failing PULs, intrauterine pregnancies (IUPs) and, most importantly, ectopic pregnancies[3]. For the detection of ectopic pregnancy, Model M1 had a sensitivity of 91.7%, a specificity of 84.2%, a positive likelihood ratio of 5.8, a positive predictive value (PPV) of 27.5% and a negative predictive value (NPV) of 99.4%[3]. The aim of the study reported here was to improve on the performance of this logistic regression model for the detection of developing ectopic pregnancies in women with a PUL. We generated and evaluated a logistic regression model from hormonal data and compared its performance with that of Model M1.

METHODS

Data collection

This was a prospective observational study. All 4698 consecutive women attending St George's Hospital early pregnancy unit (EPU), London, between 4th March 2002 and 17th July 2003, inclusive, were examined on transvaginal ultrasound for PUL. Indications for sonography included lower abdominal pain with or without vaginal bleeding, poor obstetric history and determination of gestational age. Our study group was not the same as the group used in the development of Model M1. Women classified with a PUL were eligible for inclusion unless they were clinically unstable or had an acute abdomen or blood in the pouch of Douglas according to the ultrasound images at the time of the initial scan. PULs were defined on transvaginal ultrasound as there being no sign of either an intra- or extrauterine pregnancy or retained products of conception in a woman with a positive pregnancy test. If the location of the pregnancy could not be ascertained by sonography, peripheral blood was taken at presentation to measure the levels of serum hCG (World Health Organization, Third International Reference 75/537) and progesterone (Roche Elecsys 2010 Progesterone II test, Roche Diagnostics, Lewes, UK) using automated electrochemiluminescence immunoassays (ECLIA). Serum hCG ≥ 5 U/L represented a positive pregnancy test; no reference range was used for serum progesterone. The levels of the hormones hCG and progesterone were measured again 48 h later according to the unit's protocol.

Demographic variables recorded included the woman's age and gestational age at presentation. Sonographic variables recorded were endometrial thickness, the character of the endometrium (intact or disrupted) and the presence or absence of free fluid in the pouch of Douglas. These variables were recorded at the initial scan by the sonographer.

The study group consisted of 376 consecutive women with a PUL. The data from the first 201 women classified as having a PUL collected between March 2002 and November 2002 were used as the training set. Statistical analysis and development of the logistic regression model were based on this dataset. The data from the next 175 women with a PUL recruited between November 2002 and July 2003 were taken as the test set in order to assess the performance of the model prospectively.

All women were managed expectantly until the final definitive diagnosis was made, i.e. until either the location of the pregnancy was established using transvaginal ultrasound, or the serum hCG levels declined, indicating that the pregnancy had failed and resolved spontaneously. There were three outcome groups: a failing PUL, an IUP or an ectopic pregnancy. The same primary investigator (G.C.) followed up all women and was responsible for collating demographic data, biochemical results and the final clinical outcomes. The follow-up protocol was as follows. If the serum hCG rise over the 48-h period was > 66%[1,4], the women were classified initially as having an early IUP and were rescanned 2 weeks later. The presence at follow-up scan of an intrauterine sac surrounded by a brightly echoic ring, situated eccentrically within the endometrial cavity, confirmed the diagnosis[5]. These women were rescanned 2 weeks thereafter to confirm viability. If the initial serum progesterone level was < 20 nmol/L, the women were classified initially as having a failing PUL[1,2]. 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 PULs remained unknown. In these women, serum hCG levels were repeated within 7 days to confirm the diagnosis. Women who did not fall into either category were reviewed every 48 h with serum hCG testing and/or sonography until a diagnosis was made. If an ectopic pregnancy was not visualized using transvaginal ultrasound, 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 evacuation of the uterus.

The gold standard for the diagnosis of tubal ectopic pregnancy (histological confirmation of villi in the tube) was not applied to all women in this study, because some women with an ultrasound diagnosis of tubal pregnancy were treated conservatively. A sonographic diagnosis of an ectopic pregnancy was based on one the following ultrasound findings[6]:

1) an inhomogeneous mass seen in the adnexal region adjacent to the ovary and moving separate to this (referred to as the 'blob' sign);
2) a mass with a hyperechoic ring around the gestational sac seen in the adnexal region (referred to as the 'bagel' sign)
3) a gestational sac with a fetal pole with or without cardiac activity seen in the adnexal region.

There were four women in the training set and two in the test set whose serum hCG levels plateaued and in whom no pregnancy was seen on transvaginal ultrasound at any time. These were classified as having a persisting PUL and may represent women with either persistent intrauterine trophoblast or a missed ectopic pregnancy on transvaginal ultrasound. These were treated with methotrexate and were excluded because the final outcome was unknown.

Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.

Table 1 Demographic, hormonal and sonographic characteristics of 370 women with a pregnancy of unknown location (PUL)

Variable	*Descriptive statistics for each group*			*Difference in variable between groups (P)*			
				Kruskal–Wallis test: all groups	*Mann–Whitney test*		
	Failing PUL	*IUP*	*EP*		*Failing PUL vs. IUP*	*Failing PUL vs. EP*	*IUP vs. EP*
Training set (n)	109	76	12				
hCG ratio	0.353 (0.265)	2.120 (0.571)	1.414 (0.466)	< 0.0001	< 0.0001	< 0.0001	< 0.0001
hCG average (IU/L)	89 (216)	727.3 (985.8)	620.8 (354)	< 0.0001	< 0.0001	< 0.0001	0.3429
Progesterone ratio	0.667 (0.500)	0.944 (0.272)	0.804 (0.305)	< 0.0001	< 0.0001	0.0661	0.1620
Progesterone average (nmol/L)	3.5 (4.8)	66.5 (38.3)	40.8 (61)	< 0.0001	< 0.0001	< 0.0001	0.1531
Endometrial thickness (mm)	7.9 (5.1)	12.5 (6.4)	11.4 (5.8)	< 0.0001	< 0.0001	0.0016	0.4509
Gestational age (days)	43.5 (11)	36.5 (11)	43 (23)	< 0.0001	< 0.0001	0.5452	0.0236
Maternal age (years)	30 (7)	30 (6)	32 (6.5)	0.2248			
Test set (n)	94	64	15				
hCG ratio	0.418 (0.238)	2.302 (0.611)	1.309 (0.647)				
hCG average (IU/L)	89.3 (323.5)	688 (895.8)	533.5 (1416)				
Progesterone ratio	0.706 (0.600)	0.958 (0.228)	1.067 (0.282)				
Progesterone average (nmol/L)	2.8 (5)	68.3 (32.3)	27 (34.5)				
Endometrial thickness (mm)	8.4 (6)	14.1 (6.4)	11.9 (6.4)				
Gestational age (days)	46 (18)	34 (11)	45 (13)				
Maternal age (years)	32 (8)	27.5 (11)	31 (4)				

Data are expressed as median (interquartile range). Some data were not recorded: the numbers for gestational age for the failing PUL, IUP and EP groups are 104, 72 and 9, respectively, for the training set and 88, 58 and 13, respectively, for the test set; one value for progesterone in the training set and one for endometrial thickness in the test set were also not recorded. EP, ectopic pregnancy; hCG, human chorionic gonadotropin; IUP, intrauterine pregnancy.

Predictive three-category logistic regression models

To find the optimal final model, M4, we started from the set of candidate variables listed in Table 1. However, the log(progesterone average) and gestational age were not used. There was a strong correlation between log(progesterone average) and hCG ratio ($r = 0.83$), which is an indication of multicollinearity, and information on gestational age was missing in three of the 12 ectopic pregnancies. The remaining variables (hCG ratio, log(hCG average), progesterone ratio, endometrial thickness and maternal age) were entered into the model selection process. Model M4 included the log(hCG average) (termed 'LhCGmean'), the hCG ratio ('hCGratio'), and the quadratic effect of the hCG ratio ('hCGratio^2'). The quadratic effect was added because the linearity in the logit assumption was violated for the hCG ratio. A centered version of hCG ratio ('hCG ratioc') was used, consisting of the hCG ratio minus its average (1.17 in the training set) to suppress the correlation between hCG ratio and its square. Centering the variable reduced the correlation from 0.96 to 0.59. Centering is common practice, because high correlation causes instability of the parameter estimates. Subtracting 1.17 from the hCG ratio was necessary when applying this model to other datasets. These are the resulting model equations for M4:

$$\log\left(\frac{P_{failing}}{P_{IUP}}\right) = 5.88 - 1.18 \cdot LhCGmean - 5.56 \cdot hCGratioc + 2.05 \cdot hCGratioc^2 = z_1;$$

$$\log\left(\frac{P_{ectopic}}{P_{IUP}}\right) = 0.39 - 0.06 \cdot LhCGmean - 0.26 \cdot hCGratioc - 3.93 \cdot hCGratioc^2 = z_2;$$

$$\log\left(\frac{P_{ectopic}}{P_{failing}}\right) = -5.49 + 1.12 \cdot LhCGmean + 5.30 \cdot hCGratioc - 5.98 \cdot hCGratioc^2 = z_2 - z_1.$$

The predicted probability for different classes was computed using the following formulae:

$$P_{ectopic} = \frac{e^{z_2}}{1 + e^{z_1} + e^{z_2}};$$

$$P_{failing} = \frac{e^{z_1}}{1 + e^{z_1} + e^{z_2}};$$

$$P_{IUP} = \frac{1}{1 + e^{z_1} + e^{z_2}};$$

where e is the exponential constant (approximately 2.718). All three parameters in this multicategory logit model were significant, with likelihood ratio P-values ≤ 0.01. Table 2 presents the odds ratios for LhCGmean, hCGratio and hCGratio^2 according to the outcome of the pregnancy. The quadratic effect made the odds ratios for the hCG ratio difficult to interpret, but it was clear that, at least in general, higher hCG ratios meant lower odds of a PUL being either an ectopic pregnancy or a failing PUL than it being an IUP, while higher hCG ratios meant

Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.

Ultrasound Obstet Gynecol 2007; 29: 680–687
Published online 8 May 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/uog.4015

Prediction of ectopic pregnancy in women with a pregnancy of unknown location

G. CONDOUS*†, B. VAN CALSTER‡, E. KIRK*, Z. HAIDER*, D. TIMMERMAN§, S. VAN HUFFEL‡ and T. BOURNE*

**Early Pregnancy, Gynaecological Ultrasound and MAS Unit, Department of Obstetrics and Gynaecology, St. George's Hospital Medical School, University of London, London, UK, ‡Department of Electrical Engineering (ESAT) and §Department of Obstetrics and Gynaecology, University Hospital, Gasthuisberg, K.U.Leuven, Leuven, Belgium and †Early Pregnancy and Advanced Endosurgery Unit, Nepean Clinical School, University of Sydney, Nepean Hospital, Sydney, Australia*

KEYWORDS: ectopic pregnancy; human chorionic gonadotropin ratio; logistic regression; pregnancy of unknown location; quadratic effect

ABSTRACT

Objective We have previously published on the use of mathematical Model M1 to predict ectopic pregnancy in women with no signs of intra- or extrauterine pregnancy. The aim of this study was to improve on the performance of this model for the detection of developing ectopic pregnancies in women with pregnancies of unknown location (PULs). We therefore generated and evaluated a new logistic regression model from simple hormonal data and compared it with Model M1.

Methods Data were collected prospectively from women classified as having a PUL. These women were followed until the diagnosis was established as: a failing PUL, an intrauterine pregnancy (IUP) or an ectopic pregnancy. A multinomial logistic regression model, Model M4, was developed on 201 training cases and it was tested prospectively on another 175 women with a PUL. M4 performance was evaluated using receiver–operating characteristics (ROC) curves and compared with Model M1 based on the human chorionic gonadotropin (hCG) ratio alone.

Results A total of 376 women with a PUL were recruited into this study: 201 in the training set (109 (54.2%) with a failing PUL, 76 (37.8%) with an IUP and 12 (6.0%) with an ectopic pregnancy; four with a persisting PUL were excluded from analysis) and 175 in the test set (94 (53.7%) with a failing PUL, 64 (36.6%) with an IUP and 15 (8.6%) with an ectopic pregnancy; two with a persisting PUL were excluded from analysis). The log serum hCG average ((hCG 0 h + hCG 48 h)/2) and the hCG ratio (hCG 48 h/hCG 0 h) were encoded as variables following multivariate analysis of the basic data. The new Model M4 contained the log of the hCG average, the hCG ratio and its quadratic effect. In the prediction of ectopic pregnancy, M4 gave an area under the ROC curve (AUC) of 0.900 and M1 gave an AUC of 0.842 ($P = 0.0303$).

Conclusions Although Model M4 is superior to Model M1 when comparing the AUCs for prediction of developing ectopic pregnancies in a PUL population, in real terms this model did not result in substantially more pregnancies being classified correctly as developing ectopic pregnancies. Prospective multicenter studies are needed to assess the diagnostic performance of such models in different populations. Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.

INTRODUCTION

Women having non-diagnostic early pregnancy scans or pregnancies of unknown location (PULs) can be defined as those having a positive pregnancy test with an empty uterus and with no signs of an extrauterine pregnancy on transvaginal ultrasound examination[1–3]. Logistic regression analysis has been used to predict the outcome of PULs[2,3]. A serum progesterone level of < 20 nmol/L performs as well as complex multiparameter diagnostic models in the diagnosis of spontaneously resolving PULs[2]. Our unit has published on the use of a logistic regression model to predict the outcome of PULs[3]. This logistic regression model, Model M1, based on the human chorionic gonadotropin (hCG) ratio (hCG 48 h/hCG 0 h), predicted which PULs would become failing PULs, intrauterine pregnancies (IUPs) and, most importantly, ectopic pregnancies[3]. For the detection of ectopic pregnancy, Model M1 had a sensitivity of 91.7%, a specificity of 84.2%, a positive likelihood ratio of 5.8, a positive predictive value (PPV) of 27.5% and a negative predictive value (NPV) of 99.4%[3]. The aim of the study reported here was to improve on the performance of this logistic regression model for the detection of developing ectopic pregnancies in women with a PUL. We generated and evaluated a logistic regression model from hormonal data and compared its performance with that of Model M1.

Correspondence to: Mr G. Condous, University of Sydney, Nepean Clinical School, Early Pregnancy and Advanced Endosurgery Unit, Level 5 South Block, Nepean Hospital, Penrith, Sydney, New South Wales 2750, Australia (e-mail: gcondous@hotmail.com)

Accepted: 25 January 2007

Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.

METHODS

Data collection

This was a prospective observational study. All 4698 consecutive women attending St George's Hospital early pregnancy unit (EPU), London, between 4th March 2002 and 17th July 2003, inclusive, were examined on transvaginal ultrasound for PUL. Indications for sonography included lower abdominal pain with or without vaginal bleeding, poor obstetric history and determination of gestational age. Our study group was not the same as the group used in the development of Model M1. Women classified with a PUL were eligible for inclusion unless they were clinically unstable or had an acute abdomen or blood in the pouch of Douglas according to the ultrasound images at the time of the initial scan. PULs were defined on transvaginal ultrasound as there being no sign of either an intra- or extrauterine pregnancy or retained products of conception in a woman with a positive pregnancy test. If the location of the pregnancy could not be ascertained by sonography, peripheral blood was taken at presentation to measure the levels of serum hCG (World Health Organization, Third International Reference 75/537) and progesterone (Roche Elecsys 2010 Progesterone II test, Roche Diagnostics, Lewes, UK) using automated electrochemiluminescence immunoassays (ECLIA). Serum hCG ≥ 5 U/L represented a positive pregnancy test; no reference range was used for serum progesterone. The levels of the hormones hCG and progesterone were measured again 48 h later according to the unit's protocol.

Demographic variables recorded included the woman's age and gestational age at presentation. Sonographic variables recorded were endometrial thickness, the character of the endometrium (intact or disrupted) and the presence or absence of free fluid in the pouch of Douglas. These variables were recorded at the initial scan by the sonographer.

The study group consisted of 376 consecutive women with a PUL. The data from the first 201 women classified as having a PUL collected between March 2002 and November 2002 were used as the training set. Statistical analysis and development of the logistic regression model were based on this dataset. The data from the next 175 women with a PUL recruited between November 2002 and July 2003 were taken as the test set in order to assess the performance of the model prospectively.

All women were managed expectantly until the final definitive diagnosis was made, i.e. until either the location of the pregnancy was established using transvaginal ultrasound, or the serum hCG levels declined, indicating that the pregnancy had failed and resolved spontaneously. There were three outcome groups: a failing PUL, an IUP or an ectopic pregnancy. The same primary investigator (G.C.) followed up all women and was responsible for collating demographic data, biochemical results and the final clinical outcomes. The follow-up protocol was as follows. If the serum hCG rise over the 48-h period was > 66%[1,4], the women were classified initially as having an early IUP and were rescanned 2 weeks later. The presence at follow-up scan of an intrauterine sac surrounded by a brightly echoic ring, situated eccentrically within the endometrial cavity, confirmed the diagnosis[5]. These women were rescanned 2 weeks thereafter to confirm viability. If the initial serum progesterone level was < 20 nmol/L, the women were classified initially as having a failing PUL[1,2]. 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 PULs remained unknown. In these women, serum hCG levels were repeated within 7 days to confirm the diagnosis. Women who did not fall into either category were reviewed every 48 h with serum hCG testing and/or sonography until a diagnosis was made. If an ectopic pregnancy was not visualized using transvaginal ultrasound, 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 evacuation of the uterus.

The gold standard for the diagnosis of tubal ectopic pregnancy (histological confirmation of villi in the tube) was not applied to all women in this study, because some women with an ultrasound diagnosis of tubal pregnancy were treated conservatively. A sonographic diagnosis of an ectopic pregnancy was based on one the following ultrasound findings[6]:

1) an inhomogeneous mass seen in the adnexal region adjacent to the ovary and moving separate to this (referred to as the 'blob' sign);
2) a mass with a hyperechoic ring around the gestational sac seen in the adnexal region (referred to as the 'bagel' sign)
3) a gestational sac with a fetal pole with or without cardiac activity seen in the adnexal region.

There were four women in the training set and two in the test set whose serum hCG levels plateaued and in whom no pregnancy was seen on transvaginal ultrasound at any time. These were classified as having a persisting PUL and may represent women with either persistent intrauterine trophoblast or a missed ectopic pregnancy on transvaginal ultrasound. These were treated with methotrexate and were excluded because the final outcome was unknown.

Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.

Data analysis

The data were pre-processed prior to further analysis. The serum hCG levels were analyzed through their averages and their ratio: the hCG average was defined as the mean of the two serum hCG levels over an interval of 48 h ((serum hCG at 0 h + serum hCG at 48 h)/2); the hCG ratio was defined as the serum hCG level at 48 h divided by that at 0 h. The hCG ratio has been shown to outperform single hCG levels in the management of PULs[7]. Both serum progesterone levels were analyzed analogously, i.e. through their average and their ratio. Because the hCG and progesterone averages (but not the ratios) had a highly skewed distribution, they were log-transformed.

Statistical analyses were conducted with SAS 9.1 (Statistical Analysis System Version 9.1, Cary, NC, USA). Univariate and multivariate analysis was performed retrospectively on the training set in order to highlight the most significant variables for the prediction of the outcome of PULs. To check for differences between groups in continuous variables, a non-parametric Kruskal–Wallis test was used, since most of the continuous variables were not distributed normally. If the Kruskal–Wallis test yielded a statistically significant result, the statistical significance of differences between two groups was tested using Wilcoxon's rank sum test (Mann–Whitney test).

Model building

Multicategorical logistic regression using baseline-category logits[8] was conducted to predict the outcome of the PULs. In such a model, each outcome category was paired with a baseline category, i.e. IUP. This resulted in two logit equations, one comparing the ectopic pregnancy group with the IUP group and the other comparing the failing PUL group with the IUP group. The third equation, comparing the ectopic pregnancy group with the failing PUL group can be derived from the other two. Variables were selected using automated stepwise methods in combination with manual selection procedures; criteria for selection were the Akaike information criterion, the Bayesian information criterion and to a lesser extent the likelihood ratio P-value. Attention was paid to possible multicollinearity (i.e. the existence of strong associations between the predictor variables). Checks for linearity in the logit were carried out[9] and the presence of relevant interaction terms was also checked. Multicollinearity should be avoided, because it can cause instability of the model parameter estimates.

Performance measure and classification rules

In order to determine the potential predictive power of the multicategorical logit Model M4, we first considered three binary classification problems, i.e. we used the predicted probability for a certain class to distinguish that class of PUL from the other PULs. Receiver–operating characteristics (ROC) analysis could be performed on the three binary classifications. In this study, the area under the curve (AUC) was obtained by a non-parametric method to approximate the area and its associated standard error[10,11]. This also allowed the comparison of two ROC curves from the same study population[11].

The performance of Model M4 was also evaluated in terms of sensitivity, specificity, PPV and NPV. In order to classify a case into one of the three categories we used weighted predicted probabilities. The multicategorical logistic regression model provides the predicted probability (P) for each class: $P_{failing}$, P_{IUP} and $P_{ectopic}$. The weights (W) for each class are equal to $WC_{failing}$, WC_{IUP} and $WC_{ectopic}$, respectively. The weighted predicted probabilities were then given by $WC_{failing}P_{failing}$, $WC_{IUP}P_{IUP}$, and $WC_{ectopic}P_{ectopic}$. The category with the highest weighted predicted probability became the predicted outcome. The 'optimal' weights were chosen subjectively using the training data and were tuned in order to obtain balanced sensitivity and specificity results for all three classes. Without weighting, sensitivity and specificity for ectopic pregnancy may be highly unbalanced because this outcome is much less frequent than are IUP or failing PUL.

The performance of Model M4 was compared with that of Model M1 using the test set. Because three tests are reported concerning the difference in performance between M1 and M4, a Bonferroni or related correction could have been considered to correct for multiple testing. However, objections against such corrections have been formulated, because they reduce the power of the test so that relevant findings may be missed. Therefore, such a correction was not done.

RESULTS

During the study period we scanned 4698 consecutive women. A total of 376 consecutive women with PULs were recruited, 201 in the training set and 175 in the test set. In the training set of 201 women with a PUL there were 109 (54.2%) with a failing PUL, 76 (37.8%) with an IUP, 12 (6.0%) with an ectopic pregnancy and four (2.0%) with a persisting PUL. In the test set of 175 women with a PUL, there were 94 (53.7%) with a failing PUL, 64 (36.6%) with an IUP, 15 (8.6%) with an ectopic pregnancy and two (1.1%) with a persisting PUL. The six cases with a persisting PUL were excluded. In the IUP group, 76.4% of pregnancies were viable and 23.6% were non-viable at the follow-up ultrasound examination.

Table 1 presents the demographic, hormonal and sonographic characteristics of the 370 women with a PUL included in the study. Several variables were apparently useful in discriminating between a failing PUL and an IUP. In contrast, ectopic pregnancies were more difficult to distinguish from the other groups. They were more similar to the IUP group than they were to the failing PUL group.

Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.

Table 1 Demographic, hormonal and sonographic characteristics of 370 women with a pregnancy of unknown location (PUL)

				Difference in variable between groups (P)			
	Descriptive statistics for each group			Kruskal–Wallis test: all groups	Mann–Whitney test		
Variable	Failing PUL	IUP	EP		Failing PUL vs. IUP	Failing PUL vs. EP	IUP vs. EP
Training set (n)	109	76	12				
hCG ratio	0.353 (0.265)	2.120 (0.571)	1.414 (0.466)	< 0.0001	< 0.0001	< 0.0001	< 0.0001
hCG average (IU/L)	89 (216)	727.3 (985.8)	620.8 (354)	< 0.0001	< 0.0001	< 0.0001	0.3429
Progesterone ratio	0.667 (0.500)	0.944 (0.272)	0.804 (0.305)	< 0.0001	< 0.0001	0.0661	0.1620
Progesterone average (nmol/L)	3.5 (4.8)	66.5 (38.3)	40.8 (61)	< 0.0001	< 0.0001	< 0.0001	0.1531
Endometrial thickness (mm)	7.9 (5.1)	12.5 (6.4)	11.4 (5.8)	< 0.0001	< 0.0001	0.0016	0.4509
Gestational age (days)	43.5 (11)	36.5 (11)	43 (23)	< 0.0001	< 0.0001	0.5452	0.0236
Maternal age (years)	30 (7)	30 (6)	32 (6.5)	0.2248			
Test set (n)	94	64	15				
hCG ratio	0.418 (0.238)	2.302 (0.611)	1.309 (0.647)				
hCG average (IU/L)	89.3 (323.5)	688 (895.8)	533.5 (1416)				
Progesterone ratio	0.706 (0.600)	0.958 (0.228)	1.067 (0.282)				
Progesterone average (nmol/L)	2.8 (5)	68.3 (32.3)	27 (34.5)				
Endometrial thickness (mm)	8.4 (6)	14.1 (6.4)	11.9 (6.4)				
Gestational age (days)	46 (18)	34 (11)	45 (13)				
Maternal age (years)	32 (8)	27.5 (11)	31 (4)				

Data are expressed as median (interquartile range). Some data were not recorded: the numbers for gestational age for the failing PUL, IUP and EP groups are 104, 72 and 9, respectively, for the training set and 88, 58 and 13, respectively, for the test set; one value for progesterone in the training set and one for endometrial thickness in the test set were also not recorded. EP, ectopic pregnancy; hCG, human chorionic gonadotropin; IUP, intrauterine pregnancy.

Predictive three-category logistic regression models

To find the optimal final model, M4, we started from the set of candidate variables listed in Table 1. However, the log(progesterone average) and gestational age were not used. There was a strong correlation between log(progesterone average) and hCG ratio ($r = 0.83$), which is an indication of multicollinearity, and information on gestational age was missing in three of the 12 ectopic pregnancies. The remaining variables (hCG ratio, log(hCG average), progesterone ratio, endometrial thickness and maternal age) were entered into the model selection process. Model M4 included the log(hCG average) (termed 'LhCGmean'), the hCG ratio ('hCGratio'), and the quadratic effect of the hCG ratio ('hCGratio^2'). The quadratic effect was added because the linearity in the logit assumption was violated for the hCG ratio. A centered version of hCG ratio ('hCG ratioc') was used, consisting of the hCG ratio minus its average (1.17 in the training set) to suppress the correlation between hCG ratio and its square. Centering the variable reduced the correlation from 0.96 to 0.59. Centering is common practice, because high correlation causes instability of the parameter estimates. Subtracting 1.17 from the hCG ratio was necessary when applying this model to other datasets. These are the resulting model equations for M4:

$$\log\left(\frac{P_{failing}}{P_{IUP}}\right) = 5.88 - 1.18 \cdot LhCGmean - 5.56 \cdot hCGratioc + 2.05 \cdot hCGratioc^2 = z_1;$$

$$\log\left(\frac{P_{ectopic}}{P_{IUP}}\right) = 0.39 - 0.06 \cdot LhCGmean - 0.26 \cdot hCGratioc - 3.93 \cdot hCGratioc^2 = z_2;$$

$$\log\left(\frac{P_{ectopic}}{P_{failing}}\right) = -5.49 + 1.12 \cdot LhCGmean + 5.30 \cdot hCGratioc - 5.98 \cdot hCGratioc^2 = z_2 - z_1.$$

The predicted probability for different classes was computed using the following formulae:

$$P_{ectopic} = \frac{e^{z_2}}{1 + e^{z_1} + e^{z_2}};$$

$$P_{failing} = \frac{e^{z_1}}{1 + e^{z_1} + e^{z_2}};$$

$$P_{IUP} = \frac{1}{1 + e^{z_1} + e^{z_2}};$$

where e is the exponential constant (approximately 2.718). All three parameters in this multicategory logit model were significant, with likelihood ratio P-values ≤ 0.01. Table 2 presents the odds ratios for LhCGmean, hCGratio and hCGratio^2 according to the outcome of the pregnancy. The quadratic effect made the odds ratios for the hCG ratio difficult to interpret, but it was clear that, at least in general, higher hCG ratios meant lower odds of a PUL being either an ectopic pregnancy or a failing PUL than it being an IUP, while higher hCG ratios meant

Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.

Table 2 Odds ratios with 95% CIs based on Model M4 to distinguish outcomes in women with a pregnancy of unknown location (PUL)

Variable	Odds ratio		
	EP vs. IUP	*Failing PUL vs. IUP*	*EP vs. failing PUL*
log hCG average	0.944 (0.515–1.732)	0.307* (0.138–0.687)	3.072* (1.295–7.285)
hCG ratio – 1.17	0.769 (0.053–11.116)	0.004* (< 0.001–0.030)	199.7* 10.39–> 1000)
hCG ratio – 1.17[2]	0.020 (< 0.001–1.303)	7.771* (2.904–20.797)	0.003* (< 0.001–0.161)

*One odds ratio was outside the 95% CI. EP, ectopic pregnancy; hCG, human chorionic gonadotropin; IUP, intrauterine pregnancy.

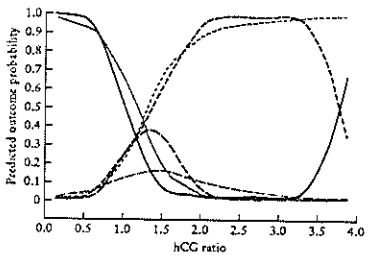

Figure 1 Probability of each outcome (failing pregnancy of unknown location (PUL) (——), intrauterine pregnancy (– – –) or ectopic pregnancy (— — —)) in women with a PUL, calculated using Model M4 versus human chorionic gonadotropin (hCG) ratio. Black lines represent the predicted probability for each outcome calculated using Model M4; gray lines represent the probabilities when the quadratic effect is discarded from Model M4.

higher odds of a PUL being an ectopic pregnancy than it being a failing PUL.

Figure 1 shows how the predicted probability for each outcome was improved when the quadratic effect was added to Model M4. The black line represents M4 and the gray line represents M4 when the quadratic effect has been discarded. By adding the quadratic effect, the background probability of an ectopic pregnancy (when the hCG ratio is between 0.75 and 1.75) was improved. Similarly, when the quadratic effect was added to Model M4, this improved the background probability of a failing PUL vs. an IUP when the hCG ratio was > 3.5. This was a result of the hCG ratio being skewed for women with a failing PUL; a few of these women had a very high hCG ratio whilst the majority had a very low one. Table 2 shows that failing PULs could be distinguished easily from both IUPs and ectopic pregnancies, but that it was more difficult to discriminate between IUPs and ectopic pregnancies. For the 12 ectopic pregnancies in the training set, the average predicted probability of a failing PUL was 0.13, while the average predicted probability of an IUP was 0.56.

When we added gestational age as a covariate to investigate post-hoc its possible importance, this variable was not thought to have a relevant influence on the outcome (likelihood ratio *P*-value = 0.2068).

ROC analysis on the three binary classifications

The ability of the logistic regression models to predict one outcome vs. the other two outcomes was evaluated by constructing ROC curves. Table 3 gives the AUCs of Model M4 for the training and test sets. The performance of Model M1, which incorporates the hCG ratio alone[3], was evaluated on the test set only and was compared with that of Model M4 (Table 3). Model M4 performed better than did Model M1 on the test set for the prediction of ectopic pregnancy. The test set AUC for Model M4 was 5.8% higher than was that for Model M1 on the same dataset.

Three-category classification rules

We derived classification rules using weighted predicted probabilities. The optimal (relative) weight values for misclassifying a failing PUL, an IUP and an ectopic pregnancy, selected based on the performance on the training set, were 1, 1 and 4, respectively. These weights were also optimal for Model M1[3]. The results of using these rules are shown in Table 4. The sensitivity and specificity with regard to ectopic pregnancy of Model M4 were fair: 80.0% and 88.6%, respectively, on the test set. The corresponding values for Model M1 were 73.3% and 87.3%, with one additional ectopic pregnancy being classified correctly by Model M4.

Prediction of the outcome of persistent PULs

The persistent PULs were excluded from model development and validation, which might impact on the model's performance when it is used in practice. In order to discover the potential influence of the model prediction on these patients, we checked the predicted probability of the six persistent PULs using Model M4. The hCG ratios of the six cases with a persistent PUL were similar to those of the 27 cases with an ectopic pregnancy (Figure 2). Five of the six women with a persistent PUL were predicted to have an ectopic pregnancy; when the same classification rules as those in Table 4 were used, one was predicted to have a failing PUL. Therefore, the cases of persistent PUL had characteristics similar to those of ectopic pregnancy. A persistent PUL was assigned as an ectopic pregnancy by the multicategory logit models with a chance of around 80% (5/6).

Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.

Table 3 Areas under the receiver–operating characteristics curves (AUC) for Models M4 and M1 to distinguish outcomes in women with a pregnancy of unknown location (PUL)

Predicted outcome/model	AUC (95% CI)		p*
	Training set	*Test set*	
Failing PUL			0.2731
Model M4	0.991 (0.981–1.000)	0.978 (0.954–1.000)	
Model M1		0.965 (0.936–0.994)	
Intrauterine pregnancy			0.5790
Model M4	0.976 (0.956–0.996)	0.974 (0.954–0.994)	
Model M1		0.969 (0.945–0.993)	
Ectopic pregnancy			0.0303
Model M4	0.941 (0.886–0.996)	0.900 (0.812–0.988)	
Model M1		0.842 (0.722–0.962)	

*Comparison of test-set AUCs.

Table 4 Performance of Models M4 and M1 to distinguish outcomes in women with a pregnancy of unknown location (PUL) based on the weighted predicted probabilities as developed on their respective training sets

Dataset and model/ outcome	Predicted category				Performance index					
	Failing PUL (n)	*IUP (n)*	*EP (n)*	*Total (n)*	*Sens. (%)*	*Spec. (%)*	*PPV (%)*	*NPV (%)*	*LR+*	*LR–*
Training data: Model M4										
Failing PUL	106	1	2	109	97.3	97.7	98.2	96.6	42.3	0.03
IUP	2	60	14	76	79.0	98.3	96.8	88.1	46.5	0.21
EP	0	1	11	12	91.7	91.4	40.7	99.4	10.7	0.09
Total	108	62	27	197						
Test set: Model M4										
Failing PUL	82	2	10	94	87.2	97.5	97.6	86.5	34.9	0.13
IUP	1	55	8	64	85.9	96.3	93.2	92.1	23.2	0.04
EP	1	2	12	15	80.0	88.6	40.0	97.9	7.02	0.23
Total	84	59	30	173						
Test set: Model M1										
Failing PUL	81	2	11	94	86.2	98.7	98.8	85.7	66.3	0.14
IUP	0	55	9	64	85.9	95.4	91.7	92.0	18.7	0.15
EP	1	3	11	15	73.3	87.3	35.5	97.2	5.77	0.31
Total	82	60	31	173						

EP, ectopic pregnancy; IUP, intrauterine pregnancy; LR+, positive likelihood ratio; LR–, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; Sens., sensitivity; Spec., specificity.

DISCUSSION

When an early pregnancy ultrasound examination is non-diagnostic, logistic regression models can be used to predict the final outcome of the pregnancy. In this study, Model M4 outperformed the previously published Model M1. Adding to the model the log of the average hCG levels and the quadratic effect of the ratio of hCG resulted in a relative increase of 6.9% of the test-set AUC for the prediction of ectopic pregnancy ($P = 0.0303$). We acknowledge that the number of ectopic pregnancies in this study was small and therefore the results should be interpreted with this in mind. M4 is a promising model, but to prove it to be clinically superior to Model M1 one would need to test its performance on a large number of women.

Both these models were developed on the same inner-city London Hospital PUL population, albeit on different model-building training data. We do not advocate the use of these models in other centers as the general applicability is yet to be established. We acknowledge that there are potential limitations in the use and uptake of these models in other centers. It is very important to assess the diagnostic performance of such models in different populations; prospective multicenter studies are now needed. As the models rely solely on the objective measurement of serum hCG (World Health Organization, Third International Reference 75/537) at presentation and at 48 h, we believe that these results should be reproducible in other scanning based units. The models do not have to be used at the same gestational period in early pregnancy provided the serum hCG levels are < 10000 U/L, because the rate of serum hCG change is linear at such hCG concentrations[12].

We have previously published on the use of the hCG ratio to predict failing PULs[7]. We argued that it is important to classify failing PULs such that this group

Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.

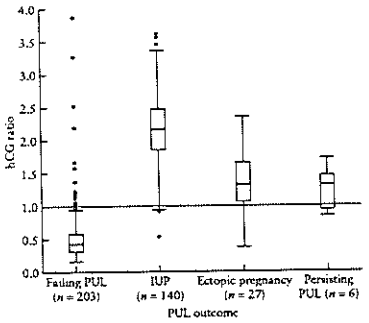

Figure 2 Box and whisker plot of the human chorionic gonadotropin (hCG) ratio versus outcome (failing or persisting pregnancy of unknown location (PUL), intrauterine pregnancy (IUP) or ectopic pregnancy) in women with a PUL (calculations made on the whole dataset of 376 women). Boxes represent the interquartile range (IQR), internal lines the median, whiskers the range of extreme observations within 1.5 × IQR and stars the extreme observations beyond this range.

can be potentially discharged at 48 h; this in turn reduced the need for follow-up. Our mathematical models could be utilized for this purpose because they identify failing PULs. PULs not classified as failing in this situation would need closer follow-up in the form of a repeat ultrasound scan in 7 days, irrespective of the risk of an ectopic pregnancy.

The performance of Model M4 depends on the accuracy of the diagnoses for each of the final clinical outcomes. All women with a failing PUL were followed up until their serum hCG levels were *< 5 U/L*. The location of these pregnancies is not known and therefore these were classified as failing PULs and not as failing IUPs or failing ectopic pregnancies. All women with an intrauterine sac confirmed on transvaginal ultrasound examination were rescanned 2 weeks after the follow-up ultrasound examination to confirm the viability of the pregnancy. Thus, the diagnosis of an IUP was scan-based; however, such a non-invasive approach is safe and did not jeopardize the viable ongoing IUPs (76.4%). We have shown previously the limitation of uterine curettage in the characterization of PUL[13]. The gold standard of laparoscopy for the diagnosis of ectopic pregnancy was not applied to all women with a PUL in this study[14,15]. In an ideal PUL study, all women should have the gold standard applied in order to eliminate any diagnostic suspicion bias. The universal use of laparoscopy would give an accurate representation of the prevalence of ectopic pregnancies in a PUL population. However, laparoscopy is invasive, is not without risks and does not confer 100% sensitivity for the detection of ectopic pregnancy[16]. Such an approach in a PUL population is difficult to justify on ethical grounds, especially when the vast majority of PULs are non-ectopic pregnancies[1,2,4,7,13,16–19]. Furthermore, we have demonstrated previously that over 90% of ectopic pregnancies will be visualized using ultrasound prior to surgical intervention[6]. The use of non-invasive diagnostic techniques including serum hormone measurement and transvaginal ultrasound examination were used for all three outcome groups as in previous PUL studies published by our unit[1,3,7,13,16,19].

As we have shown with Model M1[19], we believe that Model M4 can be usefully applied clinically in an early pregnancy setting. The improved detection of ectopic pregnancies compared with current strategies allows clinicians to triage PULs into low and high risk for ectopic pregnancy at 48 h. The concept of classifying PULs into low- and high-risk groups is not new[16]. Currently, women often have multiple visits to the EPU as part of their management. By using the new model in the clinical setting we may be able to classify the majority of PULs at 48 h. Furthermore, the interpretation of the hormone data requires little knowledge or experience of the behavior of hCG in early pregnancy. In this way, less experienced clinicians may be able to use this tool in order to reduce the overall number of visits to the EPU, with a subsequent decrease in financial and labor costs. We are currently performing a clinical interventional study, based on the use of Model M4, which evaluates not only the triaging potential of this model at 48 h but also whether there is a reduction in the number of follow-up visits.

ACKNOWLEDGMENTS

This research is sponsored by the Research Council of the K.U.Leuven (GOA-MEFISTO 666, GOA-AMBioRICS), the Flemish Government (FWO: projects G.0407.02/G.0269.02/G.0360.05, research communities ICCoS & ANMMM), the Belgian Federal Government (DWTC: IUAP IV-02 (1996-2001) and IUAP V-22 (2002-2006)), and the EU (PDT-COIL: contract NNE5/2001/887; BIOPATTERN: contract FP6-2002-IST 508803; eTUMOR: contract FP6-2002-LIFESCIHEALTH 503094).

REFERENCES

1. Condous G, Lu C, Van Huffel S, Timmerman D, Bourne T. Human chorionic gonadotrophin and progesterone levels for the investigation of pregnancies of unknown location. *Int J Gynecol Obstet* 2004; 86: 351–357.
2. 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: 158–163.
3. Condous G, Okaro E, Khalid A, Timmerman D, Lu C, Zhou Y, Van Huffel S, Bourne T. The use of a new logistic regression model for predicting the outcome of pregnancies of unknown location. *Hum Reprod* 2004; 19: 1900–1910.
4. Kadar N, Caldwell BV, Romero R. A method of screening for ectopic pregnancy and its indications. *Obstet Gynecol* 1981; 58: 162–166.

Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.

5. Warren WB, Timor-Tritsch IE, Peisner DB, Raju S, Rosen MG. Dating the early pregnancy by sequential appearance of embryonic structures. *Am J Obstet Gynecol* 1989; 161: 747–753.
6. Condous G, Okaro E, Khalid A, Lu C, Van Huffel S, Timmerman D, Bourne T. The accuracy of transvaginal ultrasonography for the diagnosis of ectopic pregnancy prior to surgery. *Hum Reprod* 2005; 20: 1404–1409.
7. Condous G, Kirk E, Van Calster B, Van Huffel S, Timmerman D, Bourne T. Failing pregnancies of unknown location: A prospective evaluation of the human chorionic gonadotrophin ratio. *BJOG* 2006; 113: 521–527.
8. Agresti A. *An introduction to categorical data analysis*. Wiley & Sons: New York, 1996.
9. Khan KS, Chien PF, Dwarakanath LS. Logistic regression models in obstetrics and gynecology literature. *Obstet Gynecol* 1999; 93: 1014–1020.
10. Hanley JA, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29–36.
11. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 1988; 44: 837–845.
12. Kadar N, Freedman M, Zacher M. Further observations on the doubling time of human chorionic gonadotropin in early asymptomatic pregnancies. *Fertil Steril* 1990; 54: 783–787.
13. Condous G, Kirk E, Lu C, Van Calster B, Van Huffel S, Timmerman T, Bourne T. There is no role for uterine curettage in the contemporary diagnostic work-up of women with a pregnancy of unknown location. *Hum Reprod* 2006; 21: 2706–2710.
14. Ankum WM, Van der Veen F, Hamerlynck JV, Lammes FB. Laparoscopy: a dispensable tool in the diagnosis of ectopic pregnancy? *Hum Reprod* 1993; 8: 1301–1306.
15. Ankum WM, Van der Veen F, Hamerlynck JV, Lammes FB. Transvaginal sonography and human chorionic gonadotrophin measurements in suspected ectopic pregnancy: a detailed analysis of a diagnostic approach. *Hum Reprod* 1993; 8: 1307–1311.
16. Condous G, Okaro E, Khalid A, Lu C, Van Huffel S, Timmerman D, Bourne T. A prospective evaluation of a single visit strategy to manage pregnancies of unknown location. *Hum Reprod* 2005; 20: 1398–1403.
17. Hahlin M, Thorburn J, Bryman I. The expectant management of early pregnancies of uncertain site. *Hum Reprod* 1995; 10: 1223–1227.
18. 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.
19. Kirk E, Condous G, Haider Z, Lu C, Van Huffel S, Timmerman D, Bourne T. The practical application of a mathematical model to predict the outcome of pregnancies of unknown location. *Ultrasound Obstet Gynecol* 2006; 27: 311–315.

Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.