05_36

PREDICTION OF PREGNANCY OF UNKNOWN LOCATION OUTCOME: THE DIFFICULT YET IMPORTANT CASE OF ECTOPIC PREGNANCY

B. Van Calster
K.U.Leuven, Belgium

G. Condous
St. Georges Hosp., UK

S. Van Huffel
K.U.Leuven, Belgium

T. Bourne
St. Georges Hosp., UK

D. Timmerman
K.U.Leuven, Belgium

ABSTRACT

Early Pregnancy Units, transvaginal ultrasound (TVS) and rapid immunoassay of human chorionic gonadotrophin (hCG) and progesterone have increased the number of unruptured ectopic pregnancies diagnosed. The increased use of first trimester TVS, however, has led to more pregnancies of unknown location (PULs). These PULs can represent failing PULs, intra-uterine pregnancies (IUPs) or ectopic pregnancies. Early prediction of the outcome is important, as the high-risk ectopic pregnancy group poses the greatest danger. Different approaches for this prediction, with a focus on ectopic pregnancy, are discussed together with their results. These approaches include threshold rules, subjective interpretation and mathematical models such as multinomial logistic regression and support vector machines. Our results confirm the usefulness of mathematical modelling. The most important variable for the prediction of the outcome of PULs is the change in the hCG level over 48 hours. Finally, difficulties in predicting the ectopic pregnancy group are highlighted. Although smallest in size, it is the most important PUL subgroup. Also, its profile is not distinct from the other groups whilst failing PULs and IUPs are clearly different. Thus, other variables are needed in addition to the hCG ratio for predicting ectopic pregnancy.

Keywords: pregnancy of unknown location (PUL), ectopic pregnancy, multinomial logistic regression, mathematical modelling.

INTRODUCTION

Ectopic pregnancy is the fourth leading cause of direct maternal deaths in the UK (Lewis and Drife, (1)). Between 2000 and 2002, 11 out of 15 first trimester deaths in the UK were due to ectopic pregnancy (1). It is the single most important cause of maternal death during the first trimester of pregnancy (Goldner et al (2)). Because the pregnancy is located outside the uterus, there is a serious risk of rupture and consequent haemorrhage that can be life threatening. An early diagnosis of ectopic pregnancy is therefore very important.

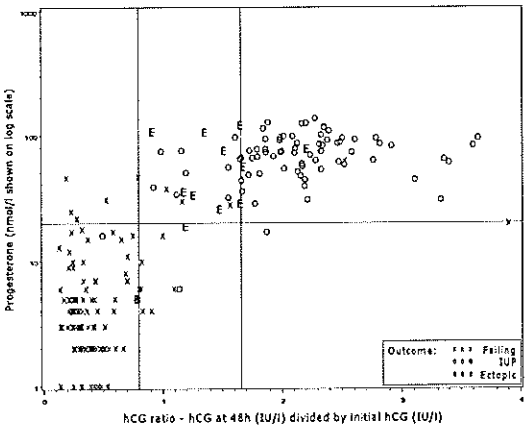
Until recent decades, concern about a possible ectopic pregnancy only rose in pregnant women with severe abdominal pain, irregular vaginal blood loss and even collapse during the first pregnancy trimester. With the introduction of serum chorionic gonadotrophin (hCG) measurements, transvaginal ultrasonography (TVS), and also the uptake of less invasive surgical techniques (e.g. laparoscopy), the prevalence of ruptured ectopic pregnancies has decreased (Kirk et al (3)). In the US, the case-fatality rate fell 90% from 1970 to 1987 (Ectopic pregnancy - US (4)).

The above-mentioned innovations resulted in an earlier detection of ectopic pregnancy and lowered the morbidity and mortality. They also, however, led to an earlier ultrasound evaluation of pregnancies, which in turn increased the number of women with a pregnancy that could not be localised. These pregnancies of unknown location (PULs) can be defined as a situation in which there is a positive pregnancy test with no signs of intra- or extra-uterine pregnancy on TVS, and remnants of miscarriage are absent (Condous et al (5)). Eight to 31% of women presenting at an Early Pregnancy Unit fulfil these criteria (Hahlin et al (6), Banerjee et al (7), Cacciatore et al (8)). Such pregnancies can evolve in four ways. The majority represent spontaneous resolution of pregnancy (failing PUL), which could be either ectopic or intra-uterine, but is never seen on TVS. Another large group of PULs are early intra-uterine pregnancies (IUPs). A third, smaller group of PULs represent ectopic pregnancies. The fourth possibility is very rare, and these can be described as persisting PULs. There is some evidence that such PULs behave biochemically like ectopic pregnancies and almost certainly represent missed ectopic pregnancies (Condous et al (9)). Condous and colleagues found 54% failing PULs, 36% IUPs, 8% ectopic PULs, and 2% persisting PULs in a sample of 388 pregnancies (9). Thus, the overwhelming majority of PULs do not represent ectopic pregnancies.

An expectant, wait-and-see approach has been demonstrated to be safe for women with a PUL that is ectopic (6-8). This approach can be a serious psychological and financial burden, however, and the unpredictable nature of ectopic pregnancies further emphasises the need for early prediction of this subgroup of PULs such that the well-being of this group of women can be guaranteed.

The prediction of the outcome in a PUL population is the greatest challenge in early pregnancy today. Research into this, with a special focus on the prediction of ectopic pregnancy, is the aim of this paper. We focus on numerical and statistical approaches that have been

FIGURE 1 - Scatter plot of serum progesterone and hCG ratio level for 196 PUL cases (9) with reference lines at thresholds suggested by (17-18)



developed and evaluated in order to predict ectopic pregnancies in a PUL population. Also, some issues that hamper an adequate prediction of ectopic pregnancy are discussed. Although the prediction of ectopic pregnancy among PULs has improved, these difficulties have had their influence but also bring about considerations for further research.

CLASSICAL PREDICTION RULES AND PUL

Historically, predictions about the presence or absence of an ectopic pregnancy were made when pregnant women presented with lower abdominal pain or vaginal bleeding. Possible criteria were investigated by several authors (e.g. Kadar et al (10), and Cacciatore et al (11)). Now, more women present at earlier gestations, with no symptoms or signs of ectopic pregnancy. The number of PULs is potentially rising. Prediction of ectopic pregnancy in case of a PUL is therefore becoming increasingly important. Old predictive criteria are not necessarily sufficient for the new group of patients and should be revised. When the population under study consists of women with abdominal pain and/or vaginal bleeding, the a priori probability of the pregnancy being ectopic is higher than when dealing with mainly asymptomatic women with a non-visualised pregnancy.

Some classic predictive rules can be used when a woman is classified with a PUL. A serum hCG level increase of less than 66% over 48 hours would indicate an ectopic pregnancy, while otherwise an IUP should be expected (10); an initial serum hCG level of at least 1000 IU/l in combination with failure to visualise a pregnancy in the uterus would indicate the presence of an ectopic pregnancy (11).

The need to revise these criteria can be shown through the 196 test set PULs of (9). Using an hCG increase of less than 66% to predict ectopic pregnancy gives a sensitivity of 83%, a specificity of 36%, a positive predictive value (PPV) of 8% and a negative predictive value (NPV) of 97%. Far too many non-ectopic PUL cases are predicted to be ectopic. Using the initial hCG-level of at least 1000 IU/l to predict ectopic pregnancy yields a sensitivity of 8%, a specificity of 90%, a PPV of 5% and a NPV of 94%. Only 1 of the 12 ectopic PULs is predicted as such. Condous et al (12) report similar results.

In a recent study by Kohn et al (13), pregnant women with lower abdominal pain or vaginal bleeding and a serum hCG level below 1500 IU/l had an increased risk of ectopic pregnancy but a decreased risk of a normal IUP. Using 1500 IU/l as a threshold on the test set data of (9) to predict ectopic pregnancy yields a sensitivity, specificity, PPV, and NPV of 92%, 4%, 6%, and 89%, respectively.

RECENT PUL OUTCOME RESEARCH

New thresholds

Banerjee et al (14) discovered that a serum progesterone level below 20 nmol/l was predictive of failing PUL. In their data set, they achieved a sensitivity of 93%, a specificity of 94%, a PPV of 97% and a NPV of 86%. Condous and colleagues also reported on the usefulness of serum progesterone to distinguish between the different PUL outcomes (12). They also, however, pointed at a new index: the hCG ratio. This is the ratio of the 48-hour hCG level divided by the initial hCG level. Note that others already looked at a 48-hour hCG increase of 66% (i.e. an hCG ratio of 1.66) (10).

The hCG ratio appeared highly fruitful for distinguishing between all three PUL outcomes, which is obviously an important step forward. Research suggested that an hCG ratio below 0.8 would indicate a failing PUL, an hCG ratio of over 1.66 would indicate an IUP, and an hCG ratio between 0.8 and 1.66 would increase the risk for an ectopic pregnancy (Condous et al (15)). On the test data of (9), this resulted in a sensitivity, specificity, PPV, and NPV of 86%, 97%, 94%, and 91%, respectively, for the prediction of IUP, of 89%, 98%, 98%, and 88% for the prediction of failing PUL, and of 75%, 89%, 31%, 98% for the prediction of ectopic pregnancy. Most ectopic pregnancies will be predicted as such, yet the majority of pregnancies predicted to be ectopic are not ectopic pregnancies. Although this simple threshold is already very good, some failing PULs and IUPs have an hCG ratio between 0.8 and 1.66. From Figure 1, using the test data of (9), one can conclude that the hCG ratio is superior to progesterone as it allows one to better distinguish between an ectopic pregnancy and an IUP.

Single visit prediction

The predictive value of the hCG ratio thresholds is very promising. Yet, it is worthwhile to investigate whether a

single visit could suffice. This would be beneficial for both patient and gynaecologist. Avoiding extra visits may have psychological and time-related advantages for the patients, as well as workload advantages for the staff (Condous et al (16)).

In 2002, a protocol was proposed by Jurkovic (unpublished). It suggested that a progesterone level between 20 and 60 nmol/l in combination with an hCG level over more than 25 IU/l, there should be an increased risk for ectopic pregnancy. Next, a progesterone level exceeding 60 nmol/l together with an hCG level of over 1000 IU/l would also indicate ectopic pregnancy. When applying these rules to the test data of (9), only 6 out of 12 ectopic pregnancies would be predicted as such, while only 14% of the pregnancies predicted to be ectopic were indeed ectopic. These results are less than encouraging.

Condous and colleagues, in a non-interventional prospective trial, also investigated whether a single visit could suffice to determine who is at low or high risk for ectopic pregnancy (16). Those deemed to be at low risk could then be discharged and exempt from unnecessary follow-up. Their training data (n = 200) suggested that patients with a serum progesterone level of at most 10 nmol/l or an hCG level of at most 25 IU/l would have a failing PUL. Women whose serum progesterone level exceeded 50 nmol/l were regarded as having an IUP. Finally, women were at high risk for ectopic pregnancy when the serum progesterone level fell between 10 and 50 nmol/l and the hCG level exceeded 25 IU/l. When tested prospectively on 318 test cases, these rules led to a sensitivity of only 33% for the prediction of a high-risk pregnancy. Specificity was 84%, PPV 9%, and NPV 96%. Again, this is clearly unacceptable. A single visit seems unsafe for predicting the PUL outcome.

Subjective assessment of the biochemical profile

As the expert is still responsible for the management of PULs and algorithms and mathematics serve as an additional aid, it is important to check the accuracy of and inter-observer variability in the subjective interpretation of a PUL. Therefore, Condous et al (17) had several experts with varying experience look at the biochemical profile of women with a PUL.

Five gynaecologists with 3, 3, 6, 12, and 15 years experience were presented with the initial and 48-hour serum progesterone and hCG levels of 185 PULs (102 failing PUL, 63 IUP, 20 ectopic). The clinicians knew that the women were stable and that they had been given the diagnosis of a PUL. They were blind to the PUL outcome and to the predictions of their colleagues. Inter-expert agreement was measured using Cohen's kappa κ and a modified kappa coefficient to measure agreement between more than two assessors (Fleiss (18)). This coefficient incorporates mere chance agreement and varies between 0 (mere chance agreement) to 1 (perfect agreement). Values over 0.60 are substantial; values over 0.81 are very good.

The five experts correctly predicted the outcome in 86 to 88% of the cases. These differences are small and non-significant. Prediction accuracy (cf. sensitivity) was outstanding for failing PUL (89 to 99%) and IUP (87 to 94%). The correct prediction of ectopic cases was poor (25 to 60% with the two most experienced gynaecologists obtaining 40 and 25%, respectively). However, these seemingly large inter-expert differences are only based on 20 ectopic pregnancies. Similarly, inter-observer agreement was very good for non-ectopic pregnancies (κ = 0.86 for failing PUL and 0.89 for IUP) but poor for ectopic pregnancies (κ = 0.50). Finally, for 85% of the non-ectopic pregnancies, all five experts were correct. This was the case for none of the ectopic pregnancies. Actually, for seven of these pregnancies (35%), all five clinicians were wrong.

These findings tell us that experience does not lead to more accurate predictions. Also, subjective assessment may help in judging the viability of a PUL (whether it is a failing PUL or an immediately viable IUP), but is not good for judging the location of the pregnancy (i.e. intra- or extra-uterine) (17). This can be interpreted as a cry for help in the prediction of ectopic pregnancy in the PUL population. The hCG ratio thresholds could serve well here. Another, more flexible aid for gynaecologists could be mathematical modelling.

Can logistic regression help to predict the outcome of a pregnancy of unknown location?

Logistic regression models may provide a very fruitful approach for predicting the outcome of a PUL as such models can use multiple predictors simultaneously. For threshold systems, using more than two measurements becomes cumbersome. Also, logistic regression models give predicted probabilities for each outcome, which is more informative than black and white decisions.

Again, Banerjee and colleagues have done research on this issue with a focus on failing PUL (7). Their sample consisted of 127 women, of which 64 failing PULs (50%), 45 IUPs (36%) and 18 ectopic pregnancies (14%). Variables measured were age, length of gestation (estimate based on last menstrual date), obstetric and gynaecological history variables (such as number of pregnancies, normal deliveries, ectopic pregnancies, abortions, miscarriages, Caesarean sections, and history of pelvic infection), presence of vaginal bleeding, presence of pelvic pain, endometrial thickness, and levels of serum progesterone and hCG.

Using stepwise logistic regression a model was found to predict failing PUL versus another outcome. Their model includes vaginal bleeding (yes/no), endometrial thickness (mm), serum progesterone level (nmol/l) and serum hCG level (IU/l). The area under the ROC curve was as high as 0.95. When using a predicted probability of 0.70 as a cut-off for the prediction of failing PUL,

sensitivity and specificity both were 92%.

No test set was used to test the performance of their model on new data. When using the test data of (9), a sensitivity of 91% and a specificity of 89% are obtained. The PPV and NPV are 91% and 89%, respectively. The usefulness of the hCG ratio, which was not considered in (7) is evident from the fact that the hCG ratio threshold for failing PUL gives values of 89, 98, 98, and 88%, which is better than the logistic regression model.

The major challenge, however, is the prediction of ectopic pregnancy. As shown above, failing PUL and IUP seem to be predicted well, whilst it is more difficult to predict ectopic pregnancies. Predictive models for ectopic pregnancies are the topic of our present research. Condous and colleagues have tried to predict all three outcomes simultaneously using logistic regression (9).

THE PRESENT RESEARCH: THREE-CATEGORY LOGISTIC REGRESSION

The three-category model

Regular logistic regression is suitable for a two-class classification problem. When there are more groups, however, an expansion of the basic model is needed. For k categories, $k-1$ model equations will be made in which one category is compared with a baseline category (Agresti (19)). General significance tests deal with the question whether a predictor is able to make any kind of distinction between the k outcome groups. If so, it can be entered into all model equations. In this way, the Type-I error for the effect of a variable on the outcome can be kept at 5% (i.e. p-value < 0.05 yields significance). If a predictor's effect is tested for each two-class problem separately, each test is performed using a significance level of 0.05 such that the general Type-I error exceeds 5%. With respect to the PUL outcome, two model equations could be constructed comparing failing PUL and ectopic PUL with IUP (reference category), using m predictors:

$$\log\left(\frac{P_{Failing}}{P_{IUP}}\right) = \beta_0 + \beta_1 * Var_1 + \ldots + \beta_m * Var_m$$

$$\log\left(\frac{P_{Ectopic}}{P_{IUP}}\right) = \beta_0' + \beta_1' * Var_1 + \ldots + \beta_m' * Var_m$$

A third equation comparing failing PUL and ectopic pregnancy can be readily derived from the first two. Predicted probabilities for each outcome can also be derived.

These three resulting probabilities, P(failing), P(IUP), and P(ectopic), should now be used to come to a final prediction of the outcome. This is less straightforward as compared to the two-class logistic regression. One approach involves the use of two cut-offs. As ectopic pregnancy is hard to predict, the first cut-off is the probability of ectopic pregnancy above which ectopic pregnancy will be predicted. Focus is then placed on whether the PUL is ectopic or not. When the first cut-off is not met, ectopic pregnancy is not predicted and the second cut-off will be decisive. The second cut-off is the failing PUL probability above which failing PUL instead of IUP is predicted. Both cut-offs are found through the minimisation of a loss function L, for example where

$$L = \sqrt{(1-\text{sensitivity})^2 + (1-\text{specificity})^2},$$

first for the prediction of ectopic pregnancy versus other (cut-off 1), then for the prediction of failing PUL versus IUP (cut-off 2). Thus, sensitivity and specificity are maximised at the same time, but of course other loss functions or decision strategies are possible.

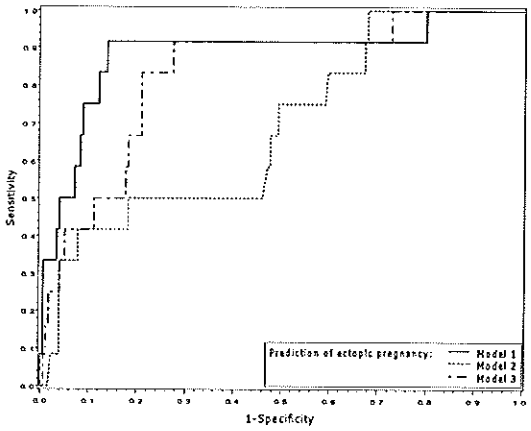
A second approach is based on weighted posterior probabilities (WPP). Outcome prediction is not done in two steps, but only in one as the outcome with the highest probability is chosen. First, however, the predicted probabilities based on the model equations are multiplied by class weights C_i: the outcome i with the highest $C_i * P_i$ will be predicted. The weights represent misclassification costs, arising from the perceived disadvantage of misclassifying a particular PUL outcome as well as from the misclassification rate when the raw (unweighted) probabilities are used. The higher the weight, the more the raw probabilities need to be boosted to avoid a high misclassification rate.

To compute sensitivity and other indices, three dual problems can be considered: failing versus other, IUP versus other, and ectopic pregnancy versus other. As a consequence, three ROC curves can be constructed: one for each class.

Results

The three-category model was applied to a set of 388 women having a PUL (9). A training set was created with 189 of the women, of which 102 with a failing PUL (54%), 63 with an IUP (33%), and 20 with an ectopic pregnancy (11%). Four pregnancies were persisting PULs (2%) and were discarded. Variables measured were the initial and 48-hour serum progesterone (nmol/l) and hCG levels (IU/l), age (years), length of gestation (days), endometrial thickness (mm), result of the midline echo, i.e. ultrasound scan of the middle of the uterine cavity, (intact vs. disrupted), and presence or absence of free fluid in the pouch of Douglas. The hCG ratio was calculated as the 48-hour level divided by the initial level. Also, because both progesterone levels usually are similar, both values were combined by the average progesterone level. As this variable has an extremely

FIGURE 2 - Test data ROC curves for the three logistic regression models from (9)



skewed distribution, the logarithm of the average was used.

The hCG ratio and (the logarithm of) the progesterone average were best able to differentiate between the groups. All three groups differed significantly ($p < 0.0001$ using Mann-Whitney tests). The hCG level of failing PULs decreased 50% over 48 hours (hCG ratio = 0.5), while that of IUPs more than doubled (hCG ratio = 2.2). For ectopic pregnancies, it modestly increased (hCG ratio = 1.1). Serum progesterone was lowest for failing PULs (mean progesterone average = 9.9; mean log average = 1.7), ectopic PULs had a higher level (25.3; 3.1) and IUPs had the highest level (68.7; 4.1). The endometrium was thicker for IUPs while the length of gestation was shorter. Fewer women with failing PULs had free fluid in the pouch of Douglas.

Three interesting models were found using a stepwise variable selection procedure. Model 1 only included the hCG ratio, Model 2 only included the logarithm of progesterone average, and Model 3 included the hCG ratio, the logarithm of progesterone average, and age. Three models were sought because the hCG ratio and the progesterone average are highly correlated while inter-predictor associations can be dangerous for the quality of the model.

Using nonparametric ROC analysis (Hanley and McNeil (20)) and a test data set, the performance of these three models was evaluated. The test data set contained 199 women of which 109 failing PULs (55%), 75 IUPs (38%), and 12 ectopic PULs (6%). Three persisting PUL cases were discarded. Based on previous results that suggested that failing PULs and IUPs can be predicted quite well, it is not surprising that the test data area under the ROC curve for predicting failing PUL varies from 0.97 to 0.99, while that for predicting IUP varies from 0.96 to 0.98. The ectopic pregnancies thus decide which model is to be preferred. Model 1, 2, and 3 achieve an area under the curve of 0.89, 0.69, and 0.84, respectively (see Figure 2). HCG ratio thus performs best while serum progesterone was not good. Model 3 cannot outperform Model 1, possibly as a negative consequence of the high association between hCG ratio and progesterone average. Despite the training data difference in progesterone, the test data and Figure 1 show that this variable is not capable to clearly separate IUPs from ectopic pregnancies.

Using the two cut-off and the WPP systems, it was found that using WPPs with weights 1, 1, and 5 for failing PUL, IUP, and ectopic pregnancy, respectively, gave the best results. The test data sensitivity, specificity, PPV, and NPV for predicting failing PULs were 86%, 99%, 99%, and 85%, respectively. When trying to predict IUPs, 77%, 98%, 95%, and 87% were obtained. Most importantly, for predicting ectopic pregnancy, 92%, 84%, 28%, and 99% were obtained. The fact that an excellent sensitivity is obtained at the cost of predicting several PULs to be ectopic, thus resulting in an insufficient PPV, remains a major limitation. This mathematical model seems to perform better than the hCG ratio threshold system for predicting ectopic pregnancy, but is worse for predicting IUP.

Mathematical models are promising, but it seems that hCG ratio, although a very good predictor of the outcome, is not enough. Other variables seem to be necessary to compensate for the flaws of hCG ratio. It seems that the progesterone level, although informative by itself, cannot fulfil this need. Table 1 shows some results for the prediction of ectopic pregnancy.

At present, we are developing mathematical models using patients for which additional measurements were done to see if the predictive abilities of hCG ratio can be improved (Van Calster (21)).

TABLE 1 - Performance indices for predicting ectopic pregnancy in a PUL population derived from various algorithms

Algorithm to predict ectopic PUL	Sens (%)	Spec (%)	PPV (%)	NPV (%)
hCG rise < 66%	83	36	8	97
hCG > 1000	8	90	5	94
hCG < 1500	92	4	6	89
0.8 ≤ hCG ratio ≤ 1.66	75	89	31	98
10 ≤ prog ≤ 50, hCG > 25	33	24	9	96
Subjective assessment	25-60			
Logistic regression	92	84	28	99

FUTURE WORK

A new approach into the prediction of the outcome of a PUL is the use of non-linear models. Logistic regression is mainly capable of creating a separation between

groups through linear hyperplanes in m-dimensional space where m is the number of predictors used in predicting the outcome. When only one predictor is used for a three-category problem, as is the case when hCG ratio is used to predict failing PUL, IUP, or ectopic pregnancy, this boils down to two lines dividing a one-dimensional space in three parts. The hCG ratio value determines into which part of the space a patient is positioned and thus what prediction is made.

Non-linear models are capable of separating groups by non-linear hyperplanes, which makes this kind of models highly flexible. Some preliminary work using least squares support vector machines (LS-SVM) is already done (Zhou (22)). Support vector machines (Suykens et al (23)) construct linear hyperplanes to separate groups by maximizing the distance of the hyperplane to the nearest points of the two classes (misclassifications are of course tolerated). To do this, only the data points or cases that lie close to the decision boundaries are needed, and these are called the support vectors. Non-linear separations are obtained by first mapping the input data x into a high dimensional feature space using a non-linear mapping $\varphi(x)$. A linear hyperplane is constructed in the feature space, resulting in a non-linear separating hyperplane for the input data.

Another very important step is to test some algorithms prospectively by means of an international multi-centre study. It is obvious that one should test the generalising capabilities of algorithms are and whether the results are reproducible in different centres. This is another step that we would like to take with respect to PUL research.

STATISTICAL PROBLEMS FOR THE PREDICTION OF ECTOPIC PREGNANCY

Centre differences in the proportion of ectopic PULs

In general, it seems that around 10% of all ectopic pregnancies cannot be visualised on ultrasound (14, Shalev et al (24), Condous et al (25)) and, as such, can be considered PULs. However, the detection rate of ectopic pregnancies varies among centres. As Banerjee and colleagues point out, "the proportion of false negative ultrasound findings in ectopic pregnancies varies between centres, depending on the expertise of the operators and the technique of ultrasound examination" (7). Centre differences in the detection of ectopic pregnancies lead to a different proportion of ectopic cases among the PULs. In (6, 7 ,9 , Ankum et al (26), Ankum et al (27)), the percentage of ectopic PULs ranges from 8 to 42%.

Having more ectopic pregnancies in the PUL database may seem a good thing because differences with other groups can be determined more accurately and reliably and contributing variables may reach significance more easily. Varying prevalence of ectopic pregnancies within a PUL population is well described. This is no doubt related to the scanning ability of a given unit. The only way to assess the generalisability of any new model is to see if its predictive power is indeed influenced by a unit's scanning ability. This is one aspect that deserves exploration when we arrive at prospective multi-centre evaluation of our algorithms.

Size of the ectopic PUL group

The number of ectopic PULs is small. In datasets with around 150 to 200 PULs, often no more than 20 ectopic cases are found (7, 9). This is a very counterproductive characteristic of the ectopic PUL group. Several problems arise.

One problem is that the variables measured need to be interpreted with caution as the results may vary from sample to sample. This is illustrated in (9). The 20 ectopic cases in the training data have an average hCG ratio of 1.07. This suggests that the hCG level rises only to a minor degree over 48 hours, i.e. 7%. The 12 test set cases have an average ratio of 1.40, suggesting a moderate increase of 40% over 48 hours. The effect of hCG ratio is computed using an average increase of only 7%. This affects the test set performance. Also, progesterone levels are much lower and more different from the IUP levels in the training set than in the test set. A solution, if feasible, would be to wait for a very large PUL dataset that can be divided into two large subsets for model training and testing.

Another effect is that differences with other groups do not easily reach significance. Variables that could separate the ectopic PULs from failing PULs and/or IUPs may therefore fail to be selected in the logistic regression model. Variables that can separate between failing PULs and IUPs will obtain smaller p-values more easily and therefore the model may be dominated by the separation of these two groups. Fortunately, hCG ratio not only separates failing PULs from IUPs but also ectopic pregnancies from both failing PULs and IUPs.

The small size of the important ectopic PUL groups also affects interpretations of the performance indices like the sensitivity, specificity, PPV, NPV and the area under the ROC curve. Of course these values will have much larger confidence intervals when they deal with the prediction of ectopic pregnancy, but clinicians primarily look at the point estimate rather than the interval estimate (i.e. the value obtained in the sample versus the interval in which the true value probably lies). The obtained values make big jumps that may look like huge improvements. Condous and colleagues obtain a sensitivity of 83% using one particular decision strategy (weighting the predicted probabilities of failing PUL, IUP, and ectopic pregnancy with 1, 1, and 4, respectively) (9). Using a very similar decision strategy (using 1, 1, and 5 as weights) gives a sensitivity of 92%, which looks a lot better. In reality, 10 out of 12 ectopic cases are correctly predicted in the first case, while 11 out of 12 are correctly predicted in the second case. The 9% difference in sensitivity can thus be mere chance. As

FIGURE 3 - Loss function for predicting ectopic pregnancy after three-category logistic regression

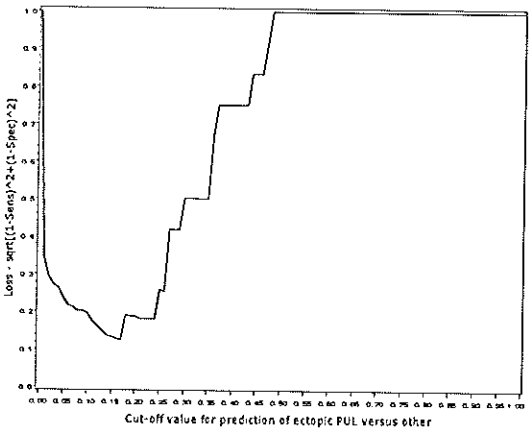
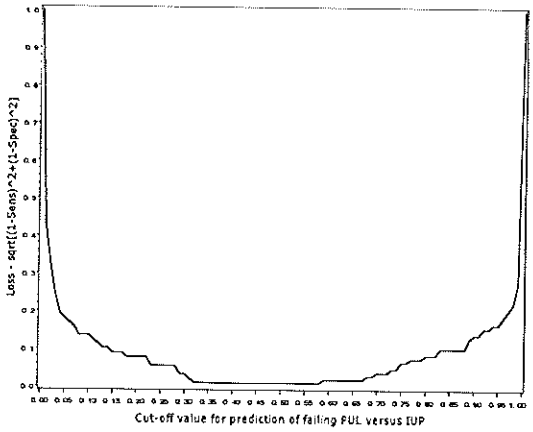


FIGURE 4 - Loss function for predicting failing PUL versus IUP after three-category logistic regression



a consequence the ROC curves for predicting ectopic pregnancy are a lot more edgy than those for predicting another outcome.

A major problem in the prediction of ectopic pregnancy is the low PPV. This is around 30%, which would indicate that 7 out of 10 women who would be informed that they probably have an ectopic pregnancy actually have another type of pregnancy. Obviously, such situations should be avoided as much as possible. It is partly related to the fact that ectopic pregnancies are an underrepresented group in the PUL population such that achieving a high sensitivity includes predicting a lot of ectopic pregnancies. Another reason is related to the problem presented in the next section.

Ectopic PUL: an in-between group?

The ectopic group seems the most difficult group to predict not merely because there are but a few women in that group. For some variables it seems that the ectopic PULs position themselves in between, i.e. between the failing PULs and the IUPs. Most importantly, this seems to be the case for hCG ratio. Failing PULs have a 50% decrease in hCG over 48 hours, IUPs more than double their hCG level, but ectopic PULs seem to be somewhere in the middle with a moderate increase (20% for the 32 ectopic PULs in (9)). The same goes for the serum progesterone level, endometrial thickness, the presence of free fluid in the pouch of Douglas, and the presence of vaginal bleeding (7, 9).

This can be illustrated using the two cut-off system described earlier. When a three-category logistic regression model is found, predicted probabilities can be computed. Then, the ectopic pregnancy probability that serves as a cut-off for predicting ectopic pregnancy is determined through minimizing the loss function L, which leads to the simultaneous maximisation of both sensitivity and specificity. The lower the value of L, the better the cut-off. Figure 3 presents a typical graph of L by the cut-off value for ectopic PUL. This graph has a clear V-shape with a steep descent rapidly followed by a steep rise. The superior cut-offs can be found within a rather small range surrounded by cut-offs with a performance that is clearly worse. Only a few cut-offs stand out. Figure 4 shows a similar graph for cut-off 2, i.e. the failing PUL probability to separate failing PUL from IUP predictions. This graph has a very broad U-shape. Only very high or very low cut-offs perform badly while the large range of intermediate cut-offs all perform very well. This suggests that failing PUL and IUP women have a large predicted probability for their own category but a small one for the other category, i.e. that they are very well separated.

These in-between characteristics may also explain the low PPV that is obtained throughout all algorithms for the prediction of ectopic pregnancy (cf. Table 1). If we want a good sensitivity and thus correctly predict the majority of ectopic pregnancies, the consequence is that quite a lot of non-ectopic pregnancies are nevertheless predicted to be ectopic pregnancies. Thus, as already mentioned before, we need other variables to join hCG ratio in a predictive model, but this is hard when only a few ectopic PULs are available as this can result in overfitting. Many variables may help in separating ectopic pregnancies from other PULs in the training data set. The ectopic group profile in the training set, however, may be biased for a lot of variables, such that a large model, although tuned very well for the training cases, may underperform on new cases as compared to simpler models.

CONCLUSION

Ectopic pregnancy is a potentially dangerous condition, which cannot be underestimated. Although women with a pregnancy of unknown location have a low risk for ectopic pregnancy, it is extremely important to classify them correctly into low and high risk for ectopic

pregnancy in order to arrange safe and appropriate management. The different PUL outcomes need different treatments and have different psychological sequelae. It is difficult to correctly classify PULs as an ectopic pregnancy, as this small subgroup of PULs is not well separated from the other groups. Moreover, centres differ with respect to the proportion of PULs that are ectopic pregnancies, which may influence the generalization of algorithms for prediction. All in all, the road to a correct prediction of ectopic pregnancy is not without his hurdles. To date, a logistic regression using hCG ratio performs best. Its predicted probability of ectopic pregnancy is rather small for the ectopic PULs, such that the ectopic pregnancy cut-off should be small enough (two cut-off system) or its weight should be large enough (WPP system). The resulting PPV will be small. More variables and larger data sets are needed, which is not always feasible in practice.

ACKNOWLEDGEMENTS

We would like to thank the following people for their important contribution to the research on pregnancies of unknown location and ectopic pregnancy: E. Okaro, A. Khalid, C. Lu, and Y. Zhou. 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; eTUMOUR: contract FP6-2002-LIFESCIHEALTH 503094).

REFERENCES

1. Lewis G, Drife J (eds), 2004, "Why mothers die 2000-2002", RCOG Press, London.

2. Goldner TE, Lawson HW, Xia Z, Atrash HK, 1993, MMWR, 42 (SS6), 73-85.

3. Kirk E, Condous G, Bourne T, 2004, Hosp Med, 65, 657-660.

4. Ectopic pregnancy – United States, 1988-1989, 1992, MMWR, 41, 591-594.

5. Condous G, Okaro E, Bourne T, 2002, In D Timmerman, J Deprest, TH Bourne, "Ultrasound and endoscopic surgery in obstetrics and gynecology". Berlin: Springer.

6. Hahlin M, Thorburn J, Bryman I, 1995, Hum Reprod, 10, 1223-1227.

7. Banerjee S, Aslam N, Zosmer N, Woelfer B, Jurkovic D, 1999, Ultrasound Obstet Gynecol, 14, 231-236.

8. Cacciatore B, Ylöstalo P, Stenman UH, Widholm O, 1988, Br J Obstet Gynaecol, 95, 497-502.

9. Condous G, Okaro E, Khalid A, Timmerman D, Lu C, Zhou Y, Van Huffel S, Bourne T, 2004, Hum Reprod, 19, 1900-1910.

10. Kadar N, Caldwell BV, Romero R, 1981, Obstet Gynecol, 58, 162-166.

11. Cacciatore B, Stenman UH, Ylöstalo P, 1990, Br J Obstet Gynaecol, 97, 904-908.

12. Condous G, Okaro E, Khalid A, Zhou Y, Lu C, Van Huffel S, Timmerman D, Bourne T, 2002, Ultrasound Obstet Gynaecol, 20 Suppl 1, 36-37.

13. Kohn MA, Kerr K, Malkevich D, O'Neil N, Kerr MJ, Kaplan BC, 2003, Acad Emerg Med, 10, 119-126.

14. Banerjee S, Aslam N, Woelfer B, Lawrence A, Elson J, Jurkovic D, 2001, Br J Obstet Gynaecol, 108, 158-163.

15. Condous G, Okaro E, Khalid A, Alkatib M, Rao S, Lu C, Van Huffel S, Timmerman D, Bourne T, 2003, Hum Reprod, 18 Suppl 1, 95-96.

16. Condous G, Okaro E, Khalid A, Lu C, Van Huffel S, Timmerman D, Bourne T, Hum Reprod, in press.

17. Condous G, Lu C, Van Huffel S, Timmerman D, Bourne T, 2004, Int J Gynecol Obstet, 86, 351-357.

18. Fleiss JL, 1981, "Statistical methods for rates and proportions", NY, Wiley.

19. Agresti A, 1996, "An introduction to categorical data analysis", NY, Wiley.

20. Hanley JA, McNeil BJ, 1982, Radiology, 143, 29-36.

21. Van Calster B, 2003, Master's thesis, K.U.Leuven, Belgium.

22. Zhou Y, 2002, Master's thesis, K.U.Leuven, Belgium.

23. Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J, 2002, "Least squares support vector machines", Singapore: World Scientific.

24. Shalev E, Yarom I, Bustan M, Weiner E, Ben-Shlomo I, 1998, Fertil Steril, 69, 62-65.

25. Condous G, Okaro E, Khalid A, Lu C, Van Huffel S, Bourne T, Hum Reprod, in press.

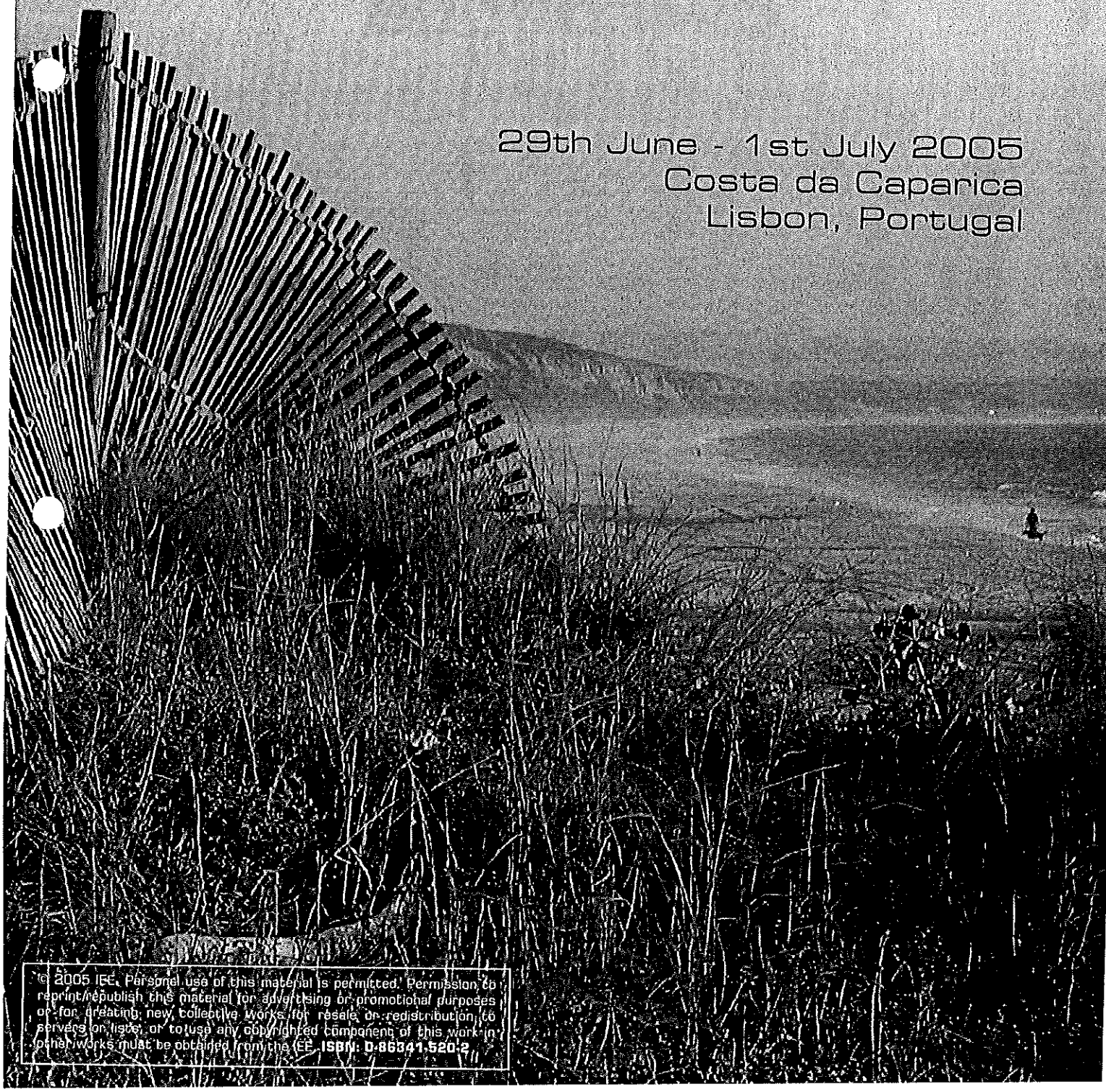
26. Ankum WM, Van der Veen F, Hamerlynck JV, Lammes FB, 1993, Hum Reprod, 8, 1301-1306.

27. Ankum WM, Van der Veen F, Hamerlynck JV, Lammes FB, 1993, Hum Reprod, 8, 1307-1311.

CIMED2005

Proceedings of the 2nd International Conference on Computational Intelligence in Medicine and Healthcare

29th June - 1st July 2005
Costa da Caparica
Lisbon, Portugal

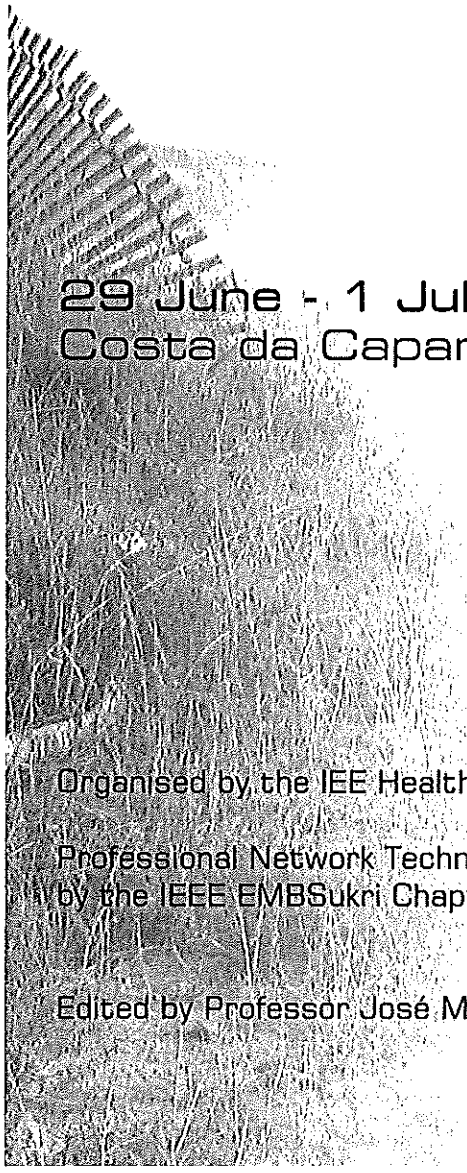


© 2005 IEEE. Personal use of this material is permitted. Permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to use any copyrighted component of this work in other works must be obtained from the IEEE. ISBN: 0-86341-520-2

Proceedings of the

2nd International Conference on Computational Intelligence in Medical and Healthcare

CIMED2005



29 June - 1 July 2005
Costa da Caparica, Portugal

Organised by the IEE Healthcare Technologies Professional Network Technically co-sponsored by the IEEE EMBSukri Chapter

Edited by Professor José Manuel Fonseca