2010	Retrospective	544	US ED Single center	M4	81.9%/93.1%/0.92	54.8%/87.7%/0.83	83.1%/83.1%/0.90
Van Calster et al.[91] *UOG* 2016	Retrospective development and validation cohorts	2753	UK EPAU Single center	M6NP validation cohort M4 validation cohort	 NA NA	 92.8%/61.5%/0.87 81.2%/76.8%/0.85	 NA NA
Nadim et al.[88] *AOGS* 2020	Retrospective	413	Australia EPAU Single center	M4	NA	80.6%/75.7%/-	NA
Valdera et al.[90] *UOG* 2021	Prospective interventional	243	Spain ARU Single center	M4	NA	60.0%/79.8%/0.72	NA

Abbreviations: hCG, human chorionic gonadotropin; AUC, area under the receiver-operating characteristic curve; IUP, intrauterine pregnancy; EP, ectopic pregnancy; PUL, pregnancy of unknown location; EPAU, early pregnancy assessment unit; NA, not available; ED, emergency department; ARU, assisted reproduction unit.

[a] Pooled sensitivity and specificity are presented for the meta-analysis, and summary AUC, representing all cutoffs for the protocol not only the cutoff specified in the protocol column.

Novel biomarkers

No single tool is diagnostic with certainty for any specific early pregnancy outcome. Novel serum biomarkers have been developed singly or in combination to discriminate between pregnancy viability or location with the aim of diagnosing an EP with a single blood test. Studies have shown conflicting results for EP diagnosis, however, and no marker has been validated in clinical practice. Some of the biomarkers are proteins involved in implantation and the formation of the fetoplacental unit and have been evaluated in women with PUL.[98, 99] One study reported that activin A, which normally increases during pregnancy until delivery, was significantly lower in EPs than in viable IUPs and miscarriages.[100] When a cutoff of 0.37 ng/mL was used, the area under the receiver-operating characteristic (ROC) curve (AUC) for prediction of EP was 1.00 and the sensitivity and specificity were 100% and 99.6%, respectively. In the same study, a single hCG or single progesterone measure had respective AUCs for EP of 0.81 and 0.62.[100] The results of activin A in the former study have not been reproducible in subsequent studies, which have yielded AUC values from 0.58 to 0.78 for predicting EP. [99, 101, 102]

Inhibin A was another promising serum biomarker, but its AUC was 0.55 for predicting EP, and it was not better than progesterone for predicting a spontaneously resolving PUL.[101, 103] Other potential serum markers of pregnancy location (PAPP-A, glycodelin, and ADAM12) have been less discriminatory, and combinations of these also seem to have limited clinical utility because of underperformance and algorithm complexity.[99, 102] The search for a novel marker remains an important area of investigation, and in recent screening for serum biomarkers using data-independent acquisition proteomics, GSTO1 and ECM-1 in combination with hCG showed promising results for discriminating between EP and IUP.[104]

In summary, hCG is the most common biomarker used for monitoring PUL and EP but there is no common strategy. The use of progesterone as a single marker or in combination with hCG as part of a prediction model have shown promising results. To find a clinically useful novel biomarker remains an important research topic.

Application

The key concept of PUL management is to identify patients at high risk of complications and in need of intensive monitoring, while surveillance of low-risk individuals can be reduced to optimize resource.[41, 91] The aim of a diagnostic protocol

is not to direct treatment or diagnose a pregnancy per se, but to provide a safe and effective strategy that minimizes adverse outcomes and unnecessary clinic visits and interventions by triaging according to risk. The NICE has published a clinical algorithm representing the standard care of women with PUL in the UK.[41] This algorithm contains specific cutoffs in hCG changes to predict if a viable IUP, spontaneously resolving PUL, or an EP is the most likely pregnancy outcome of a PUL, and the need for subsequent follow-up visits is decided accordingly.

Evaluation of prediction-based decision rules such as clinical algorithms or regression models is crucial to ensure a useful diagnostic performance in the intended population and not only in the derivation population.[105] Most prediction models will not undergo external validation, and performance may be substantially worse.[106] Performance is influenced by the characteristics of the underlying population and may not be generalizable to a different clinical context in another country, so that there is a risk of harm with implementation.[107-109]

M4 has been temporally and externally validated in the same type of clinical settings in the UK,[54] but no formal external validation assessing discrimination, calibration, and decision curve analysis has been performed outside the UK. Another important step before implementing a new prediction model is to demonstrate clinical improvement compared to standard care.[108, 110] M4 has proved to be effective and clinically safe with few adverse events in the development setting.[13] The general perception is that integrating a prediction model such as M6 would enhance PUL management but no prediction model has been tested head-to-head with the current standard of care for women with PUL.[111, 112]

In a Swedish context, evaluation of diagnostic protocols is scarce, and no common PUL strategy is available. Published guidelines from other countries or prediction models can be of great interest to Swedish patients and clinicians in gynecological emergency units, which usually manage all types of gynecology emergencies in Sweden. The diagnostic accuracy and clinical outcomes of M4 have not been prospectively compared with the algorithm published by NICE. To evaluate the clinical utility – i.e., diagnostic performance and clinical outcomes – of this model, an RCT may be feasible.[113, 114]

Psychological distress and health-related quality of life

Anxiety and depression are common symptoms following pregnancy loss and are reported in 20% to 40% of women who experience a miscarriage, with a profound impact on their wellbeing.[115-118] Although a majority of PULs represent a pregnancy loss, women with this history have not been part of the vast research conducted to learn more about the psychological aspects of miscarriage.[119] There may be important differences between having a confirmed pregnancy loss and a PUL, with potential differential psychological effects. Anxiety and depression in women with miscarriage are typically measured from the point of diagnosis when worries revolve around reasons and risk of recurrence.[120] Women with PUL likely have concerns about the health of the pregnancy and personal health, related to the awareness of possibly having an EP. The often prolonged waiting time before a diagnosis is determined can potentially cause psychological distress that is exclusively associated with uncertainty.[121] One study found higher levels of anxiety among women with PUL (n=13) compared to women with a definite but negative diagnosis in the first 72 hours after an examination.[122] Another important aspect is that up to 30% of PULs will proceed to a viable pregnancy. Women in these circumstances may suffer more from anxiety and depression than other women in antenatal care, conditions that if untreated can lead to impaired health-related quality of life that can be associated with preterm birth and low birth weight.[123-125] Research is limited on the psychological impact of PULs, with small samples, a focus on subgroups, or analyses done only once a pregnancy outcome is confirmed.[126, 127] These limitations are concerning because women with PUL are reported to perceive a lack of structured psychological support and some would prefer such support to be specifically adapted to their experience with PUL.[128] Early detection and psychological intervention may be beneficial to mitigate long term negative psychological effects.[129]

Aims of the thesis

The overall aim of this thesis was to establish a well-functioning strategy for managing women with PUL in gynecological emergency units and to investigate the psychological effect of experiencing a PUL. The specific aims were as follows:

- To evaluate the performance of four published hCG-based protocols for diagnosing the pregnancy outcome in women with PUL (Paper I);

- To externally validate and compare two risk-prediction models in women with PUL (Paper II);

- To assess whether M4 has noninferior sensitivity for EP compared with the NICE algorithm (standard of care) and higher specificity for non-EP (IUP and spontaneously resolving PUL), and to evaluate clinical outcomes in women managed either with M4 or the NICE algorithm (Paper III); and

- To (1) determine if there is a difference in psychological distress between women with a PUL and women attending their first appointment in antenatal care, (2) determine if there is a difference in psychological distress between having a nonviable or viable pregnancy starting as a PUL, (3) and to evaluate health-related quality of life in women with a PUL (Paper IV).

Patients and methods

Settings and study designs

This thesis includes three study populations. The first comprises women with PUL whose data were retrospectively assembled at the Sahlgrenska University Hospital in Gothenburg during January 2011 to December 2013. The second population consists of women with PUL prospectively recruited into an RCT at the gynecological emergency units of three teaching hospitals during 2018 to 2022, and the third population is women recruited from two antenatal care units during 2022 to 2024. All study populations were confined to the Region Västra Götaland in western Sweden, where Gothenburg is the most populous city.

Three types of research methodology were used in the work described in this thesis. Two involved cross-sectional analyses of retrospective data, and one of these two was a validation study. To enable a structured approach, the validation study followed the TRIPOD (Transparent Reporting of a multivariable for individual Prognosis or Diagnosis) guidelines.[130] The third study was a pragmatic RCT with a cross-sectional design for the primary outcomes and a randomized design for secondary outcomes. The reporting of the RCT was based on the CONSORT guidelines.[131] Originally, the RCT had a noninferiority design for the primary outcome of sensitivity for EP and a superiority design for the primary outcome of specificity for non-EP. However, after an external review on 30 March 2022, the planned parallel design regarding the primary outcomes was abandoned but was kept for secondary outcomes. This decision allowed for better use of the obtained sample size while maintaining sufficient power regarding the primary outcomes. Because of slow recruitment, the study was prematurely terminated at the end of 2022 when half of the first estimated study participants were enrolled. The fourth study examined outcomes prospectively at two timepoints in women with PUL and at one timepoint in a control group consisting of early pregnant women in antenatal care. The two latter studies are registered at www.clinicaltrials.gov (NCT 03461835).

Inclusion and exclusion criteria

Papers I and II

The first two studies (papers I and II) are based on the first study population, with extended inclusion criteria for the second study (Paper II) (**Table 4**). The first population consisted of women who had a positive pregnancy test, were evaluated for potential complications, or worries in early pregnancy and were classified as having a PUL after a TVS examination. Women were eligible if at least two consecutive hCG values were available and no exclusion criteria were met. Exclusion criteria included (1) a first hCG value ≥10,000 IU/L, (2) a hemodynamically unstable status, (2) becoming an inpatient before a second hCG was analyzed, (3) a lapse of more than 72 hours or less than 24 hours between two hCG measurements, and (4) incomplete medical records. The second study also included inpatients and patients with a first hCG value >10,000 IU/L. The rationale behind the exclusion criteria in the first study was to match the M4 development population. In the second study, the aim with the study population was to be more representative of the clinical setting for which the prediction models were tested, and this study included both in- and outpatient management of PUL, regardless of hCG levels.

Table 4. Overview of the studies referred to in this thesis.

	Paper I	Paper II	Paper III	Paper IV
Study period	2011 to 2013		2018 to 2022	2018 to 2022 (PUL group) and 2021 to 2024 (antenatal group)
Study design	Comparative study of four diagnostic protocols on retrospective data	External validation and comparison of the M6NP and M4 prediction models on retrospective data	Multicenter, randomized, controlled, triple blinded trial with a cross-sectional design for the primary outcome and by randomized groups for secondary outcomes	Multicenter prospective study
Setting	Emergency gynecology unit Single center	Emergency gynecology unit Single center	Three emergency gynecology units	Three emergency gynecology units and two antenatal care units
Sample size	915	1061 Sample size sufficient for achieving reliable estimates of outcomes	595 Sample size calculation performed	268 Sample size calculation performed. 162 women in the PUL group recruited from the RCT (Paper III) and 106 women in the antenatal care group
Procedures and interventions	NA	NA	Randomly assignment to M4 or the NICE algorithm in a 1:1 ratio with standardized follow-up based on the predicted pregnancy outcome	Women in the PUL group responded to the HADS 1 and 4 weeks after randomization, and SF-36 after 4 weeks. Women in antenatal group responded to the HADS after their first appointment in antenatal care
Data sources	Melior electronic case record		Melior and Obstetrix electronic case records and eCRF containing data from psychometric questionnaires and health questionnaires	
Outcome	Sensitivity for EP and specificity for non-EP	Sensitivity for EP, specificity for non-EP., calibration-in-the-large, calibration slope, calibration in particular regions of estimated probabilities, AUC, and net benefit	Primary outcomes were sensitivity for EP and specificity for non-EP. Secondary outcomes included time to diagnosis of pregnancy outcome, length of follow-up, successful treatment of EP, hCG resolution time, adverse events, and number of hCG samples and TVS examinations	Primary outcome was an anxiety and depression score ≥8 when first measured with the HADS. Secondary outcomes included scores of ≥8 on both subscales, Physical and Mental component summary on SF-36 and change of anxiety and depression scores between 1 and 4 weeks in the PUL group. Exploratory outcomes were anxiety and depression scores ≥11
Main group analyzed	All included women	All included women	All randomized women	Women in the PUL group vs women in the antenatal group.
Subgroups analyzed	Women with PUL exclusive of probable EP and probable IUP	NA	High-risk vs low-risk PUL and PUL with false vs PUL with true classification results	Women with a viable vs nonviable pregnancy within the PUL group and viable pregnancy within the PUL group vs women in the antenatal group

Abbreviations: PUL, pregnancy of unknown location; NA, not applicable; HADS, hospital anxiety and depression scale; SF-36, 36-item Short Form; eCRF, electronic case report form; EP, ectopic pregnancy; AUC, area under the curve, hCG, human chorionic gonadotropin; TVS, transvaginal ultrasound; IUP, intrauterine pregnancy.

Papers III and IV

Women presenting with a PUL were enrolled into the RCT (Paper III) in the absence of the following exclusion criteria: (1) hospitalization at the first visit, (2) declining to participate, or (3) not understanding the oral or written (Swedish) study information. For randomization, requirements were (1) a first hCG level ≤10,000 IU/L, (2) no hospitalization before the second hCG, and (3) a time span of 44 to 56 hours between the first and second hCG measurements. Women could be included only once during the same pregnancy but could be included again in a subsequent pregnancy. For the fourth study (Paper IV), women in the RCT and women attending their first appointment in antenatal care were eligible. Women agreeing to participate in the fourth study but who did not complete the RCT were excluded. Within antenatal care, women were excluded for (1) a gestation ≥12 weeks, (2) a TVS in the current pregnancy prior to participation, and (3) not being speaking Swedish.

Types of PUL included in papers I–IV

Three TVS findings in women with a positive pregnancy test were accepted as PUL for all studies. Given parenthetically below is the abbreviated terminology used for the specific PUL to differentiate among them:

1. No signs of a pregnancy inside or outside the uterus (true PUL)
2. A sac-like structure or suspicion of retained products of conception (heterogeneous or echogenic tissue) within the uterine cavity (probable IUP or probable miscarriage)
3. Ectopic inhomogeneous (solid) structure or saclike structure (probable EP)

Diagnostic protocols

Papers I-III

Five protocols were evaluated among the three studies, as indicated in **Table 5**. The first protocol (UK protocol) has been used in the UK and is based on cutoffs published in two studies.[80, 132] The second protocol (US protocol) is based in part on the current American College of Obstetricians and Gynecologists and American Society for Reproductive Medicine guidelines, and cutoffs were derived from studies in the United States.[78, 81, 83, 133] The third protocol (NICE) is current standard of care in the UK, but cutoffs were taken from studies performed in the United States and the Netherlands.[31, 134, 135] Two risk-prediction models developed in the UK were evaluated: M4 and M6NP.[87, 91] Both models use multinomial logistic regression to estimate the probability of a spontaneously resolving PUL, an IUP, or an EP based on the first and second hCG values. A 5% estimated probability for EP classifies a

PUL as high risk for EP in the original studies, but different cutoffs were assessed in the external validation study (Paper II). The original equations from the development studies of M4 and M6NP were used to compute probabilities as recommended instead of refitting the model.[87, 91, 108] In the M4 development sample, the time interval between the first and second hCG values was 48 hours, and patients with a first hCG value >10,000 IU/L were excluded. In the M6NP development sample, a 48-h hCG value was imputed when the time interval for a patient was more than 3 days. The covariates of both models are presented in **Table 6**.

Table 5. The specific diagnostic protocols evaluated in Paper I-III in the thesis.

	UK[a]	US[b]	NICE[c]	M4	M6NP
Outcome	**hCG change between the first and second hCG value for prediction of the pregnancy outcome**			**Estimated probability of the pregnancy outcome**	
Spontaneously resolving PUL	13% decrease	35%-50%[c] decrease	>50% decrease	Probability EP <5% and probability spontaneously resolving PUL> probability IUP	Probability EP <5% and probability spontaneously resolving PUL > probability IUP
IUP	66% increase	53% increase	>63% increase	Probability EP <5% and probability IUP >probability spontaneously resolving PUL	Probability EP <5% and probability IUP > probability spontaneously resolving PUL
EP	<13% decrease or <66% increase	<35-50% decrease or <53% increase	≤50% decrease or ≤63% increase	Probability EP ≥5%	Probability EP ≥5%

Abbreviations: hCG, human chorionic gonadotropin; PUL, pregnancy of unknown location; IUP, intrauterine pregnancy; EP, ectopic pregnancy.
[a]Refers to protocol A in the publication of Paper I.
[b]Refers to protocol B in the publication of Paper I.
[c]Refers to protocol C in the publication of Paper I.
[d]Depending on the initial hCG level.

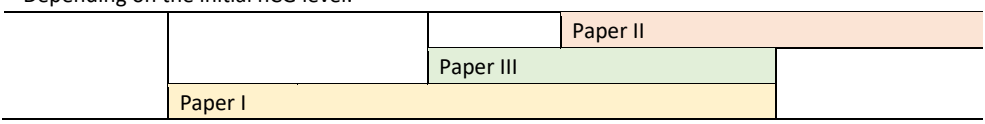


Table 6. Covariates used in the development of M4 and M6NP.

Model	Covariate	Description
M4	log(average hCG)	Log transformation of (first hCG + second hCG)/2
M4	centered[a] hCG ratio	Centering transformation of (second hCG /first hCG-1.17[b])
M4	(centered hCG ratio)^2	The quadratic effect was added because the linearity for the logit assumption was violated for hCG ratio.
M6NP	log(first hCG)	Log transformation of the first hCG
M6NP	log(hCG ratio)	Log transformation of (second hCG/first hCG)
M6NP	[log(hCG ratio)]^2	Square of the log transformed hCG ratio

Centering reduced the correlation between hCG ratio and its square from 0.96 to 0.59.
[a]Abbreviations: hCG, human chorionic gonadotropin. Log refers to the natural logarithm.
[b]Average hCG ratio in the development data set.

Reference standard

To evaluate the performance of a protocol, the predicted pregnancy outcome was compared with the observed pregnancy outcome. Four pregnancy outcomes of PUL were set as the reference standard. The reference standard was based on the combination of hCG development, TVS results (as detailed below), or evident from surgical procedures.

1. A spontaneously resolving PUL was defined as a pregnancy without confirmed location and with an uneventful decline in hCG below the level for a positive serum (the reference hCG level depended on which equipment used as detailed below) or a negative urine hCG test.

2. An IUP regardless of viability was diagnosed with TVS, histological findings of chorionic villi after uterine curettage (vacuum aspiration), medical or surgical completion of a suspected miscarriage as verified with TVS, or a negative serum or urine hCG test.

3. An EP was diagnosed when visualized with TVS or via laparoscopy.

4. A persistent PUL was defined when no pregnancy was visible on TVS and serum hCG levels plateaued (≥3 hCG intervals with a change in hCG levels <15% between two measurements). Also defining a persistent PUL was an absence or no analysis of chorionic villi in a uterine aspiration specimen and persistent hCG levels, regardless of prior hCG development. A persistent PUL either resolved spontaneously after the plateau phase or was treated with methotrexate at the discretion of the physician.

Terminology and ultrasonographic diagnosis of pregnancy outcome after the first TVS:

- Pregnancy of unknown location
  - No signs of a pregnancy inside or outside the uterus
- Intrauterine pregnancy (within the uterine cavity)
  - Nonviable
    - Fetal pole ≥7 mm and absence of heartbeat
    - Empty sac-like structure ≥25 mm
    - Heterogeneous or echogenic tissue (complete miscarriage if the anterio-posterior diameter was <15 mm)
  - Viable
    - Fetal pole with heartbeat
  - Uncertain viability
    - Fetal pole <7 mm and absence of heartbeat
    - Empty sac-like structure <25 mm
- Ectopic pregnancy (outside the uterine cavity)
  - Sac-like (empty) structure, either adnexal (partial or complete interstitial; isthmic; ampullary or ovarian), cervical, intramural or at a caesarean scar.
  - Solid structure, either adnexal (partial or complete interstitial; isthmic; ampullary or ovarian), cervical, intramural or at a caesarean scar.
  - Gestational sac with yolk sac, either adnexal (partial or complete interstitial; isthmic; ampullary or ovarian), cervical, intramural or at a caesarean scar.
  - Live, either adnexal (partial or complete interstitial; isthmic; ampullary or ovarian), cervical, intramural or at a caesarean scar.
    - Gestational sac containing fetal pole with heartbeat

The size of a sac or solid structure was based on the average of the measurement in two planes expressed in mm.

hCG assays and analyzers used in Paper III

Elecsys® HCG+β assay (upper limit for nonpregnant premenopausal women <5.3 IE/L) run on cobas® e 602 module was used most of the study period to measure serum hCG but was changed to cobas® e801 on December 20, 2022, in two of the participating hospitals. In Södra Älvsborg Hospital the Beckman Coulter Access Total ßhCG (5th IS) assay (upper limit for nonpregnant premenopausal women <5.0) was run on UniCel DxI 800 until April 10, 2019, and thereafter Elecsys® HCG+β assay (upper limit for nonpregnant premenopausal women <5.3) run on Cobas e411. The NADAL® hCG Pregnancy Test (analytical detection limit of 25 mIU/mL) in urine was used at follow-up.

Data sources

Papers I and II

All data needed for the studies were obtained from each patient's electronic case record, Melior, which is used across all emergency hospitals in Region Västra

Götaland. With the Sieview web application, Melior databases from all hospitals in the region are accessible.

Paper III

A paper case report form (CRF) was used to collect data from the first visit throughout the second hCG sampling. At the time of randomization, all data were transferred to an electronic CRF (eCRF) and randomization module. Data from a health questionnaire containing demographic information, medical history, and self-reported symptoms were collected upon entry into the study. These data were also entered into the eCRF. Complementary data were collected from Melior during follow-up. Data were also collected from Obstetrix, an electronic case record used by most antenatal and delivery units in the Region Västra Götaland, which is available through Sieview. The eCRF was locked before data were transferred to a database set up by the study statistician.

Paper IV

The eCRF constructed for the RCT contained data from the self-reporting psychometric questionnaires and a health questionnaire responded to by women in the PUL and antenatal groups. Data about women in the antenatal group were restricted to the questionnaires, but all PUL group data were obtainable from the eCRF for transfer to a database set up by the study statistician.

Procedures and interventions

No procedure or intervention was carried out in the first two studies because outcomes were measured on retrospective data.

Paper III

After providing informed consent, women had a first hCG sampling and were scheduled for a second hCG sampling. Once the second hCG level was available, the principal investigator was contacted, and both hCG values together with the sampling time were entered into the randomization software. Participants were randomly assigned in a 1:1 ratio to the M4 or NICE algorithm group. The randomization sequence was computer generated, and t-test minimization regarding the two entered hCG values was performed to ensure an even allocation ratio between groups concerning hCG values. Participants, clinicians, and investigators were blinded to randomized groups.

The pregnancy outcome of PUL was predicted to be an IUP, a spontaneously resolving PUL, or an EP by the allocated protocol, and a recommendation for follow-up was provided. This information was distributed to the responsible clinician, who either approved or dismissed the proposed action. One of the following three clinical pathways was possible for each randomized patient in both groups: (1) Participants with a PUL predicted to be an EP (high-risk classification) had a re-examination within 24 to 48 hours. (2) Participants with a PUL predicted to be a spontaneously resolving PUL (low-risk classification) had a home urine pregnancy test after 2 weeks, and if the test was positive, a re-examination was performed after 24 to 48 hours. (3) Participants with a PUL predicted to be an IUP (low-risk classification) had a re-examination after 1 week.

Paper IV

Women enrolled in the RCT were asked to participate in the fourth study evaluating psychological distress and health-related quality of life in connection with the second serum hCG sampling. They responded to two psychometric questionnaires at 1 and 4 weeks after randomization. After 4 weeks, they were contacted to be reminded about the questionnaires and asked if their pregnancy had continued unless a pregnancy loss was already confirmed. Questionnaires were returned in prepaid envelopes. Control participants were women in antenatal care invited to participate when attending their first visit in antenatal care. They responded to one psychometric questionnaire at home and returned it in a prepaid envelope. Women in antenatal care responded only once.

Instruments

Two established and validated research tools were selected to evaluate anxiety, depression, and health-related quality of life, yet none have been validated specifically in the PUL group.

The Hospital Anxiety and Depression Scale (HADS)

Anxiety and depression were assessed with the Swedish version of HADS, a self-reporting scale of 14 items that measure anxiety (7 item subscale) and depression (7 item subscale) in the past 7 days on a 4-point Likert scale (range 0–3).[136] A summary score of 0–21 is obtained for each subscale and is separated into the following classes: normal (0 to 7), mild (8 to 10), moderate (11 to 14), and severe (14 to 21) symptoms of anxiety or depression. Thus, higher scores indicate worse health. According to a systematic review with a large number of studies, a cutoff ≥8 produces the best tradeoff between true-positive and false-negative results.[137] In a

large sample from the general population in Sweden, the mean scores for females were 4.76 for anxiety and 3.76 for depression.[138] HADS is frequently used in miscarriage research, with a high internal consistency (Cronbach's alpha of 0.83 for anxiety and 0.82 for depression).[137, 139]

The 36-item Short Form Health Survey (SF-36)
Health-related quality of life was evaluated with the Swedish version of the SF-36, which has been extensively used in medical research and validated within different groups of patients, including women in early pregnancy.[140-143] The SF-36 consists of 36 questions covering eight health domains and assesses the respondent's health status for the past 4 weeks. The physical component summary (PCS) can be calculated from domains 1–4 and the mental component summary (MCS) from domains 5–8. Individual domain scores and summary scores range from 0 to 100, and higher scores signify better health.

Outcomes

Paper I

Sensitivity for EP, specificity for non-EP, PPV, and NPV.

Paper II

Sensitivity for EP, specificity for non-EP, PPV, NPV, calibration-in-the-large, calibration slope, calibration in particular regions of estimated probabilities, AUC, and net benefit.[144, 145]

Paper III

Primary and secondary outcomes
The primary outcomes were sensitivity for EP and specificity for non-EP, along with AUC, NPV, and PPV. Secondary outcomes were time from randomization to diagnosis, successful first-line treatment of EP (i.e., not needing an additional treatment modality), time from diagnosis until resolution (hCG <5.3 IU/L) of EP and persistent PUL, length of follow-up, number of hCG and TVS evaluations, interventions (laparoscopy, vacuum aspiration, methotrexate or misoprostol administration), and adverse events (ruptured EP, emergency surgery, accidental disruption of a viable IUP, negative laparoscopy, unplanned visit, side effects of methotrexate [hepatotoxicity/nephrotoxicity/pulmonary toxicity/infection/ myelosuppression], infection after surgery [tubo-ovarian abscess/salpingitis/wound

infection], reoperation of persistent trophoblast, reoperation for bleeding). A repeat dose of methotrexate was considered a successful treatment if hCG was resolved. Emergency surgery was defined as an unscheduled visit to the emergency gynecology unit because of deteriorating symptoms, and surgery followed shortly thereafter. Reasons for nonadherence to protocol were recorded and were either due to nurse planning, physician decision, or an unplanned visit by the patient. No core outcome set exists for PUL, but there is a high degree of convergence with those developed for EP, including treatment success, resolution time, number of additional interventions, and adverse events, which were analyzed in the study.[146]

Paper IV

The primary outcome was anxiety and depression score ≥8 on HADS (symptoms requiring clinical assessment) after one week. Secondary outcomes after one and four weeks included scores of ≥8 on both subscales, scores on both subscales as a continuous outcome, and SF-36 domain and summary scores (PCS and MCS). Exploratory outcomes were anxiety and depression ≥11 (moderate or severe symptoms). HADS outcomes after four weeks and SF-36 scores were only applicable in the PUL group.

Sample size

Papers I and II

No sample size estimation was performed. Conventionally when externally validating prediction models of binary outcomes a minimum of 200 outcomes in each group have been advised to make accurate and precise estimates of calibration, discrimination and clinical utility.[147] A sample size calculation may however be more reliable.[148]

Paper III

Originally, the sample size was designed to test two hypotheses. The first hypothesis was that M4 would have a noninferior sensitivity for EP compared with the NICE algorithm. A total of 541 women per randomized group was needed to test for a noninferiority margin of seven percentage points with 80% power, assuming a sensitivity of 85% by M4 and 86% by the NICE algorithm at a 2-sided significance level of 0.05. The second hypothesis was that M4 compared with the NICE algorithm would have superior specificity for non-EP. We hypothesized a difference of eight percentage points based on estimates of 78% specificity by M4 and 70% by the NICE

algorithm.[51] A total of 495 women allocated to each group was necessary to detect a clinically important difference of eight percentage points with 80% power at a 2-sided significance level of 0.05. In all, 1200 women were planned to be randomized into the study, accounting for a 10% loss to follow-up.

Paper IV

Between 20% and 40% of women experience increased levels of anxiety in the weeks following a miscarriage.[139] In a recent Dutch study of 2897 women in early pregnancy, the prevalence of anxiety was 17.9%.[149] We estimated a 30% prevalence of anxiety in women with PUL and 15% in women in antenatal care. With 177 women in the PUL group completing HADS, 105 women in the antenatal group were needed for 80% power to detect an absolute 15% difference in anxiety scores ≥8 after 1 week, at a 2-sided significance level of 0.05 using a chi-square test. For protection against a 10% loss to follow-up, the target for the antenatal group was 115 women. In a second calculation using the HADS anxiety subscale as a continuous scale, 160 women in the PUL group and 105 in the antenatal group would have been needed for a power of 80% with a two-sample t-test to detect a mean difference of 1.2 HADS units between the two groups, with a standard deviation of 3.4 in both groups at a 2-sided significance level 0.05.[150] A difference of 1.2 HADS units was regarded as a minimally important difference based on mean scores on the HADS anxiety subscale previously reported for women in early pregnancy.[149]

Statistical methods

The sensitivity for EP by any diagnostic protocol represented confirmed EP and persistent PUL correctly classified as high risk, and the specificity represented non-EP (spontaneously resolving PUL and IUP) correctly classified as low risk. The false-positive rate (1 – specificity) was presented in Paper II instead of specificity. The PPV represented EPs among all PULs categorized as high risk, and the NPV represented non-EPs among all PULs classified as low risk. This dichotomy also was applied to spontaneously resolving PULs and IUPs in Paper III. For the studies, persistent PULs were included among EPs for calculating diagnostic outcomes.

Descriptive statistics were presented by mean, standard deviation (SD), median, first quartile (Q1) and third quartile (Q3), or minimum and maximum (continuous variables), and number and percentage (categorical variables). For comparisons of continuous variables for independent groups, the t-test was used if data were

normally distributed and the Mann–Whitney U test for nonnormally distributed data. For comparison of categorical data for independent groups, a Fisher's exact test was used, and for paired observations, the sign test or McNemar's test was used. A nonparametric permutation test was used for paired observations of continuous variables. For comparison of ordered categorical variables for independent groups, a Mantel–Haenszel chi-square test was used. Further statistical methods and considerations for each paper are presented below.

Paper I

The time interval between the first and the second hCG samplings varied from 24 to 72 hours for patients in the study. For patients with missing exact 48-h hCG values, an estimated 48-h value was computed. A linear function of log-transformed values between the first hCG (baseline) and the second measurement was assumed for each patient, as in previous studies.[89, 151] Estimated 48-h values were acquired by interpolating values taken after 48 hours according to this line and extrapolating values according to this line when a second hCG value was taken within 48 hours.

Paper II

Calibration

Calibration was used to examine the agreement between the estimated probability of an EP and the observed outcome of EP among women with PUL. Calibration can be assessed on different levels, including calibration-in-the-large or mean calibration, which is the relationship between the average estimated probability of EP and the observed EP rate among all PULs. The calibration-in-the large can be assessed from a calibration curve representing its intercept. The target value is 0, and values under the curve suggest overestimation, whereas values above the curve suggest underestimation. A calibration curve represents observed EPs among PULs plotted against the whole range of estimated probabilities of EPs.[152, 153] The slope of the curve represents the association between observed and estimated probabilities of EP in different ranges. A slope below 1 indicates that estimated probabilities are too extreme – i.e., PULs with high estimated probabilities (close to 1) are too high, and PULs with low estimated probabilities (close to 0) are too low. A slope above 1 indicates the opposite: that high estimated probabilities are too low and low estimated probabilities are too high.[154] A flexible locally estimated scatterplot smoothing curve was computed for visual assessment of calibration in specific regions of estimated probabilities.

Discrimination

To quantify how well M4 and M6NP predictions separate between women with and without EP the AUC was calculated. An AUC of 0.5 means that for any randomly selected pair of PUL, the model assigns a higher probability to the woman with an actual EP than the woman without EP 50% of the time and indicates that the model is no better than chance. If the AUC is 1.0, the model has a 100% predictive ability. A ROC curve was constructed to show the diagnostic performance of the model across different cutoffs. DeLong's test was used to test the difference in AUC.

Clinical utility

M4 and M6NP are used for triaging women with PUL according to risk classification. A 5% estimated probability of EP is a commonly used threshold for high-risk classification in clinical practice.[61, 91] A lower threshold will inevitably classify more PULs as high risk, both actual EPs (true positive) and non-EPs (false positive). Consequently, more women with EP could potentially be identified early on with a lower threshold, whereas more women with a non-EP may undergo unnecessary examinations and investigations. The opposite is true if a higher threshold is used. When weighting the clinical consequences of correctly classifying (expected benefit) an EP against misclassifying (expected harm) a non-EP, a weighting factor can be appreciated that considers the overall consequences (net benefit, NB) of using a specific threshold. The NB takes this into account by using a weighting factor that is the odds (p/1-p) of EP at a chosen probability threshold (p_t).

The NB is calculated by subtracting the number of false-positive results, multiplied by the weighting factor, from the number of true-positive results relative to the total sample (True positive)/N - (False positive)/$N \cdot ((p_t/(1-p_t))$. Setting a 5% probability (odds 1:19 of EP) threshold indicates that it is clinically appropriate to value false-positive results at 1/19th of true-positive results. The NB was plotted across a range of probability thresholds in a decision curve. For the decision curve, two default strategies were plotted, one that assigned positive results to all and the other that assigned positive results to none. For clinical utility, the NB of a model must be above both default strategies at the chosen probability threshold.

Paper III

Randomized participants with a known pregnancy outcome were included in the full analysis set (FAS). Participants in the FAS population who were followed up in line with recommendations constituted the per-protocol population. Participants who did not adhere to the recommended follow-up were not a part of the per-protocol

population. A sensitivity analysis evaluating the primary outcome according to randomized groups was performed on the per-protocol population (noninferiority design), corresponding to the original study design. An exploratory analysis of the primary outcome according to randomized groups also was performed for the FAS. For secondary outcomes, the primary analysis was performed on the FAS according to randomized groups. A subgroup analysis (high-risk vs low-risk participants) of secondary outcomes was performed on the FAS and the per-protocol population.

Absolute differences in categorical and continuous variables were presented with 95% confidence intervals (CI). The CI for dichotomous variables was the unconditional exact confidence limit. If no exact limits could be computed, the asymptotic Wald confidence limits with continuity correction were calculated instead. Calculation of CIs for continuous variables assumed normality. When variances were not equal ($p<0.05$), the SD was based on Satterthwaite's approximation; otherwise, the SD was based on the pooled SDs. Risk ratios (RRs) and 95% CIs were calculated. Kaplan–Meier curves were constructed to report the time to event, and the log-rank test was used to test differences.

Paper IV

The primary study group (PUL vs antenatal group), within the PUL group (nonviable vs viable pregnancy), and a subgroup (viable pregnancy within the PUL group vs the antenatal group) were evaluated for outcomes.

Adjusted analyses of dichotomous outcome variables between two groups were performed with multivariable logistic regression, and the results were given as adjusted odds ratios ([a]ORs) with 95% CIs and p values. Adjusted analyses of continuous outcome variables between two groups were performed using analysis of covariance, and the results were given as adjusted mean differences with 95% CIs and p values. The unadjusted results were presented as percentage point differences or mean differences with 95% CIs using the Farrington–Manning method for dichotomous outcome variables and t-tests for continuous variables. The main analyses included adjustments for the following confounders: age, gestational age, previous pregnancy loss, and investigated for infertility. A sensitivity analysis was performed using multiple imputation of missing values for confounders in the primary analysis. One hundred datasets were generated using independent stochastic imputation with fully conditional specification. Running 100 logistic regressions of the 100 data sets and then pooling the results yielded a summary estimate for the primary outcome of an anxiety score ≥8. Adjusted ORs were presented. The PCS and

MCS of the SF-36 were calculated using a standardized mean score of 50 with an SD of 10 points for a Swedish population.[155]

All tests were two-tailed and conducted at the 0.05 significance level. Statistical analyses were performed using SPSS V.21.0 (SPSS, Chicago, Illinois, USA) and SAS V.9.4 (Cary, North Carolina, USA).

Methodological considerations

Papers I and II

A first evaluation of diagnostic test accuracy or performance of prediction models on retrospective data is a conventional study design.[108, 156] Retrospective data were used but handled differently in the first two studies. Estimated 48-h hCG values were calculated in the first study but not in the second. In both studies, a complete case analysis was performed, excluding a substantial number of PULs for not having an acceptable sampling interval or having only a single hCG value. This loss of information could increase the uncertainty of the results and increase bias if not missing completely at random, because then the complete analysis set is not a random sample of all PULs. In both studies, "missing data" were considered missing completely at random, making a complete case analysis appropriate; thus, bias was avoided, but the statistical power was reduced because of the reduced sample size.[157] Multiple imputation of missing 48-h hCG values would be the preferred method to increase the sample size in both studies.[157] Another consideration is the risk of verification bias because of variation in the assessment of the pregnancy outcome of PULs.[158] A pregnancy outcome could potentially be included in the wrong group and lead to overestimation or underestimation of protocol performance. Traditionally, more EPs have been managed surgically and could be verified in a different manner compared with current practice when many are treated conservatively and therefore diagnosed with TVS.[47, 159] As with most reference standards, the method is not perfect but remains the best available. Although equal for all protocols and like previous studies, it must be considered when interpreting the results.[16, 45, 91, 132] Moreover the studies are retrospective in nature, and as such, dependent on existing data.

Paper III

The M4 and the NICE algorithm were selected for the RCT because of their higher sensitivity for EP compared with the other protocols in the first study (Paper I) and

their having been evaluated in clinical practice (M4) or constituting the current standard of care (NICE algorithm).[13, 41, 51] Many studies have compared the diagnostic accuracy of different protocols which serves as a surrogate measure for important clinical outcomes.[84, 86, 160] It is recommended that diagnostic strategies and prediction models are compared to standard of care by also assessing the clinical impact.[108, 110, 160] A parallel design using randomized groups facilitated the linking of the diagnostic performance of two protocols to important patient health outcomes. A directive approach was used, meaning that the prediction of the protocol and a management recommendation were provided to the physician.[161] All management beyond the initial follow-up was individually decided.

An important aspect making it possible to transfer the design of Paper III from an RCT design to a cross-sectional design is that sensitivity and specificity (diagnostic performance) for the pregnancy outcomes, of the two protocols are not dependent on the clinical management. Thus, for evaluating diagnostic performance, the doubled sample size could be utilized to increase power. Only secondary outcomes such as time to diagnosis, treatment success of EP, adverse events and so forth might have been influenced by which protocols that was used, if the risk classification differed significantly. All these secondary outcomes were thus analyzed according to randomized groups. For various reasons, the optimal way may be to cluster (randomize) entire clinics instead of patients, although cluster randomization also has inherent limitations.[162] However, in the RCT, both patients and physicians were randomly part of the process and blinded to the allocated protocol. The RCT was a serious attempt to provide evidence for stronger recommendation of a specific PUL protocol.[163]

Paper IV

Psychometric questionnaires together with stamped return envelopes were given to participants by health professionals, and follow-up contact for reminding the participants was done to increase the response rate.[164] Women in the PUL group received questionnaires for both assessments simultaneously, and the follow-up contact was executed in time for the second assessment. This method ensured a low attrition rate between the first and second assessments in the PUL group. Using an online survey method could potentially increase the response rate without changing the frequency of reminders but may increase sampling bias if more highly educated participants are responding.[165]

Ethical permissions and considerations

Before their initiation, all studies were reviewed and given ethical approval from either the Regional Ethical Review Board, Gothenburg, or the Swedish Ethical Review Authority. Written informed consent was provided by all women included in studies described in papers III and IV on entry into the studies.

Paper I (first study population): No. 501-15 (Regional board)
Paper II (first study population): No. 501-15 (Regional board)
Paper III (second study population): No. 382-17 (Regional board)
Paper IV (second and third study populations): No. 2020-06775 (National authority) and No. 382-17 (Regional board)

Papers I and II

These studies were purely observational, and the received care was not influenced by being an aspect of the studies. No consent was obtained, which legally was not needed. Still, a patient might have objected to being part of the study if aware, which could be an integrity issue. However, analyses and publication of the results were performed on a group level on fully anonymized data without any possibility for identification. Thus, being part of the study imposed no risk of adverse consequences for the patients.

Paper III

Women in both groups were managed according to established routines evaluated in published studies or part of national guidelines. These routines are assumed to be superior to the routines formerly used in the participating clinics, as they are based on updated evidence. The first study (Paper I) demonstrated that the protocols similarly identified a vast majority of EPs, which could allow for an early diagnosis. This similarity implies that regardless of randomization group, there was probably no significant difference from a safety perspective. It cannot be certain, however, that the protocols produced similar results in the RCT. Of note, the protocols served only as a decision support for the physician, who always integrated other clinical factors before approving the recommended management or not.

Paper IV

In the fourth study, women in the PUL group reported their psychological well-being by filling out psychometric questionnaires without having an organized follow-up. Thus, women with distress may not have been sufficiently taken care of to alleviate

harm, despite the harm having been reported. This risk also applies to women in the antenatal group. The lack of acknowledgement of their need for emotional support could even have exacerbated their distress.[166] In retrospect, all participants should have been given an opportunity to share their psychological experience of having a PUL with health professionals in a standardized manner within the framework of the study.

Results

Paper I

Classification of PULs according to four different hCG-based protocols

A total of 1472 women with PUL attended the emergency gynecology unit during the study period. After exclusion criteria were applied, 950 were eligible. Of this group, 35 had no known pregnancy outcome and were excluded, so that 915 women were included in the primary analysis.

Characteristics of participants and pregnancy outcomes

At the first TVS, 187 (20%) cases were a probable IUP, 16 a probable EP (1.7%), and 712 a "true" PUL (77.8%). Baseline characteristics and hCG data according to pregnancy outcome are shown in **Table 7**. Of 146 EPs, 141 (96.6%) were in the fallopian tube and treated with laparoscopic surgery, which included 102 salpingectomies. A total of 85 of the 141 EPs (60.3%) were visualized with TVS prior to surgery, and four had a fetal pole with heartbeat. There were 21 (14.9%) ruptured EP corresponding to a rate of 2.3% (21/915 PULs) in the whole study population and there were 34 (19.4%) negative laparoscopy corresponding to a rate of 3.7% (34/915 PULs) in the whole study population. In one case of negative laparoscopy, a viable IUP was later confirmed. Of 35 persistent PULs, six were treated with methotrexate, and the remaining resolved spontaneously.

Diagnostic (classification) performance

The M4 and the NICE algorithm had better sensitivity for EPs but lower specificity for non-EPs than the US and UK protocols (**Table 8**). There was no statistically significant difference in sensitivity or specificity between M4 and NICE. The NPV ranged from 91% to 96%, the PPV from 40% to 49%, and low-risk classification from 55% to 73% (**Table 8**). The M4 mainly misclassified IUPs (54%) as high risk and less so spontaneously resolving PULs (18%). The NICE algorithm misclassified 30% of IUPs and 37% of spontaneously resolving PULs. In a subgroup analysis of 712 PULs, excluding probable IUPs and probable EPs, differences in sensitivity and specificity between protocols were generally consistent (data not shown). The M4 misclassified fewer IUPs (39%), and its specificity for non-EPs was 78%, compared to 70% for NICE (p=.010). A majority of probable IUPs were nonviable IUPs with hCG ratios more like those for EPs than for viable IUPs. The result was that estimated probabilities were affected, leading to a higher rate of false positive results when

these data were included in the analysis. The results showed that 19% of EPs misclassified by M4 had rising hCG levels, as compared to 87% for NICE (p<0.01). The initial mean (SD) hCG levels (IU/L) of these EPs misclassified by M4 and NICE were 310 (453) IU/L and 770 (792), respectively (p<.05).

Table 7. Baseline characteristics of 915 women with a PUL and hCG data within each pregnancy outcome.

Characteristics	Spontaneously resolving PUL	Intrauterine pregnancy[a]	Ectopic pregnancy[b]
n (%)	**425 (46.4%)**	**309 (33.8%)**	**181 (19.8%)**
Age, years, mean (SD)	31.0 (6.5)	28.9 (6.0)	31.3 (5.6)
Gestational age, days, mean (SD)	45.7 (9.9)	39.2 (9.3)	43.7 (10.9)
Symptoms, n (%)			
Bleeding only	80 (19%)	27 (9%)	34 (19%)
Abdominal pain no bleeding	38 (9%)	201 (65%)	39 (21%)
Abdominal pain and bleeding	304 (71%)	67 (21%)	101 (56%)
None	3 (1%)	14 (5%)	7 (4%)
Previous ectopic pregnancy, n (%)	21 (5%)	34 (11%)	16 (8%)
Time (h) between hCG measures, mean (SD)	49.4 (9.3)	49.4 (8.9)	47.9 (10.1)
First hCG value (IU/L), mean (SD)	804 (1220)	2139 (2291)	1471 (1881)
Second hCG value (IU/L)[c], mean (SD)	330 (464)	3529 (3762)	1634 (2179)
Ratio of the second and first hCG value, mean (SD)	0.52 (0.35)	1.90 (0.56)	1.14 (0.41)
No. of hCG samples, mean (SD)	2.8 (1.2)	2. 6 (1.1)	3.8 (1.6)
No. of ultrasounds performed, mean (SD)	1.7 (0.9)	2.5 (1.1)	3.2 (1.2)

Abbreviations: hCG, human chorionic gonadotropin; PUL, pregnancy of unknown location.
Percentages may not total 100 due to rounding.
[a]Including 85 nonviable intrauterine pregnancies.
[b]Including 35 persistent PUL.
[c]Unadjusted hCG.

Table 8. Performance of the four protocols diagnosing ectopic and non-ectopic pregnancies in 915 women with PUL and the misclassification of non-ectopic pregnancies as high risk.

	% (95% CI)				*P*-value UK vs US	*P*-value UK vs NICE	*P*-value UK vs M4	*P*-value US vs NICE	*P*-value US vs M4	*P*-value M4 vs NICE
Outcome	**UK**	**US**	**M4**	**NICE**						
Low-risk classification, n (%)	663 (73%)	601 (66%)	511 (56%)	507 (55%)	NA	NA	NA	NA	NA	NA
Sensitivity for EP[a]	68% (61 to 75)	81% (74 to 86)	88% (83 to 93)	87% (82 to 92)	.010	.000	.000	.000	.016	.86
Specificity for non-EP[b]	82% (80 to 85)	77% (74 to 80)	67% (63 to 70)	66% (62 to 69)	.000	.000	.000	.000	.000	.70
Negative predictive value	91%	94%	96%	95%	NA	NA	NA	NA	NA	NA
Positive predictive value	49%	47%	39%	40%	NA	NA	NA	NA	NA	NA
Spontaneously resolving PUL misclassified as high risk	11% (8 to 14)	24% (20 to 28)	18% (15 to 22)	37% (32 to 42)	NA	NA	NA	NA	NA	NA
Intrauterine pregnancy misclassified as high risk	27% (22 to 32)	22% (17 to 27)	54% (48 to 60)	30% (25 to 35)	NA	NA	NA	NA	NA	NA

Abbreviations: EP, ectopic pregnancy; PUL, pregnancy of unknown location; NA, not applicable.
Data presented as % (95% CI) if not stated otherwise.
[a]Including 35 persistent PUL.
[a]Including Spontaneously resolving PUL and intrauterine pregnancy.

Paper II

PUL: external validation of the hCG-based M6NP and M4 prediction models in an emergency gynecology unit

Of 1208 women with PULs and at least two hCG samples eligible for inclusion, 125 had a sampling interval >3 days or <1 day and were excluded. Another 16 women lost to follow-up and another six diagnosed with a gestational trophoblastic disease were excluded from analysis.

Characteristics of participants and pregnancy outcomes

A total of 1061 women with a PUL were included in the analysis (**Table 9**). Pregnancy outcomes were 461 (43.4%) spontaneously resolving PULs, 264 (24.9%) viable IUPs, 98 (9.2%) nonviable IUPs, 192 (18.1%) EPs, and 46 (4.3%) persistent PULs.

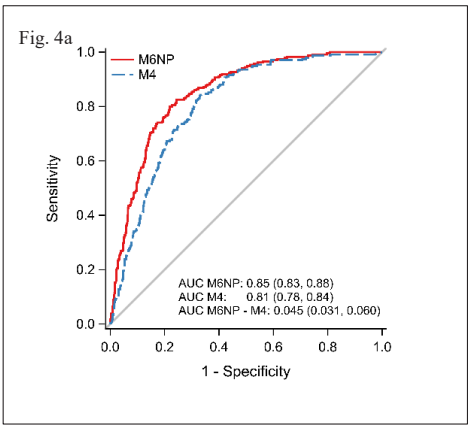
Table 9. Baseline characteristics of 1061 women with PUL in the validation sample and hCG data for each pregnancy outcome.

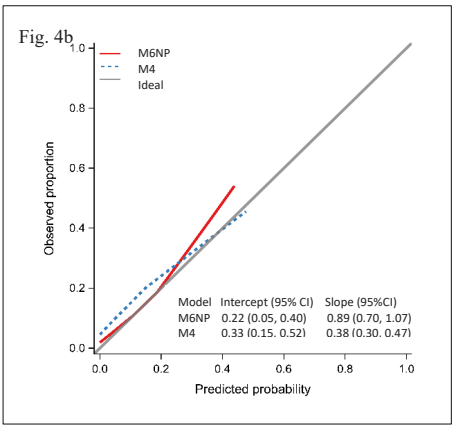
Characteristics	Total (n=1061)	Spontaneously resolving PUL (n=461)	Intrauterine pregnancy (n=362)	Ectopic pregnancy[a] (n=238)
Age, years, median (Q1;Q3)	31 (24;35)	31 (27;35)	29 (24;33)	32 (27:36)
Type of PUL, n (%)				
True PUL	793 (74.7)			
Probable IUP	239 (22.5)			
Probable EP	29 (2.7)			
Vaginal bleeding, n (%)	723 (68.1)			
Previous ectopic pregnancy, n (%)	76 (7.2)			
First hCG value (IU/L), median (Q1;Q3)	703 (210;2312)	334 (120;1078)	1521 (464;5125)	810 (252;2315)
Second hCG value (IU/L), median (Q1;Q3)	635 (150;2700)	158 (53;460)	2765 (983;7900)	920 (265:2735)
hCG ratio, median (Q1;Q3)	0.94 (0.45;1.59)	0.42 (0.27;0.64)	1.81 (1.39;2.21)	1.05 (0.86:1.31)
Hours between first and second hCG sample, n (%)				
<24	0 (0)			
24—39	141 (13.3)			
40—56	751 (70.8)			
57—72	169 (15.9)			
>72	0 (0)			

Abbreviations: Q1, first quartile; Q3, third quartile; IUP, intrauterine pregnancy; EP, ectopic pregnancy.
[a]Including 46 persistent PUL.

Discrimination, calibration, and NB

A summary of key performance measures of M6NP and M4 is shown in **Figure 4**. For EPs (including persistent PULs) vs non-EPs (spontaneously resolving PULs and IUPs), M6NP had an AUC of 0.85 and M4 had an AUC of 0.81 (p<.0.001). For M6NP, the calibration intercept (calibration-in-the-large) was 0.22 for the estimated probability of EP (including persistent PUL), and the calibration slope was 0.89. The intercept and slope for M4 were 0.33 and 0.38, respectively. As shown in the calibration plot, M6NP was well calibrated up to estimated probabilities of ~0.2, but above this threshold, estimated probabilities of EP were too low, as also can be seen for M4 up to estimates of 0.4 (**Figure 4B**). M6NP had a NB at the 5% threshold, but M4 did not. The decision curves (**Figure 4C**) represent NB for M6NP and M4 across a range of probability thresholds.





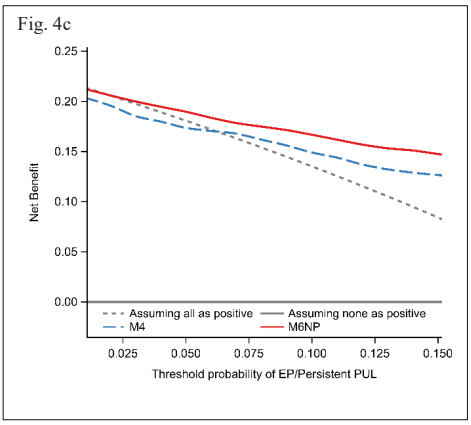


Figure 4. Summary of key outcomes and interpretation of the M6NP and M4 comparisons

(a) The area under the receiver-operating characteristic curves (AUC) of the estimated probabilities for ectopic pregnancy (EP), including persistent pregnancy of unknown location (PUL) versus non-EP. M6NP had a higher AUC than M4, meaning better discriminatory ability.

(b) Calibration curves of the estimated probability for EP should be on the ideal line for perfect calibration. The intercept of the calibration curve is >0 for both models, meaning they are underestimating the total number of EP among PUL, but not as much for M6NP. The calibration slope <1 indicates that estimated probabilities of EP by M6NP and M4 were on average, too high for estimates close to 1 and too low close to 0, but more so for M4. However, M6NP estimates of EP are well calibrated in the region for estimates up to 0.20, where the curve is on the ideal line whereas the curve of M4 is above the ideal line, making underestimations of EP up to circa 0.4.

(c) The decision curves based on the estimated probabilities of EP show that M6NP has a net benefit (NB), clinical utility, at the 5% threshold and beyond, being over the NB of assuming all or none at high risk of EP. The M4 had no NB until the 7.5% threshold.

Diagnostic (classification) performance

For M6NP, the sensitivity for EP was 95%, the false-positive rate of non-EP was 50%, and the NPV was 97% at the 5% probability threshold (**Table 10**). Compared with M6NP, M4 had a lower sensitivity, false-positive rate, and NPV at the 5%, and 10% thresholds (**Table 10**). The NICE algorithm classification performance was analyzed in the validation sample for comparison (**Figure 5**).

Table 10. Performance of M6NP and M4 diagnosing ectopic and non-ectopic pregnancies in 1061 women with PUL

Outcome	M6NP	M4	*P*-value
5% threshold			
Sensitivity for EP[a]	95% (92 to 98)	85% (81 to 90)	<.001
False positive rate of non-EP[b]	50% (47 to 54)	37% (34 to 40)	<.001
Negative predictive value	97% (96 to 99)	94% (92 to 96)	<.001
Low risk classification	40% (37 to 42)	52% (50 to 55)	<.001
10% threshold			
Sensitivity for EP	91% (87 to 94)	80% (75 to 85)	<.001
False positive rate of non-EP	39% (36 to 43)	30% (34 to 37)	<.001
Negative predictive value	96% (94 to 98)	92% (90 to 94)	<.001
Low risk classification	49% (46 to 52)	58% (56 to 61)	<.001

Abbreviations: EP, ectopic pregnancy.
Data presented as % (95% CI).
[a]Including persistent PUL.
[b]Spontaneously resolving PUL and intrauterine pregnancy.

Fig. 5a

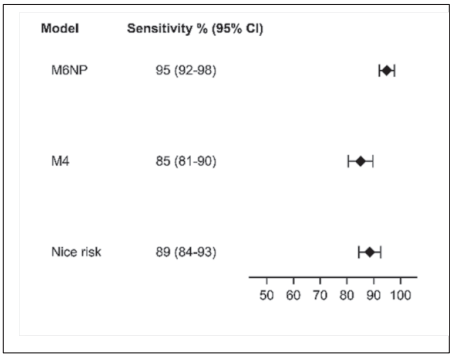


Fig. 5b

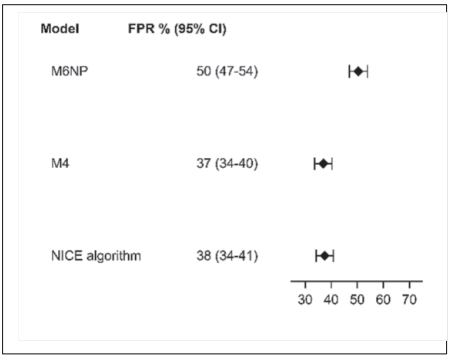


Fig. 5c

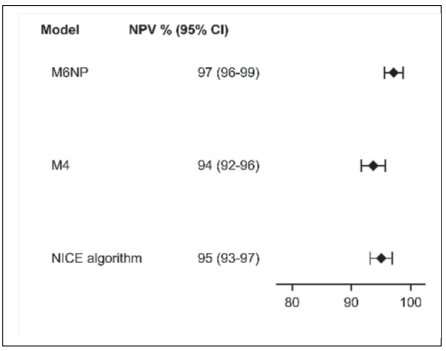


Fig. 5d

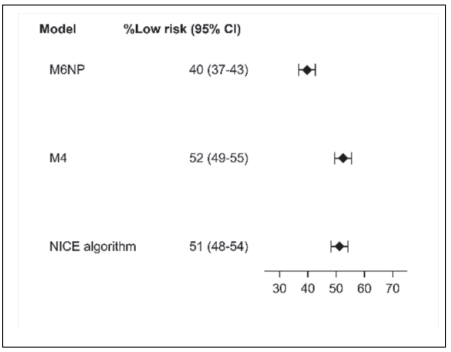


Figure 5. Classification performance for ectopic pregnancy (EP)/persistent pregnancy of unknown location (PUL) and non-EP in 1061 PULs of M6NP, M4 and the NICE algorithm, using the 5% threshold for the prediction models. (a) Sensitivity for EP (b) False positive rate (FPR) of non-EP (percentage of spontaneously resolving PUL and intrauterine pregnancy classified as high risk. (c) Negative predictive value (NPV), percentage of non-EP among PUL classified as low risk. (d) The percentage of PUL classified as low risk.

Paper III

Managing PULs with the M4 prediction model or the NICE algorithm: an RCT

Of 883 women with complaints in early pregnancy who were assessed for trial inclusion, 649 were recruited (**Figure 6**). Of this latter group, 42 women did not meet randomization criteria after the first hCG value was analyzed, leaving 607 women randomly assigned to M4 (305 women) or the NICE algorithm (302 women). Additionally, three women were lost to follow-up in the M4 group and eight in the NICE group. Thus, 302 women in the M4 group and 293 in the NICE group were available for the cross-sectional analysis of the primary outcome and by randomized groups of secondary outcomes. Overall, 513 women managed according to protocol were included in the per-protocol population (256 in the M4 group and 257 in the NICE group).

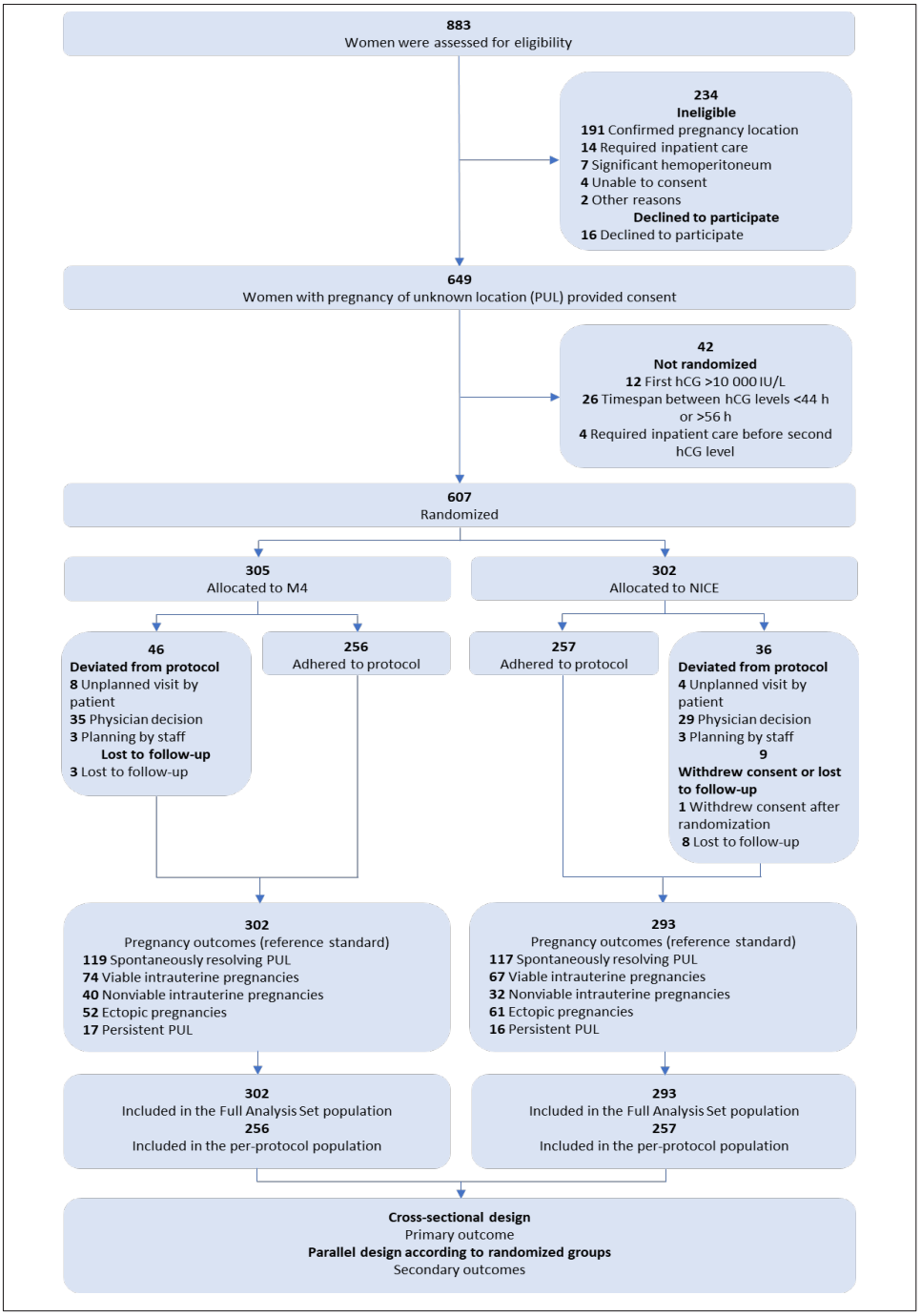


Figure 6. Patient flow in the trial. Allocated protocol (M4 or NICE) predicted if the pregnancy outcome was a spontaneously resolving PUL, an intrauterine or an ectopic pregnancy, and provided a follow-up recommendation.

Characteristics of participants and pregnancy outcome

The baseline and clinical characteristics of the participants were similar between the two study groups (**Table 11**). The pregnancy outcomes in the M4 group were 119 (38.4%) spontaneously resolving PULs, 74 (24.5%) viable IUPs, 40 (13.2%) nonviable IUPs, 52 (17.2%) EPs, and 17 (5.6%) persistent PULs. In the NICE group, there were 117 (38.7%) spontaneously resolving PULs, 67 (22.9) viable IUPs, 32 (10.9) nonviable IUPs, 61 EPs (20.8), and 16 (5.5) persistent PULs (**Table 12**).

Table 11. Baseline and clinical characteristics of participants at the first visit in full analysis set.

Characteristics	**M4 (n=302)**	**NICE (n=293)**
Age (years), mean (SD)	31.3 (5.3)	31.4 (5.6)
BMI (kg/m^2), mean (SD)	24.7 (4.2)	24.6 (4.0)
Gestational age (weeks), mean (SD)	6.00 (1.83)	5.81 (1.51)
Conceived with IVF, n (%)	11 (3.6%)	15 (5.1%)
Gravida, n (%)		
1	90 (29.8%)	83 (28.3%)
2	69 (22.8%)	77 (26.7%)
≥3	143 (47.3.%)	135 (45.7%)
Para, n (%)		
0	176 (58,2%)	168 (56.3%)
1	63 (20.9%)	74 (25.2%)
≥2	63 (20.9%)	51 (17.4%)
≥1 prior miscarriage ≤ 12 weeks, n (%)	84 (27.8%)	86 (29.4%)
Prior ectopic pregnancy, n (%)	25 (8.3%)	26 (8.9%)
≥1 prior caesarean section, n (%)	20 (6.6%)	25 (8.5%)
Symptoms, n (%)		
Abdominal pain and bleeding	178 (58.9%)	175 (59.7%)
Abdominal pain only	87 (28.8%)	81 (27.6%)
Bleeding only	23 (7.6%)	25 (8.5%)
None/Worries	14 (4.6%)	12 (4.1%)
Type of PUL, n (%)		
True PUL	209 (69.2%)	191 (65.2%)
Probable intrauterine pregnancy	68 (22.5%)	70 (23.9%)
Probable ectopic pregnancy	17 (5.6%)	23 (7.8%)
Probable miscarriage	8 (2.6%)	9 (3.1%)
First hCG value, IU/L		
Mean (SD)	1311 (1889)	1336 (1928)
Median (min; max)	540 (7; 9780)	540 (10; 9600)
<500, n (%)	146 (48.3%)	141 (48.1%)
500-999, n (%)	44 (14.6%)	46 (15.7%)
1000-1999, n (%)	47 (15.6%)	42 (14.3%)
≥2000, n (%)	65 (21.5%)	64 (21.8%)

Abbreviations: BMI, body mass index; hCG, human chorionic gonadotropin.
Percentages may not total 100 due to rounding

Table 12. Classification, predicted outcome and observed pregnancy outcome according to randomized groups.

		Pregnancy outcome, n (%) within classification and predicted outcome				
Classification and predicted outcome, n (%)		**Viable IUP**	**Nonviable IUP**	**Spontaneously resolving PUL**	**Ectopic pregnancy**	**Persistent PUL**
M4						
High risk	130 (43.0%)	24 (18.5%)	35 (26.9%)	17 (13.1%)	42 (32.3%)	12 (9.2%)
Low risk	172 (57.0%)	50 (29.1%)	5 (2.9%)	102 (59.3%)	10 (5.8%)	5 (2.9%)
Intrauterine pregnancy	52 (17.2%)	50 (96.2%)	1 (1.9%)	-	1 (1.9%)	-
Spontaneously resolving PUL	120 (39.7%)	-	4 (3.3%)	102 (85%)	9 (7.5%)	5 (4.2%)
NICE						
High risk	116 (40.0%)	4 (3.4%)	23 (19.8%)	25 (21.6%)	49 (42.2%)	15 (12.9%)
Low risk	177 (60.0%)	63 (35.6%)	9 (5.1%)	92 (52.0%)	12 (6.8%)	1 (0.6%)
Intrauterine pregnancy	80 (27.1%)	63 (78.8%)	8 (10.0%)	-	9 (11.3%)	-
Spontaneously resolving PUL	97 (32.9%)	-	1 (1.0%)	92 (94.5%)	3 (3.1%)	1 (1.0%)

Abbreviations: IUP, intrauterine pregnancy; PUL, pregnancy of unknown location.

Primary outcome (diagnostic performance)

In the cross-sectional analysis of the FAS population, no statistically significant difference was observed between M4 and NICE in sensitivity for EP (79% vs 85%, *p*=.15). The NICE algorithm had a higher specificity than M4 for non-EP (74% vs 67%, *p*=.0003) (**Table 13**). The AUC for EP vs non-EP was higher for NICE compared to M4 (0.80 vs 0.73, *p*=.0014) (**Table 13**).

Sensitivity analysis and exploratory outcomes (diagnostic performance)

M4 had better sensitivity for spontaneously resolving PUL, but the NICE algorithm had higher specificity and AUC for spontaneously resolving PUL and higher sensitivity, specificity, and AUC for IUP versus other pregnancy outcomes in the cross-sectional analysis (**Table 13**). The sensitivity of M4 for EP was noninferior to the NICE algorithm (96% vs 91%; difference, 5.0 percentage points; 95% CI, -5 to 15) when analyzed according to randomized groups in the per-protocol population. The specificity for non-EP was not statistically significantly different between M4 and NICE (68% vs 76%; difference, -8 percentage points; 95% CI, -8.0 to 82). In the FAS population, M4 and NICE did not differ significantly in sensitivity for EP (78% vs 83%; difference, -5.0 percentage points; 95% CI, -18 to 8) or in specificity for non-EP (67% vs 76%; difference, -9.0 percentage points; 95% CI, -17 to 0).

Table 13. Cross-sectional analysis of the diagnostic performance of M4 and NICE in the full analysis set (n=595).

Outcome	% (95% CI) n/total M4	NICE	*P*-value
Primary			
Sensitivity for ectopic pregnancy[a]	79% (72 to 85) 115/146	85% (79 to 91) 124/146	.15
Specificity for non-ectopic pregnancy	67% (62 to 71) 300/449	74% (70 to 78) 334/449	.0003
AUC (95% CI) for ectopic vs non-ectopic pregnancy	0.73 (0.69 to 0.77)	0.80 (0.76 to 0.83)	.0014
Positive predictive value	44% (38 to 50) 115/264	52% (46 to 58) 124/239	NA
Negative predictive value	91% (87 to 94) 300/331	94% (91 to 96) 334/356	NA
Exploratory			
Sensitivity for intrauterine pregnancy	47% (40 to 54) 100/213	68% (61 to74) 144/213	<.0001
Specificity for other than intrauterine pregnancy	40% (35 to 44) 151/382	95% (93 to 97) 363/382	<.0001
AUC (95% CI) for intrauterine pregnancy vs other pregnancy	0.57 (0.53 to 0.61)	0.81 (0.78 to 0.85)	<.0001
Sensitivity for spontaneously resolving PUL	85% (80 to 89) 200/236	75% (69 to 81) 177/236	.0003
Specificity for other than spontaneously resolving PUL	64% (59 to 68) 228/359	96% (93 to 98) 343/359	<.0001
AUC (95% CI) for spontaneously resolving PUL vs other pregnancy	0.74 (0.71 to 0.78)	0.85 (0.82 to 0.88)	<.0001

Abbreviations: AUC, area under the curve; NA, not applicable; hCG, human chorionic gonadotrophin; PUL, pregnancy of unknown location.
[a]Including 33 persistent PUL

Secondary outcomes (clinical outcomes)

The hCG data did not differ significantly between the M4 and NICE groups within each pregnancy outcome (**Table 14**). EPs in the NICE group had a numerically higher mean. M4 misclassified 10 EPs as low risk, of which all but one were predicted to be a spontaneously resolving PUL (**Figure 7**). Nine of twelve EPs misclassified by NICE were predicted to be an IUP. Details of clinical outcomes for misclassified EPs are presented in **Table 15**. A higher percentage of women in the NICE group underwent laparoscopic treatment of EP compared to the M4 group (**Table 16**). The rates of EP diagnosed prior to surgery and of methotrexate or expectant management as the only treatment were not significantly different between groups (**Table 16**). In the M4 group, 30.5% (92 of 302) of all PULs were predicted to be spontaneously resolving PULs and were managed in line with protocol, compared to 28.3% (83 of 293) in the NICE group. Four women in the M4 group had positive urine hCG after

2 weeks, and all were managed expectantly, as were two women with positive tests in the NICE group.

Table 14. First and second hCG values and hCG ratio for all pregnancy outcomes according to randomized groups.

Outcome	M4 (n= 302)	NICE (n=293)	Difference Mean (95% CI)
Ectopic pregnancy, n	**52**	**61**	
First hCG value (IU/L)			
Mean (SD)	862 (1514)	1302 (1945)	-439.1 (-1097.3 to 219.2)
Median (min; max)	405 (36; 8000)	610 (40; 9500)	
Second hCG value (IU/L)			
Mean (SD)	913 (1631)	1360 (1979)	-446.5 (-1130.1 to 237.1)
Median (min; max)	370 (28; 9100)	550 (45; 8830)	
Ratio of second/first hCG value (IU/L)			
Mean (SD)	1.06 (0.42)	1.13 (0.45)	-0.07 (-0.23 to 0.10)
Median (min; max)	1.03 (0.26; 2.5)	1.12 (0.27; 2.18)	
Intrauterine pregnancy, n	**114**	**99**	
First hCG value (IU/L)			
Mean (SD)	2182 (2375)	2070 (2069)	111.6 (-494.4 to 717.7)
Median (min; max)	1200 (22; 9780)	1420 (52; 9000)	
Second hCG value (IU/L)			
Mean (SD)	3315 (3511)	3263 (3662)	52.0 (-918.0 to 1022.0)
Median (min; max)	2200 (42; 17000)	2100 (57; 22000)	
Ratio of second/first hCG value (IU/L)			
Mean (SD)	1.80 (0.64)	1.77 (0.70)	0.03 (-0.15 to 0.21)
Median (min; max)	1.91 (0.17; 2.93)	1.89 (0.11; 3.38)	
Spontaneously resolving PUL, n	**119**	**117**	
First hCG value (IU/L)			
Mean (SD)	772 (1177)	844 (1704)	-71.8 (-448.3 to 304.7
Median (min; max)	220 (7; 6500)	190 (10; 9600)	
Second hCG value (IU/L)			
Mean (SD)	290.8 (431.6)	308.3 (547.2)	-17.6 (-144.1 to 109.0)
Median (min; max)	130 (1.2; 2600)	74 (0.6; 2700)	
Ratio of second/first hCG value (IU/L)			
Mean (SD)	0.47 (0.29)	0.45 (0.35)	0.02 (-0.06 to 0.10)
Median (min; max)	0.39 (0.08; 1.88)	0.35 (0.007; 1.74)	
Persistent PUL, n	**17**	**16**	
First hCG value (IU/L)			
Mean (SD)	617 (1003)	607 (959)	9.97. (-687.41 to 707.34)
Median (min; max)	180 (25; 3900)	210 (64; 4000)	
Second hCG value (IU/L)			
Mean (SD)	587 (969)	639 (910)	-52.3 (-720.8 to 616.2)
Median (min; max)	220 (36; 3800)	335 (45; 3800)	
Ratio of second/first hCG value (IU/L)			
Mean (SD)	1.00 (0.35)	1.10 (0.33)	-0.11 (-0.35 to 0.13)
Median (min; max)	1.00 (0.536; 1.933)	1.06 (0.7; 2.04)	

Abbreviations: hCG, human chorionic gonadotropin; PUL, pregnancy of unknown location.

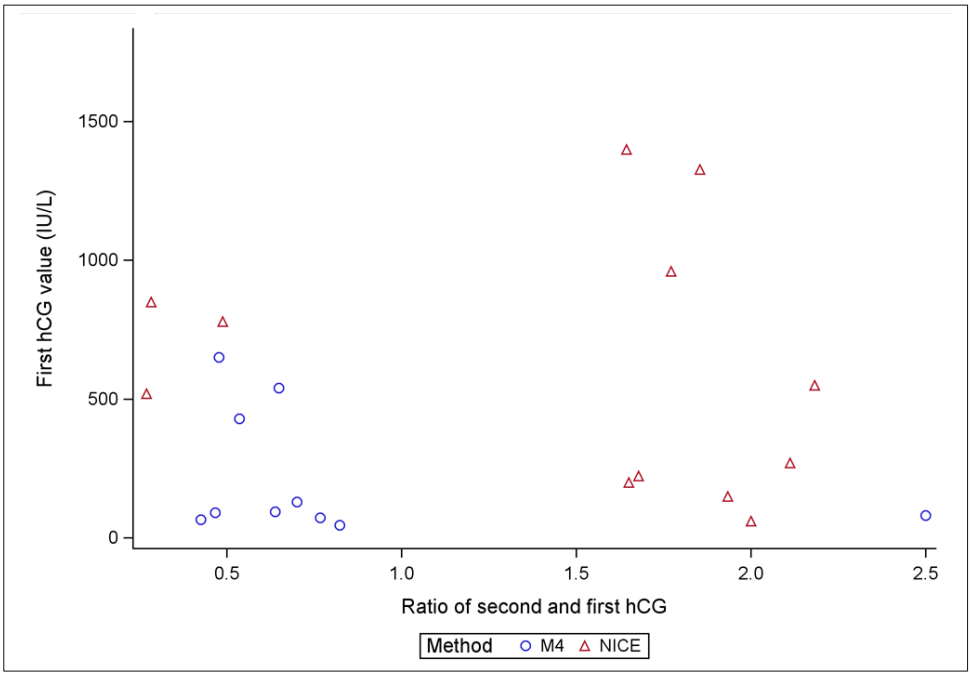


Figure 7. Distribution of first hCG values and ratios of the second/first hCG value of misclassified ectopic pregnancies in the M4 (n=10) and NICE (n=12) groups.

Table 15. Clinical outcomes of 22 misclassified ectopic pregnancies according to randomized groups.

Protocol/ No. of patients	Predicted outcome	Adhered to protocol/primary reason if not	Suspicion of EP first ultrasound	Intervention (s)	Adverse event (s)
M4					
1	Spontaneously resolving PUL	No/physician	Yes	Expectant management	No
2	Spontaneously resolving PUL	No/patient pain	No	Methotrexate	Unplanned visit
3	Spontaneously resolving PUL	No/patient pain	No	Methotrexate	Unplanned visit
4	Spontaneously resolving PUL	No/physician	Yes	Expectant management	Unplanned visit
5	Spontaneously resolving PUL	No/physician	Yes	Expectant management	Unplanned visit
6	Spontaneously resolving PUL	No/physician	Yes	Laparoscopic surgery	Unplanned visit
7	Spontaneously resolving PUL	No/physician	No	Methotrexate	No
8	Spontaneously resolving PUL	No/patient pain	No	Methotrexate	Unplanned visit
9	Intrauterine pregnancy	Yes	No	Methotrexate/laparoscopic surgery	Unplanned visit
10	Spontaneously resolving PUL	No/physician	No[a]	Methotrexate/laparoscopic surgery	Emergency surgery/ ruptured ectopic pregnancy
NICE					
1	Spontaneously resolving PUL	No/physician	Yes	Expectant management	No
2	Spontaneously resolving PUL	No/physician	Yes	Laparoscopic surgery	No
3	Intrauterine pregnancy	No/physician	Yes	Laparoscopic surgery	No
4	Intrauterine pregnancy	Yes	No	Laparoscopic surgery	No
5	Spontaneously resolving PUL	No/patient pain	No	Laparoscopic surgery	No
6	Intrauterine pregnancy	No/physician	Yes	Methotrexate/laparoscopic surgery	Unplanned visit
7	Intrauterine pregnancy	Yes	No	Laparoscopic surgery	No
8	Intrauterine pregnancy	No/physician[b]	No	Laparoscopic surgery	Emergency surgery/ruptured ectopic pregnancy
9	Intrauterine pregnancy	No/physician[a]	No	Methotrexate	No
10	Intrauterine pregnancy	No/patient pain	No	Laparoscopic surgery	Unplanned visit
11	Intrauterine pregnancy	Yes	No	Laparoscopic surgery	No
12	Intrauterine pregnancy	Yes	No	Methotrexate/Laparoscopic surgery	No

[a]Previous ectopic pregnancy (EP).
[b]hCG follow-up 1 week, progression as a viable intrauterine pregnancy, monitoring stopped.

Table 16. Women with ectopic pregnancy, their treatments and ultrasonographic morphology prior to surgery.

Outcome	Total (n=113)	M4 (n=52)	NICE (n=61)	Difference (95% CI)
Laparoscopic surgery, n (%)	51 (45.1)	17 (32.7%)	34 (55.7%)	-23.0 (-42.7 to -3.4)
EP visualized before surgery, n (%)	46 (90.2%)	16 (94.1%)	30 (88.2%)	5.9 (-18.1 to 23.0)
Morphology, n (%)				
Adnexal solid structure	29 (63.0%)	9 (56.3%)	20 (66.7%)	-
Adnexal sac-like structure	13 (28.3%)	5 (31.3%)	8 (26.7%)	-
Adnexal gestational sac with yolk sac	2 (4.3%)	0 (0.0%)	2 (6.7%)	-
Adnexal live pregnancy	1 (2.2%)	1 (6.3%)	0 (0.0%)	-
Interstitial sac-like structure	1 (2.2%)	1 (6.3%)	0 (0.0%)	-
Methotrexate as only treatment, n (%)	27 (23.9%)	14 (26.9%)	13 (21.3%)	5.6 (-12.0 to 23.2)
Expectant management as only treatment, n (%)	24 (21.2%)	12 (23.1%)	12 (19.7%)	3.4 (-13.6 to 20.4)

Abbreviations: EP, ectopic pregnancy.
Differences are given in percentage points.

Results of secondary outcomes according to randomized groups in the FAS population are presented in **Tables 17–19**. M4 and NICE did not differ significantly in time to diagnosis of EP, IUP, or spontaneously resolving PUL; mean number of hCG samplings or TVS performed (**Table 17**); success rate of first-line treatment and hCG resolution time of an EP (**Table 18**); or the combined rate of surgical and medical interventions (**Table 19**).

The total successful first-line treatment of persistent PUL was not significantly different for M4 vs NICE (58.8% vs 81.3%; difference, -22.5 percentage points; 95% CI, -52.4 to 10.6) or within the group of women managed expectantly (45.5% vs 50.0%; difference, -4.5 percentage points; 95% CI, -52.2 to 44.1) or receiving methotrexate (83.3% vs 100%; difference, -16.7 percentage points; 95% CI, -64.1 to 25.5). The mean number of days for hCG resolution did not differ significantly between M4 and NICE (23.6 vs 24.1; difference, -0.5 percentage points; 95% CI, -7.2 to 6.3). One woman with a persistent PUL in the NICE group underwent laparoscopy without evidence of a pregnancy, and two women with a persistent PUL were treated with uterine curettage without a positive specimen. All three women were successfully treated with methotrexate administered because of sustained hCG levels.

In the M4 group, five additional women with a miscarriage underwent uterine curettage in addition to five women having a concomitant laparoscopy, resulting in an overall surgical intervention rate of 7.9%. In the NICE group, one additional woman with a miscarriage underwent uterine curettage, two women had concomitant laparoscopy, two women with a persistent PUL had uterine curettage, and one woman underwent laparoscopy, for an overall surgical intervention rate of 14.3%.

Table 17. Secondary outcomes according to randomized groups in the full analysis set.

Outcome	Total	M4	NICE	Difference Mean (95% CI)	*P*-value
All pregnancies, n	**595**	**302**	**293**		
No. of serum hCG samples					
Mean (SD)	3.6 (2.7)	3.6 (2.7)	3.6 (2.6)	0.05 (-0.38 to 0.47)	.84
Median (min; max)	2 (2; 18)	2 (2; 18)	2 (2; 15)	NA	NA
No. of ultrasounds					
Mean (SD)	2.27 (1.26)	2.28 (1.24)	2.24 (1.28)	0.04 (-0.16 to 0.24)	.70
Median (min; max)	2 (1; 11)	2 (1; 8)	2 (1; 11)	NA	NA
Ectopic pregnancy, n	**113**	**52**	**61**		
Days to diagnosis from randomization					
Mean (SD)	3.8 (4.5)	3.4 (4.2)	4.1 (4.8)	-0.59 (-2.31 to 1.02)	.46
Median (min; max)	2 (0; 26)	2 (1; 26)	2 (0; 22)	NA	NA
No. of serum hCG samples					
Mean (SD)	6.2 (3.4)	6.4 (3.3)	5.9 (3.6)	0.49 (-0.80 to 1.81)	.42
Median (min; max)	5 (2; 18)	6.5 (2; 18)	5 (2; 15)	NA	NA
No. of ultrasounds					
Mean (SD)	3.2 (1.3)	3.2 (1.1)	3.2 (1.5)	0.01 (-0.29 to 0.51)	.96
Median (min; max)	3 (1; 11)	3 (2; 6)	3 (1; 11)	NA	NA
Spontaneously resolving PUL, n	**236**	**119**	**117**		
Days to diagnosis from randomization					
Mean (SD)	13.2 (5.8)	13.3 (5.7)	13.1 (5.9)	0.30 (-1.21 to 1.72)	.72
Median (min; max)	14 (1; 35)	14 (1; 35)	14 (1; 35)	NA	NA
Days of follow-up from diagnosis					
Mean (SD)	13.8 (5.4)	13.8 (5.3)	13.8 (5.4)	0.01 (-1.36 to 1.38)	.99
Median (min; max)	14 (1; 35)	14 (1; 35)	14 (1; 35)	NA	NA
No. of serum hCG samples					
Mean (SD)	2.5 (1.1)	2.5 (1.1)	2.5 (1.1)	-0.12 (-0.31 to 0.29)	.70
Median (min; max)	2 (2; 8)	2 (2; 8)	2 (2; 7)	NA	NA
No. of ultrasounds					
Mean (SD)	1.3 (0.6)	1.3 (0.6)	1.3 (0.7)	-0.01 (-0.18 to 0.2)	.89
Median (min; max)	1 (1; 4)	1 (1; 4)	1 (1; 4)	NA	NA
Intrauterine pregnancy, n	**213**	**114**	**99**		
Days to diagnosis from randomization					
Mean (SD)	7.9 (6.2)	8.11 (6.6)	7.6 (5.6)	0.52 (-1.18 to 2.21)	.55
Median (min; max)	7 (1; 34)	7 (1; 34)	7 (1; 30)	NA	NA
Days of follow-up from diagnosis					
Mean (SD)	11.6 (8.9)	11.9 (9.0)	11.1 (8.8)	0.81 (-1.62 to 3.23)	.52
Median (min; max)	8 (1; 45)	8 (1; 40)	8 (1; 45)	NA	NA
No. of serum hCG samples					
Mean (SD)	2.7 (1.4)	2.7 (1.3)	2.7 (1.5)	-0.01 (-0.4 to 0.4)	.96
Median (min; max)	2 (2; 10)	2 (2; 9)	2 (2; 10)	NA	NA
No. of ultrasounds					
Mean (SD)	2.7 (1.0)	2.7 (1.0)	2.6 (1.0)	0.06 (-0.21 to 0.40)	.58
Median (min; max)	2 (1; 6)	2 (1; 6)	2 (1; 6)	NA	NA
Persistent PUL, n	**33**	**17**	**16**		
No. of serum hCG samples					
Mean (SD)	8.4 (2.7)	9.0 (2.9)	7.7 (2.2)	1.31 (-0.51 to 3.21)	.16
Median (min; max)	9 (4; 16)	9 (4; 16)	8 (4; 11)	NA	NA
No. of ultrasounds					
Mean (SD)	3. (1.2)	3.6 (1.5)	3.2 (0.8)	0.39 (-0.50 to 1.29)	.35
Median (min; max)	3 (1; 8)	4 (1; 8)	3 (2; 5)	NA	NA

Abbreviations: hCG, human chorionic gonadotropin; PUL, pregnancy of unknown location. NA, not applicable.
Differences are reported as M4-NICE.

Table 18. Successful first-line treatment in 113 women with ectopic pregnancy, their hCG data and hCG resolution time.

Outcome	M4 (n=52)	NICE (n=61)	Difference (95% CI)	*P*-value
Total successful first-line treatment, n (%)	36 (69.2%)	48 (78.7%)	-9.5 (-27.5 to 8.5)	.35
hCG resolution time[a], d				
Mean (SD)	17.4 (15.0)	14.6 (15.6)	-2.81 (-8.51 to 42.12)	.33
Median (min; max)	15 (0; 61)	8 (0; 56)	NA	NA
First-line treatment, n (%)				
Laparoscopic surgery	13 (25.0%)	24 (39.3%)	-14.3 (-36.5 to 0.1)	.072
Methotrexate	16 (30.8%)	16 (26.2%)	4.6 (-12.7 to 21.7)	.74
Expectant management,	23 (44.2%)	21 (34.4%)	9.8 (-10.0 to 29.6)	.38
Successful first-line treatment by modality				
Laparoscopic surgery, n (%)	11 (84.6%)[b]	24 (100%)	-15.4 (-45.4 to 2.3)	.23
hCG resolution time, d				
Mean (SD)	3.09 (5.03)	1.92 (4.60)	1.17 (-2.33 to 4.68)	.50
Median (min; max)	1 (0; 15)	1 (0, 23)	NA	NA
First hCG value IU/L				
Mean (SD)	1929 (2201)	1924 (2419)	4.72 (-1713.8 to 1723.3)	1.00
Median (min; max)	1100 (310; 8000)	1300 (43; 9500)	NA	NA
Ratio of second and first hCG				
Mean (SD)	1.16 (0.35)	1.17 (0.46)	-0.01 (-0.31 to 0.29)	.93
Median (min; max)	1.16 (0.68; 1.78)	1.09 (0.27; 2.18)	NA	NA
Methotrexate, n (%)	13 (81.3%)	12 (75.0%)	6.3 (-24.7 to 36.3)	1.00
hCG resolution time, d				
Mean (SD)	27.7 (18.2)	27.9 (9.6)	-0.20 (-11.2 to 10.8)	.97
Median (min; max)	19 (10, 61)	19.5 (11, 43)	NA	NA
First hCG value IU/L				
Mean (SD)	433.3 (399.7)	1318 (2117)	-885 (-2026 to 256)	.12
Median (min; max)	330 (51; 1300)	510 (60; 8300	NA	NA
Ratio of second and first hCG				
Mean (SD)	1.17 (0.47)	1.23 (0.43)	-0.06 (-0.41 to 0.28)	.69
Median (min; max)	1.14 (0.64; 2.5)	1.32 (0.53; 2)	NA	NA
Expectant management, n (%)	12 (52.2%)	12 (57.1%)	-5.0 (-34.7 to 24.9)	.98
hCG resolution time, d				
Mean (SD)	18.7 (9.8)	19.2 (16.5)	-0.50 (-8.93 to 7.93)	.90
Median (min; max)	19 (5; 38)	16 (4; 56)	NA	NA
First hCG value IU/L				
Mean (SD)	704 (1497)	577 (492)	127.1 (-550.0 to 804.3)	.70
Median (min; max)	310 (36; 7300)	380 (40; 1900)	NA	NA
Ratio of second and first hCG				
Mean (SD)	0.96 (0.42)	1.00 (0.44)	-0.04 (-0.29 to 0.21)	.74
Median (min; max)	0.929 (0.261; 1.846)	1.04 (0.28; 1.77)	NA	NA

Abbreviations: hCG, human chorionic gonadotropin; NA, not applicable.
Differences are reported as M4-NICE.
Differences between percentages are given in percentage points; differences between other values are given in the unit for that value.
[a]Calculated from the time of diagnosis until hCG <5.3/5.0 IU/L, data were missing for one woman in the M4 group.
[b]Persistent trophoblast treated with methotrexate after two cases of salpingotomy.

Table 19. Combined rate of surgical and medical interventions according to randomized groups.

	n/total (%)		
Outcome	**M4 (n=302)**	**NICE (n=293)**	**Difference (95% CI)**
Ectopic pregnancy	40/52 (76.9%)	49/61 (80.3%)	-3.4 (-20.4 to 13.6)
Intrauterine pregnancy	22/114 (19.3%)	15/99 (15.2%)	4.1 (-6.9 to 15.2)
Persistent PUL	12/17 (70.6%)	13/16 (81.3%)	-10.7 (-40.4 to 21.7)
Spontaneously resolving PUL	0/119 (0.0%)	0/117 (0.0%)	NA

Abbreviations: PUL, pregnancy of unknown location; NA, not applicable.
Differences are reported as M4-NICE and are given in percentage points.

Adverse events

Forty-five women (14.9%) had any adverse event in the M4 group and 34 (11.6%) in the NICE group with no statistically significant difference (**Table 20**). Among 52 EPs in the M4 group there were two (3.8%) verified ruptures and four (6.6%) in the NICE group. Two (10.5%) of 19 laparoscopies performed were negative in the M4 group and six (15.0%) of 40 laparoscopies in the NICE group. Two women in the NICE group with a subsequent confirmed IUP were given methotrexate prior to laparoscopy and viability was not excluded for sure before treatment. One woman in the M4 group was given misoprostol although a slow-growing gestational sac was seen reaching 16 mm once treatment started. Details of seven women that underwent laparoscopy with subsequent diagnosis of a miscarriage are showed in **Table 21**.

Table 20. Adverse events according to randomized groups.

Outcome	**M4 (n=302)**	**NICE (n=293)**	**Difference (95% CI)**	***P*-value**
Women with any adverse event[a], n (%)	45 (14.9%)	34 (11.6%)	3.3 (-2.5 to 9.1)	.29
Ruptured ectopic pregnancy	2 (0.7%)	4 (1.4%)	-0.7 (-2.7 to 1.2)	.66
Emergency surgery	4 (1.3%)	3 (1.0%)	0.3 (-1.8 to 2.4)	1.00
Methotrexate or misoprostol given to a potentially viable intrauterine pregnancy	1 (0.3%)	2 (0.7%)	-0.3 (-1.8 to 1.1)	1.00
Negative laparoscopy	2 (0.7%)	6 (2.0%)	-1.4 (-3.6 to 0.8)	.27
Unplanned visit by patient	42 (13.9%)	27 (9.2%)	4.7 (-0.8 to 10.1)	.096
Side effect of methotrexate[b]	2 (0.7%)	1 (0.3%)	0.3 (-1.1 to 1.8)	1.00
Infection after surgery	0 (0.0%)	1 (0.3%)	-0.3 (-1.3 to 0.7)	.98

Differences are reported as M4-NICE.
Differences are given in percentage points.
[a]Women could have >1 recorded adverse event.
[b]All women had mild transient elevations in alanine transaminase levels of which two with an ectopic pregnancy had received a repeat dose of methotrexate.

Table 21. Details of women with an intrauterine pregnancy undergoing laparoscopy according to randomized groups.

Protocol/ No. of patients	Ultrasound finding	Classification	First hCG (IU/L)	Second hCG (IU/L)	hCG ratio	Intervention (s)	Primary indication for intervention
M4							
1	No sign of pregnancy	High risk	1300	1200	0,92	Laparoscopy and uterine curettage	Diagnostic
2	Intracavitary sac-like structure, 2 mm	High risk	1200	1100	0.92	Laparoscopy and uterine curettage[a]	Pain
NICE							
1	Intracavitary sac-like structure	High risk	800	870	1.09	Methotrexate and laparoscopy	Diagnostic
2	No sign of pregnancy	Low risk	52	110	2.16	Laparoscopy and uterine curettage	Pain
3	Intracavitary sac-like structure 1.8 mm	High risk	650	860	1.32	Methotrexate and laparoscopy	Diagnostic
4	Adnexal solid structure	High risk	2000	3000	1.5	Laparoscopy and uterine curettage	Ultrasound finding and hCG results
5	Adnexal solid structure	High risk	3100	3500	1.13	Laparoscopy	Ultrasound finding and hCG results

Abbreviations: hCG, human chorionic gonadotropin.
[a]Emergency surgery.

Emergency laparoscopic surgery was carried out in five women with an EP, where three had ruptured two in the NICE group and one in the M4 group. One of the women in the NICE group returned shortly after the second hCG requiring laparoscopic surgery and transfusion with one unit of blood. The woman in the M4 group was classified as low risk but was assessed as high risk and was under treatment with methotrexate. The third woman with a ruptured EP was predicted to be an IUP by NICE but instead of planning a TVS after seven days a hCG was taken. Levels of hCG indicated a normally progressing IUP whereby monitoring was ended but the patient returned with ambulance nine days after randomization needing emergency laparoscopy. Additionally, two women with an EP had emergency laparoscopy after either attempting a repeat dose of methotrexate or expectant management without having a ruptured EP.

Three of six cases of ruptured EP underwent planned laparoscopic surgery and no suspicion of rupture preceded two cases, while the third had hemoperitoneum. All women had high risk assessment, were diagnosed prior to surgery and one had received methotrexate. Out of six women with a ruptured EP, one was classified as high risk by M4 and three were classified as high risk by NICE. Three of these women were managed according to protocol and diagnosed shortly after.

Results of secondary outcomes for the per-protocol population were generally consistent with those in the FAS population, data not shown.

Subgroup analysis (high- vs low-risk classification)
In the FAS population, a total 246 PULs were classified as high risk and 349 were classified as low risk when aggregating the classification of the M4 and NICE groups (**Table 12**). Secondary outcomes were compared between high and low risk classification groups within pregnancy outcome or totally.

The times to diagnosis of EPs and IUPs were not significantly different for PULs classified as high vs low risk (**Figure 8**) but were longer for spontaneously resolving PULs classified as high risk. The hCG resolution time was longer for EPs classified as high vs low risk (17.2 vs 10.7 days; mean difference, 6.5 days; 95% CI, 1.0 to 12.0 days), as was the length of follow-up (20.8 vs 15.0 days; mean difference, 5.8 days; 95% CI, 0.3 to 11.7 days). The median time to diagnosis of EPs in the per-protocol population was shorter for EPs classified as high risk (two days) compared to EPs classified as low risk (8.5 days) (**Figure 8**).

Fig. 8a

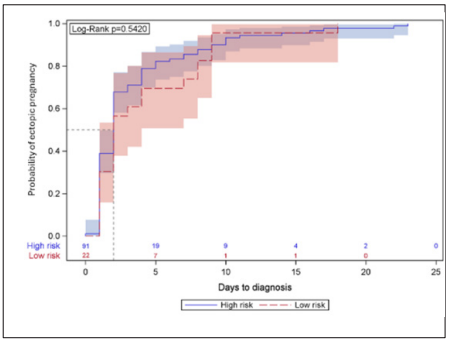


Fig. 8b

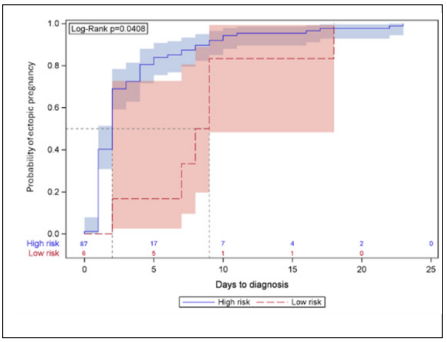


Fig. 8c

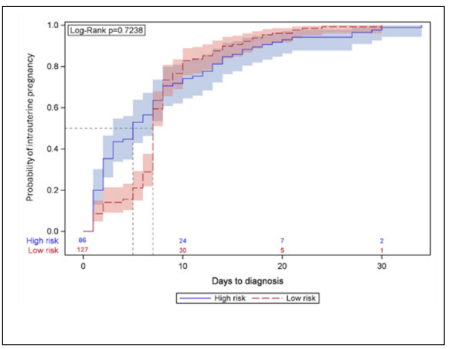


Fig. 8d

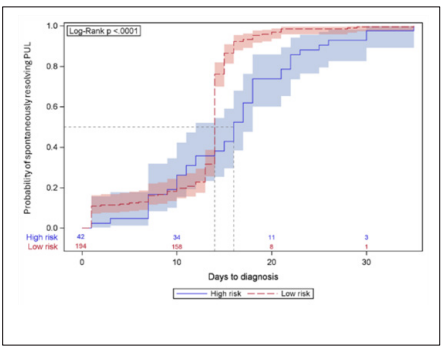


Figure 8. Kaplan-Meier curves for days from randomization to diagnosis of ectopic pregnancy (EP), intrauterine pregnancy (IUP) or spontaneously resolving pregnancy of unknown location (PUL) according to high and low risk classification. (a) The median time to diagnosis of EP classified as high risk or low risk was 2 days in the full analysis set population. (b) The median time to diagnosis for EP classified as high risk was 2 days vs 8.5 days for EP classified as low risk in the per-protocol population. The median time to diagnosis of IUP classified as high risk was 5 days vs 7 days for IUP classified as low risk in the full analysis set population. (c) The median time to diagnosis of spontaneously resolving PUL classified as high risk was 16 days vs 14 days for those classified as low risk in the full analysis set population. (d) Shaded areas show 95% CIs.

Overall successful first line treatment (77.8% vs 60.9%; difference, 16.9 percentage points; 95% CI, -4.7 to 39.8) and the mean number of hCG measures (6.4 vs 5.1; difference, 1.3; 95% CI, -0.3 to 2.9, p=.11) and TVS 3.2 vs 3.2; difference, -0.04; 95% CI, -0.7 to 0.6, p=.90) did not differ significantly between EPs classified as high risk vs low risk. In total, thirty-nine of 91 (42.9%) EPs classified as high risk and 12

of 22 (54.5%) classified as low risk underwent laparoscopic surgery. IUPs and spontaneously resolving PULs classified as high risk involved more hCG testing and ultrasound examinations than those classified as low risk (Table 22). IUPs classified as high risk involved surgical and medical procedures more often than those classified as low risk (39.5% vs 2.3%; difference, 37.2 percentage points; 95% CI, 25.6 to 48.8; p<.0001), but EPs and persistent PULs did not differ significantly in this regard. Spontaneously resolving PUL with false positive results had longer follow-up period compared to those classified as low risk (16.6 vs 13.2 days; mean difference, 3.4; 95% CI, 1.2 to 5.7, p=.0035). The rate of any adverse event was higher for PUL classified as high risk but the only adverse event category that was significantly higher was having a negative laparoscopy (**Table 23**).

Table 22. Number of serum hCG test and transvaginal ultrasound performed in women with a spontaneously resolving PUL or an intrauterine pregnancies classified as either high or low risk in the full analysis set.

Outcome	High risk	Low risk	Difference Mean (95% CI)	*P*-value
Intrauterine pregnancy, n	**86**	**127**		
No. of serum hCG samples				
Mean (SD)	3.3 (1.8)	2.3 (0.9)	1.00 (0.59 to 1.41)	.0009
Median (min; max)	3 (2; 10)	2 (2; 9)	NA	NA
No. of ultrasounds				
Mean (SD)	3.0 (1.2)	2.4 (0.8)	0.56 (0.26 to 0.85)	<.0001
Median (min; max)	3 (1; 6)	2 (1; 5)	NA	NA
Spontaneously resolving PUL, n	**42**	**194**		
No. of serum hCG samples				
Mean (SD)	3.9 (1.5)	2.2 (0.7)	1.75 (1.26 to 2.23)	<.0001
Median (min; max)	4 (2; 8)	2 (2; 7)	NA	NA
No. of ultrasounds				
Mean (SD)	2.1 (0.9)	1.1 (0.4)	0.96 (0.68 to 1.23)	<.0001
Median (min; max)	2 (1; 4)	1 (1; 3)	NA	NA

Abbreviations: hCG, human chorionic gonadotropin; PUL, pregnancy of unknown location; NA, not applicable.
Differences are reported as M4-NICE.

Table 23. Adverse events according to risk classification in the full analysis set (n=595).

Outcome	High risk (n=246)	Low risk (n=349)	Difference (95% CI)	*P*-value
Women with any adverse event[a], n (%)	45 (18.3%)	34 (9.7%)	8.6 (2.5 to 14.6)	.0039
Ruptured ectopic pregnancy	4 (1.6%)	2 (0.6%)	1.1 (-1.1 to 3.2)	.39
Emergency surgery	5 (2.0%)	2 (0.6%)	1.5 (-0.8 to 3.7)	.22
Methotrexate or misoprostol given to a potentially viable intrauterine pregnancy	3 (1.2%)	0 (0.0%)	1.2 (-0.5 to 2.9)	.14
Negative laparoscopy	7 (2.8%)	1 (0.3%)	2.6 (0.1 to 5.1)	.02
Unplanned visits initiated by patient	35 (14.2%)	34 (9.7%)	4.5 (-1.2 to 10.2)	.12
Side effect of methotrexate[b]	3 (1.2%)	0 (0.0%)	1.2 (-0.5 to 2.9)	.14
Infection after surgery	1 (0.4%)	0 (0.0%)	0.4 (-0.7 to 1.5)	.83

Differences are reported as high risk-low risk.

Differences are given in percentage points.

[a]Women could have >1 recorded adverse event.

[b]All were mild transient elevations in alanine transaminase levels of which two women with an ectopic pregnancy had received a repeat dose of methotrexate.

Paper IV

Psychological aspects and health-related quality of life in women with a PUL: a prospective multicenter cohort study

In the PUL group, 59.4% (177 of 298) of recruited women responded to the HADS at 1 week and the HADS and the SF-36 at 4 weeks. One participant was lost to follow-up, and 14 did not complete the RCT and were thus excluded from the analysis. The response rate for the HADS at 1 week in the antenatal group was 56.7% (114 of 201). From this group, eight participants were excluded upon review of health questionnaires revealing noneligible criteria (gestational age ≥12 weeks or had undergone a viability scan). Study outcomes were determined for 162 participants in the PUL group and 106 participants in the antenatal group. In the PUL group, the pregnancy outcomes were as follows: 30 viable pregnancies, 29 EPs, 13 miscarriages, 83 spontaneously resolving PULs, and 7 persistent PULs.

Three groups were evaluated: (1) the primary study group (the entire PUL vs antenatal group), (2) the PUL group (nonviable vs viable pregnancy), and (3) a subgroup (viable pregnancy within the PUL group vs antenatal group). Outcomes were adjusted for gestational age, previous pregnancy loss, planned pregnancy, and investigation for infertility.

Baseline characteristics

The mean age was similar among participants with a nonviable pregnancy within the PUL group (31.9), a viable pregnancy (27.8), and the antenatal group (33.1). The mean gestational age was lower in the PUL group compared to the antenatal group (6.1 vs 7.7 weeks; p<.0001), vaginal bleeding was more common (76% vs 13%; p<.0001), and the rate of a previous EP was higher (8% vs 0%; p=.0024). Participants in the antenatal group had a higher self-reported general health (p=.027) than participants in the PUL group. Compared to participants with a nonviable pregnancy, participants with a viable pregnancy within the PUL group had a shorter mean gestational age (4.7 vs 6.4 weeks; p <.0001), less frequent vaginal bleeding (20% vs 88%; p <.0001), and a lower rate of infertility evaluation (7% vs 26%; p=.030). The gestational age was shorter in the antenatal group (p<.0001) and previous EP was less common (p=.0040) than in participants with a viable pregnancy within the PUL group.

Primary, secondary, and exploratory outcomes

At 1 week, an anxiety score ≥8 was more frequent in the PUL group than in the antenatal group (58.6% vs 29.2%; aOR, 3.20; 95% CI, 1.74 to 5.87), as were all other outcomes (**Table 24**). Within the PUL group anxiety and depression outcomes were not statistically significantly different at 1 or 4 weeks (**Table 24**). Compared to women in the antenatal group, women with a viable IUP within the PUL group had a higher adjusted mean ($mean_{adj}$) score on the depression subscale (4.11 vs 2.63: $mean_{adj}$ difference, 1.48; 95% CI, 0.01 to 2.96), but other HADS outcomes did not differ significantly.

Results of the sensitivity analysis (multiple imputation of missing confounders, 36 missing gestational age values and 4 missing planned pregnancies) were generally consistent with the primary analysis of an anxiety score ≥8 at 1 week for all three study groups: aOR 3.10 (95% CI, 1.73 to 5.56) for the primary (PUL group vs antenatal group) study group; aOR 0.98 (95% CI, 0.37 to 2.61) for the PUL (nonviable vs viable) group; and aOR 3.13 (95% CI, 0.97 to 10.1) for the subgroup (viable pregnancy within the PUL group vs antenatal group).

Change in HADS scores between 1 and 4 weeks in the PUL group

For women with a nonviable pregnancy, all classes on both subscales of the HADS were significantly lower at 4 weeks compared to the 1-week assessment. Frequencies of scores at different cutoffs for the HADS at 1 vs 4 weeks were as follows: anxiety ≥8, 58.0% vs 48.1% (p=.015); anxiety ≥11, 35.1% vs 20.6% (p<.0001); depression

≥8, 32.1% vs 17.6%; (p<.0001); and depression ≥11, 29.8% vs 16.8% (p=.0005). For scores ≥8 on both subscales, values for 1 vs 4 weeks were 12.2% vs 1.5% (p=.018). For women with a viable pregnancy, most anxiety and depression scores were not statistically significantly lower at 4 weeks compared to 1 week. Frequencies of scores at different cutoffs for the HADS at 1 vs 4 weeks were as follows: anxiety ≥8, 60.0% vs 30.0% (p=.012); anxiety ≥11, 33.3% vs 16.7% (p=.063); depression ≥8, 16.7% vs 23.3% (p=.56); and depression ≥11, 10.0% vs 10.0% (p=1.00). In this group, frequencies of a score ≥8 on both subscales were 13.3% at 1 week vs 16.7% at 4 weeks (p=1.00). There was a statistically significant change of mean anxiety score from 1 to 4 weeks (8.53 vs 6.57; mean difference, -1.97; 95% CI, -2.92 to -1.01) but not for the mean depression score (4.50 vs 4.87; mean difference, 0.37; 95% CI, -0.78 to 1.51).

SF-36

The mean_{adj} PCS score was below the norm average of 50 for women with a viable pregnancy (47.7), which was significantly lower than for women with a nonviable pregnancy (52.2; mean_{adj} difference, -4.5; 95% CI, -8.3 to -0.7), mainly because of significantly lower scores on the physical functioning domain (80.8 vs 94.0; mean_{adj} difference, -13.2; 95% CI, -19.6 to -6.9). The mean_{adj} MCS scores for both groups were below the norm average of 50, but no significant difference was observed between women with a viable vs a nonviable pregnancy (38.0 vs 33.5; mean_{adj} difference, 4.5; 95% CI, -2.4 to 11.3).

Table 24. Anxiety and depression (HADS) at 1 week in the three study groups and also at 4 weeks for the PUL group.

Outcome	n (%) PUL group (n=162)	Antenatal group (n=106)	Difference[a] (95% CI)	OR (95% CI)	*P*- value	aOR[b] (95% CI)	*P*- value$_{adj}$
Primary study group							
Primary							
Anxiety score ≥8	95 (58.6%)	31 (29.2%)	29.4 (17.1 to 41.7)	3.43 (2.03 to 5.78)	<.0001	3.20 (1.74 to 5.87)	.0002
Depression score ≥8	47 (29.0%)	8 (7.5%)	21.5 (12.1 to 30.9)	5.01 (2.26 to 11.10)	<.0001	6.18 (2.43 to 15.73)	.0001
Secondary							
Score on both subscales ≥8	43 (26.5%)	6 (5.7%)	20.9 (12.0 to 29.8)	6.02 (2.46 to 14.73)	<.0001	7.17 (2.52 to 20.42)	.0002
Anxiety score, mean$_{adj}$ (SD)	8.70 (4.65)	5.63 (3.43)	3.06 (1.86 to 4.26)	NA	NA	NA	<.0001[c]
Depression score, mean$_{adj}$ (SD)	5.32 (3.85)	2.62 (2.63)	2.70 (1.76 to 3.64)	NA	NA	NA	<.0001
Exploratory							
Anxiety score ≥11	56 (34.6%)	9 (8.5%)	26.1 (16.3 to 35.9)	5.69 (2.67 to 12.12)	<.0001	6.20 (2.60 to 14.79)	<.0001
Depression score ≥11	19 (11.7%)	2 (1.9%)	9.8 (3.5 to 16.2)	6.91 (1.57 to 30.31)	.010	6.13 (1.26 to 29.92)	.025
PUL group							
Primary	**Nonviable pregnancy[d] (n=132)**	**Viable pregnancy (n=30)**					
Anxiety score ≥8	76 (58.0%)	18 (60.0%)	-2.0 (-23.5 to 19.5)	0.92 (0.41 to 2.07)	.84	1.28 (0.46 to 3.55)	.64
Depression score ≥8	42 (32.1%)	5 (16.7%)	15.4 (-2.2 to 33.0)	2.36 (0.84 to 6.60)	.10	3.14 (0.77 to 12.75)	.11
Secondary							
Score on both subscales ≥8	39 (29.8%)	4 (13.3%)	16.4 (-0.1 to 33.0)	2.76 (0.90 to 8.42)	.075	2.85 (0.69 to 11.76)	.15
Anxiety score, mean$_{adj}$ (SD)	9.09 (4.59)	7.79 (5.02)	1.30 (-1.12 to 3.72)	NA	NA	NA	.29
Depression score, mean$_{adj}$ (SD)	5.64 (3.94)	3.90 (3.44)	1.74 (-0.16 to 3.65)	NA	NA	NA	.22
4-weeks assessment							
Anxiety score ≥8	63 (48.1%)	9 (30.0%)	18.1 (-2.5 to 38.6)	2.16 (0.92 to 5.07)	.077	2.58 (0.87 to 7.62)	.087
Depression score ≥8	23 (17.6%)	7 (23.3%)	-5.8 (-24.3 to 12.7)	0.70 (0.27 to 1.82)	.47	0.69 (0.18 to 2.63)	.59
Score on both subscales ≥8	22 (16.8%)	5 (16.7%)	0.1 (-16.7 to 17.0)	1.01 (0.35 to 2.92)	.99	1.38 (0.30 to 6.28)	.68
Anxiety score, mean$_{adj}$ (SD)	7.22 (4.12)	6.18 (4.17)	1.04 (-1.11 to 3.19)	NA	NA	NA	.34
Depression score, mean$_{adj}$ (SD)	3.83 (2.99)	4.89 (4.09)	-1.06 (-2.75 to 0.63)	NA	NA	NA	.22
Exploratory							
Anxiety score ≥11	46 (35.1%)	10 (33.3%)	1.8 (-19.0 to 22.6)	1.08 (0.47 to 2.51)	.85	1.12 (0.37 to 3.40)	.12
Depression score ≥11	16 (12.2%)	3 (10.0%)	2.2 (-11.9 to 16.4)	1.25 (0.34 to 4.61)	.062	4.14 (0.45 to 38.14)	.21
Subgroup							
Primary	**Viable pregnancy from PUL (n=30)**	**Antenatal group (n=106)**					
Anxiety score ≥8,	18 (60.0%)	31 (29.2%)	30.8 (9.1 to 52.4)	3.63 (1.56 to 8.42)	.0027	2.42 (0.75 to 7.86)	.14
Depression score ≥8	5 (16.7%)	8 (7.5%)	9.1 (-7.3 to 25.5)	2.45 (0.74 to 8.14)	.14	1.91 (0.24 to 14.91)	.54

Cont.

Table 24 continues

Outcome	Viable pregnancy from PUL (n=30)	Antenatal group (n=106)	Difference[a] (95% CI)	OR (95% CI)	*P*- value	aOR[b] (95% CI)	*P*- value$_{adj}$
Secondary							
Score on both subscales ≥8	4 (13.3%)	6 (5.7%)	7.7 (-7.4 to 22.7)	2.56 (0.67 to 9.76)	.17	2.56 (0.28 to 23.29)	.40
Anxiety score, mean$_{adj}$ (SD)	7.21 (5.02)	5.63 (3.43)	1.58 (-0.50 to 3.65)	NA	NA	NA	.14
Depression score, mean$_{adj}$ (SD)	4.11 (3.44)	2.63 (2.63)	1.48 (0.01 to 2.96)	NA	NA	NA	.049
Exploratory							
Anxiety score ≥11	10 (33.3%)	9 (8.5%)	24.8 (5.0 to 44.7)	5.39 (1.94 to 14.96)	.0012	3.73 (0.71 to 19.68)	.12
Depression score ≥11	3 (10.0%)	2 (1.9%)	8.1 (-5.1 to 21.3)	5.78 (0.92 to 36.33)	.062	0.36 (0.01 to 14.70)	.59

Abbreviations: HADS, hospital anxiety and depression scale; PUL, pregnancy of unknown location, OR, odds ratio, aOR, adjusted odds ratio; NA, not applicable.

Data presented as n (%) if not stated otherwise.

[a]Differences between percentages are given in percentage points; differences between other values are given in the unit for that value. Adjusted difference is presented for means and unadjusted for percentage points.

[b]Adjusted for age (years), gestational age, previous pregnancy loss, planned pregnancy and investigated for infertility using logistic regression. P-values and OR are based on original values and not classes within HADS. OR is the ratio for the odds for an increase of the predictor of one unit.

[c]Adjusted for age (years), gestational age (weeks), previous pregnancy loss, planned pregnancy and investigated for infertility using Analysis of Covariance.

[d]Nonviable pregnancies were ectopic pregnancy, miscarriage, spontaneously resolving PUL and persistent PUL.

Discussion

Paper I

The EP rate among PUL cases was 19.8%, which is at the higher end of what has been reported.[13, 67, 167] Laparoscopic intervention was undertaken in one out of five women, and most of the EPs were diagnosed with TVS before surgery. The M4 and NICE algorithms have similar and better sensitivity than the two other protocols but misclassified more non-EPs. In a subgroup of PULs, fewer non-EPs were misclassified, especially by M4, yielding a higher specificity than NICE. The M4 sensitivity for EP was similar but specificity was lower compared to findings from a study performed in six EPAUs in the UK.[54] hCG profiles of misclassified EPs differed between M4 and NICE in the present study. EP misclassification by M4 may be less harmful because of the lower mean hCG levels and primarily declining hCG levels compared to those misclassified by NICE.[168]

Paper II

Compared to M4, M6NP has better predictive performance, gives better calibrated estimated probabilities of EP, and offers superior NB. The estimates of AUC and the calibration slope were precise for M6NP and M4, suggesting an adequate sample size for externally validating the prediction models.[148] The calibration-in-the-large showed that both models on average underestimated the prevalence of EP among the cases, especially M4. The lower rate of EP in the development cohorts could have contributed to this miscalibration.[87, 91] The flexible calibration curve revealed that unlike M4, M6NP offered accurate estimated probabilities of EP in the region where clinical decision-making is made. It can be argued that less well-calibrated estimates outside this region may be irrelevant for triaging PUL, making no prognostic statements.[152]

A main difference between the models from a clinical perspective is that M6NP had a higher sensitivity for EP and false-positive rate for non-EP compared to M4. This advantage could lead to a shorter time to EP diagnosis, but it also could lead to more women with PUL undergoing close monitoring and potentially unnecessary diagnostic procedures. The 10% threshold may be clinically more optimal for M6NP because of a better tradeoff between true- and false-positive results, maintaining a high sensitivity for EP. The results of applying M6NP and M4 in the present study were highly consistent with those reported from eight EPAUs in the UK.[167] An

updated version of M6NP was recently validated in two EPAUs in the UK, yielding an AUC of 0.82–0.86 for EP versus non-EP, a sensitivity of 93%–94%, and a false-positive rate of 47%–52%.[92] These results align well with those in the present study. The findings of Paper I and Paper II suggest that M4, M6NP, and the NICE algorithm are useful for classifying PULs as high and low risk in a gynecological emergency unit. However, when evaluating key measures of model performance, M4 was inferior to M6NP in every aspect. M4 was developed in a smaller number of participants with more strict inclusion criteria than M6NP. Consequently, M4 may not be as robust as M6NP when applied to a more diverse validation population.

Paper III

In this RCT, most women were managed in line with the prediction and recommended follow-up in their respective randomized groups. The diagnostic and predictive performance of the NICE algorithm was generally better than that of M4 when evaluated on cross-sectional data. Analyses of the performance according to randomized groups, both in the FAS population and the per-protocol population, showed no difference, probably because the numbers of participants were too low in each group. The performance estimates by randomized groups were in the same direction in the FAS population as those of the cross-sectional analysis in the FAS population.

The findings did not demonstrate any statistically significant between-group differences in number of hCGs or TVS undertaken, time to diagnosis of pregnancy outcomes, length of follow-up, first-line treatment success for EPs, or adverse events. The rate of adverse events was almost twice as high for women classified as being at high risk compared to those at low risk. The time to diagnosis of EPs classified as high risk was not shorter than for EPs classified as low risk (false-negative results), but hCG resolution time and follow-up time were longer in the FAS population. However, in the per-protocol population time to diagnosis was shorter for EPs classified as high risk than for EPs classified as low risk. Spontaneously resolving PULs with false-positive results involved more hCG and TVS examinations and a prolonged time to diagnosis and length of follow-up compared to those correctly classified as low risk in the FAS population. There were more TVS exams and hCG measurements for women with an IUP classified as high risk than for women with an IUP classified as low risk.

The sensitivity of M4 for EPs was similar to that reported in studies from the UK and Australia.[13, 36] Compared to the results from Paper I, M4 had a sensitivity for EPs

that was eight percentage points lower but similar specificity for non-EPs.[51] High hCG levels increase estimates for the probability of an EP, and in PUL populations with lower hCG levels, the sensitivity is impaired, as previously reported.[90] The mean hCG level for EP was 1471 IU/L in Paper I vs 862 IU/L in this study, whereas other pregnancy outcomes involved similar hCG levels. These findings suggest that there were no significant assay-related variations between study populations, which potentially could have been substantial, as was demonstrated in an evaluation of the congruity of seven major assays.[169]

Unlike M4, the diagnostic performance of the NICE algorithm was similar between Paper I and Paper II. This diagnostic performance was highly consistent with the results of a recently published retrospective study from Australia of 724 PULs, although the NICE algorithm was not used in clinical practice.[49] In that study, the EP rate was approximately 10%, and emergency surgery was employed for five (7%) women when expectant or methotrexate treatment failed compared to 6 (5%) of 113 EPs in the present study. In the Australian study, the success rate of single-dose methotrexate (95.8%) or expectant management (84%) was higher compared to our results; EP cases in that study had a lower first hCG average within treatment groups, which could have contributed to this difference.[168] In a Dutch RCT comparing methotrexate to expectant management in EPs (n=15) with hCG <1500 IU/L or persistent PUL (n=58) with plateauing hCG <2000 IU/L, pregnancy resolution was comparable between the groups (76% vs 59%), similar to the current results.[71] In a similar RCT from the UK, the success rates were higher for both single-dose methotrexate (89%) and expectant management (74%) when EPs with hCG <1500 IU/L at the first visit were included.[170] In the current study, women with hCG >1500 but <3000 IU/L could be considered for methotrexate, potentially leading to fewer successful treatments in the M4 and NICE groups. Also, EPs in the RCT from the UK were diagnosed at the first visit, whereas in the current study, potentially different types of EP were included because of the focus on PUL management. Still, the methotrexate success rate in both groups in the present study was within the range of 70%-90% reported in other studies.[171]

In the implementation of M6 in eight EPAUs in the UK, 2625 women with PUL with either a true PUL (50.0%), probable IUP (12.9%), probable miscarriage (31.7%) or probable EP (4.9%) were evaluated.[61] In a first step, PULs with progesterone ≤2 nmol/L were triaged to a follow-up pregnancy test after 2 weeks. The remaining patients were classified using M6, which includes the first progesterone value and two hCG values, as opposed to M6NP, which does not include progesterone for

estimating probabilities. The study population included 320 (12.1%) EPs, of which an unknown number were persistent PULs. In the two steps of the process, 43% of PULs were classified as low risk and 16 (5%) of EPs were misclassified as low risk. The M6P had lower accuracy for low-risk versus high-risk PUL and differs in this respect from the protocols used in the present study. The authors of that study reported an EP rupture frequency of 0.3% (7 of 2625 PULs) compared to 0.7% in the M4 group and 1.4% in the NICE group in the current work. When M4 was implemented in three EPAU in the United Kingdom, 18% (17 of 92) of EPs were misclassified as low risk, comparable to the results in the present study.[13] There were no reports of ruptured EP, but one case of methotrexate was given to a viable IUP, compared to two in our study and none in the M6 study. It is important to be familiar with ultrasonographic criteria to safely rule out a viable IUP and not rely on too slow incremental change in hCG levels.[26, 75]

With little doubt, none of the ruptured EPs can be attributed to the specific protocol in the current study. Only two cases with a ruptured EP were undiagnosed prior to surgery, and in one of these, involving a woman in the NICE group, the case was not managed according to the high-risk prediction associated with it. The woman in the second case also was in the NICE group and did not have TVS at 7 days, as recommended, as her monitoring was prematurely halted. Furthermore, most of the ruptured EPs in the present study were already under close surveillance or had been treated with methotrexate and thus were not apparently associated with a delay in TVS confirmation. Careful selection of women with EPs for expectant management and identification of EPs at increased risk of failing methotrexate are means for limiting rupture rates.[168, 172]

In the M6 implementation study, 124 (38%) EPs were treated with laparoscopy, more than in the M4 group (32%) and fewer than in the NICE group (55%) in the present study. If persistent PULs were to be included among EPs, the proportion having surgical treatment would be lower in the M4 and NICE groups. The higher proportion of surgery in the NICE group likely reflects the numerically higher hCG values for EPs in that group compared to the M4 group. The difference in hCG levels for EPs between groups is likely to have been random, as such differences were not seen for other PUL outcomes. The average hCG levels for EPs treated with laparoscopic surgery as first line were highly similar between the M4 (1929 IU/L) and NICE (1924 IU/L) groups. This similarity suggests that laparoscopy indeed was necessary, but for a higher proportion of EPs in the NICE group, resulting in a similar overall treatment success between groups and no significant difference in ruptured EPs.

A major shift in the treatment of EP emerged between the studies represented in Paper I and Paper III because of a policy change.[51] The rupture rate of EPs was 2.3% in Paper I, but it is difficult to discern whether this higher rate resulted from treatment being initiated too late or because ruptures could be confirmed visually. Fewer EPs were detected with TVS prior to surgery, which could indicate that women were not closely monitored, leading to a postponed diagnosis. The EPs in the previous study also had numerically higher hCG levels, possibly associated with an increased risk for rupture. As EPs were mainly treated surgically without the need for further hCG monitoring, the number of hCG samplings was almost doubled in the present study while the numbers appear to be lower for spontaneously resolving PULs and similar for IUPs.[51]

The reported number of hCG samplings and TVS performed in the M6 implementation study was lower than in the present study except for spontaneously resolving PULs.[61] The two-step approach allowing for a substantial proportion of women to undergo only one blood test and the fact that recording stopped once a diagnosis was made may have contributed to a reduced number of hCG and TVS assessments. Compared to the M4 implementation study, TVS and blood tests were fairly similar for IUPs and spontaneously resolving PULs in this study, but were considerably lower for EPs, probably due to being recorded only prior to diagnosis. The rate of medical intervention for non-EPs was comparable between the present study and the M6 study. The overall surgical intervention rate was lower (7.9%) in the M4 group but higher in the NICE group (14.1%) compared to a recent study reporting a 14.0% surgical intervention rate in a population with an 11.9% EP rate.[19]

There were 13.6% (8 of 59) laparoscopies that were negative in the present study and 15.2% and 7.5% in the M4 and M6 implementation studies respectively. Most laparoscopies were performed due to urgent severe pain or when hCG levels were high and there was an ultrasonographic suspicion of EP. There is no consensus on an acceptable rate of negative laparoscopy which is unavoidable for safe management of PUL. In a study conducted in 44 EPAUs in the UK 18.0% of women undergoing laparoscopy of a suspected EP were not diagnosed with an EP.[37] This was of great concern given the economic costs, but also because a significant proportion of EPs diagnosed with TVS were false negative findings and striving to lower these rates was recommended. The frequency of unplanned visits in the M4 (13.9%) and NICE (9.2%) groups was much higher compared with the M4 (1.8%) and M6 (0.5%) implementation studies.

This RCT did not demonstrate statistically significant differences in any secondary outcomes for M4 vs the NICE algorithm. There are several reasons for this which are also limitations of the study. First, the study was not powered to detect differences in these outcomes, and the premature termination of the study made this limitation more obvious. Interestingly, the analysis according to risk classification showed that EPa classified as high risk were not diagnosed earlier than EPs classified as low risk. However, in the per-protocol population which disregarded the dynamics of patient risk factors for EP, deteriorating symptoms and the clinical judgment of the physician led to a shorter time to diagnosis of EP. Most of the EPs classified as high risk (95%) were manged according to protocol, but this was not the case for low risk EPs (23%). The study also showed that spontaneously resolving PULs and IUPs classified as low risk had a reduced number of TVS and blood tests. Second, even if the projected sample size was achieved and a statistically significant difference in the diagnostic performance was observed, it is unlikely that secondary outcomes would be affected in a significant way. The protocols were diagnostically too similar requiring a much larger sample size for not diluting potential differences of clinically important outcomes. Third, the protocols direct the initial management of PUL in a noninterventional manner and many decision points exist after this point that could influence the outcome. Both groups were followed up according to a standardized set of rules keeping most women within a 1week evaluation timeframe. In any circumstance it would be difficult to determine if clinical outcomes, at least safety outcomes, depended on the application of a protocol per se or a standardized monitoring system. If a protocol triaged a greater proportion of PULs to a 2-week follow-up safety and efficiency outcomes may be significantly affected and measurable. Accepting a 21% hCG level decline as previously proposed or utilize a progesterone cutoff <10 nmol/L for allowing no further clinic visits are protocols that potentially could demonstrate significant clinical differences even in a sample of similar size to the present study.[35, 85, 132] Fourth, the number of EPs between the M4 group and the NICE group was unevenly distributed which could result from the small study sample. This important difference needs to be considered when interpreting the results. Other pregnancy outcomes, baseline characteristics or hCG data did not differ between groups which implies successful randomization. Fifth, the M6NP risk-prediction model was not used in the trial because no external validation study had yet been carried out.[91] The M6 and M6NP have better predictive ability of EP but also a lower predictive ability of foremost IUP than M4.[91, 167, 173] A significant difference in sensitivity for EPs and specificity for non-EPs between the M6NP and the NICE algorithm is likely to be observed if tested in the present study sample. A much larger sample size could potentially also show significant

differences in clinical outcomes. As M6NP are skewed towards sensitivity for EP compared to the NICE algorithm, such a study is desirable from a patient safety perspective. Sixth, this study included women with a true PUL but also women with a probable IUP or a probable EP. The study results may thus not be generalizable to clinical settings where only patients with no evidence of intra-or extrauterine pregnancy are managed as PUL.

This study showed that the NICE algorithm had better specificity for non-EPs and discriminatory ability for EP vs non-EP than M4 in a cross-sectional analysis. No significant difference of diagnostic performance was, however, observed when analyzed as a parallel design, which did not utilize the sample size optimally. Secondary outcomes did not differ significantly between groups but there was a higher number of EP in the NICE group and laparoscopic surgery was performed more frequently compared to the M4 group. Women with a IUP and a spontaneously resolving PULs seem to benefit from being correctly classified as low risk, while EPs classified as high risk did not ensure a faster diagnosis compared to EPs classified as low risk. A limitation of the study was the small sample size. Interpretation of subgroup findings and secondary outcomes necessitates caution since the study was not powered for these analyses. There is a need for large well-designed studies to compare protocols with substantial dissimilarities, such as single-visit approaches using progesterone versus serial hCG levels. The development of a core outcome set for PUL could guide future research while the development of a risk-prediction model that use the type of PUL as a predictor variable may also be a fruitful future effort.

Paper IV

In this study, the prevalence and magnitude of anxiety and depression were profound in women with PUL compared to women in antenatal care. Levels of anxiety were more pronounced than depression at both assessment time points for women with a nonviable pregnancy and for women with a viable pregnancy within the PUL group. This pattern also has been reported in miscarriage studies and for women in early pregnancy in general.[139, 174, 175] This study demonstrated no significant differences within the PUL group, but women with a viable pregnancy did not show the overall decrease in psychological distress over the course of 4 weeks that women with a nonviable pregnancy had. Nevertheless, almost 50% of women with a nonviable pregnancy had symptoms of anxiety at 4 weeks. These rates are like those reported for women at 1 month after a verified pregnancy loss.[127]

At 4 weeks, the prevalences of anxiety and depression were 30% and 23%, respectively, for women with a viable pregnancy. The whole PUL group had poor mental health quality of life, as made evident by low scores on the MCS of the SF-36. Women with a viable pregnancy also scored lower than the norm on the PCS. Depressive symptoms are independently associated with low health-related quality of life in pregnant women, as well as with anxiety in the general population.[123, 176] In women with a viable pregnancy, anxiety and depression are prevalent, but common pregnancy symptoms such as fatigue and nausea also could contribute to impaired life quality.[119]

To date, this is the only study to investigate the psychological impact of experiencing a PUL. Psychological distress related to a PUL may affect women in the same way that miscarriage does. The populations experiencing PULs and miscarriages are of comparable size in some clinical settings.[37] This similarity should raise concerns given the gap between the need for well-organized psychological support that this study highlights and the reported perceived lack of this support among women with PUL.[128] Negative experiences with health personnel perceived as being unsupportive and the shortage of structured emotional support in hospital settings can exacerbate psychological distress for couples dealing with miscarriage.[166] In the present study it is reasonable to assume that patient satisfaction with care could influence psychological well-being in both positive and negative directions.

The combined findings of this study moreover suggest that women with a viable pregnancy that is first identified as PUL are a particularly vulnerable group. They had numerically higher anxiety and depression scores than women in the antenatal group and a significantly higher mean depression score. There was neither any measurable significant differences in anxiety or depression between women with a viable pregnancy and women with a nonviable pregnancy within the PUL group. This lack of difference could be associated with the general effects of the PUL experience or with the high degree of uncertainty associated with early pregnancy even after viability has been verified. This group of women is an important one to recognize after transfer to antenatal care as needing distinct attention regarding their mental health. Symptoms of anxiety and depression during pregnancy may be associated not only with increased healthcare consumption but also with adverse perinatal outcomes such as preterm birth and low birthweight, with short- and long-term consequences for infant health.[177-180] The National Board of Health and Welfare in Sweden has published evidence-based methods for midwifery to identify women in early pregnancy with possible anxiety and depression, and screening is

recommended by the American College of Obstetricians and Gynecologists and NICE.[181, 182 183] A study from the UK, however, showed that 90% of women undergoing clinical evaluation for pain and/or bleeding in early pregnancy had no formal subsequent referral, even though one-third had symptoms of significant distress.[184]

There are several limitations to consider in this study. First, the anxiety prevalence was higher than estimated for the sample size calculation, both for women in antenatal care and among those with PULs. The psychological impact in the immediate aftermath of experiencing a PUL was higher than for women with miscarriage after 1 month, which was used as reference.[185] In the antenatal group, no viability scan was performed, possibly leading to a high degree of uncertainty regarding the health of a pregnancy. The point prevalence of anxiety in our antenatal group (29.2%) could be compared with that of the Swedish general population (14.7%) and that for 2167 women in early pregnancy in a Dutch study (17.9%).[149, 186] Second, the HADS does not account for the type of anxiety being measured. Differentiating among trait, state, and pregnancy-specific anxiety (for women with a viable pregnancy) may have been optimal, especially given the small numbers in the groups.[125] The results of the subgroup analysis, also with small numbers of participants, need cautious interpretation since the study was not powered to detect differences between these groups. Third, recruitment and response rates were low, and how representative the study population is of the broader population is unclear. Participant inclusion was strictly based on entry criteria, which were broad, with a limited number of exclusion criteria to ensure generalizability; nevertheless, volunteer bias must be considered. Fourth, health-related quality of life was not evaluated in the antenatal group, which would have enabled comparison with the PUL group. The mental component of the SF-36 was below the norm for both subgroups of PUL. The mental component of quality of life has been described as stable throughout pregnancy, while the physical component decreases.[187] A recent study found both components to be lower in women in the first trimester than for a nonpregnant control group,[188] which calls for careful interpretation of the SF-36 results for women with a viable pregnancy within the PUL group.

This study showed that self-reported psychological distress is highly prevalent among women with PUL, and that health-related mental quality of life is low. Both at the early stage of PUL and after 4 weeks, anxiety and depression were as common as has been reported for women experiencing miscarriage. We observed no difference in psychological well-being between women with a viable pregnancy and

those with a nonviable pregnancy starting as a PUL. Our study highlights the need for robust psychological support for women with PUL and the need for more research on this topic. Given the large proportion of women with a viable pregnancy within the PUL group and the possibility that they may suffer more psychological distress than other women in antenatal care, this is a special area of interest. Studies with a larger number of participants and a longer follow-up period are required to gain a full understanding of all aspects of mental health in this patient population.

Summary, conclusions, and future perspective

Papers I-III

- The NICE algorithm showed consistent diagnostic performance across two study populations (first the retrospective cohort in Paper I and then the second prospective cohort in Paper III) with high discriminatory ability for EPs vs non-EPs which was better than for M4 when analyzed prospectively on cross-sectional data. The M4 sensitivity for EPs was worse in the second study population potentially because of EPs with lower initial hCG levels. This suggest that the M4 diagnostic performance was affected unlike the NICE algorithm which only depends on hCG changes for prediction. As evident from other research the underlying study populations is a determinant of the performance of a predictions model which is why external validation studies are important before considering implementation in new clinical settings.

- The M6NP had higher discriminatory ability, NB, and calibration than the M4. The results from the studies in this thesis are consistent with that reported from EPAUs in the UK both risk-prediction models were developed. M4 was established in a smaller number of participants with more strict inclusion criteria than M6NP. Consequently, M4 may not be as robust as M6NP when applied on a more diverse validation population. A main difference between the models from a clinical perspective is that M6NP had a higher sensitivity for EPs and false-positive rate for the non-EPs compared to M4. This could lead to a shorter time to EP diagnosis, but it also could lead to more women with PUL undergoing close monitoring and potentially unnecessary diagnostic procedures.

- Clinical outcomes for M4 and the NICE algorithm did not differ significantly in the RCT because of being too similar regarding diagnostic performance in addition to a small sample size. Furthermore, the prediction by any protocol is not the only entity driving the initial management of PUL in the pursuit of a diagnosis. The judgment of the physician and patient's symptoms are other integral parts of handling a complex clinical situation. To some extent this offsets the differences between protocols diagnostic performance. This was demonstrated in the RCT (Paper III) by EPs classified as high risk not being diagnosed earlier than low risk classified EPs. Moreover, it cannot be taken for granted that early diagnosis of EPs would reduce complications thereof as

there are many decisions points ahead influencing the outcome, and rupture is an unprecedented event occurring in both low and high risk classified EPs. It is always important to adopt evidence based criteria when selecting an EP for conservative treatments and to uphold close communication with a well-informed patient.

- The clinical outcomes in the RCT were similar to those reported from larger studies from EPAUs in the UK evaluating M4 and M6, but the rate of unplanned visits was much higher and the EP rupture rate somewhat higher although rare in all studies. Given M6NPs safety profile with better sensitivity for EPs it may be preferable to the NICE algorithm, especially in a clinical setting with an elevated proportion of PULs being EPs. Large well-designed studies, ideally comparing the M6/M6NP or a single-visit approach using progesterone, with the NICE algorithm, guided by a core outcome set for PUL are needed. The development of a risk-prediction model that use the type of PUL as a predictor variable would also be of interest given the large proportion of PUL being probable IUPs in the studies of this thesis and as reported in other studies.

- The proportion of EPs among PUL was at high levels in the first and second study population in this thesis, compared with that reported in other studies. A major shift in the treatment of EP took place between study populations, being much more diversified in the second. The rate of TVS diagnosis of EPs before laparoscopic surgery increased and there was a lower rate of confirmed rupture of EPs in the second study population. Further analyses of the data sets are needed to clarify these differences. The development of clinical key performance indicators would be helpful to enhance the quality of PUL practice in general and to improve effectiveness and patients' safety.

Paper IV

- Symptoms of psychological distress measured with the HADS are substantial in women experiencing PUL and health-related mental quality of life is low. Both at the early stage of PUL and after four weeks, anxiety and depression were as common as reported for miscarrying women. This aspect of PUL is novel findings and more research on this topic is of importance. The study emphasizes a need for a structured psychological support adapted for women with PUL which previously has been described to be lacking.

- Moreover, the study demonstrated no significant difference in anxiety and depression between women with a viable pregnancy and those with a nonviable pregnancy starting as a PUL at 1 or 4 weeks. The worries for the health of the pregnancy and personal health may contribute to psychological distress at the early stage for all women with PUL. After a viable pregnancy has been confirmed for some of the women, the high degree of uncertainty associated with early pregnancy and the experience of PUL may continue to cause psychological distress. Since this is subgroup results for which the study was not powered careful interpretation is necessary. Given the large proportion of women with a viable pregnancy within the PUL group and the possibility that they may suffer more than other women in antenatal care larger studies are necessary to gain a full understanding of psychological distress in this group of patients.

Acknowledgements

This thesis have been realized with the help and guidance of several individuals, who contributed in different ways, not least all the women participating in the studies. It is my pleasure to thank those who made it possible.

First and foremost, I´m deeply grateful to my main supervisor, Assoc. Prof. Annika Strandell, for her efforts to initiate my PhD candidature, and her patience and steadfast support to finalize it. Your brilliant scientific knowledge has been invaluable for the thesis, but also an inspiration that has shaped my interest in research and encouraged future endeavors. Secondly, I would like to express my gratitude to my co-supervisor Prof. Christina Bergh, for her guidance and constructive feedback. Your meticulous, insightful, and systematic inputs have been vital for this thesis and truly an aspirational motivation to me. Thirdly, I wish to thank the co-author, Assoc. Prof. Stina Järvholm for valuable, interesting, and enjoyable discussions leading to the formation of Paper IV, and co-authors Helen Garbergs and Katja Bergman for your support as principal site investigators of the RCT (Paper III).

My gratitude extends to the Sahlgrenska University Hospital for the funding opportunity to undertake my studies at the Department of Obstetrics and Gynecology, and to Livio where I currently resides. Additionally, I would like to thank former and present managers for enabling my research at Sahlgrenska, Lotta Wassén, Corinne Pedroletti, and Ali Khatibi, and at Livio, Göran Westlander, AnnaKarin Lind and Ali Khatibi. I´m also grateful to the Institute of Clinical Sciences at the University of Gothenburg and Kliniska forskarskolan for providing the foundation of my education, and in-depth teaching about clinical research.

I´m grateful of the continual help and personal support from my former colleagues at the gynecological emergency unit at Sahlgrenska University Hospital. I wish to acknowledge Kajsa Sandberg Kedfors and Malin Ströby for the responsibility of recruiting women in the antenatal care units, and Lilian Engström, Susanna Hedelin and Iris Wikström who participated in data collection and site monitoring in the RCT. I would also like to thank Annika Söderström and Sara Lindström for the relentless struggling with fitting my schedule into the daily work at Livio, and to all my co-workers over there.

I want to thank Annette Nattland for making the well-designed layout of this thesis. Without your help the compilation of the thesis would have been even more hectic.

I also wish to thank Mattias Molin, Nils-Gunnar Pehrsson, Henrik Imberg, Henrik Albrektsson, Hussein Hamoodi and Christopher Backström at Statistiska Konsultgruppen, Gothenburg, for valuable statistical support. The construction of the randomization software and its maintenance was crucial to the RCT, for which I´m grateful to Artemis Mårtensson.

I´m also thankful for the generous financial assistance from the Healthcare Board, Region Västra Götaland, the Hjalmar Svensson's Research Foundation and the Swedish state under the agreement between the Swedish government and the county councils.

Finally, without the support of my family, achieving a PhD would not have seemed possible. Despite the challenges the past years, my sweet wife Lisa has not only coped, but also shown

understanding and patience, particularly the last few months during the final intense preparation of my thesis. My beloved children, Wilhelm, Mathilda, and Greta have made this journey more bearable and worthwhile, simply by being there. I dedicate a heartfelt appreciation to my dear parents, Linda and Michalis, and my brother Nicola. Without you, this accomplishment would not have been attainable.

References

1. Harville EW. Vaginal bleeding in very early pregnancy. Human Reproduction. 2003;18(9):1944-7.
2. DeVilbiss EA, Naimi AI, Mumford SL, Perkins NJ, Sjaarda LA, Zolton JR, et al. Vaginal bleeding and nausea in early pregnancy as predictors of clinical pregnancy loss. Am J Obstet Gynecol. 2020;223(4):570 e1- e14.
3. Hasan R, Baird D, Herring A, Olshan A, Funk Jonsson M, Hartmann K. Association between first-trimester vaginal bleeding and miscarriage. Obstet Gynecol. 2009;114(4):860-7.
4. Kurinczuk JJ, Draper ES, Field DJ, Bevan C, Brocklehurst P, Gray R, et al. Experiences with maternal and perinatal death reviews in the UK--the MBRRACE-UK programme. BJOG. 2014;121 Suppl 4:41-6.
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(4):231-6.
6. Kamwendo F, Forslin L, Bodin L, Danielsson D. Epidemiology of ectopic pregnancy during a 28 year period and the role of pelvic inflammatory disease. Sex Transm Infect. 2000 76(1):28-32.
7. Raine-Bennett T, Fassett MJ, Chandra M, Armstrong MA, Xie F, Shi JM, et al. Disparities in the Incidence of Ectopic Pregnancy in a Large Health Care System in California, 2010-2019. Perm J. 2022;1426(3):61-8.
8. Hoover KW, Tao G, Kent CK. Trends in the diagnosis and treatment of ectopic pregnancy in the United States. Obstet Gynecol. 2010;115(3).
9. Marion LL, Meeks GR. Ectopic pregnancy: History, incidence, epidemiology, and risk factors. Clin Obstet Gynecol. 2012 55(2):376-86.
10. Qin G, Zhang S, Liu J, Yang L, Li H, Tang J, et al. Global burden and trends of ectopic pregnancy: An observational trend study from 1990 to 2019. Plos One. 2023;18(10).
11. MBRRACE-UK: Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK.
12. Crochet J, Bastian L, Chireau M. Does this woman have an ectopic pregnancy?: the rational clinical examination systematic review. JAMA. 2013;309(16):1722-29.
13. Bobdiwala S, Guha S, Van Calster B, Ayim F, Mitchell-Jones N, Al-Memar M, et al. 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(7):1425-35.
14. MM-ARG. Kvinnoläkaren 2021(5):8. Available at: https://www.sfog.se/start/kvinnolaekaren/kvinnolaekaren/.
15. Seeber BE, Sammel MD, Guo W, Zhou L, Hummel A, Barnhart KT. Application of redefined human chorionic gonadotropin curves for the diagnosis of women at risk for ectopic pregnancy. Fertil Steril. 2006;86(2):454-9.
16. 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(6):1744-50.
17. Barnhart K, van Mello NM, Bourne T, Kirk E, Van Calster B, Bottomley C, et al. Pregnancy of unknown location: a consensus statement of nomenclature, definitions, and outcome. Fertil Steril. 2011;95(3):857-66.
18. Expert Panel on Women's I, Brown DL, Packard A, Maturen KE, Deshmukh SP, Dudiak KM, et al. ACR Appropriateness Criteria((R)) First Trimester Vaginal Bleeding. J Am Coll Radiol. 2018;15(5S):S69-S77.
19. Dooley WM, de Braud LV, Wong M, Platts S, Ross JA, Jurkovic D. Development of a single-visit protocol for the management of pregnancy of unknown location following in vitro fertilization: a retrospective study. Hum Reprod. 2024;39(3):509-15.
20. Benson CB, Doubilet PM, Peters HE, Frates MC. Intrauterine fluid with ectopic pregnancy: a reappraisal. J Ultrasound Med. 2013;32(3):389-93.

21. Nyberg DA, Laing FC, Filly RA, Uri-Simmons M, Jeffrey RB, Jr. Ultrasonographic differentiation of the gestational sac of early intrauterine pregnancy from the pseudogestational sac of ectopic pregnancy. Radiology. 1983;146(3):755-9.
22. Phillips CH, Benson CB, Durfee SM, Heller HT, Doubilet PM. "Pseudogestational Sac" and Other 1980s-Era Concepts in Early First-Trimester Ultrasound: Are They Still Relevant Today? J Ultrasound Med. 2020;39(8):1547-51.
23. Lee IT, Rubin ES, Wu J, Koelper N, Barnhart KT. The incidence and importance of the pseudogestational sac revisited. Am J Obstet Gynecol. 2022;226(4):537 e1- e7.
24. Kirk E, Ankum P, Jakab A, Le Clef N, Ludwin A, Small R, et al. Terminology for describing normally sited and ectopic pregnancies on ultrasound: ESHRE recommendations for good practice. Hum Reprod Open. 2020(4):hoaa055.
25. Doubilet PM, Benson CB, Bourne T, Blaivas M. Diagnostic criteria for nonviable pregnancy early in the first trimester. N Engl J Med. 2013;369(15):1443-51.
26. Preisler J, Kopeika J, Ismail L, Vathanan V, Farren J, Abdallah Y, et al. Defining safe criteria to diagnose miscarriage: prospective observational multicentre study. BMJ. 2015;351:h4579.
27. Tsakiridis I, Giouleka S, Mamopoulos A, Athanasiadis A, Dagklis T. Diagnosis and Management of Ectopic Pregnancy: A Comparative Review of Major National Guidelines. Obstet Gynecol Surv. 2020;75(10):611-23.
28. Valentin L. High-quality gynecological ultrasound can be highly beneficial, but poor-quality gynecological ultrasound can do harm. Ultrasound Obstet Gynecol. 1999;13(1):1-7.
29. Pinto A, Pinto F, Faggian A, Rubini G, Caranci G, Macarini L, et al. Sources of error in emergency ultrasonography. Crit Ultrasound J. 2013;5 Suppl 1(Suppl 1):S1.
30. Hahlin M, Thorburn J, Bryman I. The expectant management of early pregnancies of uncertain site. Hum Reprod. 1995;10(5):1223-7.
31. Mol B, Hajenius P, Engelsbel S, Ankum W, Van der Veen F, Hemrika D, et al. Serum human chorionic gonadotropin measurement in the diagnosis of ectopic pregnancy when transvaginal sonography is inconclusive. Fertil Steril. 1998;70(5):972-81.
32. Barnhart KT, Fay CA, Suescum M, Sammel MD, Appleby D, Shaunik A, et al. Clinical factors affecting the accuracy of ultrasonography in symptomatic first-trimester pregnancy. Obstet Gynecol. 2011;117(2 Pt 1):299-306.
33. Condous G, Timmerman D, Goldstein S, Valentin L, Jurkovic D, Bourne T. Pregnancies of unknown location: consensus statement. Ultrasound Obstet Gynecol. 2006;28(2):121-2.
34. Condous G, Kirk E, Lu C, Van Huffel S, Gevaert O, De Moor B, et al. Diagnostic accuracy of varying discriminatory zones for the prediction of ectopic pregnancy in women with a pregnancy of unknown location. Ultrasound Obstet Gynecol. 2005;26(7):770-5.
35. 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(6):693-7.
36. Ooi S, De Vries B, Ludlow J. How do the M4 and M6 models perform in an Australian pregnancy of unknown location population? Aust N Z J Obstet Gynaecol. 2021;61(1):100-5.
37. Memtsa M, Goodhart V, Ambler G, Brocklehurst P, Keeney E, Silverio S, et al. Variations in the organisation of and outcomes from Early Pregnancy Assessment Units: the VESPA mixed-methods study. Health Serv Deliv Res. Health Services and Delivery Research. 8. Southampton (UK)2020. p. 1-138.
38. Wilcox AJ, Baird D, Weinberg CR. Time of implantation of the conceptus and loss of pregnancy. N Engl J Med. 1999;340(23):796-9.
39. Bottomley C, Van Belle V, Mukri F, Kirk E, Van Huffel S, Timmerman D, et al. The optimal timing of an ultrasound scan to assess the location and viability of an early pregnancy. Hum Reprod. 2009;24(8):1811-7.
40. Condous G, Okaro E, Khalid A, Bourne T. Do we need to follow up complete miscarriages with serum human chorionic gonadotrophin levels? BJOG. 2005;112(6):827-9.
41. NICE guideline [NG126]. 2019. https://wwwniceorguk/guidance/ng126 (Accessed November 14, 2024).
42. Fossum GT, Davajan V, Kletzky OA. Early detection of pregnancy with transvaginal ultrasound. Fertil Steril. 1988;49(5):788-91.

43. Doubilet PM, Benson CB. Double sac sign and intradecidual sign in early pregnancy: interobserver reliability and frequency of occurrence. J Ultrasound Med. 2013;32(7):1207-14.
44. Richardson A, Gallos I, Dobson S, Campbell BK, Coomarasamy A, Raine-Fenning N. Accuracy of first-trimester ultrasound in diagnosis of intrauterine pregnancy prior to visualization of the yolk sac: a systematic review and meta-analysis. Ultrasound Obstet Gynecol. 2015;46(2):142-9.
45. Kirk E, Papageorghiou AT, Condous G, Tan L, Bora S, Bourne T. The diagnostic effectiveness of an initial transvaginal scan in detecting ectopic pregnancy. Hum Reprod. 2007;22(11):2824-8.
46. Nadim B, Infante F, Lu C, Sathasivam N, Condous G. Morphological ultrasound types known as 'blob' and 'bagel' signs should be reclassified from suggesting probable to indicating definite tubal ectopic pregnancy. Ultrasound Obstet Gynecol. 2018;51(4):543-9.
47. Dooley WM, Chaggar P, De Braud LV, Bottomley C, Jauniaux E, Jurkovic D. Effect of morphological type of extrauterine ectopic pregnancy on accuracy of preoperative ultrasound diagnosis. Ultrasound Obstet Gynecol. 2019;54(4):538-44.
48. Bignardi T, Condous G, Kirk E, Van Calster B, Van Huffel S, Timmerman D, et al. Viability of intrauterine pregnancy in women with pregnancy of unknown location: prediction using human chorionic gonadotropin ratio vs. progesterone. Ultrasound Obstet Gynecol. 2010;35(6):656-61.
49. Jin CS, Uzuner C, Condous G. Safety of methotrexate administration in women with pregnancy of unknown location at high risk of ectopic pregnancy. Ultrasound Obstet Gynecol. 2024;64(1):97-103.
50. Bignardi T, Condous G, Alhamdan D, Kirk E, Van Calster B, Van Huffel S, et al. The hCG ratio can predict the ultimate viability of the intrauterine pregnancies of uncertain viability in the pregnancy of unknown location population. Hum Reprod. 2008;23(9):1964-7.
51. Fistouris J, Bergh C, Strandell A. Classification of pregnancies of unknown location according to four different hCG-based protocols. Hum Reprod. 2016;31(10):2203-11.
52. 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;14:231-6.
53. Nadim B, Leonardi M, Stamatopoulos N, Reid S, Condous G. External validation of risk prediction model M4 in an Australian population: Rationalising the management of pregnancies of unknown location. Aust N Z J Obstet Gynaecol. 2020.
54. Van Calster B, Abdallah Y, Guha S, Kirk E, Van Hoorde K, Condous G, et al. Rationalizing the management of pregnancies of unknown location: temporal and external validation of a risk prediction model on 1962 pregnancies. Hum Reprod. 2013;28(3):609-16.
55. Kirk E, Condous G, Haider Z, Lu C, Van Huffel S, Timmerman D, et al. The practical application of a mathematical model to predict the outcome of pregnancies of unknown location. Ultrasound Obstet Gynecol. 2006;27(3):311-5.
56. Banerjee S, Aslam N, Woelfer B, Lawrence A, Elson J, Jurkovic D. Expectant management of early pregnancies of unknown location: a prospective evaluation of methods to predict spontaneous resolution of pregnancy. BJOG. 2001;108:158-63.
57. Kirk E, Daemen A, Papageorghiou AT, Bottomley C, Condous G, De Moor B, et al. Why are some ectopic pregnancies characterized as pregnancies of unknown location at the initial transvaginal ultrasound examination? Acta Obstetricia et Gynecologica Scandinavica. 2010;87(11):1150-4.
58. Lattouf I, Lu C, Pixton S, Reid S, Condous G. Is there a difference in the behaviour and subsequent management of ectopic pregnancies seen at first scan compared to those ectopic pregnancies which commence as pregnancies of unknown location? Aust N Z J Obstet Gynaecol. 2016;56(1):107-12.
59. 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(3):351-7.
60. Barnhart KT, Hansen KR, Stephenson MD, Usadi R, Steiner AZ, Cedars MI, et al. Effect of an Active vs Expectant Management Strategy on Successful Resolution of Pregnancy Among Patients With a Persisting Pregnancy of Unknown Location: The ACT or NOT Randomized Clinical Trial. JAMA. 2021;326(5):390-400.

61. Bobdiwala S, Christodoulou E, Farren J, Mitchell-Jones N, Kyriacou C, Al-Memar M, et al. Triaging women with pregnancy of unknown location using two-step protocol including M6 model: clinical implementation study. Ultrasound Obstet Gynecol. 2020;55(1):105-14.
62. Kolte AM, Bernardi LA, Christiansen OB, Quenby S, Farquharson RG, Goddijn M, et al. Terminology for pregnancy loss prior to viability: a consensus statement from the ESHRE early pregnancy special interest group. Hum Reprod. 2015;30(3):495-8.
63. Barnhart K, Menutti M, Benjamin I, Jacobson S, Goodman D, Coutifaris C. Prompt diagnosis of ectopic pregnancy in an emergency department setting. Obstet Gynecol. 1994;84(6):1010-5.
64. Hahlin M, J. T, I. B. The expectant management of early pregnancies of uncertain site. Hum Reprod. 1995;10(5):1223–7.
65. Ankum W, Van der Veen F, Hamerlynck J, Lammes F. Laparoscopy: a dispensable tool in the diagnosis of ectopic pregnancy? Hum Reprod. 1993;8(8):1301-6.
66. Thorburn J, Bryman I, Hahlin M. Distinction between early normal intrauterine pregnancies and pathological pregnancies by means of a logistic model. Hum Reprod. 1992;7(1):120-2.
67. Morse CB, Sammel MD, Shaunik A, Allen-Taylor L, Oberfoell NL, Takacs P, et al. Performance of human chorionic gonadotropin curves in women at risk for ectopic pregnancy: exceptions to the rules. Fertil Steril. 2012;97(1):101-6 e2.
68. Kadar N. Early recourse to laparoscopy in the management of suspected ectopic pregnancy. Accuracy and morbidity. Journal of reproductive medicine, . 1990;35(12):153-6.
69. Li T, Tristram A, Hill A, Cooke D. A review of 254 ectopic pregnancies in a teaching hospital in the Trent Region, 1977-1990. Hum Reprod. 1991;6(7):1002-7.
70. Wilasrusmee C, Sukrat B, McEvoy M, Attia J, Thakkinstian A. Systematic review and meta-analysis of safety of laparoscopic versus open appendicectomy for suspected appendicitis in pregnancy. Br J Surg. 2012;99(11):1470-8.
71. van Mello NM, Mol F, Verhoeve HR, van Wely M, Adriaanse AH, Boss EA, et al. Methotrexate or expectant management in women with an ectopic pregnancy or pregnancy of unknown location and low serum hCG concentrations? A randomized comparison. Hum Reprod. 2013;28(1):60-7.
72. Barnhart K, Franasiak J. ACOG Practice Bulletin No. 191: Tubal Ectopic Pregnancy. Obstetrics and Gynecology. 2018;131(191):409-11.
73. Elson C, Salim R, Potdar N, Chetty M, Ross J, Kirk E, et al. Diagnosis and Management of Ectopic Pregnancy: Green-top Guideline No. 21. BJOG. 2016;123(13):e15-e55.
74. Condous G, Kirk E, Lu C, Van Calster B, Van Huffel S, Timmerman D, et al. There is no role for uterine curettage in the contemporary diagnostic workup of women with a pregnancy of unknown location. Hum Reprod. 2006;21(10):2706-10.
75. Bobdiwala S, Kyriacou C, Christodoulou E, Farren J, Mitchell-Jones N, Al-Memar M, et al. Evaluating cut-off levels for progesterone, beta human chorionic gonadotropin and beta human chorionic gonadotropin ratio to exclude pregnancy viability in women with a pregnancy of unknown location: A prospective multicenter cohort study. Acta Obstet Gynecol Scand. 2022;101(1):46-55.
76. Nurmohamed L, Moretti ME, Schechter T, Einarson A, Johnson D, Lavigne SV, et al. Outcome following high-dose methotrexate in pregnancies misdiagnosed as ectopic. American Journal of Obstetrics and Gynecology. 2011;205(6):533.e1-.e3.
77. Huchon C, Deffieux X, Beucher G, Capmas P, Carcopino X, Costedoat-Chalumeau N, et al. Pregnancy loss: French clinical practice guidelines. Eur J Obstet Gynecol Reprod Biol. 2016;201:18-26.
78. Barnhart K, Franasiak J. Tubal ectopic pregnancy. ACOG Practice Bulletin No. 193. American College of Obstetricians and Gynecologists. Obstet Gynecol. 2018;131(3):e91-e103.
79. Romero R, Kadar N, Jeanty P, Copel JA, Chervenak FA, DeCherney A, et al. Diagnosis of ectopic pregnancy: value of the discriminatory human chorionic gonadotropin zone. Obstet Gynecol. 1985;66(3):357-60.
80. Kadar N, Caldwell BV, Romera R. A method of screening for ectopic pregnancy and its indications. Obstet Gynecol. 1981;58(2):162-6.

81. Barnhart KT, Sammel MD, Rinaudo PF, Zhou L, Hummel AC, Guo W. Symptomatic patients with an early viable intrauterine pregnancy: HCG curves redefined. Obstet Gynecol. 2004;104(1):50-5.
82. Silva C, Sammel MD, Zhou L, Gracia C, Hummel AC, Barnhart K. Human chorionic gonadotropin profile for women with ectopic pregnancy. Obstet Gynecol. 2006;107(3):605-10.
83. Butts SF, Guo W, Cary MS, Chung K, Takacs P, Sammel MD, et al. Predicting the decline in human chorionic gonadotropin in a resolving pregnancy of unknown location. Obstet Gynecol. 2013;122(2 Pt 1):337-43.
84. van Mello NM, Mol F, Opmeer BC, Ankum WM, Barnhart K, Coomarasamy A, et al. Diagnostic value of serum hCG on the outcome of pregnancy of unknown location: a systematic review and meta-analysis. Hum Reprod Update. 2012;18(6):603-17.
85. Condous G, Van Calster B, Kirk E, Timmerman D, Van Huffel S, Bourne T. Prospective cross-validation of three methods of predicting failing pregnancies of unknown location. Hum Reprod. 2007;22(4):1156-60.
86. Bobdiwala S, Saso S, Verbakel JY, Al-Memar M, Van Calster B, Timmerman D, et al. Diagnostic protocols for the management of pregnancy of unknown location: a systematic review and meta-analysis. BJOG. 2019;126(2):190-8.
87. Condous G, Van Calster B, Kirk E, Haider Z, Timmerman D, Van Huffel S, et al. Prediction of ectopic pregnancy in women with a pregnancy of unknown location. Ultrasound Obstet Gynecol. 2007;29(6):680-7.
88. Nadim B, Leonardi M, Infante F, Lattouf I, Reid S, Condous G. Rationalizing the management of pregnancies of unknown location: Diagnostic accuracy of human chorionic gonadotropin ratio-based decision tree compared with the risk prediction model M4. Acta Obstet Gynecol Scand. 2020;99(3):381-90.
89. Barnhart KT, Sammel MD, Appleby D, Rausch M, Molinaro T, Van Calster B, et al. Does a prediction model for pregnancy of unknown location developed in the UK validate on a US population? Hum Reprod. 2010;25(10):2434-40.
90. Valdera Simbron CJ, Hernandez Rodriguez C, Llanos Jimenez L, Perez Garcia L, Plaza Arranz J, Albi Gonzalez M. Management of early gestations with low beta-human chorionic gonadotropin conceived by assisted reproductive technologies: performance of M4 predictive model. Ultrasound Obstet Gynecol. 2021;58(4):616-24.
91. Van Calster B, Bobdiwala S, Guha S, Van Hoorde K, Al-Memar M, Harvey R, et al. 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(5):642-9.
92. Kyriacou C, Ledger A, Bobdiwala S, Ayim F, Kirk E, Abughazza O, et al. Updating M6 pregnancy of unknown location risk-prediction model including evaluation of clinical factors. Ultrasound Obstet Gynecol. 2024;63(3):408-18.
93. Mol B, Lijmer J, Ankum W, van der Veen F, Bossuyt P. The accuracy of single serum progesterone measurement in the diagnosis of ectopic pregnancy: a meta-analysis. Hum Reprod. 1998;13(11):3220-7.
94. Csapo AI, Pulkkinen M. Indispensability of the human corpus luteum in the maintenance of early pregnancy. Luteectomy evidence. Obstet Gynecol Surv. 1978;33(2):69-81.
95. Kratzer PG, Taylor RN. Corpus luteum function in early pregnancies is primarily determined by the rate of change of human chorionic gonadotropin levels. Am J Obstet Gynecol. 1990;163(5 Pt 1):1497-502.
96. Verhaegen J, Gallos ID, van Mello NM, Abdel-Aziz M, Takwoingi Y, Harb H, et al. Accuracy of single progesterone test to predict early pregnancy outcome in women with pain or bleeding: meta-analysis of cohort studies. BMJ. 2012;345:e6077.
97. Guha S, Ayim F, Ludlow J, Sayasneh A, Condous G, Kirk E, et al. Triaging pregnancies of unknown location: the performance of protocols based on single serum progesterone or repeated serum hCG levels. Hum Reprod. 2014;29(5):938-45.
98. Cartwright J, Duncan WC, Critchley HO, Horne AW. Serum biomarkers of tubal ectopic pregnancy: current candidates and future possibilities. Reproduction. 2009;138(1):9-22.

99. Senapati S, Sammel MD, Butts SF, Takacs P, Chung K, Barnhart KT. Predicting first trimester pregnancy outcome: derivation of a multiple marker test. Fertil Steril. 2016;106(7):1725-32 e3.
100. Florio P, Severi FM, Bocchi C, Luisi S, Mazzini M, Danero S, et al. Single serum activin a testing to predict ectopic pregnancy. J Clin Endocrinol Metab. 2007;92(5):1748-53.
101. Kirk E, Papageorghiou AT, Van Calster B, Condous G, Cowans N, Van Huffel S, et al. The use of serum inhibin A and activin A levels in predicting the outcome of 'pregnancies of unknown location'. Hum Reprod. 2009;24(10):2451-6.
102. Bollig KJ, Senapati S, Sammel MD, Takacs P, Robins JC, Haisenleder DJ, et al. Validation of a multiple marker test for early pregnancy outcome prediction. J Assist Reprod Genet. 2023;40(4):837-44.
103. Chetty M, Sawyer E, Dew T, Chapman AJ, Elson J. The use of novel biochemical markers in predicting spontaneously resolving 'pregnancies of unknown location'. Hum Reprod. 2011;26(6):1318-23.
104. Ma D, Yang R, Chen Y, Huang Z, Shen Y, He C, et al. Identification of noninvasive diagnostic biomarkers for ectopic pregnancy using data-independent acquisition (DIA)proteomics: a pilot study. Sci Rep. 2022;12(1):19992.
105. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ. 2024;384:e074819.
106. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol. 2015;68(1):25-34.
107. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-38.
108. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-8.
109. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.
110. Sachs MC, Sjolander A, Gabriel EE. Aim for Clinical Utility, Not Just Predictive Accuracy. Epidemiology. 2020;31(3):359-64.
111. McCarthy CM, O'Leary A, Meaney S, Russell NE. Follow-up and outcomes of patients with a pregnancy of unknown location: A comparison of two prediction models. Eur J Obstet Gynecol Reprod Biol. 2023;287:126-9.
112. Maheut C, Panjo H, Capmas P. Diagnostic accuracy validation study of the M6 model without initial serum progesterone (M6(NP)) in triage of pregnancy of unknown location. Eur J Obstet Gynecol Reprod Biol. 2024;296:360-5.
113. Reilly B, Evans A. Translating Clinical Research into Clinical Practice Impact of Using Impact of Using Prediction Rules To Make Decisions. Ann Intern Med. 2006;44:201-9.
114. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338:b606.
115. Neugebauer R, Kline J, O'Connor P, Shrout P, Johnson J, Skodol A, et al. Depressive symptoms in women in the six months after miscarriage. Am J Obstet Gynecol. 1992;166(1 Pt 1):104-9.
116. Janssen HJ, Cuisinier MC, Hoogduin KA, de Graauw KP. Controlled prospective study on the mental health of women following pregnancy loss. Am J Psychiatry. 1996;153(2):226-30.
117. Bergner A, Beyer R, Klapp BF, Rauchfuss M. Pregnancy after early pregnancy loss: a prospective study of anxiety, depressive symptomatology and coping. J Psychosom Obstet Gynaecol. 2008;29(2):105-13.
118. Koert E, Hartwig TS, Hviid Malling GM, Schmidt L, Nielsen HS. 'You're never pregnant in the same way again': prior early pregnancy loss influences need for health care and support in subsequent pregnancy. Hum Reprod Open. 2023;2023(3):hoad032.
119. Quenby S, Gallos ID, Dhillon-Smith RK, Podesek M, Stephenson MD, Fisher J, et al. Miscarriage matters: the epidemiological, physical, psychological, and economic costs of early pregnancy loss. Lancet. 2021;397(10285):1658-67.

120. Nikcevic AV, Tunkel SA, Nicolaides KH. Psychological outcomes following missed abortions and provision of follow-up care. Ultrasound Obstet Gynecol. 1998;11(2):123-8.
121. Grupe DW, Nitschke JB. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. Nat Rev Neurosci. 2013;14(7):488-501.
122. Richardson A, Raine-Fenning N, Deb S, Campbell B, Vedhara K. Anxiety associated with diagnostic uncertainty in early pregnancy. Ultrasound Obstet Gynecol. 2017;50(2):247-54.
123. Nicholson W, Setse R, Hill-Briggs F, Strobino D, Powe N. Depressive Symptoms and Health-Related Quality of Life in Early Pregnancy. Obstet Gynecol. 2006;107(4):798-806.
124. Lobel M, Cannella DL, Graham JE, DeVincent C, Schneider J, Meyer BA. Pregnancy-specific stress, prenatal health behaviors, and birth outcomes. Health Psychol. 2008;27(5):604-15.
125. Dunkel Schetter C, Rahal D, Ponting C, Julian M, Ramos I, Hobel CJ, et al. Anxiety in pregnancy and length of gestation: Findings from the healthy babies before birth study. Health Psychol. 2022;41(12):894-903.
126. van Mello NM, Mol F, Hajenius PJ, Ankum WM, Mol BW, van der Veen F, et al. Randomized comparison of health-related quality of life in women with ectopic pregnancy or pregnancy of unknown location treated with systemic methotrexate or expectant management. Eur J Obstet Gynecol Reprod Biol. 2015;192:1-5.
127. Farren J, Jalmbrant M, Falconieri N, Mitchell-Jones N, Bobdiwala S, Al-Memar M, et al. Posttraumatic stress, anxiety and depression following miscarriage and ectopic pregnancy: a multicenter, prospective, cohort study. Am J Obstet Gynecol. 2020;222(4):367 e1- e22.
128. Wu JK, Sadecki EN, Kyweluk MA, Senapati S, Flynn AN, Steider E, et al. Patient attitudes and preferences for the management of pregnancy of unknown location. F S Rep. 2022;3(3):246-52.
129. Barbe C, Boiteux-Chabrier M, Charillon E, Bouazzi L, Maheas C, Merabet F, et al. Utility of early, short psychological care for women who experience early miscarriage: protocol for the randomized, controlled MisTher trial. BMC Psychol. 2023;11(1):368.
130. 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;350:g7594.
131. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c869.
132. 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(5):521-7.
133. Practice Committee of American Society for Reproductive M. Medical treatment of ectopic pregnancy: a committee opinion. Fertil Steril. 2013;100(3):638-44.
134. Daus K, Mundy D, Graves W, Slade B. Ectopic pregnancy. What to do during the 20-day window. J Reprod Med. 1989;34(2):162-6.
135. Dart R, Mitterando J, Dart L. Rate of change of serial beta-human chorionic gonadotropin values as a predictor of ectopic pregnancy in patients with indeterminate transvaginal ultrasound findings. Ann Emerg Med. 1999;34(6):703-10.
136. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67(6):361-70.
137. Bjelland I, Dahl A, Haug T, T., Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res. 2002;52(2):69-77.
138. Lisspers J, Nygren A, Soderman E. Hospital Anxiety and Depression Scale (HAD): some psychometric data for a Swedish sample. Acta Psychiatr Scand. 1997;96(4):281-6.
139. Lok IH, Neugebauer R. Psychological morbidity following miscarriage. Best Pract Res Clin Obstet Gynaecol. 2007;21(2):229-47.
140. Sullivan M, Karlsson J, Ware Jr JE. The Swedish SF-36 Health Survey--I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. Soc Sci Med. 1995;41(10):1349-58.
141. MacHorney C, A., Ware Jr JE, Lu J, F., Sherbourne C, D. The MOS 36-ltem Short-Form Health Survey (SF-36)_ III. Tests of Data Quality, Scaling Assumptions, and Reliability Across Diverse Patient Groups. Med Care. 1994;32(1):40-66.

142. Sullivan M, Karlsson J. The Swedish SF-36 Health Survey III. Evaluation of criterion-based validity: results from normative population. J Clin Epidemiol. 1998;51(11):1105-13.
143. Jomeen J, Martin CR. The factor structure of the SF-36 in early pregnancy. J Psychosom Res. 2005;59(3):131-8.
144. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. . BMJ. 2009;338:b605.
145. Fitzgerald M, Saville B, Lewis R. Decision Curve Analysis JAMA. 2015;313(4):409-10.
146. Chong KY, Solangon S, Barnhart K, Causa-Andrieu P, Capmas P, Condous G, et al. A core outcome set for future research in ectopic pregnancy: an international consensus development study. Fertil Steril. 2023;119(5):804-12.
147. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med. 2016;35(2):214-26.
148. Snell KIE, Archer L, Ensor J, Bonnett LJ, Debray TPA, Phillips B, et al. External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. J Clin Epidemiol. 2021;135:79-89.
149. van de Loo KFE, Vlenterie R, Nikkels SJ, Merkus PJFM, Roukema J, Verhaak CM, et al. Depression and Anxiety during Pregnancy: The Influence of Maternal Characteristics. Journal of Mood Disorders and Therapy. 2019;1(1).
150. Do NC, Secher AL, Cramon P, Ringholm L, Watt T, Damm P, et al. Quality of life, anxiety and depression symptoms in early and late pregnancy in women with pregestational diabetes. Acta Obstet Gynecol Scand. 2017;96(2):190-7.
151. Zee J, Sammel MD, Chung K, Takacs P, Bourne T, Barnhart KT. Ectopic pregnancy prediction in women with a pregnancy of unknown location: data beyond 48 h are necessary. Hum Reprod. 2014;29(3):441-7.
152. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. JAMA. 2017;318(14):1377-84.
153. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. Stat Med. 2014;33(3):517-35.
154. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016;74:167-76.
155. Ware Jr JE, Gandek B, Kosinski M, Aaronson NK, Apolone G, Brazier J, et al. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol. 1998;51(11):1167-70.
156. Mathes T, Pieper D. An algorithm for the classification of study designs to assess diagnostic, prognostic and predictive test accuracy in systematic reviews. Syst Rev. 2019;8(1):226.
157. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol. 2017;17(1):162.
158. Leeflang MM, Deeks JJ, Gatsonis CA, Bossuyt P. Systematic Reviews of Diagnostic Test Accuracy. Annals of Internal Medicine. 2008;149(12):889-97.
159. Solangon SA, Van Wely M, Van Mello N, Mol BW, Ross JA, Jurkovic D. Methotrexate vs expectant management for treatment of tubal ectopic pregnancy: An individual participant data meta-analysis. Acta Obstet Gynecol Scand. 2023;102(9):1159-75.
160. Schünemann H, Oxman A, Brozek J, Glasziou P, Jaeschke R, Vist G, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ. 2008;336(7653):1106-10.
161. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagn Progn Res. 2018;2:2-11.
162. Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol. 2008;61(11):1085-94.

163. Brozek JL, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. Allergy. 2009;64(8):1109-16.
164. Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wetnz R, et al. Increasing response rates to postal questionnaires: systematic review. BMJ. 2002;324.
165. Smith AB, King M, Butow P, Olver I. A comparison of data quality and practicality of online versus postal questionnaires in a sample of testicular cancer survivors. Psychooncology. 2013;22(1):233-7.
166. Galeotti M, Mitchell G, Tomlinson M, Aventin A. Factors affecting the emotional wellbeing of women and men who experience miscarriage in hospital settings: a scoping review. BMC Pregnancy Childbirth. 2022;22:270.
167. Christodoulou E, Bobdiwala S, Kyriacou C, Farren J, Mitchell-Jones N, Ayim F, et al. External validation of models to predict the outcome of pregnancies of unknown location: a multicentre cohort study. BJOG. 2021;128(3):552-62.
168. Mavrelos D, Nicks H, Jamil A, Hoo W, Jauniaux E, Jurkovic D. Efficacy and safety of a clinical protocol for expectant management of selected women diagnosed with a tubal ectopic pregnancy. Ultrasound Obstet Gynecol. 2013;42(1):102-7.
169. Lorde N, Elgharably A, Kalaria T. Impact of Variation between Assays and Reference Intervals in the Diagnosis of Endocrine Disorders. Diagnostics (Basel). 2023;13(22).
170. Jurkovic D, Memtsa M, Sawyer E, Donaldson ANA, Jamil A, Schramm K, et al. Single-dose systemic methotrexate vs expectant management for treatment of tubal ectopic pregnancy: a placebo-controlled randomized trial. Ultrasound in Obstetrics & Gynecology. 2017;49(2):171-6.
171. Alur-Gupta S, Cooney LG, Senapati S, Sammel MD, Barnhart KT. Two-dose versus single-dose methotrexate for treatment of ectopic pregnancy: a meta-analysis. Am J Obstet Gynecol. 2019;221(2):95-108 e2.
172. Dudley PS, Heard MJ, Sangi-Haghpeykar H, Carson SA, Buster JE. Characterizing ectopic pregnancies that rupture despite treatment with methotrexate. Fertil Steril. 2004;82(5):1374-8.
173. Fistouris J, Bergh C, Strandell A. Pregnancy of unknown location: external validation of the hCG-based M6NP and M4 prediction models in an emergency gynaecology unit. BMJ Open. 2022;12(11):e058454.
174. Lee AM, Lam SKSMSM, Lau SM, Chong CSY, Wai H, Chui HW, et al. Prevalence, course, and risk factors for antenatal anxiety and depression. Obstet Gynecol. 2007;110(5):1101-12.
175. Brunton RJ, Dryer R, Saliba A, Kohlhoff J. Pregnancy anxiety: A systematic review of current scales. J Affect Disord. 2015;176:24-34.
176. Liao W, Luo Z, Dong X, Wu X, Mei Y, Cui N, et al. Associations between depressive symptoms, anxiety symptoms, their comorbidity and health-related quality of life: a large-scale cross-sectional study. BMC Public Health. 2021;21(1):1911.
177. Bussières E-L, Tarabulsy GM, Pearson J, Tessier R, Forest J-C, Giguère Y. Maternal prenatal stress and infant birth weight and gestational age: A meta-analysis of prospective studies. Developmental Review. 2015;36:179-99.
178. Ding XX, Wu YL, Xu SJ, Zhu RP, Jia XM, Zhang SF, et al. Maternal anxiety during pregnancy and adverse birth outcomes: a systematic review and meta-analysis of prospective cohort studies. J Affect Disord. 2014;159:103-10.
179. Venkatesh KK, Riley L, Castro VM, Perlis RH, Kaimal AJ. Association of Antenatal Depression Symptoms and Antidepressant Treatment With Preterm Birth. Obstet Gynecol. 2016;127(5):926-33.
180. Andersson L, Sundstrom-Poromaa I, Wulff M, Astrom M, Bixo M. Implications of antenatal depression and anxiety for obstetric outcome. Obstet Gynecol. 2004;104(3):467-76.
181. Graviditet, förlossning och tiden efter: Metodbeskrivning och kunskapsunderlag. 2022. https://www.socialstyrelsen.se/kunskapsstod-och-regler/regler-och-riktlinjer/nationella-riktlinjer/nationella-riktlinjer-graviditet-forlossning-och-tiden-efter/rekommendationer/oversikt-psykisk-ohalsa/identifiera-gravida-med-trolig-depression-eller-angest/ (Accessed November 14, 2024).

182. Antenatal and postnatal mental health: clinical management and service guidance. Clinical guideline [CG192]. 2014. https://www.nice.org.uk/guidance/cg192 (Accessed November 24, 2024).
183. Screening and Diagnosis of Mental Health Conditions During Pregnancy and Postpartum. ACOG Clinical Practice Guideline No. 4. Obstetrics & Gynecology 2023;141(6):1232-61.
184. Moscrop A, Harrison S, Heppell V, Heneghan C, Ward A. Primary care follow-up and measured mental health outcomes among women referred for ultrasound assessment of pain and/or bleeding in early pregnancy: a quantitative questionnaire study. BMJ Open. 2013;3(4).
185. Farren J, Mitchell-Jones N, Verbakel JY, Timmerman D, Jalmbrant M, Bourne T. The psychological impact of early pregnancy loss. Hum Reprod Update. 2018;24(6):731-49.
186. Johansson R, Carlbring P, Heedman A, Paxling B, Andersson G. Depression, anxiety and their comorbidity in the Swedish general population: point prevalence and the effect on health-related quality of life. PeerJ. 2013;1:e98.
187. Lagadec N, Steinecker M, Kapassi A, Magnier AM, Chastang J, Robert S, et al. Factors influencing the quality of life of pregnant women: a systematic review. BMC Pregnancy Childbirth. 2018;18(1):455.
188. Boutib A, Chergaoui S, Azizi A, Saad EM, Hilali A, Youlyouz Marfak I, et al. Health-related quality of life during three trimesters of pregnancy in Morocco: cross-sectional pilot study. EClinicalMedicine. 2023;57:101837.