International Journal of Gynecology and Obstetrics 86 (2004) 351–357

International Journal of
GYNECOLOGY
& OBSTETRICS

www.elsevier.com/locate/ijgo

Article

# Human chorionic gonadotrophin and progesterone levels in pregnancies of unknown location

G. Condous[a,*], C. Lu[b], S.V. Van Huffel[b], D. Timmerman[c], T. Bourne[a]

[a]*Early Pregnancy, Gynaecological Ultrasound and MAS Unit, Department of Obstetrics & Gynaecology, St George's Hospital Medical School, Cranmere Terrace, London SW17 0RE, UK*
[b]*Department of Electrical Engineering (ESAT), K.U.Leuven, Belgium*
[c]*Department of Obstetrics and Gynaecology, University Hospital Gasthuisberg, K.U.Leuven, Belgium*

Received 26 January 2004; received in revised form 12 April 2004; accepted 12 April 2004

**Abstract**

*Objective*: To evaluate accuracy, user variability and impact of experience on the use of serum hCG and progesterone in women who have a pregnancy of unknown location (PUL's). *Materials and methods*: This was a retrospective study. Presenting 1932 consecutive women to an Early Pregnancy Unit had a transvaginal scan. The location of the pregnancy could not be found in 189 women (Pregnancy of unknown location, PUL), and so blood was taken to measure serum hCG and progesterone at presentation and subsequently after 48 h, according to the protocol. All women were monitored at regular intervals until the final outcome was known, which was a failing PUL, a viable or failing intra-uterine pregnancy, an ectopic pregnancy or a persisting PUL. The final study group comprised 185 PUL, as four cases of persisting PUL were treated and excluded from the analysis. Five investigators assessed the hormonal data independently. The investigator's experience as defined by the number of years working in obstetrics and gynecology ranged from 2 to 15 years. Each investigator knew the women were clinically stable and that the scan result was consistent with a PUL, i.e. there were no signs of intra- or extra-uterine pregnancy, and there was no hemoperitoneum on TVS. When assessing the PUL's, each investigator was given the hormonal results at time 0 and 48 h for serum hCG and progesterone and asked to classify the PUL's as failing PUL's, immediately viable intra-uterine PUL's and ectopic PUL's. No other clinical information about the women was made available. *Results*: Complete data 185 women (89%): 102 failing PUL's, 63 immediately viable intra-uterine PUL's and 20 ectopic PUL's (total 185). The most experienced investigator obtained the best accuracy 163/185 (88.1%); not significantly different from those obtained by less experienced investigators (range 85.9–87.6%). Mean correct classification of failing PUL and immediately viable intra-uterine PUL's was 93% (range 89–95%); corresponding value for ectopic PUL's was 42% (range 25–60%). Agreement between observers for classification of failing PUL's and immediately viable intra-uterine PUL's was almost perfect (Cohen's kappa 0.86–0.90), whereas the value for ectopic PUL's group was fair to moderate (Cohen's kappa 0.39–0.67). All 5 investigators misdiagnosed same 35% of ectopic PUL's. *Conclusions*: Serum hCG and progesterone levels at defined times can be used to predict the immediate viability of

*Corresponding author. Tel.: +44-208-725-0050; fax: +44-208-725-0094.
*E-mail address:* gcondous@hotmail.com (G. Condous).

0020-7292/04/$30.00 © 2004 International Federation of Gynecology and Obstetrics. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijgo.2004.04.004

a PUL, but cannot be used reliably to predict its location. Clinical experience does not significantly improve the ability to assess PUL outcome.
© 2004 International Federation of Gynecology and Obstetrics. Published by Elsevier Ireland Ltd. All rights reserved.

*Keywords:* Ectopic pregnancy; Human chorionic gonadotrophin; Progesterone; Interuser variability; Transvaginal sonography; Pregnancy of unknown location

## 1. Introduction

The diagnosis of ectopic pregnancy should be based on the positive visualization of an extra-uterine pregnancy outside the uterus, rather than the absence of an intrauterine pregnancy. Between 87 and 93% of ectopic pregnancies may be visualized using transvaginal sonography (TVS) [1–3]. Standard algorithms for the diagnosis of ectopic pregnancy in the absence of the positive visualization of an extra-uterine pregnancy utilize serum hCG and progesterone measurements with ancillary aids that include ultrasound imaging, laparoscopy and diagnostic dilatation and curettage. In our unit we rely primarily on ultrasound, with 90.8% of ectopic pregnancies visualized using transvaginal sonography (TVS) prior to surgery [3]. This means that only a few ectopic pregnancies will fall into the pregnancy of unknown location (PUL) category. It is this group that we have concentrated on in this study.

With the introduction of Early Pregnancy Units and the use of high-resolution transvaginal probes, ectopic pregnancies are detected at a relatively asymptomatic stage and more treatment options have become available. However, as women have been encouraged to present earlier in pregnancy, the number of women presenting where a pregnancy is not seen either inside or outside the uterus has increased. These cases are classified as pregnancies of unknown location (PUL). The established hormonal criteria for the diagnosis of ectopic pregnancy have been based on data collected from pregnancies associated with abdominal pain and abnormal transvaginal bleeding and not from relatively asymptomatic women. Consequently new diagnostic criteria are being developed and tested in order to detect ectopic pregnancies in women with a PUL whilst avoiding intervention in early intra-uterine pregnancies. A PUL is found in 8–31% of women who present to an Early Pregnancy Unit (EPU) [4–7].

Traditionally the discriminatory zone or level of serum hCG above which an intra-uterine pregnancy should be visualized has been used to predict the likelihood of an ectopic pregnancy in women with a PUL [2]. The trend in serum hCG levels over a 48 h period is also utilized to predict the clinical outcome in these women [8]. Measurements of serum progesterone have been shown to be useful in evaluating the chances of early pregnancy failure. According to previously published data, a baseline serum progesterone level of $<20$ nmol/l can be used to identify a failing PUL with a PPV of $\geq 95\%$ [7]. Approximately 10% of PUL are ectopic pregnancies and it is the detection of women in this group that poses the greatest challenge. To our knowledge there are no data to examine interuser variation of the interpretation of these hormonal indices and observer experience on the accurate classification of PUL. Hence the objective of our study was to evaluate the accuracy, interuser variation and impact of clinical experience on the interpretation of measurements of serum hormones for the assessment of women with a PUL.

## 2. Materials and methods

Presenting 1932 consecutive women presenting to the EPU at St George's Hospital, London between June 2001 and March 2002 were studied. All women had a transvaginal scan with a 5 MHz probe (Aloka SSD 900 or 2000, Keymed Ltd., Southend, UK and Aloka Co. Ltd., Tokyo, Japan).

The location of the pregnancy could not be found in 189 women, and so blood was taken to measure serum human chorionic gonadotrophin (hCG, World Health Organization, Third International Reference Preparation 75/537) and proges-

terone (Roche Elecsys 2010 Progesterone II test) levels using an automated electrochemiluminescence immunoassay ('ECLIA'). These samples were measured at presentation and subsequently after 48 h, according to the protocol.

A PUL was defined on the basis of a serum hCG level $>5$ U/l and the absence of signs of either an intra- or extra-uterine pregnancy or retained products of conception by TVS. All women were monitored at regular intervals until the final outcome was known, which was a failing PUL, a viable or failing intra-uterine pregnancy, an ectopic pregnancy or a persisting PUL. The final study group comprised 185 PUL, as four cases of persisting PUL were treated and excluded from the analysis.

Indications for sonography included non-specific lower abdominal pain, with or without vaginal bleeding, poor obstetric history (previous miscarriage or ectopic pregnancy) or to determine gestational age.

The ultimate diagnosis of each woman was made in the following way. If the initial serum progesterone level was $<20$ nmol/l, the women were classified as having a failing PUL. Spontaneous resolution of the pregnancy was defined as a decrease in the serum hCG level to $<5$ U/l with the disappearance of symptoms. The location of these failing pregnancies remained unknown. Serum hCG levels were repeated within 7 days to confirm the diagnosis.

If the serum hCG rise over the 48 h period was $\geq 66\%$, for the purposes of this study, the women were considered to have an immediately viable intra-uterine PUL and were rescanned 2 weeks later to confirm the diagnosis. When a gestation sac was seen, a further scan was performed 2 weeks later to confirm the presence or absence of fetal cardiac activity. Those with no cardiac activity at the rescan were defined as being non-viable pregnancies.

Women who did not fall into either category were reviewed every 48 h until a diagnosis was made by sonography. If an EP was visualized on TVS, the women were counselled appropriately and offered either expectant management, medical management in the form of parenteral methotrexate or surgery. If an EP was not visualized, but there was a high index of suspicion based on symptomatology, clinical findings and sub-optimal rises of serial serum hCG levels, a laparoscopy was performed with or without an evacuation of the uterus. Those women who had negative findings on TVS and negative findings on laparoscopy, but their serum hCG levels had reached a plateau were given methotrexate. All women were followed up until a final diagnosis was established.

## 3. Study design

The study was retrospective. Five investigators assessed the hormonal data independently. The investigator's experience as defined by the number of years working in obstetrics and gynecology ranged from 2 to 15 years. Each investigator knew the women were clinically stable and had a scan result consistent with a PUL, i.e. there were no signs of intra- or extra-uterine pregnancy, and there was no haemoperitoneum on TVS. When assessing the PUL, each investigator was given the hormonal results at time 0 and 48 h for serum hCG and progesterone. No other clinical information about the women was made available.

Each investigator used accepted current criteria for the prediction of failing PUL, immediately viable intra-uterine PUL and ectopic PUL. A low serum progesterone at time 0 h was used to predict failing PUL's; a serum hCG rise $\geq 66\%$ over 48 h was used to predict immediately viable intra-uterine PUL's; and either a discriminatory zone $\geq 1500$ U/l and/or a sub-optimally rising serum hCG over 48 h was used to predict ectopic PUL's. Each investigator was asked to apply these algorithms to the raw hormonal data as they would in normal clinical practice. The investigators were blind to the final classification of the PUL and the opinion of the other users.

### *3.1. Statistical analysis*

Accuracy was defined as the sum of the correctly diagnosed ectopic PUL's, failing PUL's and immediately viable intra-uterine PUL's divided by the total number of cases studied ($n=185$), expressed as a percentage. The interuser agreement was evaluated with kappa statistics, which give

the chance-corrected measures of agreement. A kappa statistic of 1.0 suggests complete agreement. Kappa statistics of 0.81–1.0 indicate almost perfect agreement, 0.61–0.8 substantial agreement, 0.41–0.6 moderate agreement and 0.21–0.4 fair agreement. A kappa statistic of 0 suggests that the same degree of agreement would be expected by chance alone [9]. Cohen's kappa was used to analyze the agreement between two observers [10]. The modified method of Fleiss was used to assess the agreement among multiple observers at the one time [11]. An unweighted kappa statistic was computed individually for each PUL subclass and an overall composite kappa statistic across all subclasses. A $P$ value of less than 0.05 was regarded as significant. Statistical analysis was performed using the SAS software package, version 8.01 (SAS, Haasrode, Belgium).

## 4. Results

### 4.1. Subclassification of PUL

The final clinical outcome for the 185 PUL were: 165 non-ectopic pregnancies (102 failing PUL (55.1%) and 63 intrauterine pregnancies (34.1%) and 20 ectopic pregnancies (10.8%).

Of the 63 immediately viable intra-uterine pregnancies, 51 were viable and 12 were non-viable on rescanning.

### 4.2. Diagnostic accuracy of different observers

The most experienced observer (A) obtained an overall accuracy of 88.1%, but this value was not significantly different from the other observers (range 85.9–87.6%). The mean classification of non-ectopic pregnancies by all observers was very high for failing PUL's (93.4%, range 89.2–99%) and immediately viable intra-uterine PUL's (90.8%, range 87.3–93.7%). Conversely, the mean accuracy for the classification of ectopic PUL's was poor by all observers (42%, range 25–60) (see Table 1).

Table 2
Agreements between observer A and observers B–E

| Observers | Kappa | Standard error (S.E.) | 95% Confidence limits | |
|---|---|---|---|---|
| Ectopic PUL's | | | | |
| A,B | 0.664 | 0.139 | 0.391 | 0.937 |
| A,C | 0.438 | 0.160 | 0.124 | 0.752 |
| A,D | 0.394 | 0.159 | 0.082 | 0.705 |
| A,E | 0.672 | 0.141 | 0.395 | 0.949 |
| Failing PUL's and immediately viable intra-uterine PUL's | | | | |
| A,B | 0.878 | 0.034 | 0.810 | 0.945 |
| A,C | 0.859 | 0.037 | 0.787 | 0.932 |
| A,D | 0.858 | 0.037 | 0.785 | 0.932 |
| A,E | 0.899 | 0.032 | 0.836 | 0.961 |

### 4.3. Interuser agreement

The performance of the most experienced observer (A) was compared to the other observers for agreement. Interuser agreement between observer A and the others for the diagnosis of non-ectopic PUL's is shown in Table 2. Similarly the agreement between observer A and the others for the diagnosis of ectopic PUL's is shown in Table 2. There was almost perfect agreement between observer A and the other observers in the classification of the group containing failing PUL's and immediately viable intra-uterine PUL's (Cohen's kappa 0.858–0.899). This finding is in contrast to the classification of the ectopic PUL group,

Table 1
Diagnostic accuracy of each observer

| Accuracy (%) | Observer/years experience | | | | |
|---|---|---|---|---|---|
| | A/15 | B/12 | C/6 | D/3 | E/3 |
| Ectopic PUL's | 40.0 | 25.0 | 60.0 | 60.0 | 25.0 |
| Failing PUL's and immediately viable intra-uterine PUL's | 93.9 | 94.5 | 90.9 | 89.1 | 94.5 |
| Overall | 88.1 | 87.1 | 87.6 | 85.9 | 87.0 |

Table 3
Agreements between all observers

| Category | Frequency (%) | Kappa | S.E. |
|---|---|---|---|
| Ectopic PUL's | 11 (6.0)–25 (13.5) | 0.498 | 0.023 |
| Failing PUL's | 99 (53.5)–117 (63.2) | 0.861 | 0.023 |
| Immediately viable intra-uterine PUL's | 56 (30.3)–64 (34.6) | 0.892 | 0.023 |
| Overall | 185 (100) | 0.814 | 0.019 |

in which there was only fair to substantial agreement between the observers (Cohen's kappa 0.394–0.672). In 75.1% of PUL all five observers made a correct classification. However, whilst all five correctly classified 88.3% of failing PUL's and 80.7% of immediately viable intra-uterine PUL's, in no cases of ectopic pregnancy did all five observers agree. In 4.9% of the PUL, all five observers made an incorrect classification. In 35% of ectopic PUL's all five observers missed the diagnosis.

Similar information was also given by the kappa statistics for agreement among multiple observers. Table 3 demonstrates the proportion of PUL subclasses assigned by each observer and the kappa statistics for individual and overall PUL subclasses. Kappa statistics were highest (0.861 and 0.892) for the high-frequency subclasses of failing PUL's and immediately viable intra-uterine PUL's. The kappa statistic for the low-frequency category of ectopic PUL's was only 0.498.

## 5. Discussion

The distinction between PUL's that are developing ectopic pregnancies, early intra-uterine pregnancies and failing PUL's based on the interpretation of hormonal markers is the most difficult diagnostic problem seen in Early Pregnancy Units (EPU). Although the vast majority will be failing PUL's and early intra-uterine pregnancies, it is the group of women with an early ectopic pregnancy (approx. 10%) too early to visualize that pose the greatest concern [12]. To date, there are no published data examining interuser variation of the interpretation of hormonal indices and observer experience on the accurate classification of PUL. The primary aim of this study was to assess whether interpretation of recorded serum hCG and progesterone levels could be used to classify PUL as ectopic PUL's or failing PUL's or immediately viable intra-uterine PUL's. In women with a PUL these are the criteria on which management is based. A further aim was to evaluate the influence of experience on overall test performance.

Gestational age and endometrial thickness have not been shown to be useful in the diagnosis of ectopic pregnancy in women with a PUL [12,13]. Thus, in this study, the investigators were not provided with additional demographic or ultrasonographic information.

According to our data, current algorithms for the diagnosis of failing PUL's (low initial serum progesterone) and immediately viable intrauterine PUL's ('doubling' serum hCG) are extremely accurate. The almost perfect interuser agreement in the non-ectopic pregnancy group probably means that most cases of failing PUL and immediately viable intra-uterine PUL's are in fact quite easy to characterize on the basis of serum hormone levels. Conversely, a discriminatory zone $>1500$ U/l and/or sub-optimally rising serum hCG levels for the diagnosis of ectopic PUL's are poor diagnostic tests. The fair to substantial agreement between the observers demonstrates the difficulty in classifying the ectopic group of PUL's. This highlights the need for improved diagnostic tests for the prediction of ectopic pregnancy.

If we consider the ectopic PUL group (20/185, 10.8%), which pose the greatest threat to women in the first trimester, the results are in fact quite discouraging. All clinicians failed to diagnose ectopic PUL's in a high percentage of cases. On average, 58% (range 40–75%) of the women with early ectopic PUL's were misclassified as failing PUL and would have been managed inappropriately with a possible adverse outcome.

The ectopic PUL's were misclassified in the following way. Observer A classified 45% of the ectopic PUL's as failing PUL and 15% as immediately viable intra-uterine PUL's. Observer B classified 55% of the ectopic PUL's as failing PUL and 20% as immediately viable intra-uterine PUL's. Observer C classified 35% of the ectopic PUL's as failing PUL and 5% as immediately viable intra-uterine PUL's. Observer D classified

30% of the ectopic PUL's as failing PUL and 10% as immediately viable intra-uterine PUL's. Observer E classified 70% of the ectopic PUL's as failing PUL and 5% as immediately viable intra-uterine PUL's. The majority of ectopic PUL's were misclassified as failing PUL's. This is not surprising as 10/20 ectopic PUL's had initial serum progesterone $<20$ nmol/l. This misclassification is a potential clinical problem as seven is this group required surgical intervention. Although the present report is of a retrospective analysis, it highlights the need for the development of newer models that are not only well accepted universally, but more importantly improve the diagnostic accuracy in this ectopic pregnancy group. A newer mathematical model has been developed for the prediction of failing PUL's [7], however, its performance is no better than an initial serum progesterone $<20$ nmol/l.

One interpretation of the poor prediction of ectopic pregnancy may be poor application of the algorithms by clinicians. Alternatively, the current criteria for the prediction of ectopic pregnancy in a PUL population are poor. The latter is more likely, as hormone trends in women with early undetectable ectopic PUL's on TVS are highly variable. This is not surprising as they represent a range of failing and viable pregnancies, albeit in the wrong location. Although the overall interuser agreement was almost perfect, this can be attributable to the high prevalence of failing PUL's and immediately viable intra-uterine PUL's in this study group. As the number of ectopic PUL's increases, the interuser agreement decreases.

Experience per se, did not confer a statistically significant higher accuracy amongst the observers. In fact, the two most experienced observers misclassified 60 and 75% of ectopic PUL's.

Evaluation of serum hormone levels at defined times in pregnancies of unknown location can be used reliably to predict the immediate viability of a PUL, but cannot be used to predict its location. Experience plays no role in the accuracy of diagnosis. This study highlights the need for newer, more sensitive and reproducible models for the distinction between ectopic and non-ectopic pregnancies.

## Acknowledgments

This research was supported by interdisciplinary research grants of the research council of the Katholieke Universiteit Leuven, Belgium (IDO/99/03 and IDO/02/009 projects), by the Belgian Programme on Interuniversity Poles of Attraction (IUAP Phase V-22) and the Concerted Action Project MEFISTO-666 of the Flemish Community. Chuan Lu is supported by a KU Leuven Ph.D. scholarship.

## References

[1] Shalev E, Yarom I, Bustan M, Weiner E, Ben-Shlomo I. Transvaginal sonography as the ultimate diagnostic tool for the management of ectopic pregnancy: experience with 840 cases. Fertil Steril 1998;69:62–65.

[2] Cacciatore B, Stenman UH, Ylostalo P. Diagnosis of ectopic pregnancy by vaginal ultrasonography in combination with a discriminatory serum hCG level of 1000 IU/l (IRP). Br J Obstet Gynaecol 1990;97:904–908.

[3] Rosello N, Condous G, Okaro E, Khalid A, Alkatib M, Rao S, et al. Does transvaginal ultrasonography accurately diagnose ectopic pregnancy? Hum Reprod 2003;18(Supp 1):160.

[4] Hahlin M, Thorburn J, Bryman I. The expectant management of early pregnancies of uncertain site. Hum Reprod 1995;10:1223–1227.

[5] Banerjee S, Aslam N, Zosmer N, Woelfer B, Jurkovic D. The expectant management of women with pregnancies of unknown location. Ultrasound Obstet Gynecol 1999;14:231–236.

[6] Cacciatore B, Ylostalo P, Stenman UH, Widholm O. Suspected ectopic pregnancy: ultrasound findings and hCG levels assessed by an immunofluorometric assay. Br J Obstet Gynaecol 1988;95:497–502.

[7] 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. Br J Obstet Gynaecol 2001;108:158–163.

[8] Romero R, Kadar N, Copel JA, Jeanty P, DeCherney AH, Hobbins JC. The value of serial human chorionic gonadotropin testing as a diagnostic tool in ectopic pregnancy. Am J Obstet Gynecol 1986;155:392–394.

[9] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.

[10] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.

[11] Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley, 1981.

[12] Condous G, Okaro E, Khalid A, Zhou Y, Lu C, Van Huffel S, et al. Role of biochemical and ultrasonographic indices in the management of pregnancies of unknown location. Ultrasound Obstet Gynecol 2002; 20(Supp. 1):36–37.

[13] Mol BW, Hajenus PJ, Engelsbel S, Ankum WM, van der Veen F, Hemrika DJ, et al. Are gestational age and endometrial thickness alternatives for serum human chorionic gonadotropin as criteria for the diagnosis of ectopic pregnancy? Fertil Steril 1999;72:643–645.