Differentiation between brain metastases and glioblastoma multiforme based on MRI, MRS and MRSI

Jan Luts, Johan A.K. Suykens, Sabine Van Huffel
Katholieke Universiteit Leuven
Department of Electrical Engineering (ESAT), Research Division SCD
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
jan.luts@esat.kuleuven.be

Teresa Laudadio
Istituto Applicazioni Calcolo, CNR, sez. Bari
via G. Amendola 122/D, I-70126 Bari, Italy

Sofie Van Cauter, Uwe Himmelreich
Katholieke Universiteit Leuven
University Hospitals, Radiology Section
Herestraat 49, B-3000 Leuven, Belgium

Enrique Molla
Hospital de La Ribera, Radiology Service
Carretero Alzira-Corbera, Km 1, Alzira 46600 Valencia, Spain

Jose Piquer
Hospital de La Ribera, Neurosurgery Service
Carretero Alzira-Corbera, Km 1, Alzira 46600 Valencia, Spain

M.Carmen Martinez-Bisbal, Bernardo Celda
University of Valencia, Department of Physical Chemistry
CIBER-BBN, ISC-III, Valencia, Spain
C/Dr Moliner, 50, Burjassot, 46100 Valencia, Spain

Abstract

Brain metastases and glioblastoma multiforme are the most aggressive and common brain tumours in adults and they require a different clinical management. Anatomical magnetic resonance imaging (MRI) or clinical history, cannot always clearly distinguish between them. This study describes and verifies the use of magnetic resonance spectroscopy (MRS) and magnetic resonance spectroscopic imaging (MRSI) in combination with MRI for differential diagnosis of glioblastomas and metastases. Feature selection methods are applied to the magnetic resonance (MR) spectra of 121 patients and relevant features are detected. Different classification methods are used to distinguish glioblastoma multiforme and metastasis based on the single-voxel MR spectra, but no reliable differentiation is obtained: the accuracy varies from 50 to 78%. Next, MRSI and MRI data from 10 patients (5 glioblastomas, 5 solitary metastases) are used for differentiation purposes. The combination of multivoxel MR data and MRI data suggests a more clear differentiation between glioblastoma multiforme and brain metastasis. The results are visualized based on nosologic images, which are generated by including spectroscopic information in the segmented MR image. The methodology offers a new way that may support clinicians in decision making.

1. Introduction

Anatomical magnetic resonance imaging (MRI) is widely used in oncology as a non-invasive diagnostic tool to localize tumours and visualize their morphology. However, the ability to differentiate between tumour types is still very limited and, a biopsy is often required for an adequate diagnosis and therapy planning. An important example of the difficult differentiation is represented by brain metastases and glioblastoma multiforme, being the most common

978-1-4244-4878-4/09/$25.00 ©2009 IEEE

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on September 24, 2009 at 05:51 from IEEE Xplore. Restrictions apply.

brain neoplasms in adults. Clinical history and the number of lesions are two important factors to distinguish metastases from glioblastoma. Although solitary brain metastases and glioblastomas are sometimes similar from the radiological point of view, they require different therapeutic decisions.

In recent years, magnetic resonance spectroscopy (MRS) is increasingly being used for more detailed and specific non-invasive evaluation of brain tumours, since it can provide metabolic information on the lesion tissues. By examining the changes in metabolite concentrations, one can better understand the pathology. However, individual inspection and analysis of many spectral patterns is not practical in a clinical setting since it is extremely time-consuming and requires specific spectroscopic expertise. Automated processing and evaluation of the spectroscopic data as well as easy and rapid display of the results are needed for routine clinical interpretation of an exam. As such, different studies used automated pattern recognition methods to distinguish single-voxel magnetic resonance (MR) data from glioblastomas and metastases. Areas under the receiver operating characteristic curve of 0.60 and 0.64 were obtained [4, 13]. The study in [10] considered the lipid peak area ratio (total peak area at 1.3 parts per million (ppm) to that at 0.9 ppm) and obtained a sensitivity of 80% and a specificity of 80%. The absence of creatine (Cr) in the tumoural area was found to most likely indicate metastasis [6]. The area surrounding the enhancing area in glioblastoma multiforme and metastases has also been investigated, since it is known that glioblastomas diffusely infiltrate into the normal brain parenchyma. As such, abnormal choline-to-N-acetylaspartate (Cho/NAA) ratios have been found in the non-enhancing/non-T2 hyperintense area of glioblastomas and not in metastases [3]. The study in [5] found elevated Cho in the tumoural and surrounding region of gliomas. Myo-Inositol (mI) was increased in the enhancing foci of gliomas [5]. Compared to single-voxel MR spectroscopy, magnetic resonance spectroscopic imaging (MRSI) provides metabolite maps of the brain. Elevated Cho/Cr ratios have been found in the peri-enhanced and enhancing region in glioblastomas compared to metastases using MRSI [12].

In this study, we further investigate the differentiation between glioblastomas and metastases using MRS, MRSI and MRI information. Different automated feature selection techniques are applied to the MRS data. Afterwards, the biological relevance of the selected features is verified. In a next step, automated pattern recognition methods are used to classify the MRS data as glioblastoma or metastasis. Furthermore, MRI and MRSI are combined to more reliably differentiate glioblastomas from metastasis. Different metabolite ratios are investigated for differentiation purposes and a procedure to visualize tumour infiltration is established. The results are presented in a nosologic image, which combines the segmented MRI result with the processed spectroscopic data.

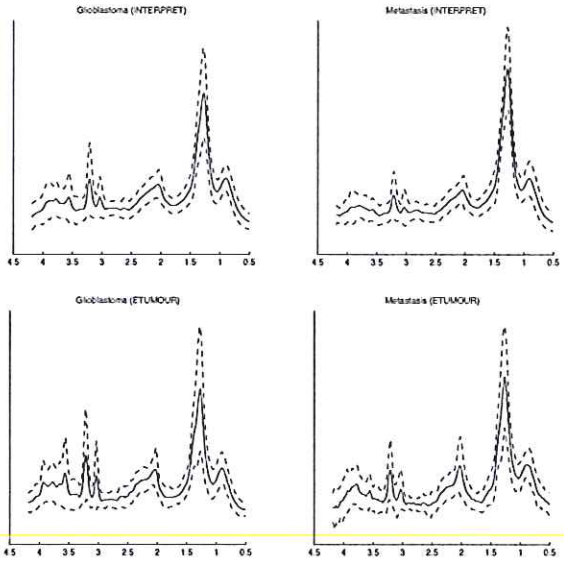
2. Data

The MRS data set contains 121 short echo time (TE) single-voxel 1H MR spectra from untreated patients (84 glioblastomas, 37 metastases), selected from the INTERPRET (2000-2002) database [1] (Figure 1). Data were acquired by 6 contributing centers at 1.5T by stimulated-echo acquisition mode (STEAM) or point resolved spectroscopy (PRESS) sequence with TE between 20 and 32 ms. Repetition time (TR) varied from 1600 to 2020 ms, spectral width was 1000 or 2500 Hz and number of points equaled 512 or 2048. Each spectrum was subjected to a strict validation process set up by the INTERPRET Quality Control committee. Tumour assignment was performed according to the rules of the World Health Organization. Preprocessing included zero order phase correction using Klose's method, removal of residual water resonance by Hankel Lanczos singular value decomposition (HLSVD), baseline correction and L2-norm normalization. Real MR spectra were calculated and used for verifying the biological relevance of the features that were obtained during automated feature selection. These spectra were also used for training classifiers. To verify the performance of the constructed classifiers on independent data, additional test data were acquired by eight international institutions in the framework of eTUMOUR project [2] (2004-2009). Acquisition, processing and validation of the untreated test cases (28 glioblastomas, 32 metastases) is similar to the one of the training set (Figure 1). More details about the MRS data sets can be found in [9].

MRSI and MRI data from 10 untreated patients (5 glioblastomas, 5 metastases) were selected from the eTUMOUR database for this study, all confirmed by histology. Data were acquired at the Radiology Service of Clinica Quirón, Valencia, and the Hospital de La Ribera, Alzira (Spain), as part of the routine clinical preoperative MRI and spectroscopy protocol, before biopsy or surgical resection. Data were acquired in 1.5 T Philips Achieva scanners. T1-weighted images (slice thickness = 5 mm, matrix size = 256 × 256) before and after gadolinium (Gd) injection were acquired prior to the MRSI study. A T2-weighted image package, with the same field-of-view and orientation in plane location of the MRSI slice, was acquired before MRSI. MRSI data were two-dimensional turbo spectroscopic imaging (2D-TSI) data (including water-suppressed and -unsuppressed MR spectra) and were acquired using a 90° - 180° - 180° pulse sequence. The multivoxel spectroscopic images consisted of a grid of 24 × 24 voxels with a field-of-view of 230 × 230 mm and a slice thickness of 20

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on September 24, 2009 at 05:51 from IEEE Xplore. Restrictions apply.

Figure 1. Mean real MR spectra (± standard deviation) for glioblastomas (left) and metastases (right). First row: MRS training data from INTERPRET. Bottom row: MRS test data from ETUMOUR.



mm. Each volume unit dimension was 9.6 × 9.6 × 20 mm (1.8 ml). 2D-TSI data were acquired with long TE (272 ms) and TR = 2000 ms. 256 complex data points were acquired and zero filled to 512 points. Data processing included eddy current correction, frequency alignment and removal of the residual water resonance by using HLSVD-PRO. Magnitude MR spectra were computed and peak integration (PI) was used to calculate an estimate of the concentration of Cho (3.1-3.3 ppm), Cr (2.9-3.1 ppm) and NAA (1.9-2.1 ppm). The concentration of the metabolite corresponds to the area under the metabolite's peak.

3. Methods

3.1. Feature selection analysis

Prior to feature selection, all spectra from the INTERPRET MRS data set were processed by simple PI. Thirteen different metabolite estimates were extracted from the spectra: Cr (3.845-3.995 ppm), glutamine and glutamate + alanine (Glx + Ala) (3.665-3.815 ppm), taurine + glucose (Tau + Glc) (3.345-3.495 ppm), mI + glycine (Gly) (3.450-3.680 ppm), mI + Tau (3.195-3.345 ppm), Cho (3.115-3.265 ppm), Cr (2.955-3.105 ppm), Glx (2.270-2.510 ppm), NAA (1.935-2.085 ppm), Ala (1.395-1.545 ppm), lactate (Lac) (1.235-1.385 ppm), lipids (Lip1) (1.125-1.275 ppm) and Lip2 (0.825-0.975 ppm). Three filter feature selection methods were used to select the most relevant metabolites from the total set of 13: Fisher's criterion, the Kruskal-Wallis test and Relief-F.

Fisher's criterion is a simple measure that is based on the mean of the groups and the within class scatter $\frac{(\mu_1-\mu_2)^2}{(\sigma_1^2+\sigma_2^2)}$, where μ represents the mean, σ indicates the standard deviation, and the subscripts denote the two different classes. All 13 features are sorted based on these scores and ranks are assigned.

The Kruskal-Wallis test is a non-parametric alternative to the well-known one factor independent-samples analysis of variance model (ANOVA). The test statistic is

$$H = \frac{12}{N(N+1)}\left(\sum_{i=1}^{n_C} n_i \overline{R_i}^2\right) - 3(N+1), \overline{R_i} = \frac{1}{n_i}\sum_{j=1}^{n_i} R_{ij}, \quad (1)$$

where n_i denotes the number of samples in group i, n_C is the number of groups, R_{ij} is the rank of case j in group i after ranking all the cases of all groups and N is the total number of samples. The value for the Kruskal-Wallis test is compared with the chi-square statistic with $n_C - 1$ degrees of freedom if the sample size within each group is large enough (e.g. > 5). All features are sorted according to their chi-square values and the ranks are obtained.

Compared to most feature ranking methods, Relief-F does not assume conditional independence of the features [16]. The significance of the features is obtained based on how well they discriminate between samples that are close. The Relief-F algorithm for two classes is structured as explained below.

Input:
r = number of random examples to draw
f = number of features
k = number of near neighbors
x = training data
y = labels of the training data

for $a = 1, \ldots, f$ **do**
  $W_a = 0$
end for

for $t = 1, \ldots, r$ **do**
  Randomly select an instance (x_t, y_t)
  Compute the k nearest h_j with similar y_t value
  Compute the k nearest m_j with different y_t value
  for $a = 1, \ldots, f$ **do**
    $W_a := W_a - \sum_{j=1}^{k} \frac{\text{diff}(a, x_t, h_j)}{rk} + \sum_{j=1}^{k} \frac{\text{diff}(a, x_t, m_j)}{rk}$
  end for
end for

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on September 24, 2009 at 05:51 from IEEE Xplore. Restrictions apply.

20:
21: Output:
22: W = weights indicating quality of the features

If the selected case has different values for an attribute compared to the neighbors of the same class, the weight of that attribute decreases. On the other hand when the attribute value for the specific case is different from the ones of the opposite class neighbors, the weight of the attribute increases. The contribution of all the hit and missing neighbors is averaged in the weight computation rule. The difference measure is defined as $\text{diff}(a, x_1, x_2) = \frac{|x_{1a} - x_{2a}|}{\text{range}(a)}$, where x_{1a} and x_{2a} are respectively the value for feature a for instance x_1 and instance x_2. The feature weights are the output of the Relief-F algorithm and these are used to calculate a rank for each of the variables in the data set.

All three feature selection methods were applied in a bagging analysis. The data set is resampled 50 times and each time the methods assign a rank number to each feature. Afterwards, these ranks are averaged over the 50 runs for each of the methods, resulting into 3 values for each feature. A feature was considered to be relevant if at least two out of three filter methods obtained an averaged rank of less than 6.

3.2. Classification

Simple linear discriminant analysis (LDA) and least squares support vector machines (LS-SVMs) were employed for classification purposes. The least squares support vector machine classifier optimizes $\min_{w,e,b} \mathcal{J}(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2$ subject to $y_i(w^T \varphi(x_i) + b) = 1 - e_i,\ i = 1, ..., N$, with x_i a training case with corresponding class label $y_i \in \{-1, +1\}$, $e = [e_1 \dots e_N]^T$ a vector of error variables to tolerate misclassifications, $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^{d_h}$ a mapping from the input space into a high-dimensional (e.g. infinite dimensional) feature space of dimension d_h, w a vector of the same dimension as φ, γ a positive regularization constant and b a bias term [8]. The primal problem is expressed in terms of the feature map, the dual problem in terms of the kernel function. The resulting classifier in the dual space is given by $y(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\right)$, with K the kernel function (linear or radial basis function (RBF)) with elements of the kernel matrix $K(x_i, x) = \varphi(x_i)^T \varphi(x)$ and α_i the Lagrange multipliers.

The peak-integrated features that were considered to be relevant based on the feature selection analysis from the previous section were used as input for the classifiers. In addition, for some classifiers all 13 peak-integrated features were used as input, while independent component analysis (ICA) was used to extract features from the real spectra for others. The INTERPRET training set was split in a set for training (2/3) and a set for testing (1/3) based on stratified repeated random sampling. The area under the receiver operating characteristic curve was used for performance evaluation (e.g. selection of the number of independent components for ICA) on the set for testing. Furthermore, decision thresholds were computed based on stratified repeated random sampling using the error rate (ER) and balanced error rate (BER). Finally, the performance of all constructed classifiers was verified based on independent eTUMOUR data.

3.3. Metabolite analysis and nosologic imaging

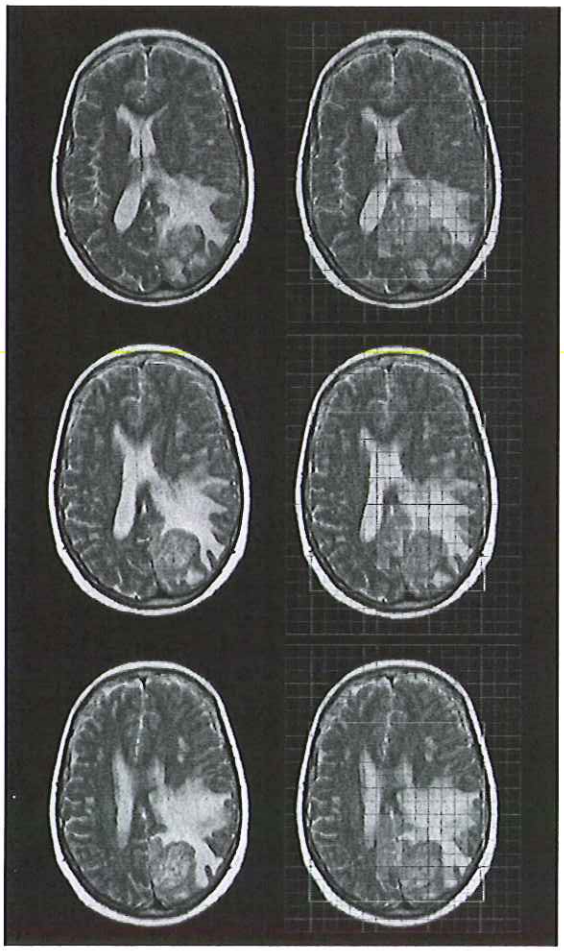
The analysis of the MRSI and MRI data to distinguish between the tumour types involves two approaches: the analysis and comparison of the metabolite ratios and the creation of nosologic images. For the analysis of the metabolites the Cho/Cr and Cho/NAA ratios were computed for every voxel of the MRSI grid. In addition for each patient, the Cho/Cr and Cho/NAA ratios belonging to the voxels of the enhancing tumour, peri-enhancing area (i.e. hyperintense on T2-weighted images, but not enhancing on post-Gd T1-weighted images) and the non-enhancing/non-T2 hyperintense region were averaged within each of those regions. Afterwards, these values were compared in order to differentiate glioblastomas and metastases.

A nosologic image summarizes the presence of different tissues and lesions in a single image by color coding each voxel according to the tissue type it is assigned to. The creation of the nosologic images in the present study involves two steps: abnormal tissue recognition based on MRSI data and MRI segmentation based on the approach in [14]. For abnormal tissue recognition the method by Otsu was applied to the Cho/NAA map of each patient [15]. This automatic threshold method calculated an optimal threshold, such that voxels with high Cho/NAA ratios (originating from abnormal tissue) were separated from voxels with low Cho/NAA ratios (originating from normal tissue). The goal of the method is to find a threshold that minimizes the within-class variance, which has been shown to be equivalent to maximizing the between-class variance. Using σ_T^2 for the total variance and σ_B^2 for the between-class variance, the function $t^* = \arg\max_t \frac{\sigma_B^2}{\sigma_T^2}$ is maximized, where t^* is the optimal threshold. Once t^* is fixed, the final output of the threshold method is a binary map, indicating voxels with normal and abnormal Cho/NAA ratios (Figure 2).

In the second step an MRI segmentation method is used to detect the abnormal tissue, visible on the T2-weighted MR image (i.e. the peri-enhancing area) [14]. First, a registered digital brain atlas and a subject-specific abnormal tissue prior, obtained by applying canonical correlation analysis to the MRSI data [11], were used for the seg-

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on September 24, 2009 at 05:51 from IEEE Xplore. Restrictions apply.

Figure 2. Left: T2-weighted MR images included inside the MRSI volume for patient with metastasis. Right: Binary map created by application of Otsu threshold to the Cho/NAA ratio map. The MRSI grid is superimposed to the MR images. Voxels with normal and abnormal Cho/NAA ratio are indicated in blue and red, respectively. The red voxels fall inside the hyperintense area on the T2-weighted MR images, suggesting no infiltration.



mentation of the MRI data. The segmentation procedure consists of two stages (i.e. detection of abnormality and reclassification with spatial and geometric constraints):

1: {Detection of abnormality}
2: **loop**
3: Threshold posterior (or atlas probabilities in first iteration) probabilities to sample from high confidence regions
4: Outlier detection based on minimum covariance determinant estimator to remove samples from normal tissue
5: Compute probability density functions for all tissue classes using a non-parametric model (i.e. kernel expansions)
6: Apply Bayes' rule to calculate posterior probabilities
7: Perform bias field correction
8: **end loop**
9: {Reclassification with spatial and geometric constraints}
10: **loop**
11: Threshold posterior probabilities to sample from high confidence regions
12: Tumour samples are pruned using region competition snakes
13: Compute probability density functions for all tissue classes using a non-parametric model (i.e. kernel expansions)
14: Apply Bayes' rule to calculate posterior probabilities
15: Perform bias field correction
16: **end loop**

As a result, the segmentation step produced a segmented MR image that separated abnormal tissue from gray matter, white matter and cerebrospinal fluid(CSF) [14].

Finally, to create the nosologic image the results from using Otsu's method and the MRI segmentation are combined. The binary map, obtained by applying Otsu's threshold method to the Cho/NAA ratio estimated from the MRSI data, was superimposed to the segmented abnormal tissue region from MRI. Abnormal voxels of the binary map which fell outside the lesion obtained by the segmentation step, corresponded to infiltration, which is expected for glioblastomas but not for metastases.

4. Results

Table 1 presents the averaged ranks, which are assigned by feature selection methods, for each feature. Averaged ranks below 6 and features, denoted as relevant by the methods for distinguishing glioblastomas and metastases are printed in boldface in Table 1: mI + Gly, Cho, Cr, Glx, Lac and Lip1.

The performance of ten different classifiers (including three classifiers that take the six relevant extracted features as input), trained on the INTERPRET MRS data and tested on the independent eTUMOUR MRS data is depicted in Table 2. The error rate on the independent test data varies from 0.5 to 0.22, indicating no reliable separation of glioblastomas and metastases.

Table 3 summarizes the Cho/Cr and Cho/NAA ratios, which were averaged over all patients with the same type

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on September 24, 2009 at 05:51 from IEEE Xplore. Restrictions apply.

Table 1. Averaged ranks obtained by feature selection methods (R-F=Relief-F, FC=Fisher's criterion, KW=Kruskal-Wallis test).

	Cr	Glx + Ala	Tau + Glc	mI + Gly	mI + Tau	Cho	Cr	Glx	NAA	Ala	Lac	Lip1	Lip2
R-F	11.080	9.0200	11.160	9.2800	7.6800	4.4800	7.5000	5.9200	5.6200	3.6600	4.3400	5.3600	5.9000
FC	10.680	8.9200	8.1400	3.9800	6.0400	4.2800	5.5000	3.3600	8.9800	9.3200	3.7400	6.5400	11.520
K-W	8.6200	9.5600	9.7000	4.7400	7.7200	5.6200	2.7400	3.6000	7.5600	10.540	4.9200	5.1400	10.540

Table 2. Error rate (ER) and balanced error rate (BER) on training (T) data and independent (I) test data for different classifiers (FE = feature extraction method and number of extracted features, C = classifier, CO = cutoff probability, L = linear kernel, R = RBF kernel).

FE	C	CO	ER(T)	BER(T)	ER(I)	BER(I)
PI6	LDA	0.66	0.28	0.42	0.43	0.41
PI6	LSSVM(L)	0.48	0.30	0.46	0.45	0.42
PI6	LSSVM(R)	0.43	0.26	0.27	0.4	0.39
ICA1	LDA	0.59	0.31	0.5	0.5	0.48
ICA1	LDA	0.65	0.31	0.5	0.45	0.42
PI13	LS-SVM(R)	0.5	0.22	0.28	0.4	0.38
PI13	LS-SVM(R)	0.55	0.23	0.33	0.4	0.37
PI13	LDA	0.55	0.37	0.40	0.22	0.21
PI13	LDA	0.6	0.33	0.40	0.22	0.21
PI13	LDA	0.65	0.35	0.45	0.25	0.24

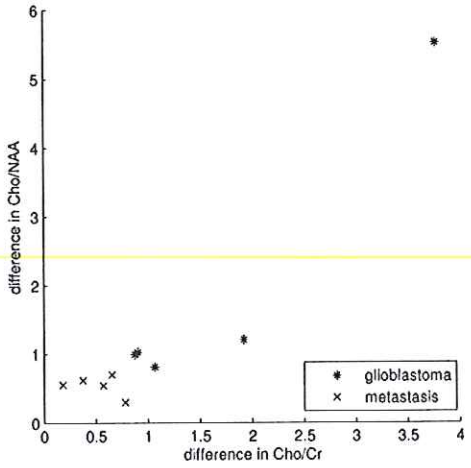
Table 3. Mean and standard deviation of Cho/Cr and Cho/NAA in non-enhancing/non-T2 hyperintense region (I), peri-enhancing area (II), enhancing tumour (III) for glioblastomas (GBM) and metastases (MET).

		I	II	III
Cho/Cr	GBM	1.42±0.19	2.02±0.73	3.13±1.22
	MET	1.28±0.18	1.18±0.15	1.80±0.24
Cho/NAA	GBM	0.48±0.06	1.08±0.89	2.39±2.05
	MET	0.43±0.08	0.54±0.08	0.97±0.13

of tumour, for the three regions. The averaged Cho/Cr and Cho/NAA ratios of glioblastomas are higher in the peri-enhancing area and the enhancing tumour as compared to metastases. In addition, for Figure 3 the difference (i.e. a inter-patient normalization) between the averaged ratio from the enhancing tumour and the non-enhancing/non-T2 hyperintense area was calculated for each patient. A clear separation between metastases and glioblastomas was obtained when plotting this difference measure for Cho/Cr versus Cho/NAA.

Figure 4 presents the nosologic images for two cases (glioblastoma and metastasis). Abnormal tissue from MRI is indicated in red, green represents CSF, light blue white matter and dark blue gray matter. Voxels with abnormal Cho/NAA ratios inside the normal appearing area on T2-

Figure 3. Difference between enhancing tumour and the non-enhancing/non-T2 hyperintense area for Cho/Cr and Cho/NAA. Glioblastomas and metastases are separated.



weighted MRI, are colored in white. A white, infiltrated, area is found in the glioblastoma and not in the metastasis (where the remaining white pixels are due to partial volume effects from voxels partially situated in enhancing or peri-enhancing areas), allowing a visual differentiation. Spectra obtained with MRSI are depicted and show higher Cho/NAA for glioblastoma and smaller Cho/NAA for the metastasis.

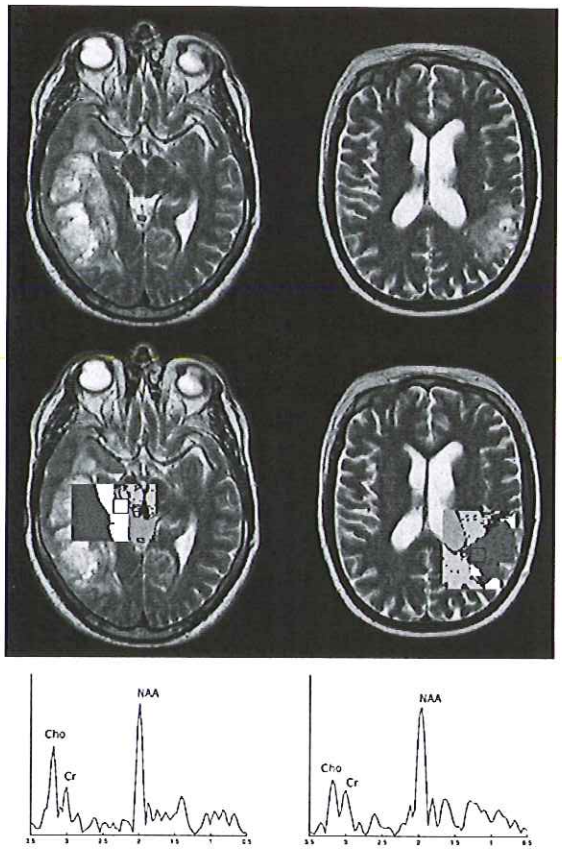
5. Discussion and conclusion

Features that were denoted as relevant by the filter methods, based on the single-voxel MR spectra, are known to be important characteristics for distinguishing glioblastomas and metastases. It is known that lipids have characteristics that might allow discrimination between metastases and glioblastomas [10]. Lac is frequently elevated for glioblastomas. The same holds for Cho and mI [5], whereas Glx was reduced in [5]. Absence of creatine (Cr) indicates metastasis [6]. Similarly to the studies in [4, 13], no reliable differentiation between glioblastomas and metastases

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on September 24, 2009 at 05:51 from IEEE Xplore. Restrictions apply.

was obtained based on pattern recognition using the single-voxel MR spectra. The error rate on the independent test set varied from 0.5 to 0.22.

Figure 4. Left: glioblastoma. Right: metastasis. First row: T2-weighted MRI. Middle row: nosologic image. Bottom row: MR spectra from voxels indicated by a black box on the nosologic image. White, infiltrated voxels are detected for the glioblastoma, and not for the metastasis.



While single-voxel MR spectroscopy provides metabolic information from a single location in the brain, MRSI enables to generate metabolite maps of the whole brain. The advantage of the 2D-TSI approach is that the entire tumour, the peri-enhancing region and the non-enhancing/non-T2 hyperintense brain can be sampled rapidly ($\pm$ 6 minutes) with relatively high spatial resolution. This enables to identify the metabolically most active portion of the tumour with higher likelihood compared to single-voxel MR spectroscopy, where voxel position is essential, but often subjective.

The combination of MRI and MRSI (i.e. multivoxel data) was studied and a more reliable separation of glioblastomas and metastases was obtained. Differences in averaged Cho/Cr and averaged Cho/NAA were found in the enhancing tumour and peri-enhancing area between metastases and glioblastomas. The increase of the ratios in the peri-enhancing area is a result of the infiltrative nature of glioblastomas. These findings agree with the conclusions in [5, 12]. By combining the information of both ratios a clear separation between the glioblastomas and metastases was obtained. Compared to the previous approaches, this study also focussed on the visualization of tumour infiltration, which is typical for glioblastomas but not for metastases. Nosologic images were created, offering a more realistic view of abnormal tissue regions compared to previous approaches, and provide support for clinicians in decision making. Future work will focus on the extension to 3D-MRSI and combination with diffusion-weighted MRI, diffusion tensor imaging and perfusion-weighted MRI [7, 12, 17].

Acknowledgements

This research is funded by a PhD grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen); The work of TL is supported by the Short-Term Mobility Program 2008 of the National Research Council of Italy; Research supported by Research Council KUL: GOA-AMBioRICS, PhD/postdoc & fellow grants, Center Of Excellence Optimisation, IDO 05/010 EEG-fMRI, GOA/2004/05; FWO: G.0360.05, G.0519.06, G.0341.07, G.0321.06, G.0302.07, G.0566.06; ICCoS; ANMMM; UH acknowledges the support by the KUL Center Of Excellence MoSAIC; The Medical Imaging Center (KUL) and Biomedical Informatics Group (IBIME, ITACA Institute) of the Technical University of Valencia are gratefully acknowledged; Ministerio de Educación y Ciencia del Gobierno de España (SAF2007-6547, SAF2007-29455-E, SAF2007-29394-E, SAF2007-29393-E); Belgian Federal Government: IUAP IV-02, IUAP V-22, IUAP P6/04; EU: eTUMOUR (contract no. FP6-2002-LIFESCIHEALTH 503094), HealthAgents (contract no. FP6-2005-IST 027213), FAST (contract no. FP6-019279-2), ESA: Cardiovascular Control (Prodex-8 C90242).

References

[1] http://azizu.uab.es/interpret/ (accessed: 3 april 2009).
[2] http://www.etumour.net/ (accessed: 3 april 2009).
[3] A. Di Costanzo, T. Scarabino, F. Trojsi, T. Popolizio, D. Catapano, G.M. Giannatempo, S. Bonavita, M. Portaluri, M. Tosetti, V.A. d'Angelo, U. Salvolini, and G. Tedeschi. Proton MR spectroscopy of cerebral gliomas at 3T: spatial heterogeneity, and tumor grade and extent. *European Radiology*, 18:1727–1735, 2008.
[4] A. Devos, L. Lukas, J.A.K. Suykens, L. Vanhamme, A.R. Tate, F.A. Howe, C. Majos, A. Moreno-Torres, M. van der Graaf, C. Arus, and S. Van Huffel. Classification of brain

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on September 24, 2009 at 05:51 from IEEE Xplore. Restrictions apply.

tumours using short echo time 1H MR spectra. *Journal of Magnetic Resonance*, 170:164–175, 2004.

[5] G. Fan, B. Sun, Z. Wu, Q. Guo, and Y. Guo. In vivo single-voxel proton MR spectroscopy in the differentiation of high-grade gliomas and solitary metastases. *Clinical Radiology*, 59:77–85, 2004.

[6] H. Ishimaru, M. Morikawa, S. Iwanaga, M. Kaminogo, M. Ochi, and K. Hayashi. Differentiation between high-grade glioma and metastatic brain tumor using single-voxel proton MR spectroscopy. *European Radiology*, 9:1784–1791, 2001.

[7] I.C. Chiang, Y.-T. Kuo, C.-Y. Lu, K.-W. Yeung, W.-C. Lin, F.-O. Sheu, and G.-C. Liu. Distinction between high-grade gliomas and solitary metastases using peritumoral 3-T magnetic resonance spectroscopy, diffusion, and perfusion imagings. *Neuroradiology*, 46:619–627, 2004.

[8] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002.

[9] J.M. Garcia-Gomez, J. Luts, M. Julia-Sape, P. Krooshof, S. Tortajada, J. Vicente Robledo, W. Melssen, E. Fuster-Garcia, I. Olier, G. Postma, D. Monleon, A. Moreno-Torres, J. Pujol, A.-P. Candiota, M.C. Martinez-Bisbal, J.A.K. Suykens, L. Buydens, B. Celda, S. Van Huffel, C. Arus, and M. Robles. Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 22:5–18, 2009.

[10] K.S. Opstad, M.M. Murphy, P.R. Wilkins, B.A. Bell, J.R. Griffiths, and F.A. Howe. Differentiation of metastases from high-grade gliomas using short echo time 1H spectroscopy. *Journal of Magnetic Resonance Imaging*, 20:187–192, 2004.

[11] T. Laudadio, M.C. Martinez-Bisbal, B. Celda, and S. Van Huffel. Fast nosological imaging using canonical correlation analysis of brain data obtained by two-dimensional turbo spectroscopic imaging. *Nuclear Magnetic Resonance in Biomedicine*, 21:311–321, 2008.

[12] M. Law, S. Cha, E.A. Knopp, G. Johnson, J. Arnett, and A.W. Litt. High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. *Radiology*, 222:715–721, 2002.

[13] L. Lukas, A. Devos, J.A.K. Suykens, L. Vanhamme, F.A. Howe, C. Majos, A. Moreno-Torres, M. van der Graaf, A.R. Tate, C. Arus, and S. Van Huffel. Brain tumor classification based on long echo proton MRS signals. *Artificial Intelligence in Medicine*, 31:73–89, 2004.

[14] J. Luts, T. Laudadio, A.J. Idema, A.W. Simonetti, A. Heerschap, D. Vandermeulen, J.A.K. Suykens, and S. Van Huffel. Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. *Nuclear Magnetic Resonance in Biomedicine*, 22:374–390, 2009.

[15] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, 1979.

[16] M. Robnik-Sikonja and I. Kononenko. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*, 53:23–69, 2003.

[17] S.J. Price, R. Jena, N.G. Burnet, P.J. Hutchinson, A.F. Dean, A. Pena, J.D. Pickard, T.A. Carpenter, and J.H. Gillard. Improved delineation of glioma margins and regions of infiltration with the use of diffusion tensor imaging: An image-guided biopsy study. *American Journal of Neuroradiology*, 27:1969–1974, 2006.

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on September 24, 2009 at 05:51 from IEEE Xplore. Restrictions apply.

SISTA's Internal Report Wizard 6/6

Warning: No topics?

Print this page, and give the result to Ida Tassens, together with a hardcopy or reprint of your publication.

Publist Entry

Change existing entry

Number : 08-217
Type : J
Authors : Kirk E., Papageorghiou A.T., Van Calster B., Condous G., Cowans N., Van Huffel S., Timmerman D., Spencer K., Bourne T.
Title : The use of serum inhibin A and activin A levels in predicting the outcome of 'pregnancies of unknown location'
Submitted:
Accepted :
Published: \emph{Human Reproduction}, vol. 24, no. 10, Oct. 2009
Pages : 2451-2456
Filename :
Topics :
Sponsors : GOA-AMBioRICS, CoE-EF/05/006(OPTEC), IDO-05/010 EEG-fMRI, IOF-KP06/11, FWO-G.0407.02, FWO-G.0360.05, FWO-G.0519.06, FWO-G.0321.06, FWO-G.0341.07, FWO-ICCoS, FWO-ANMMM, IAP-P6/04, EU-BIOPATTERN(FP6-2002-IST508803), EU-e-TUMOUR(FP6-2002-LIFESCIHEALTH-503094), EU-Healthagents (IST2004-27214), EU-FAST (FP6-MC-RTN-035801), ESA-Prodex-8-C90242,

This data looks like this if formatted:

08-217

Kirk E., Papageorghiou A.T., Van Calster B., Condous G., Cowans N., Van Huffel S., Timmerman D., Spencer K., Bourne T., ``The use of serum inhibin A and activin A levels in predicting the outcome of 'pregnancies of unknown location'", *Human Reproduction*, vol. 24, no. 10, Oct. 2009, pp. 2451-2456.

The program will now attempt to send an email to ida.tassens@esat.kuleuven.be. If no confirmation is displayed below, the attempt failed!! Always send in a hardcopy of this page, even if the mail got sent!!!

CONFIRMED: The email was succesfuly sent to ida.tassens@esat.kuleuven.be. Please keep handing in your hard-copies to Ida Tassens to make sure that nothing gets lost.

Comments can be mailed to SISTA publications master. This page was generated on Sep 23, 2009.

Human Reproduction, Vol.24, No.10 pp. 2451–2456, 2009
Advanced Access publication on June 23, 2009 doi:10.1093/humrep/dep066

human reproduction ORIGINAL ARTICLE *Early pregnancy*

The use of serum inhibin A and activin A levels in predicting the outcome of 'pregnancies of unknown location'

Emma Kirk[1,8], Aris T. Papageorghiou[2], Ben Van Calster[3], George Condous[4], Nicholas Cowans[5], Sabine Van Huffel[3], Dirk Timmerman[6], Kevin Spencer[5], and Tom Bourne[1,6,7]

[1]Early Pregnancy and Gynaecological Unit, St George's University of London, Cranmer Terrace, London SW17 0RE, UK [2]Fetal Medicine Unit, St George's University of London, London, UK [3]Department of Electrical Engineering (ESAT-SISTA), K.U. Leuven, Leuven, Belgium [4]Early Pregnancy and Advanced Endosurgery Unit, Nepean Hospital, Nepean Clinical School, University of Sydney, Sydney, Australia [5]Department of Clinical Biochemistry, King George Hospital, Goodmayes, Essex, UK [6]Department of Obstetrics and Gynaecology, University Hospital Gasthuisberg, K.U. Leuven, Leuven, Belgium [7]Imperial College London, Hammersmith Campus, Du Cane Road, London, W12 0HS, UK

[8]Correspondence address. E-mail: ejork@hotmail.co.uk

BACKGROUND: The aim was to evaluate the role of serum inhibin A and activin A in the prediction of the outcome of women with 'pregnancies of unknown location' (PULs).

METHODS: Serum human chorionic gonadotrophin (hCG), progesterone, inhibin A and activin A levels were measured at 0 and 48 h. Differences in the mean levels and the change in levels over 48 h expressed as a ratio (48/0 h) were examined between the three outcome groups—failing PUL, intrauterine pregnancy (IUP) or ectopic pregnancy. Variables were incorporated into logistic regression models to predict the pregnancy outcomes, which were evaluated using receiver operator characteristic curves.

RESULTS: One hundred and forty-one women were classified as PULs: 67 failing PULs (47.5%), 58 IUPs (41.1%) and 16 ectopic pregnancies (11.4%). Activin A levels were not significantly different between the three outcome groups. Inhibin A levels were significantly lower in failing PULs. The logistic regression model based on serum inhibin levels gave an area under the curve (AUC) of 0.88 for failing PUL, 0.87 for IUP and 0.60 for ectopic pregnancy. The model based on serum activin levels gave an AUC of 0.61 for failing PUL, 0.64 for IUP and 0.51 for ectopic pregnancy, and the model based on serum hCG levels gave an AUC of 0.95 for failing PUL, 0.97 for IUP and 0.67 for ectopic pregnancy.

CONCLUSIONS: Serum activin A levels at 0 and 48 h are not helpful in predicting the outcome of PULs. Although serum inhibin A levels may be of use in the prediction of failing PULs and IUPs in the PUL populations, they do not perform as well as serum hCG levels.

Key words: activin A / inhibin A / pregnancy of unknown location / human chorionic gonadotrophin / ectopic pregnancy

Introduction

A 'pregnancy of unknown location' (PUL) is defined by the finding of an empty uterus with no signs of an intrauterine or extrauterine pregnancy on a transvaginal ultrasound scan (TVS) in a woman with a positive pregnancy test. There are three possible final clinical outcomes—a failing PUL [serum human chorionic gonadotrophin (hCG) levels spontaneously decrease to <5 IU/l], an intrauterine pregnancy (IUP) or an ectopic pregnancy. At present there is no universally accepted method for predicting the final pregnancy outcome. Previous studies have shown that PULs occur in about 10% of women attending an early pregnancy unit (EPU). Due to the uncertainty of the outcomes and the risk of an underlying ectopic pregnancy, these women require serial TVS or serum tests in order to establish a diagnosis; they therefore represent a disproportionate amount of workload for EPUs, and women with this diagnosis are at risk of unnecessary diagnostic laparoscopy.

Measurements of serum progesterone and serum hCG have been used to predict the pregnancy outcome. A serum progesterone of <20 nmol/l has been shown to be predictive of failing pregnancy (Banerjee *et al.*, 2001). Single serum hCG levels have been used to describe a discriminatory zone—a level of serum hCG above which it is thought that if present, an IUP should be visualized on TVS (Romero *et al.*, 1985). Serial measurements of serum hCG are also

© The Author 2009. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org

often used to help predict the final pregnancy outcome. Kadar *et al.* (1981) first described the minimal rate of increase in serum hCG for an IUP to be 66% in 2 days. Since then various cut-offs have been suggested to predict the different pregnancy outcomes (Barnhart *et al.*, 2004a, b; Seeber *et al.*, 2006; Silva *et al.*, 2006). The change in serum hCG over 48 h has also been expressed as the hCG ratio (hCG 48 h/hCG 0 h) and used to predict the pregnancy outcome, both by itself and when incorporated into mathematical models (Condous *et al.*, 2004, 2006, 2007a, b; Kirk *et al.*, 2006).

Although much of the work on the prediction of outcome of PULs has centred on the use of serum hCG and progesterone levels, some other markers have been studied. These include creatine kinase and cancer antigen 125 (CA125). Although CA125 levels and creatine kinase levels alone were not able to predict the pregnancy outcome, the CA125 ratio was found to be able to help in distinguishing failing PULs from IUPs, but was not able to detect those at risk of ectopic pregnancy (Condous *et al.*, 2005).

Inhibin A is a glycoprotein secreted by the feto-placental unit. It has been suggested that it may be a specific marker of early pregnancy loss (Muttukrishna *et al.*, 2002). Both single and serial measurements have been used to predict subsequent pregnancy loss in women with recurrent miscarriage (Muttukrishna *et al.*, 2002; Al-Azemi *et al.*, 2003; Prakash *et al.*, 2005). Inhibin A has also been used to predict the pregnancy outcome in women undergoing *in vitro* fertilization where it has been shown to be more accurate than hCG in predicting preclinical pregnancy loss (Hauzman *et al.*, 2004). In addition, serum inhibin levels have been shown to be lower in spontaneously conceived ectopic pregnancies compared with intrauterine pregnancies (Seifer *et al.*, 1996).

Activin A is another glycoprotein secreted by the feto-maternal unit, and it is thought to be a marker of trophoblastic invasion with the potential to identify patients at risk of miscarriage and ectopic pregnancy. Both single and serial measurements of activin A levels have been used to predict miscarriage in women with recurrent miscarriage (Prakash *et al.*, 2005). Levels were significantly lower at 7–8 weeks of gestation in those who subsequently miscarried compared with those with ongoing pregnancies. Serum activin A levels have been used to predict ectopic pregnancy in women with PULs (Florio *et al.*, 2007). Serum levels were found to be significantly lower in women with ectopic pregnancies compared with those with IUPs or failing PULs. Using a cut-off value of 0.37 ng/ml gave a sensitivity and specificity of 100 and 99.6%, respectively, for the prediction of ectopic pregnancy (Florio *et al.*, 2007).

The aim of this study was, therefore, to evaluate the role of single and serial measurements of inhibin A and activin A in the prediction of the outcome of PULs.

Materials and Methods

This was a prospective observational study on women attending the EPU of an inner London teaching hospital. Women attended the unit for a number of reasons including: pain, vaginal bleeding, pain with vaginal bleeding, unsure menstrual dates, history of a previous pregnancy complication and maternal anxiety. Women could self-refer or were referred by their general practitioner, midwife, the accident and emergency department or another hospital department.

Table I Diagnostic criteria used for pregnancy location on transvaginal ultrasound scan

Pregnancy location	Transvaginal ultrasound scan: diagnostic criteria
Intrauterine pregnancy	
Viable intrauterine pregnancy (IUP)	An intrauterine gestational sac containing a fetal pole with visible cardiac activity
Pregnancy of uncertain viability (IPUV)	An empty intrauterine gestational sac measuring <20 mm in diameter or an intrauterine gestational sac containing a fetal pole with a crown rump length (CRL) <6 mm with no fetal cardiac activity
Delayed miscarriage	An intrauterine gestational sac containing a fetal pole with a CRL of more than 6 mm with no cardiac activity
Empty sac	An empty intrauterine gestational sac with a mean diameter >20 mm
Incomplete miscarriage	The presence of heterogeneous, hyperechoic irregular tissues within the uterine cavity
Complete miscarriage	An empty endometrial cavity with no evidence of retained products of conception (hyperechoic irregular tissues within the uterine cavity) in a woman who had previously undergone a TVS which showed a viable intrauterine pregnancy, pregnancy of uncertain viability, delayed miscarriage, empty sac or incomplete miscarriage
Ectopic pregnancy	
Tubal ectopic pregnancy	An empty endometrial cavity with: (i) an inhomogeneous adnexal mass or (ii) an empty extrauterine gestational sac seen as hyperechoic ring or (iii) an extrauterine gestational sac with a yolk sac and/or fetal pole with or without cardiac activity
Pregnancy of unknown location	An empty endometrial cavity, with no evidence of an intrauterine gestational sac or retained products of conception and no extrauterine pregnancy visualized

At the time of presentation, all women attending the unit underwent a TVS examination using a 5 MHz transducer for B mode imaging (Aloka SSD 900, 2000 or 4000, Aloka Co., Japan). They were diagnosed with an IUP (viable or non-viable) or an ectopic pregnancy on the basis of this TVS examination, using the sonographic criteria in Table I. If no pregnancy could be visualized on TVS, they were classified as a PUL. All haemodynamically stable, pain-free women classified as having a PUL were invited to participate in this study, which was granted research ethics approval.

Women were given an information sheet about the study and were asked for written consent. Blood samples were taken at the time of the TVS examination (time 0 h) and 48 h later to measure serum hCG (World Health Organization, Third International Reference 75/537) and progesterone (Roche Elecsys 2010 Progesterone II test) levels. Serum was aliquoted and stored at −20°C, and levels of serum inhibin A and activin A (Serotec, Kidlington, UK) were also measured at 0 and 48 h. All samples were analysed in duplicate over eight batches. The detection limit for activin was 0.05. The average coefficient of variation (CV) of the duplicates was 3.48% (range 2.92–4.43). For inhibin, the detection limit was 1.00. The average CV of the duplicates was 4.47% (range 3.54–5.64).

Table II Serum hCG, progesterone, inhibin A and activin A levels in women with a PUL

	Failing PUL (n = 67)		IUP (n = 58)		Ectopic pregnancy (n = 16)		P*
	Median	Range	Median	Range	Median	Range	
hCG ratio	0.43	0.13–1.62	2.11	1.17–4.77	1.09	0.16–2.73	<0.001
Prog avg	2.5	0.5–57.5	67	16.5–188	24.5	2–121	<0.001
Inhibin 0 h	5.3	0.7–136	41.9	3.1–595	15.9	3.6–305	<0.001
Inhibin 48 h	4.3	0.2–183	34.9	2.6–471	8.2	1.4–47	<0.001
Inhibin avg	4.8	0.8–160	39.0	2.8–533	11.1	3.8–154	<0.001
Inhibin ratio	1.00	0.07–4.47	0.79	0.01–2.07	0.77	0.01–1.78	0.211
Activin 0 h	0.57	0.21–1.31	0.47	0.16–1.21	0.60	0.35–0.85	0.008
Activin 48 h	0.57	0.33–0.98	0.51	0.13–1.10	0.49	0.38–0.71	0.563
Activin avg	0.61	0.31–1.13	0.51	0.24–1.12	0.50	0.38–0.74	0.051
Activin ratio	1.00	0.55–2.01	1.07	0.20–2.79	1.04	0.72–1.41	0.571

Inhibin = pg/ml; activin = ng/ml; PUL, pregnancy of unknown location; IUP, intrauterine pregnancy; hCG, human chorionic gonadotrophin; Prog, progesterone; avg, average.
*Based on univariate multicategory logistic regression.

The women were followed up until the final pregnancy outcome was known—failing PUL, IUP or ectopic pregnancy, according to the unit protocol. In general, if the serum hCG level increased by >66% in 48 h (Kadar *et al.*, 1981; Hahlin *et al.*, 1995), women were classified as an IUP and rescanned in 7–14 days to confirm the diagnosis. If the initial serum progesterone level was <20 nmol/l, the women were classified as a failing PUL (Banerjee *et al.*, 2001). They were also classified as a failing PUL if the serum hCG decreased by more than 20% in 48 h. Serum hCG levels were repeated at 7–14 days to confirm the diagnosis. Women who did not fall into either of these categories were thought to be at risk of ectopic pregnancy and were reviewed every 48–72 h, until a pregnancy (IUP or ectopic) could be visualized on TVS (using the criteria in Table I) or had a spontaneous decrease in serum hCG levels to <5 IU/l, confirming a failed PUL.

Statistical analysis

Missing values were observed in 1% of the patients for serum hCG levels at 0 h and in 6% of the patients for levels at 48 h. For progesterone, inhibin and activin, missing values were observed in 6 and 25%, 9 and 24%, and 9 and 25%, respectively. Using multiple imputation (Schafer, 1999; Little and Rubin, 2002), missing values were imputed five times, resulting in five completed data sets. To impute missing values of one marker (e.g. inhibin at 48 h), information of the other hormonal markers as well as information of so-called auxiliary variables was used. The auxiliary variables were age, gestational age, endometrial thickness, level of vaginal bleeding, nulliparity, history of miscarriages, history of termination of pregnancies, history of ectopic pregnancies and final pregnancy outcome. Each completed data set was analysed using standard statistical techniques, and the results were combined to obtain the final results.

The change in serum hCG, inhibin A and activin A levels over 48 h was expressed as a ratio (level at 48 h/level at 0 h). Univariate multicategory logistic regression was used to detect differences in hormonal variables between the three outcome groups—failing PUL, IUP and ectopic pregnancy. In addition, we investigated how well different hormonal indices were able to detect the different pregnancy outcomes by computing the area under the receiver operating characteristic (ROC) curve (AUC) for each outcome separately.

To evaluate the value of inhibin and activin levels for predicting the three pregnancy outcome groups, multivariate multicategory logistic regression was used. Log transformations were performed where the variables were highly skewed. The reference model contained hCG information: log of the hCG ratio and log of the average hCG. The AUC for each pregnancy outcome was computed to determine the effectiveness of the models to predict the pregnancy outcome. The models' AUC performances were estimated using bootstrapping (Efron and Tibshirani, 1993; Steyerberg *et al.*, 2001). This method starts with computing the data set AUCs, which are too optimistic because the model is fitted and evaluated on the same data. Then, using 200 bootstrap samples, an estimate of each AUC's optimism is derived and subtracted from the AUCs to obtain corrected AUCs. Statistical analyses were conducted with SAS version 9 (SAS Institute Inc., Cary, USA).

Results

Serum hCG, progesterone, inhibin A and activin A levels were available from 141 women with a classification of PUL. The samples were collected between 1 July 2004 and 30 November 2004. The final pregnancy outcomes were: 67 failing PULs (47.5%), 58 IUPs (41.1%) and 16 ectopic pregnancies (11.4%).

The median hCG, progesterone, inhibin A and activin A levels with ranges are presented in Table II. The hCG ratio and average progesterone levels were significantly different between the three pregnancy outcome groups. The activin A level was different between the outcome groups at 0 h but not at 48 h (Table II). However, there seemed to be differences for both inhibin A levels, with failing PULs having the lowest levels and IUPs the highest. Changes in activin A or inhibin A levels over time (the ratios) did not differ across pregnancy groups.

Table III shows the separate AUCs for each pregnancy outcome to assess the effectiveness of individual hormonal variables in predicting a particular pregnancy outcome. The progesterone levels at 0 and 48 h were good predictors of a failing PUL or an IUP, but were poor at predicting ectopic pregnancy. A decrease in inhibin A levels over 48 h appears to be the 'best' predictor of ectopic pregnancy (AUC = 0.61). Ectopic pregnancies had the highest median activin level at presentation. Using a cut-off of 0.37 ng/ml on the serum activin A level at presentation to predict ectopic pregnancy (Florio *et al.*, 2007) gave an average sensitivity over the completed data sets of 93% (14.8/16) and an average specificity of 13% (16.4/125).

Table III Results of AUC analysis for the use of different hormonal variables to predict each pregnancy outcome

Variable	AUCs*		
	Failing PUL	IUP	Ectopic pregnancy
hCG 0 h	0.77 −	0.75	0.56
hCG 48 h	0.88 −	0.89	0.52
Prog 0 h	0.9 −	0.94	0.51
Prog 48 h	0.95 −	0.96	0.53
Inhibin 0 h	0.88 −	0.87	0.55
Inhibin 48 h	0.82 −	0.85	0.55 −
Activin 0 h	0.62	0.66 −	0.58
Activin 48 h	0.59	0.58 −	0.52 −
hCG ratio	0.96 −	0.97	0.50
Inhibin ratio	0.60	0.56 −	0.61 −
Activin ratio	0.54 −	0.57	0.57 −

PUL, pregnancy of unknown location; IUP, intrauterine pregnancy; hCG, human chorionic gonadotrophin; Prog, progesterone.
*A minus sign indicates that lower instead of higher values are related to higher probability of the outcome.

The multicategory logistic regression model based on serum activin A levels, using the log of the average level and the log of the ratio as predictors, was not helpful in predicting the pregnancy outcome (see Table IV for the corrected AUCs for each pregnancy outcome). Although a model using serum inhibin A levels (using the ratio and the log of the average level as predictors) was of use in the prediction of failing PULs and IUPs in the PUL population, it did not perform as well as the model using serum hCG levels (Table IV). The corrected AUCs for failing PUL, IUP and ectopic pregnancy using the hCG model were 0.95, 0.97 and 0.67, respectively. Addition of both inhibin variables to the hCG model did not improve its performance. The corrected AUCs for failing PUL, IUP and ectopic pregnancy using a model with hCG and inhibin information (log of the average hCG, hCG ratio, hCG ratio square, log of the average inhibin, inhibin ratio) were 0.96, 0.97 and 0.69, respectively.

Table IV Results of multivariate multicategory logistic regression analysis for the use of different hormonal information to predict the pregnancy outcome

Model	Corrected AUCs		
	Failing PUL	IUP	Ectopic pregnancy
hCG information	0.95	0.97	0.67
Inhibin information	0.88	0.87	0.60
Activin information	0.61	0.64	0.51
hCG + inhibin information	0.96	0.97	0.69

PUL, pregnancy of unknown location; IUP, intrauterine pregnancy; hCG, human chorionic gonadotrophin; hCG information: log of the average hCG, hCG ratio, hCG ratio square; inhibin information: log of the average inhibin A level, inhibin ratio; activin information: log of the average activin A level, log of the activin ratio; corrected AUCs: AUCs corrected for optimism using a bootstrapping procedure.

Discussion

In this study, we have shown that in women with PUL maternal serum inhibin A levels can help predict pregnancy loss, whereas levels of activin A correlate weakly with the pregnancy outcome.

Inhibin A

To the best of our knowledge, this is the first study to examine the use of inhibin A in the prediction of PUL outcome. Previous studies have shown that inhibin A can be used to predict pregnancy loss in women with recurrent miscarriages (Muttuknishna *et al.*, 2002; Al-Azemi *et al.*, 2003; Prakash *et al.*, 2005), those with threatened miscarriages (Johns *et al.*, 2007) and those who have undergone *in vitro* fertilization (Treetampinich *et al.*, 2000; Hauzman *et al.*, 2004). Serum inhibin A has also been found to be a predictor of complete miscarriage in women with miscarriage managed expectantly (Elson *et al.*, 2005). We have found that inhibin A levels are different in the three pregnancy outcome groups. There was, however, no significant difference in the change in serum inhibin A levels over time between the different outcome groups. A low inhibin A level at 0 h was found to be a good predictor of pregnancy loss (failing PUL, AUC = 0.88, Table III). Conversely, we found high levels of inhibin A at 0 h to be a good predictor of an IUP (AUC = 0.87, Table III). However, inhibin levels were no better at predicting PUL outcome than progesterone (Table III).

A previous study found that inhibin A levels were lower in spontaneously conceived ectopic pregnancies compared with IUPs (Seifer *et al.*, 1996). However, this was a case–control study measuring inhibin A levels in women with ectopic pregnancies undergoing surgery using gestation-matched controls with IUPs. No comparison was made with women with failed pregnancies. Although we found that inhibin A levels were significantly different among the three possible PUL outcome groups, with levels lower in ectopic pregnancies compared with IUPs, levels were still higher in ectopic pregnancies than in failing PULs (Table II). We, therefore, found that a single inhibin A measurement was not a good predictor of ectopic pregnancy (inhibin A 0 h AUC = 0.55, Table III). Using serial inhibin A levels, the prediction of ectopic pregnancy was not much better (AUC = 0.60, Table IV), and did not perform as well as the model based on just hCG measurements (AUC = 0.67, Table IV).

As it is secreted by trophoblast, serum inhibin A has been evaluated as a possible surveillance marker after conservative surgical management of ectopic pregnancy since its half-life is significantly shorter than that of hCG (D'Antona *et al.*, 1998). However, interpretation of results was complicated as inhibin A is also secreted by the ovary and increases in levels may occur with the return of normal ovarian function. Inhibin A was, therefore, concluded not to be a useful marker for persistent trophoblast after conservative surgical treatment for ectopic pregnancy. However, the study did suggest that pretreatment inhibin A levels may provide an accurate assessment of the viability of trophoblastic tissue within ectopic pregnancies as some ectopic pregnancies with very high hCG levels had very low inhibin A levels. D'Antona *et al.* (1998), therefore, proposed that inhibin

A levels may be useful in helping assess whether ectopic pregnancy management should be expectant, medical or surgical. If correct, this may explain why the women with ectopic pregnancies in our study had varying serum inhibin A levels (3.6–305 pg/ml at presentation). Some ectopic pregnancies will have active trophoblast and behave like IUPs and others will be failing and behave like failing PULs. A previous study examining serial hCG levels in women with ectopic pregnancies initially classified as PULs found that the changes in serum hCG mimicked an IUP in 20% and a failing pregnancy in 8% of the cases in their series (Silva *et al.*, 2006).

Activin A

Activin A levels have been found to be significantly lower in women with a threatened miscarriage who subsequently miscarried compared with those who had live births (Johns *et al.*, 2007). It has also been suggested that activin A may be used as a marker to predict pregnancies that are likely to miscarry in women with recurrent miscarriage (Prakash *et al.*, 2005). However, other studies have shown that activin A levels are similar in women with miscarriages and controls (Muttukrishna *et al.*, 2002; Wallace *et al.*, 2004).

We found activin A levels are not a good predictor of PUL outcome (Table III). In contrast to previous studies suggesting lower levels in women who subsequently suffer a miscarriage, in our study activin A levels were higher in failing PULs compared with IUPs (Table II). Using the activin A level at presentation gave an AUC of 0.62 to predict failing PUL, 0.66 to predict IUP and 0.58 to predict ectopic pregnancy (Table III). This is in contrast to results obtained in a recent study evaluating the use of a single measurement of activin A in the diagnosis of ectopic pregnancy in women classified with a PUL (Florio *et al.*, 2007). The study found that activin A levels were significantly lower in women with ectopic pregnancies compared with IUPs and failing PULs. In the study by Florio, ROC analysis suggested that activin A, at a cut-off of 0.37 ng/ml, gave a sensitivity and specificity of 100 and 99.6%, respectively, for the prediction of ectopic pregnancy. When we evaluated this cut-off on our data, we found the sensitivity to be 93%, but with a very poor specificity of 13%. Activin A levels in the women with ectopic pregnancies in our study ranged from 0.35 to 0.85 ng/ml at presentation compared with 0.27 ± 0.06 ng/ml in the study by Florio *et al.* (2007). They suggested that the lower levels of activin A in ectopic pregnancies compared with those with failing PULs may be due to the difficulty of the ectopic trophoblast to correctly implant. Our results do not agree with their findings, possibly because our ectopic pregnancies may have been more of a heterogeneous group. Certainly, in our study, there were ectopic pregnancies that had low progesterone levels or low hCG ratios as often seen in failing PULs while others had high progesterone levels or high hCG ratios as seen in IUPs. The median progesterone level in the ectopic pregnancies was 24.5 nmol/l (range 2–121 nmol/l) and the median hCG ratio 1.09 (range 0.16–2.73) (Table II). Florio *et al.* (2007) did not examine changes in serum hCG over time in their PULs, but progesterone levels in their ectopic pregnancies were significantly lower (14.7 ± 0.9 ng/ml) compared with their IUPs (31.8 ± 0.4 ng/ml), suggesting many of them were failing in any event.

In this study, we have therefore shown that neither single nor serial measurements of inhibin A or activin A can be used to reliably predict all three possible final outcomes of PULs. The challenge still remains to identify a universal marker of PUL outcome, which combines accuracy, reproducibility and simplicity. However, it may be that a combination of markers may be needed in order to maximize sensitivity and specificity for the various outcomes. For the moment, our data suggest that a model based on serum hCG levels at 0 and 48 h is the optimal approach to the prediction of PUL outcome until such a marker can be identified.

Funding

This research was supported by /Research Council KUL/: GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering, /Belgian Federal Science Policy Office/ IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modeling'), /EU/: BIOPATTERN (FP6-2002-IST 508803), ETUMOUR (FP6-2002-LIFESCIHEALTH 503094), Healthagents (IST-2004-27214).

References

Al-Azemi M, Ledger WL, Diejomaoh M, Mousa M, Makhseed M, Omu A. Measurement of inhibin A and inhibin pro-alphaC in early human pregnancy and their role in the prediction of pregnancy outcome in patients with recurrent pregnancy loss. *Fertil Steril* 2003;**80**:1473–1479.

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.

Barnhart KT, Sammel MD, Rinaudo RF, Zhou L, Hummel AC, Guo W. Symptomatic patients with an early viable intra-uterine pregnancy: hCG curves redefined. *Obstet Gynecol* 2004a;**104**:50–55.

Barnhart KT, Sammel MD, Chung K, Zhou L, Hummel AC, Guo W. Decline of serum human chorionic gonadotropin and spontaneous complete abortion: defining the normal curve. *Obstet Gynecol* 2004b; **104**:975–981.

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.

Condous G, Kirk E, Syed A, Van Calster B, Van Huffel S, Timmerman D, Bourne T. Do levels of serum cancer antigen 125 and creatine kinase predict the outcome in pregnancies of unknown location? *Hum Reprod* 2005;**20**:3348–3354.

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.

Condous G, Van Calster B, Kirk E, Haider Z, Timmerman D, Van Huffel S, Bourne T. Prediction of ectopic pregnancy in women with a pregnancy of unknown location. *Ultrasound Obstet Gynecol* 2007a;**29**:680–687.

Condous G, Van Calster B, Kirk E, Timmerman D, Van Huffel S, Bourne T. Prospective cross-validation of three methods of predicting failing pregnancies of unknown location. *Hum Reprod* 2007b;**22**:1156–1160.

D'Antona D, Mamers PM, Lowe PJ, Balazs N, Groome NP, Wallace EM. Evaluation of serum inhibin A as a surveillance marker after conservative management of tubal pregnancy. *Hum Reprod* 1998;**13**:2305–2307.

Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

Elson J, Tailor A, Salim R, Hillaby K, Dew T, Jurkovic D. Expectant management of miscarriage-prediction of outcome using ultrasound

and novel biochemical markers. *Hum Reprod* 2005;**20**:2330–2333.

Florio P, Seven FM, Bocchi C, Luisi S, Mazzini M, Danero S, Torricelli M, Petraglia F. Single serum activin A testing to predict ectopic pregnancy. *J Clin Endocrinol Metab* 2007;**92**:1748–1753.

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

Hauzman E, Fedorcsák P, Klinga K, Papp Z, Rabe T, Strowitzki T, Urbancsek J. Use of serum inhibin A and human chorionic gonadotropin measurements to predict the outcome of *in vitro* fertilization pregnancies. *Fertil Steril* 2004;**81**:66–72.

Johns J, Muttukrishna S, Lygnos M, Groome N, Jauniaux E. Maternal serum hormone concentrations for prediction of adverse outcome in threatened miscarriage. *Reprod Biomed Online* 2007;**15**:413–421.

Kadar N, Caldwell BV, Romero R. A method of screening for ectopic pregnancy and its indications. *Obstet Gynecol* 1981;**58**:162–166.

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 (PULs). *Ultrasound Obstet Gynecol* 2006;**27**:311–315.

Little RJA, Rubin DB. *Statistical Analysis with Missing Data*, 2nd edn. New York: Wiley, 2002.

Muttukrishna S, Jauniaux E, Greenwold N, McGarrigle H, Jivraj S, Carter S, Elgaddal S, Groome N, Regan L. Circulating levels of inhibin A, activin A and follistatin in missed and recurrent miscarriages. *Hum Reprod* 2002; **17**:3072–3078.

Prakash A, Laird S, Tuckerman E, Li TC, Ledger WL. Inhibin A and activin A may be used to predict pregnancy outcome in women with recurrent miscarriage. *Fertil Steril* 2005;**83**:1758–1763.

Romero R, Kadar N, Jeanty P, Copel JA, Chervenak FA, DeCherney A, Hobbins JC. Diagnosis of ectopic pregnancy: value of the discriminatory human chorionic gonadotropin zone. *Obstet Gynecol* 1985;**66**:357–360.

Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999; **8**:3–15.

Seeber BE, Sammel MD, Guo W, Zhou L, Hummel A, Barnhart KT. Application of redefined human chorionic gonadotrophin curves for the diagnosis of women at risk for ectopic pregnancy. *Fertil Steril* 2006;**86**:454–459.

Sefer DB, Lambert-Messerlian GM, Canick JA, Frishman GN, Schneyer AL. Serum inhibin levels are lower in ectopic than intrauterine spontaneously conceived pregnancies. *Fertil Steril* 1996; **65**:667–669.

Silva C, Sammel MD, Zhou L, Gracia C, Hummel AC, Barnhart KT. Human chorionic gonadotrophin profile for women with ectopic pregnancy. *Obstet Gynecol* 2006;**107**:605–610.

Steyerberg EW, Harrel FE Jr, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;**54**:774–781.

Treetampinich C, O'Connor AE, MacLachlan V, Groome NP, de Kretser DM. Maternal serum inhibin A concentrations in early pregnancy after IVF and embryo transfer reflect the corpus luteum contribution and pregnancy outcome. *Hum Reprod* 2000; **15**:2028–2032.

Wallace EM, Marjono B, Tyzack K, Tong S. First trimester levels of inhibins and activin A in normal and failing pregnancies. *Clin Endocrinol (Oxf)* 2004;**60**:484–490.

Submitted on September 28, 2008; resubmitted on December 5, 2008; accepted on December 18, 2008