Université catholique de Louvain
Secteur des Sciences de la Santé
IREC – Institut de Recherche Expérimentale et Clinique
EPID – Epidémiologie et Biostatistique
Bruxelles, Belgique

Mathematical models integrating an ultrasound-based technology improve the diagnosis of ovarian cancer

Thesis submitted in fulfilment of the requirements for the degree of "Docteur en Sciences Médicales"

Evelien Vaes

Promotor: Prof. Annie Robert

2010-2011

Members of the Jury

Promotor:

Prof. Annie Robert, MSc, PhD
Pôle de recherche en Epidémiologie et Biostatistique
Institut de Recherche Expérimentale et Clinique
Université catholique de Louvain
Brussels, Belgium

President of the jury:

Prof. Dominique Lison, MD, PhD
Louvain Centre for Toxicology and Applied Pharmacology
Institut de Recherche Expérimentale et Clinique
Université catholique de Louvain
Brussels, Belgium

UCL members:

Prof. Martine Berlière, MD, PhD
Pôle de recherche en Gynécologie
Institut de Recherche Expérimentale et Clinique
Université catholique de Louvain
Brussels, Belgium

Prof. Marie-Christine Closon, MSc, PhD
Institut de recherche Santé et Société
Université catholique de Louvain
Brussels, Belgium

Prof. Catherine Legrand, MSc, PhD
Institut de Recherche multidisciplinaire
pour la modélisation et l'analyse quantitative
Université catholique de Louvain
Louvain-la-Neuve, Belgium

External members:

Prof. Usha Menon, MD, PhD
Gynaecological Cancer Research Centre
Elizabeth Garrett Institute for Women's Health
University College London
London, United Kingdom

Dr. Dror Nir, MSc, PhD
Advanced Medical Diagnostics
Waterloo, Belgium

Dr. Elizabeth Van Eycken, MD
Belgian Cancer Registry
Brussels, Belgium

Acknowledgements

Writing this part of my thesis makes me very happy, because, as I always imagined, it is the last section that I add and so this means that I finally finalized so many years of hard work. All these years and all this work would not have been possible without the never ending support of so many people whom I would really like to thank here.

In the first place, special thanks to my promoter, Professor Annie Robert, who has welcomed me very kindly into the 'EPID' unity and has been extremely supportive over the past 4,5 years. Even in times of bad luck, she has never lacked the energy to inspire me to persevere. This is really a great quality! I am very grateful for the motivation that she gave me, for her scientific support and insights, for the numerous opportunities that she gave me and for always believing in me.

I also particularly thank Dror and Rina Nir, the founders of the HistoScanning technology and of AMD. From the very first minute that I met them, I was infected by their enthusiasm for our project. They have welcomed me very warmly into their company, gave me the opportunity to meet so many interesting people and were always there for me when I needed help. Thanks Dror and Rina for your vision, your guidance, and your support and also for the truly wonderful experience of letting me take part in the HistoScanning project.

I am very grateful to all the jury members who accepted to judge this work with their particular expertise: Professor Dominique Lison, Professor Martine Berlière, Professor Marie-Christine Closon, Professor Catherine Legrand, Professor Usha Menon, Dr. Dror Nir, Dr. Elizabeth Van Eycken and Professor William D'Hoore. Their comments were truly valuable and helped me to improve this manuscript in the final stage.

Working both at UCL and AMD was a great pleasure to me and not the least a very interesting experience. I would therefore like to thank all my colleagues that I had the pleasure to work with: Merci aux épidiennes (et épidiens rare) pour tout votre aide, vos sourires, pour toutes les conversations qui remontaient le morale, pour corriger mon français et surtout pour être des gens si chouette ;-) and thanks to all colleagues at AMD for welcoming me so warmly, for your kind help and support and for the fun activities and lunch time conversations.

Special thanks also to some people who made important contributions to the scientific work: I very much appreciated the contributions and support of Dr. Ranjit

Manchanda and Professor Usha Menon at UCL London, Dr. Philippe Autier and Professor Harry Bleiberg. I am grateful for your guidance. Thanks to Mrs. Fletcher and Mrs. Ford for reviewing the Ultrasound images. And thanks to Dror Givon, who kindly accepted to proofread part of this thesis.

Being a PhD-student is a peculiar experience, one that makes you grow scientifically but also as a person. There were so many days on which my hesitations took the upper-hand and I felt that I would never finish this PhD. On any of these days, I've had a close friend to turn too. So, I am very grateful for the everlasting support of my friends, for the fun pass-time activities that made me forget for a while and for the 'cheering-me-up' conversations. Thanks to all of you (you know who you are!) you have my love and gratitude.

I would like to specially thank one of the bravest and yet sweetest persons that I know, my mother-in-law, Cherry, for her support and wise words! And also Babak, Sarah, Lila and Arian: you have been such a wonderful and loving family, thank you for everything.

My dearest parents, to my deepest sadness today only represented by my father, who cared for me my whole life, I owe this work to them. Their love, the warm home they have always created for me and their education have made me into who I am today, it's thanks to all of that and so much more that I persevered. Most importantly: 'lieve papa' thank you so much for being so supportive all these years, for trying to take up your own and mama's role, you have my love!

Finally, to you, my dearest Kiavash, even if I write another 1000 pages, I will never find enough words to thank you. You are the only person who understands me without words, who knows me to the bottom of my heart and beyond. We've shared so many beautiful things together; I cannot image a life without you. I am so grateful for your unlimited love, for cheering me up, for truly believing in me, for inspiring me every day and especially for our beautiful son Nima who made our lives a true pleasure and gives us so much joy every day. All these past years, every single day, you really believed in me, much more than I believed in myself and that, Kia, was a very generous and loving gift. You are my love, always!

Evelien Vaes, January 2011.

List of abbreviations

AICC	Akaike's Information Corrected Criterion
AMD	Advanced Medical Diagnostics
AUC	Area Under the (Receiver Operating Characteristics) Curve
BRCA	Breast Cancer Gene
CA125	Ovarian Cancer Antigen, a protein
CRF	Case Report Form
FIGO	International Federation of Gynaecology & Obstetrics
HSCU	HistoScanning cut-off
IDI	Integrated Discrimination Improvement
LC	Lower cut-off
LR	Logistic Regression
NN	(Artificial) Neural Network
RF	RadioFrequency
RMI	Risk of Malignancy Index
ROC	Receiver Operating Characteristics
SNR	Signal to Noise Ratio
TVS	Transvaginal ultrasonography
UC	Upper cut-off

US	Ultrasound
VOI	Volume Of Interest

CONTENTS

Part I: Novel tests and multivariate mathematical models for the differential diagnosis of adnexal masses

Part II: Ovarian Sonohistology: Statistical models for ultrasound data and their ability to differentiate tissue types

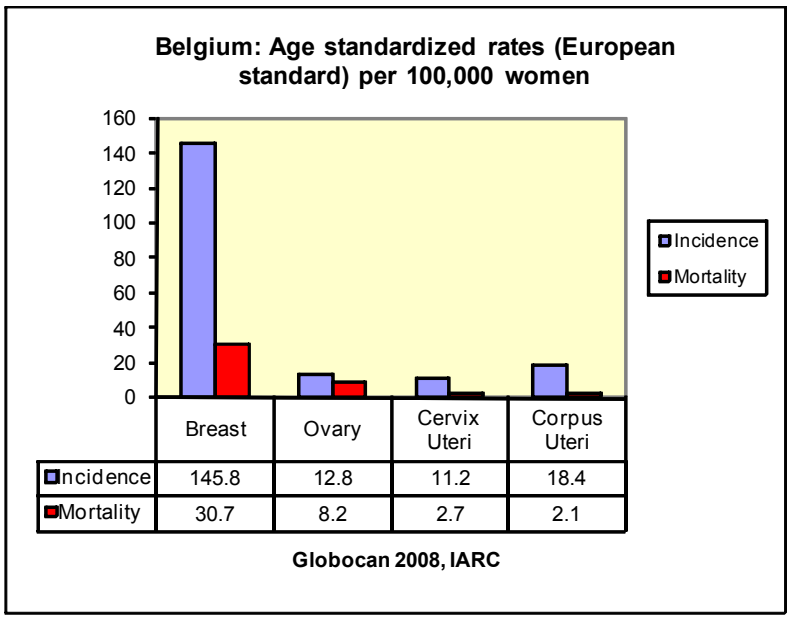
PART I

NOVEL TESTS AND MULTIVARIATE MATHEMATICAL MODELS FOR THE DIFFERENTIAL DIAGNOSIS OF ADNEXAL MASSES

OUTLINE

In the European countries, age standardized incidence rates (European standard) for ovarian cancer vary between 7.2 and 19.3/100,000 while mortality rates are ranging between 2.8 and 12.2/100,000.[1] In Belgium, ovarian cancer is not as frequent as breast cancer since breast cancer presents with very high incidence rates (similar for other European countries). However, age standardized breast cancer mortality rates in 2008 were less than one fourth of the age standardized incidence rates whereas for ovarian cancer, mortality rates were two thirds of the incidence rates (see figure 1). And, unlike for breast cancer, mortality rates for ovarian cancer were not decreasing over the past years.[2] Indeed, ovarian cancer is one of the leading causes of death from gynaecological malignancies.[3] This is explained by the fact that in general, ovarian cancer is detected at too advanced stages. Early diagnosis of ovarian cancer is thus the key for improving outcomes for the disease.

FIGURE 1: INCIDENCE AND MORTALITY RATES FOR GYNAECOLOGICAL CANCERS IN BELGIUM (2008)



Medical imaging techniques have revolutionised medicine during the last decades. Ultrasound (US) in particular gives access to vital data in a non-invasive way and is effective for imaging soft tissues of the body. Compared to other medical imaging modalities, US has the following positive attributes:

- US is a real-time, easy operation medical imaging technique
- US has a non-invasive and radiation free nature
- US is relatively low-priced,

Hence US has become widely used as a diagnostic technique in general clinical practice.

In gynaecology, US is one of the most important and primary diagnostic tools. Its use continues to increase and it is now an essential part of the diagnostic procedure in examining the female pelvis. Indeed, the use of US morphology to characterize adnexal masses and thus diagnose ovarian cancer is well established.[4]

As a part of patient management, gynaecologists use US morphology to differentiate between malignant and non-malignant ovarian masses that come to their attention. In addition, a large number of indexes and mathematical prediction models that assess the likelihood of malignancy for an ovarian mass have been developed and they all incorporate US.

A novel computer-aided technology, HistoScanning™, makes use of US data for characterising ovarian tissue suspicious of being malignant.

In part I of this research work, we investigated whether ovarian HistoScanning could improve the performance of existing prediction models for the differential diagnosis of ovarian masses and we also explored what place could be granted to this new technology in clinical practise.

Part I is organised as follows: Chapter 1 concerns the epidemiology of ovarian cancer and discusses the importance of differential diagnosis of ovarian masses. Detailed aims are described in chapter 2. Chapter 3 introduces the general methods used for this work. In chapter 4, two publications that present the results of this work are presented. Finally, a discussion, regarding the results presented in part I, concludes this work in chapter 5.

CHAPTER 1: GENERAL INTRODUCTION

In this chapter, some background notions are presented. Section 1 defines what adnexal masses and ovarian tumours are and how these are classified and staged. In section 2, the epidemiology, diagnosis and treatment of ovarian cancer are discussed. Finally, section 3 reports on the importance of differentiating benign from malignant masses and introduces the differential diagnosis techniques that were developed over the years.

1. ADNEXAL MASSES

1.1. TYPES OF ADNEXAL MASSES

Left and right to the uterus, there is a region containing an ovary and a fallopian tube together with associated vessels, ligaments, and connective tissue which is referred to as the 'adnexal region'. Adnexal mass thus indicates any mass originating from the ovary or fallopian tube but also from e.g. the uterus, bowel, or any metastatic disease from another site. Myers et al[5] defined the following list of conditions that can present as an adnexal mass:

- Benign primary ovarian tumours
- Borderline and malignant ovarian tumours
- Metastatic malignant tumours – most commonly breast and gastrointestinal tract
- Masses arising from the fallopian tube
- Masses arising from an organ 'close' to the adnexae (e.g. uterus, pelvic kidneys,…)
- Endometriosis
- Pelvic inflammatory disease
- Cysts arising from normal ovarian functions, such as follicular cysts (development of eggs) or corpus lutheum cysts (ovulation)

Adnexal masses can be discovered/visualized during bimanual pelvic examination (if they are palpable) or also by radiographic imaging.

1.2. OVARIAN TUMOURS

1.2.1. THE DIFFERENT OVARIAN TUMOURS

The normal ovaries are an oval-shaped pair of solid organs of approximately 3 cm in diameter. After the menopause though, they will decrease in size.

When tumours arise in the ovary, they are called 'primary' ovarian tumours and these primary tumours can originate from 3 broad types of cells:[5,6]

- Epithelial cells: cells on the surface of the ovary
- Germ cells: cells that form the eggs
- Sex cord-stromal cells: cells surrounding the eggs, including cells that produce ovarian hormones

Most epithelial tumours are of serous or mucinous histological subtype and all 3 abovementioned 'broad' types of ovarian tumours present with more detailed histological sub-classifications that can be found in the World Health Organisation's Handbook.[6]

Tumours are *malignant* if they possess the ability to metastasize and *benign* if they don't, regardless of the histological class they belong to. However, in the case of epithelial tumours in which there is no invasion into the ovarian stroma, some are classified somewhat in between as *borderline.* This happens when only little histological evidence of cell proliferation exists (as e.g. increased cell division, changes in the appearance of the cell nucleus). In this work, borderline ovarian tumours were considered as malignant tumours but note that quite some controversies exist over which class (benign or malignant) they really belong to.

As listed before, malignant ovarian tumours can also be the result of metastasis from another site which is often breast.

1.2.2. INTERNATIONAL FEDERATION OF GYNAECOLOGY & OBSTETRICS (FIGO) STAGING OF OVARIAN CANCER

Ovarian cancer is mainly staged surgically but histology and cytology are also to be considered at final staging. Evaluations that might affect final staging are clinical evaluation and imaging results. Table 1 Shows the classification of ovarian carcinoma's into FIGO stages as defined by the FIGO committee on Gynaecological Oncology in the paper by Benedet et al.[7]

TABLE 1: FIGO STAGING OF OVARIAN CARCINOMA'S

FIGO Stages		
I		Tumour **limited** to the **ovaries**
	IA	Tumour limited to one ovary; capsule intact, no tumour on ovarian surface; no malignant cells in ascites or peritoneal washings
	IB	Tumour limited to both ovaries; capsule intact, no tumour on ovarian surface; no malignant cells in ascites or peritoneal washings
	IC	Tumour limited to one or both ovaries with any of the following: capsule ruptured, tumour on ovarian surface, malignant cells in ascites or peritoneal washings
II		Tumour involves one or both ovaries with **pelvic extensions**
	IIA	Extension and/or implants on uterus and/or tube(s); no malignant cells in ascites or peritoneal washings
	IIB	Extension to other pelvic tissues; no malignant cells in ascites or peritoneal washings
	IIC	Pelvic extension (IIA or IIB) with malignant cells in ascites or peritoneal washings
III		Tumour involves one or both ovaries with **microscopically confirmed peritoneal metastases** outside the pelvis and/or **regional lymph node metastases**
	IIIA	Microscopic peritoneal metastasis beyond pelvis
	IIIB	Macroscopic peritoneal metastasis beyond pelvis ≤ 2 cm in greatest dimension
	IIIC	Peritoneal metastasis beyond pelvis > 2 cm in greatest dimension and/or regional lymph node metastasis
IV		**Distant metastasis** (excludes peritoneal metastasis)

2. EPIDEMIOLOGY AND TREATMENT OF OVARIAN CANCER

2.1. INCIDENCE AND MORTALITY

Annual age standardized incidence rates (European standard) for ovarian cancer vary between 7.2 and 19.3/100,000 in European countries.[1] These go together with age standardized mortality rates ranging between 2.8 and 12.2/100,000 for the European countries.[1] Worldwide, the age standardized (world standard) incidence and mortality of ovarian cancer are of 6.3 and 3.8 per 100,000, respectively.[8] This means that in 2008 worldwide, there were 225,000 new cases of ovarian cancer and 140,000 deaths from ovarian cancer.[8] This high case-fatality rate of ovarian cancer is often assigned to its 'silent' progression which leads to a late detection of this disease when the cancers are already in advanced stages and survival is poor. Indeed, 3-yr[A] survival of Figo stage III and IV cancers is of 30% and 25% respectively, whereas the 3-yr survival of Figo stage I and II cancers is of 89% and 87%, respectively.[9]

Incidence rates of ovarian cancer are the highest in industrialized countries (with the exception of Japan) and have slowly increased over the last years.[10]

2.2. AETIOLOGY

The cause of ovarian cancer remains unknown but the risk of ovarian cancer is increased by a family history of ovarian or breast cancer, personal history of breast, uterus, colon or rectal cancer, nulliparity, infertility, and ever usage of hormone replacement therapies.[11-15] Protective factors include high parity, history of tubal ligation or hysterectomy, and use of oral contraceptives.[11,13,16] Other 'ambiguous' factors with suggestive evidence of association to ovarian cancer are the usage of talc and some dietary habits.[17,18]

[A] Due to the high mortality rate (especially for FIGO stages III and IV), 5-yr survival data are difficult to find.

Genetic factors such as the breast-ovarian cancer syndrome, linked to an inherited mutation in the BRCA1 gene or the BRCA2 genes, account for approximately 5% of all ovarian cancers.[7]

2.3. SYMPTOMS

Unfortunately ovarian cancer does not cause many 'early' signs until it reaches advanced stages (FIGO III and IV). Moreover, symptoms are very unspecific and this implies that they do not always alert clinicians to the diagnosis of ovarian cancer. As a consequence, two thirds of ovarian tumours are stage III or IV at the time of diagnosis.[11] The most frequent symptoms are increased abdominal size, bloating, urinary urgency and pelvic pain.[11,19] Recent studies have shown that these symptoms, however unspecific, are more frequently and more severely experienced in women with ovarian cancer prior to diagnosis.[20]

The detection of an adnexal mass may happen incidentally, at a routine health maintenance examination or at imaging done for another indication. Symptomatic women, on the other hand, are intentionally evaluated by a physical examination and/or radiographic imaging.

2.4. DIAGNOSIS – INTRODUCTION

When ovarian cancer is suspected, the first steps in diagnosing usually include a complete physical and abdominal examination (i.e. bimanual pelvic examination). Blood samples will be taken too, to assess the overall physical condition of the patient, but they also allow assessing tumour markers as serum CA125 antibody concentrations. Transvaginal ultrasonography (TVS) (or transabdominal) is then used to visualise the ovaries, the fallopian tubes and the uterus; It is always recommended as part of the initial workup of a complex adnexal/ovarian mass.[19] Other imaging techniques (transabdominal ultrasonography, CT-scans, MRI-scans) might be performed for further investigations. Given the large variety of possible causes (section 1.1), precise diagnosis of a mass is important for selecting the most appropriate treatment option. More on this will be presented in section 3.

2.5. TREATMENT

Surgery is the first and most important part in ovarian cancer treatment. A complete staging surgery by laparotomy is nearly always performed as it is a cornerstone in the management of ovarian cancer.[7] Recently, researchers have also suggested the use of laparoscopy for the staging of early ovarian cancer.[21] Whichever choice is made for early stage ovarian cancer, the staging surgery should always include full exploration of entire pelvic and abdominal contents, omentectomy (because the omentum is a common side of spread and should thus be removed even if clinically uninvolved) and random peritoneal biopsies. If the contralateral ovary is unaffected and when the patient is young and desirous of fertility preservation, conservative surgery should be considered. Otherwise, total hysterectomy with bilateral salphingo-oophorectom and lymphadenectomy is recommended. Then, only patients with stage IA, grade 1 tumour do not require adjuvant chemotherapy. All other early stage cancers usually receive an adjuvant treatment that consists of a platinum analogue, alone or in combination with Paclitaxel.[7,11,22]

In case of advanced stage ovarian cancer, the surgeon should aim at removing as much of the tumour as possible during primary (staging) surgery. This is often referred to as debulking or cytoreductive surgery. The importance of this surgery is reflected by the fact that maximum cytoreductive surgery (often residual tumour < 2cm) is a key prognostic factor that affects survival and recurrence of late stage ovarian cancers.[23,24] Postoperatively, a variety of chemotherapy treatments can be used in combination for adjuvant treatment; the most commonly used are Cisplatin and Paclitaxel. When patients are unable to tolerate surgery due to medical conditions, a neoadjuvant chemotherapy treatment eventually followed by intermediate debulking is often recommended.[7,25] However, note that in the near future, neoadjuvant chemotherapy will play a more important role in general for patients with advanced ovarian cancer.[7,25]

3. PRE-OPERATIVE DIFFERENTIAL DIAGNOSIS

3.1. THE IMPORTANCE OF DIFFERENTIATION BETWEEN BENIGN AND MALIGNANT ADNEXAL MASSES

For symptomatic women, surgical management is the treatment of choice in order to relief symptoms. Surgeons' specialized training and experience in managing malignancy has been strongly associated with improved outcomes for patients with ovarian malignancy.[26-30] This implies that referral to and management by experienced gynaecological oncologists is important.

Secondly, in the group of asymptomatic women, an increasing number of adnexal masses, of which only a minority will be malignant, come to the attention of physicians. For these patients, it is important to propose an appropriate management in the case of malignancy while at the same time avoiding unnecessary procedures (surgery, anxiety, etc.) for women with benign conditions. Indeed, an improved patient outcome will be ensured when the most appropriate referral and management is chosen for malignancy while minimizing invasive diagnostic procedures for benign lesions that are not likely to transform into cancer.[23,26,31]

Based on these arguments, it is clear that pre-operative differential diagnosis techniques are important to facilitate these matters of correct referral and treatment choice.

In the next section, an overview of the differential diagnostic techniques and their development will be given. We will also introduce a novel computerized tissue characterization technique named 'ovarian HistoScanning'.

3.2. A HISTORY OF DEVELOPMENT OF TECHNIQUES FOR THE DIFFERENTIAL DIAGNOSIS OF ADNEXAL MASSES

The use of grey-scale US morphology to characterize adnexal masses is well established[4] and TVS has been found to perform better than trans-abdominal scanning, particularly for small masses. Over the years, a number of scoring

systems based on US-determined morphological characteristics (such as cystic volume, cyst wall structure,…) have been developed.[32-35] However, there was no standardized index.[5] Later, the addition of a variety of second line tests such as serum CA125 levels, Doppler blood flow indices, power Doppler, and clinical data to TVS morphological indices have been shown to improve test performance and to outperform individual tests.[36] The Risk-of-malignancy index (RMI)[37] is such an index which is routinely used in the United Kingdom to triage women with ovarian masses along appropriate referral pathways.[38-41] As a logical extension of these indices, researchers have used multivariate mathematical models to determine the most important variables for differential diagnosis and improve test performance.[42-56]

3.2.1. The Risk of Malignancy Index (RMI)

The RMI is a standardized index commonly used for clinical evaluation of patients with an adnexal mass. In the United Kingdom, it is routinely used to triage women with ovarian masses along appropriate referral pathways.[38-41] The RMI combines serum CA125 values (typically, CA125 < 35 U/ml is considered as normal), menopausal status, and an US-morphology score and improves on the individual performance of each of these indices.[37] The US-morphology score is defined for each patient based on the presence of the following key morphological features, each of which score 1 point: presence of a multilocular cyst, presence of solid areas, bilaterality of lesions, presence of ascites, and evidence of metastases.

The RMI compares well with most logistic regression models or artificial neural networks[57] and was advocated to be the prediction model of choice in preoperative assessment of adnexal masses.[58] The RMI has been validated in many European countries[38,40,59,60] with some prospective studies reporting slightly lower sensitivity and specificity. Nevertheless, it has been advocated that RMI is the simplest and best validated approach to differential diagnosis.[61]

Currently there are 3 versions of RMI, each varying slightly in the method of scoring from the other. The original RMI was described by Jacobs et al (RMI_1)[37] and 2 modifications (RMI_2 and RMI_3) by Tingulstad et al[40,59] exist. All three RMI are defined as the product of menopausal status (M), serum CA125 level and an US-score (U): RMI = U x M x serum CA125. The specific values of the terms U and M for the three RMI in use are given in table 2 and were developed in the original research papers based on simplified regression equations.

TABLE 2: RISK OF MALIGNANCY INDICES: TERMS AND CORRESPONDING CODES FOR EACH DEFINITION OF THE PRODUCT RMI=U $_x$ M $_x$ SERUM CA125

Indices	Menopausal status	**M**	US-score	**U**
RMI_1	Pre-menopausal	M = 1	Us-score = 0	U = 0
	Post-menopausal	M = 3	Us-score = 1	U = 1
			Us-score ≥ 2	U = 3
RMI_2	Pre-menopausal	M = 1	Us-score ≤ 1	U = 1
	Post-menopausal	M = 4	Us-score ≥ 2	U = 4
RMI_3	Pre-menopausal	M = 1	Us-score ≤ 1	U = 1
	Post-menopausal	M = 3	Us-score ≥ 2	U = 3

Because serum CA125 measurements are positive, the RMI takes only positive values in [0, +∞[. When the RMI is used in routine clinical practise, cut-offs of 200 or 250 are mostly applied for triaging: if RMI<250, the patient may be operated in a cancer unit by a general gynaecologist and if RMI≥250, the patient is to be referred to a gynaecological cancer centre to be treated by an expert gynaecological oncologists.[39,41]

3.2.2. MATHEMATICAL MODELS FOR PRE-OPERATIVE DIAGNOSIS

Various multivariate mathematical models to determine the most important variables for differential diagnosis and improve test performance have been developed in the literature.[42-56] In this section, we emphasize them shortly. The difference between the numerous mathematical models for the differential diagnosis of adnexal masses is firstly (1) the type and number of variables that have been used and secondly (2) the mathematical prediction method used for the model development:

1. Variables can be patient related, ultrasound related (including morphological features, measurements and quantitative assessment of morphology), or blood-flow related (Doppler US). Table 3 lists five references and shows the different variables used to build the final model(s).

2. Mathematical methods used include logistic regression, artificial neural networks, support vector machines, and relevance vector machines (see also table 3).

TABLE 3: DESCRIPTION OF SOME MULTIVARIATE MATHEMATICAL MODELS

Reference	Type of model	Variables in model
Tailor et al.[43]	LR	**Age,** *maximum lesion diameter, lesion volume, unilocularity, papillary projections, random echogenicity,* peak systolic velocity (PSV), time-averaged maximum velocity (TAMXV), pulsatility index (PI), resistance index (RI)
Tailor et al.[47]	NN	**Age, menopausal status,** *maximum lesion diameter, lesion volume, locularity, papillary projections, random echogenicity,* blood flow velocity waveforms, PSV, TAMXV, PI, RI
Timmerman et al.[42]	LR	**Personal history of ovarian cancer, hormonal therapy, age, pain,** *maximum lesion diameter, ascites, prescence of entirely solid tumour, maximal diameter of solid component, irregular cyst walls, acoustic shadows,* blood flow within a solid papillary projection, color score†
Van Calster et al.[48]	SVM and RVM	**Age, personal history of ovarian cancer, hormonal therapy,** *maximum diameter of the solid component, maximum ovary diameter, presence of a multilocular-solid lesion, ascites, acoustic shadows, irregular cyst walls, tumour of ovarian origin,* blood flow within a papillary projection, color score† of 4
Lu et al.[56]	SVM	**Menopausal status, CA125,** *papillations, solid tumor, bilaterality, ascites, acoustic shadows, irregular wall,* color score† of at least 3, color score† 4

Abbreviations: LR = logistic regression, NN = Artificial Neural Network, SVM = Support Vector Machine, RVM = Relevance Vector Machine

Patient related variables are in **bold**, US-based morphological features, measurements and quantitative assessments of morphology are in *italic* and finally vascular features (assessed by Doppler) are

underlined

†this color score is a semiquantitative assessment for the tumour as a whole, only based on color Doppler and was defined by a consensus report of the IOATA group.[62]

A number of validation studies have been carried out to assess the performance of mathematical models on independent data and sometimes inconsistencies remained.[5,63,64] Moreover, most mathematical models do not offer much advantage over the RMI[57,65] nor do they outperform subjective evaluation of grey scale and Doppler images by an experienced examiner.[64,66-68] Still, even the most experienced operators can sometimes find it difficult to classify ovarian masses (in particular the smaller lesions).[64] In other words, despite of the inconsistencies seen at validation, the less-experienced examiner will benefit from such models and in some cases also the more experienced examiner.

3.2.3 OVARIAN HISTOSCANNING

HistoScanning™ (see also www.histoscanning.com) is a novel US-based technique that uses tissue characterisation algorithms. These algorithms are combined in large classification schemes and the results of these analyses are used to visualise the position and extent of tissue suspected of being malignant. HistoScanning analysis is based on the fact that cancerous processes induce structural changes in the tissue texture which result in physical changes in the reflected US-waves. Different tissue characterization algorithms were developed using mathematical methods that are optimized to quantify such changes as measured in the US-data. The selection of the appropriate algorithms for a specific organ and their training is an empirical process:

- Prostate HistoScanning is the most advanced technology for which an application is currently already available on the market. The characterization algorithms and classification schemes for prostate HistoScanning were developed and validated in different research studies.[69,70] Other studies regarding prostate HistoScanning are still ongoing (see also www.histoscanning.com). The input to the prostate HistoScanning application is raw Radio-Frequency (RF) data.

- The Breast and Thyroid HistoScanning technologies are in their proof of concept phase. Both these applications also use raw RF data.

- For Ovarian HistoScanning, three different characterization algorithms were selected and a classification scheme was trained during clinical development studies in 2001-02 (unpublished data). The three characterization algorithms are typically applied to units of 20 by 20 voxels of US-data (elementary units) equivalent to tissue volumes ranging from 0.6 to 1.6 mm^3. A threshold is then applied to each characterization algorithm and these 3 results are then combined to give a 0/1 (cancer unlikely/likely) output for every unit. Finally, a 3D analysis of consecutive voxel-units is undertaken before yielding a decision for every unit. The final report of this analysis is then printed on the 2D US-image by means of colored circles, each representing an elementary unit. When the circle is green, the elementary unit is considered as non-malignant. A red circle indicates that an elementary unit is considered as malignant.[71]

  This ovarian HistoScanning application focussed on the management of patients with adnexal masses, suspected of being at risk for ovarian cancer. These three characterization algorithms were optimized on raw gray-scale voxel data. It is this ovarian HistoScanning application that is used in part I of this thesis. However note that in parallel, an ovarian HistoScanning application that uses raw RF data is currently in its proof of concept phase.

CHAPTER 2: AIMS OF OUR WORK

In chapter 1, it was discussed how survival of ovarian cancer patients is improved when they are managed by expert gyneacological surgeons. This would mean that they have to be referred to a 'cancer centre' in a so called teaching (or university) hospital. Table 4 adds some evidence to this finding; three papers that studied the survival of ovarian cancer patients are summarized in Table 4. A comparsion is made between a study located at a teaching hospital[9] (grey column in table 4) and 2 studies that englobe a complete health region with the majority of patients coming from non-teaching hospitals.[72,73] For early stage cancer (FIGO I and II) in particular, a big difference in 3-year survival between the teaching hospital and non-teaching hospitals can be seen. The low survival in the non-teaching studies was attributed to suboptimal surgical treatment. This emphasizes once again the importance of correct referral/management and pre-operative diagnosis.

TABLE 4: 3-YEAR SURVIVAL OF OVARIAN CANCER PATIENTS

	Teaching hospital (Level 3 Referral)	Complete Health region (Majority (70%) in non-teaching hospitals)	
FIGO	**Board et al.[9#]**	**Tingulstad et al.[72$]**	**Weide et al.[73£]**
FIGO I	90%	88%	88%
FIGO II	88%	56%	58%
FIGO III	35%	33%	58%
FIGO IV	24%	21%	30%

[#]Patients registered between 1997 and 2002, cancer-specific survival was reported
[$]Patients registered between 1987 and 1996, crude survival was reported
[£]Patients registered between 1995 and 2003, crude survival was reported

Chapter 1 also shows how, over the years, researchers have developed different pre-operative techniques and have tried to improve performance. Large validation studies have been carried out to validate the performance of the developed techniques on independent data. The Risk of Malignancy Index (RMI) is one of the earlier techniques that maintained its performance throughout the different

validation studies. It is widely adopted in the United Kingdom and in some Scandinavian countries for differential diagnosis and triage of ovarian masses. The literature often defined the RMI to be the simplest and best validated approach to differential diagnosis[61] and a recent systematic review of 83 different prediction models concluded that the RMI was the model of choice in the preoperative assessment of the adnexal masses.[58]

However, interpreting the RMI may sometimes be difficult, mainly because a number of women with an adnexal mass have intermediate RMI values. These intermediate values are lying in an interval around the RMI cutoffs of 200 or 250 that are used in clinical practise. We will refer to this intermediate range as the RMI 'grey zone'. The addition of other imaging techniques as e.g. MRI, CT and additional expert US for patients in the RMI grey zone was investigated by Van Trappen et al.[61] To assess the added value of further cross-sectional imaging as MRI/CT to the RMI, Van Trappen et al. predefined the RMI grey zone to be 25≤RMI<1000.

Ovarian HistoScanning is a new US-based technology that, based on its computer-aided algorithms, tries to distinguish between malignant and non-malignant tissues. A first validation study has shown the potential of HistoScanning for the assessment of adnexal masses.[71] Nevertheless, it remained a challenge to know what place to grant to Ovarian HistoScanning in clinical practise and for the work-up of adnexal masses. If one considers the number of techniques and models that exist today for the differential diagnosis of adnexal masses that all compete to be adopted by clinicians, it becomes clear that this is a difficult task. This research project has therefore adopted a design able to show how HistoScanning may be useful for clinical practise.

In the light of this background, our first aim was to study whether applying ovarian HistoScanning for the patients in the RMI grey zone would improve triage of women with ovarian masses. Our second aim was to assess the optimal cutoffs that define the RMI grey zone in which HistoScanning would further sort out the ovarian masses in order to improve RMI performance. This work is presented in chapter 4.

As mentioned in chapter 1, various mathematical prediction models have been developed. Despite of inconsistencies at validation studies, these models are very useful to assist the less experienced US-operators as their performance (in terms of sensitivity and specificity) doesn't beat the performance of most models. Our third aim was to investigate whether adding ovarian HistoScanning to multivariate

mathematical prediction models for ovarian masses could improve model performance. As a fourth aim, we were interested in finding a final model that was easy to apply and could therefore be useful for the less-experienced US-operators. These results can be found in chapter 4.

CHAPTER 3: GENERAL METHODS

1. ASSESSING THE QUALITY OF DIAGNOSTIC SYSTEMS

In this section, some basic measurements for assessing the quality of a diagnostic test are introduced. These measurements form the methodological background for the results presented in chapter 4 but also for part II of this manuscript. More detailed materials and methods used for specific purposes are fully described in chapter 4.

1.2. PERFORMANCE MEASURES AT A SPECIFIC CUT-OFF[74,75]

A gold standard diagnostic test is an accepted diagnostic test for a particular disease that serves as the reference to which all other tests are compared. Ideally, gold standards should identify all individuals with the disease and should not falsely identify someone that does not have the disease. In practise there is sometimes no true gold standard test. Throughout this work, we have used the histopathology results as the gold standard.

Most diagnostic systems don't automatically give a dichotomous answer (positive or negative). In such cases, a dichotomous outcome is obtained by cutting the outcome of the test which is a quantitative measure, x, in 2 parts by a threshold or cut-off point. When the cut-off point (c), is fixed, the dichotomous test outcome for a patient is for example defined as follows:

- A positive test outcome: $x \geq c \quad \Rightarrow \quad T+$
- A negative test outcome: $x < c \quad \Rightarrow \quad T-$

In this example, high values are indicative of a higher risk of having the disease. For some measures, like e.g. HDL-cholesterol, lower values are indicative of a higher risk and in such cases a positive test (T+) would therefore be defined as: $x \leq c$. For simplicity, we will use $x \geq c$ to define a positive test result (T+) in our text. It is

straightforward though to adapt this chapter to the case where T+ is defined by $x \leq c$.

When considering a novel diagnostic system at a specific cut-off point, c, in relation to the gold standard, any single decision of the diagnostic system for an individual yields one of the following four options:

1. True positive (TP): A correctly diagnosed malignant case
2. True Negative (TN): A correctly diagnosed non-malignant (benign or normal) case
3. False Positive (FP): A non-malignant case incorrectly diagnosed as a malignant case
4. False Negative (FN): A malignant case incorrectly diagnosed as non-malignant

TABLE 5: THE FOUR CONSEQUENCES OF APPLYING A DIAGNOSTIC TEST AT A CUT-OFF C

		Gold Standard		
		D+	D-	
Diagnostic Test	T(c)+	TP(c) True Positives	FP(c) False Positives	TP(c)+FP(c)
	T(c)-	FN(c) False Negatives	TN(c) True Negatives	FN(c)+TN(c)
		TP(c)+FN(c)	FP(c)+TN(c)	N

A whole population (N) submitted to a diagnostic test at cut-off c can be divided into those four categories as represented in table 5. In table 5, T(c)+ and T(c)- stand for the number of cases diagnosed by the test as positive or negative, respectively. Also, D+ represents the overall amount of malignant cases in the population according to the gold standard and D- the overall amount of non-malignant cases, respectively.

Assessment of the accuracy of a diagnostic system uses the probabilities that the test correctly classified a diseased subject as diseased (that is, the true positive rate or sensitivity), and that the test correctly classified a healthy subject as healthy (that is, the true negative rate or specificity). Given the notations of table 5, the

sensitivity and the specificity at the cut-off c are defined as follows in a population of size N:

Sensitivity: $SE(c) = P(T(c)+ \mid D+) = \frac{TP(c)}{TP(c) + FN(c)}$

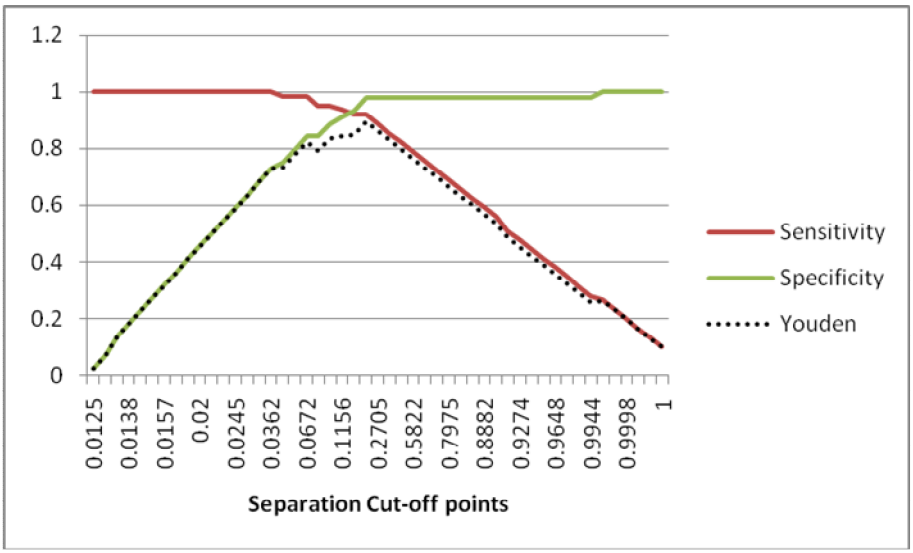
Specificity: $SP(c) = P(T(c)- \mid D-) = \frac{TN(c)}{TN(c) + FP(c)}$

The sensitivity of the test T for the disease D is the likelihood of a positive test outcome (T+) among patients with the disease (D+). The specificity is the likelihood of a negative test outcome (T-) among the patients without the disease (D-).

The Youden Index[76] translates these 2 performance measures of a diagnostic test (sensitivity and specificity) into one single index , that characterizes performance, ranging between 0 and 1: Y(c) = SE(c) + SP(c) -1. In his paper, Youden advocates that his index will give you an adequate statistic to compare the performance (at a cut-off c) of 2 different diagnostic tests.[76]

The fact that the sensitivity and the specificity are independent of the underlying disease prevalence of the population under study makes them the preferred measurements for assessing the quality of the test. This also implies that they can be estimated from a case-control study. Figure 2 gives an example of how the trade-off between sensitivity and specificity against all possible values of the cut-off point c can be visualized. The Youden Index for all possible cut-off points is also plotted on figure 2.

FIGURE 2: TYPICAL CURVES FOR SENSITIVITY, SPECIFICITY AND THE YOUDEN INDEX DEPENDING ON THE CUT-OFF



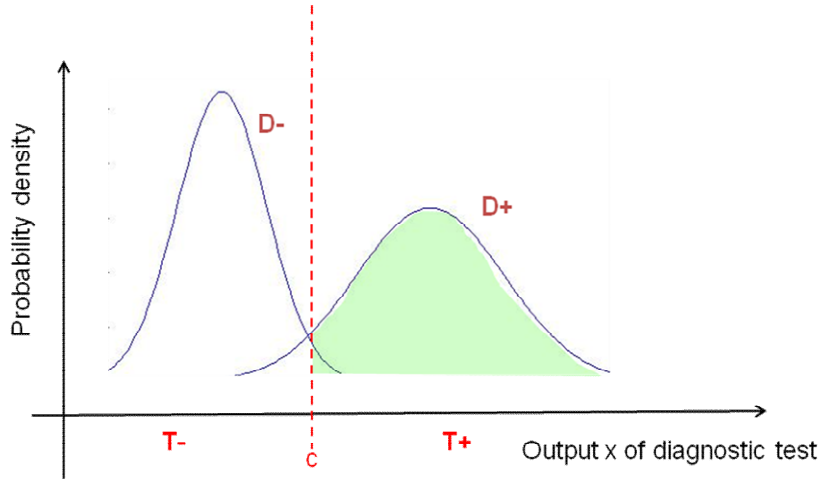
When the output value of a diagnostic test is represented as a continuous value (x is used here) ranging from $-\infty$ to $+\infty$ then the sensitivity and the specificity at a specific cut-off c are actually defined by integrating probability curves as follows:

Sensitivity: $SE(c) = \int_{c}^{+\infty} p(x \mid D+)dx$

Specificity: $SP(c) = \int_{-\infty}^{c} p(x \mid D-)dx$

Note that in the equations stated above, the direction of the integration is given for the special case where the probability density curve of the diagnostic outcomes (x) for the malignant population has higher values for the diagnostic outcomes than (is placed to the right of the density curve for) the non-malignant population. For the sensitivity, this calculation is shown in figure 3.

FIGURE 3: SENSITIVITY CALCULATED AS AREA UNDER THE PROBABILITY CURVE



Two other measurements, that are dependent on the prevalence and thus have to be interpreted with caution, are the positive and negative predictive value given by:

Positive predictive value: $PPV(c) = P(D+ \mid T(c)+) = \frac{TP(c)}{TP(c) + FP(c)}$

Negative predictive value: $NPV(c) = P(D- \mid T(c)-) = \frac{TN(c)}{TN(c) + FN(c)}$

For a diagnostic test to be interesting as a screening modality for instance, it needs to have very high specificity and PPV.

Most multivariate mathematical models applied for diagnosing malignancy produce an output that is (quasi-) continuous. Only measuring the sensitivity and the specificity at a single cut-off point doesn't give a representative assessment of the overall quality of the diagnostic test. A global estimation of the discriminating ability of a diagnostic test over all possible cut-off points is introduced in the following section.

1.2. CUT-OFF INDEPENDENT PERFORMANCE MEASURES

1.2.1 RECEIVER OPERATING CHARACTERISTICS (ROC)

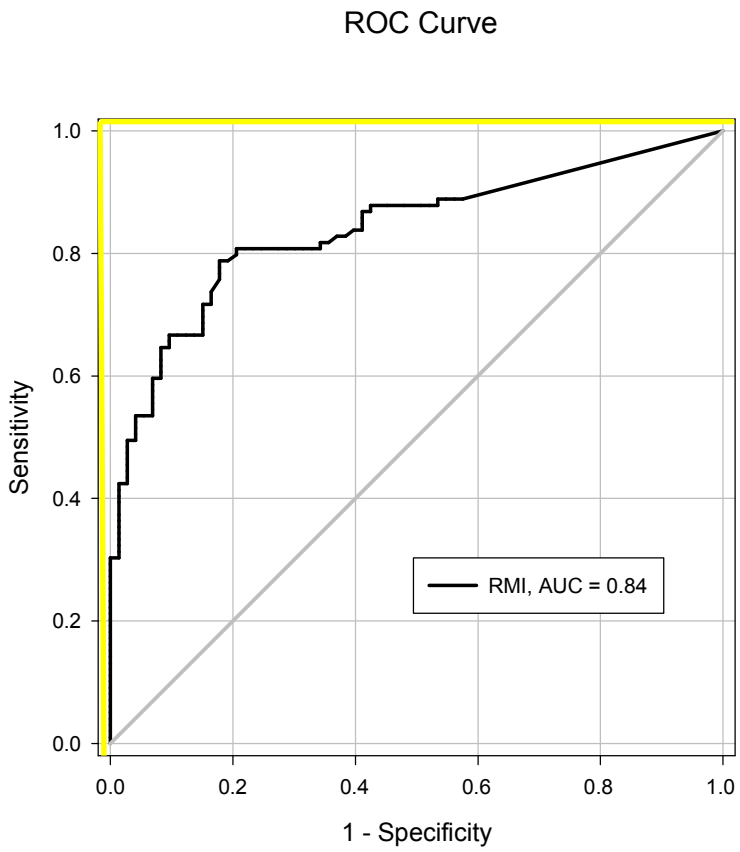
Only considering the sensitivity and the specificity at a specific cut-off point for assessing the quality of a diagnostic model or test may be misleading.

The Receiver Operator Characteristics (ROC) plot is a useful way to display the performance of a medical diagnostic test. Both sensitivity and specificity depend on the cut-off c; as c varies, so do sensitivity and specificity. By considering various possible values of the cut-off c one can construct an ROC curve as a continuous function of sensitivity against '1 –specificity'. In this way, the ROC curve permits to analyze the overall differentiating capabilities of a diagnostic test. Several methods exist to create an ROC curve based on continuous test outcomes. In this work, we have used the empirical (non-parametric) method which involves plotting the pairs of sensitivity versus 1-specificity that are calculated based on the empirical distributions (the histograms) for the diseased (malignant) and healthy (non-malignant) patients. This method is robust as no (possibly wrong) assumptions about the underlying distribution of the 2 patient groups need to be made. The resulting ROC curve will typically look jagged and non-smooth. Other non-parametric methods to create ROC curves that are smooth can be found in Zou et al[77] and Zweig et al[78] discuss various parametric methods.

For statistical analysis, a recommended summary index of the diagnostic test associated with the ROC curve is the Area Under the ROC-Curve, abbreviated as AUC. The AUC has been shown to represent the probability that, when the outcome of the diagnostic test is observed for a randomly selected individual from the diseased population and a randomly selected individual from the non-diseased population, the resulting values will be in the correct order (e.g., diseased value higher than the non-diseased value).[79,80] In our work, the AUCs, falling under the points comprising an empirical ROC curve, were calculated by the trapezoidal rule which has been shown to equal the Wilcoxon- or Mann-Whitney U-statistic.[79,81] An example of an ROC curve and AUC are given in figure 4. The AUC can range from 0.5 (useless test, represented by the grey diagonal line, also called chance line) to 1 (perfect test, represented by the yellow line). Note that if AUC is lower than 0.5, the interpretation of the test results has been reversed, i.e. a positive test outcome (T+) that was defined by $x \geq c$, should rather have been defined by $x < c$ and the other way round for a negative test outcome.

The AUC is the most frequently used summary index for comparing different diagnostic tests. A variety of parametric, semi-parametric and non-parametric tests have been proposed for evaluating differences in the AUC.[81-84] When comparing AUC indices of diagnostic models that were estimated on the same patients (paired-sample), statistical analyses must take into account the correlated nature of the data. As this was the case for the diagnostic tests and models that were developed in this work, the non-parametric Delong test[81] was used to compare AUC indices.

FIGURE 4: EXAMPLE OF A ROC-CURVE WITH AREA UNDER THE CURVE



1.2.2. FINDING THE OPTIMAL CUT-OFF POINT FOR A DIAGNOSTIC TEST WITH CONTINUOUS OUTCOME

An ROC curve, as given by figure 4, is a graph of sensitivity versus '1 – specificity' at all possible values of the cut-off point c and ROC curves can therefore be used to identify the cut-off point at which the best performance is achieved for the diagnostic system. Indeed, maximizing sensitivity corresponds to some large value on the y-axis of the ROC curve. Similarly, maximizing specificity corresponds to a small value on the x-axis of the ROC curve. Thus, a good first choice for a test cut-off value is that value which corresponds to a point on the ROC curve closest to the upper left corner of the graph (the so called 'closest to the (0,1) point' criterion). Mathematically, this implies finding a point c that satisfies the following equation: $\min\{ (1\text{-}SE(c))^2 + (1\text{-}SP(c))^2 \}$.[85] An alternative criterion is given by finding the point that maximises the Youden Index[76], which is the point c that satisfies the following equation: $\max\{ SE(c) + SP(c) - 1\}$. Intuitively, the point that maximises the Youden Index will give you the point on the ROC curve that is the farthest away from chance (the chance line being represented by the grey diagonal line in figure 4).[85] The point that maximised the Youden Index was found to have a more sound clinical meaning[85] and was thus used for finding optimal cut-off points in this work. By using the Youden criterion as defined above, we have given equal weights to both sensitivity and specificity (they have been considered equally important). This might not always be true, however. For example, in some diagnostic applications, it is important not to miss any diseased cases and therefore, it is more important to maximize sensitivity (minimize false negatives) than to maximize specificity. On the other hand, for example in prostate cancer screening with PSA, false positives are common and undesirable (expensive biopsy, emotional impact). In this case, maximizing specificity is important. A modification of the Youden criterion for such cases requiring the attribution of a certain cost (weighing) to a misclassification rate (FP or FN) is given by Perkins and Schisterman.[85]

1.2.3. EVALUATING THE ADDED PREDICTIVE ABILITY OF A NEW VARIABLE IN A DIAGNOSTIC MODEL

In the previous section, it was discussed how AUC is used to summarize the performance of a diagnostic test or model. When a new diagnostic model is developed, the improvement in AUC of this model compared to an 'old' existing model is often used to demonstrate the 'added' value of this 'new' model. As new model often implies adjusting an existing model, that already presents with reasonably good AUC, by adding one new variable (or biomarker), it became

apparent that this new variable needed to have a very strong association to the studied disease before statistically significant improvements in AUC were observed.[86,87] In other words, when evaluating the 'added' predictive ability of a new variable, other measures might be more meaningful.[87]

New measures of evaluating the added predictive ability of a new variable in an existing diagnostic model were recently proposed by Pencina et al.[87] These measures offer incremental information over the AUC.

For the multivariate mathematical models that were developed in this work, we have calculated one of these measures: the Integrated Discrimination Improvement (IDI) which is defined as follows:

$$IDI = (IS_{new} - IS_{old}) - (IP_{new} - IP_{old})$$

In this above expression, IS stands for the integral of sensitivity over all possible cut-off values c and similarly IP corresponds to the integral of '1-specificity'. The subscripts 'new' and 'old' refer to the model with the new variable and the model without the new variable, respectively.

Through this definition the IDI can be seen as a difference between improvement in average sensitivity (of the new versus the old model) and any potential increase in average '1-specificity' or also as an integrated difference in Youden's Indices[76] for all possible cut-off values.[87]

It was shown by Pencina et al. that the IDI can be estimated by:

$$\hat{IDI} = (\bar{\hat{p}}_{new,D+} - \bar{\hat{p}}_{old,D+}) - (\bar{\hat{p}}_{new,D-} - \bar{\hat{p}}_{old,D-})$$

where $\bar{\hat{p}}_{new,D+}$ is the mean of the new model-based predicted probabilities of malignancy for the diseased, $\bar{\hat{p}}_{old,D+}$ is the corresponding quantity but then based on the old model, $\bar{\hat{p}}_{new,D-}$ is the mean of the new model-based predicted probabilities of malignancy for the non-diseased and $\bar{\hat{p}}_{old,D-}$ is the corresponding quantity for the old model.

Standard deviations of $(\bar{\hat{p}}_{new,D+} - \bar{\hat{p}}_{old,D+})$ and $(\bar{\hat{p}}_{new,D-} - \bar{\hat{p}}_{old,D-})$ can be calculated as the standard error of the paired differences of new and old model-based predicted probabilities for all diseased and non-diseased subjects, respectively. The corresponding estimators are denoted by $\hat{se}_{D+}$ and $\hat{se}_{D-}$, respectively. Then, when independence is assumed between diseased and non-diseased, an asymptotic test for the null hypothesis of 'IDI = 0' is given by:

$$z = \frac{\hat{IDI}}{\sqrt{(\hat{se}_{D+})^2 + (\hat{se}_{D-})^2}}$$

which is close to the standard normal for a sufficiently large sample size.

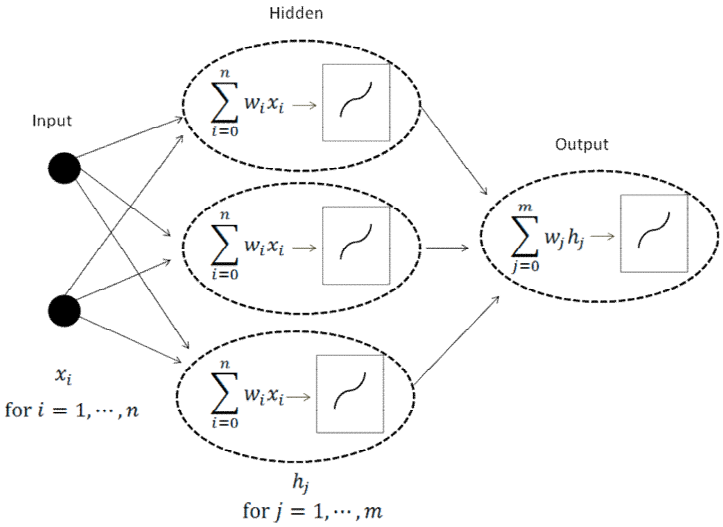
2. A SMALL INTRODUCTION TO ARTIFICIAL NEURAL NETWORK MODELS[41]

Artificial Neural Networks are networks of artificial neurons connected together that constitute a crude approximation to parts of real brains. In this way, a Neural Network (NN) model can be seen as a system that constitutes of many simple elements (the neurons or units that are grouped in layers) that are connected together to perform a particular task, for instance predicting the probability of malignancy(output of the system) based on a number of variables (inputs of the system).

Different NN architectures exist but we will focus on the multi-layer feed-forward NN that consists out of the following architecture (see figure 5):

- One input layer with one unit (neuron) for every variable of interest
- One output layer with one unit representing the outcome of the model (i.e. the predicted probability of malignancy)
- One or more hidden layers with one or more hidden processing units, which will sit between the input and output layers
- No feed-back connections

FIGURE 5: SCHEMATIC REPRESENTATION OF A MULTI-LAYER FEED-FORWARD NEURAL NETWORK



As represented in figure 5, the units of hidden and output layers take a weighted sum of their inputs and put this value trough an activation function. Weights of the inputs of each unit have to be estimated. This is often called the learning or training of a NN. A number of techniques exist to train the weights of a NN and here a parametric 'error-minimization' procedure has been used which is an algorithmic procedure that adjusts the network weights in a number of iterations. At every iteration step, the network weights are adjusted to minimise the difference (error) between the actual outputs (of the NN) and the desired outputs (or observed outputs). The function that characterises this 'error' is often called objective function and thus the objective function needs to be minimised. Often, in statistics, the weights of a NN are estimated through the maximum likelihood criterion which also means that the negative log likelihood is minimised. Whichever criterion or objective function is chosen, if on represents this function by $E(w_i)$ (or $E(w_j)$) then the following steps have to be undertaken to train a NN (learning the network):

- On a training set, generate random weights to initiate the procedure
- Update all the weights by the following weight update: $\Delta w_i = -\eta \frac{\partial E(w_i)}{\partial w_i}$

- Repeat the previous step until the network error is 'small enough' (convergence)

Note that, the more complex (i.e. high number of hidden layers and units) the NN model, the higher the risk of overfitting. Therefore, the generalizability of a trained network should always be assessed on an independent data set (testing of the network).

More information on Neural Networks can be found in Bullinaria[41] and Bishop.[88]

CHAPTER 4: RESULTS

Our challenge was to assess the place to be granted to Ovarian HistoScanning in clinical practise and for the work-up of adnexal masses. This research project has therefore adopted a design able to show how HistoScanning may be useful for clinical practise

It was investigated whether applying ovarian HistoScanning for the patients in the RMI grey zone would improve triage of women with ovarian masses. In addition, optimal cutoffs that define the RMI grey zone in which HistoScanning would further sort out the ovarian masses in order to improve RMI performance were assessed. This work is presented in section 1.1. of this chapter.

As the development of the RMI is in itself the result of a multivariate mathematical modelling exercise,[37] it seemed interesting to investigate whether we could develop multivariate models incorporating HistoScanning. It was thus investigated whether adding ovarian HistoScanning to multivariate mathematical prediction models for ovarian masses could improve model performance. And we were interested in finding a final model that was easy to apply (just like the RMI is simple and robust) and could therefore be useful for the less-experienced US-operators. These results can be found in section 1.2. of this chapter.

1. PUBLICATIONS

1.1.

DIFFERENTIAL DIAGNOSIS OF ADNEXAL MASSES: A SEQUENTIAL USE OF THE RISK OF MALIGNANCY INDEX AND A NOVEL COMPUTER AIDED DIAGNOSTIC TOOL

Evelien Vaes[1], Ranjit Manchanda[2], Philippe Autier[3], Rina Nir[4], Dror Nir[4], Harry Bleiberg[5], Annie Robert[1] and Usha Menon[2]

From [1]Université catholique de Louvain, Institut de Recherche Experimentale et Clinique, Epidemiology and Biostatistics Unit, Brussels, Belgium; [2]Gyneacoligical Oncology, UCL EGA Institute for Women's Health, London W1T 7DN, UK; [3]Advanced Medical Diagnostics, Waterloo, Belgium; [4]Jules Bordet Institute, Brussels, Belgium; [5]International Agency for Research on Cancer, Lyon, France

Submittted

1.1.1. ABSTRACT

Objective: To evaluate whether combining the Risk of Malignancy Index (RMI) and Ovarian HistoScanning, a novel computerised technique to interpret ultrasound data, can improve triage of women with adnexal masses.

Design: Retrospective analysis of prospectively collected data

Setting: Five European centres (Ovarian Histoscanning Clinical study Group) and EGA Institute for Women's Health, UCL, United Kingdom.

Population: 199 women with CA125, Ultrasound and HistoScanning data.

Methods: Ultrasound scores for RMI were obtained by blinded analysis of archived images of the 199 women. The following sequential test was developed: HistoScanning was modelled as a second-line test for RMI between a lower cut-off and an upper cut-off. The optimal combination of these cut-offs that together maximized the Youden index (Sensitivity + Specificity -1) was determined.

Main outcome measure: Sensitivities, specificities, Receiver Operating Characteristic (ROC) curves and Areas under the ROC curves.

Results: Using RMI at the standard cut-off value of 250 resulted in a sensitivity of 74% and a specificity of 86%. When RMI was combined with HistoScanning, the highest accuracy was achieved by using HistoScanning as a sequential second-line test for patients with RMI values between 105 and 2100. At these cut-off values, sequential use of RMI and HistoScanning resulted in mean sensitivity and specificity estimates of 88% and 95%, respectively.

Conclusions: Our data suggest that HistoScanning may have the potential to improve the diagnostic accuracy of RMI, which could result in a better triage of women with adnexal masses. Further prospective validation is warranted.

1.1.2. INTRODUCTION

Ovarian cancer is the leading cause of deaths from gynaecological malignancies. Annual age standardised incidence rates (European standard) across Europe vary between 7.2 and 19.3/100,000 while mortality rates are between 2.8 and 12.2/100,000.{Ferlay, 2010 219 /id This is often attributed to its 'silent' progression which leads to diagnosis in advanced stages (FIGO Stage III/IV) that have 5-yr survival rates of 15-27% only.
(http://info.cancerresearchuk.org/cancerstats/types/ovary/survival/)
There is increasing evidence that outcomes can be improved if cancer patients are managed by trained gynecological oncologists in multidisciplinary teams.{Le, 1997 78 /id;Eisenkop, 1992 79 /id;Chan, 2007 6 /id} This depends on accurate differential diagnosis of adnexal masses and appropriate timely referral of cancer patients to cancer centres, while general gynaecologists at district hospitals deal with benign lesions.
Pre-operative diagnosis of malignant ovarian neoplasms usually involves pelvic examination followed by TVS and assessment of the serum tumour marker CA125. Serum CA125 and TVS alone have high sensitivity but limited specificity for the differential diagnosis of adnexal masses.[5] Given the high prevalence of benign adnexal masses, even in postmenopausal women,[89] this results in low positive predictive values. Morphology based indices together with a variety of second line tests such as serum CA125 levels; Doppler blood flow indices and power Doppler have been investigated to improve performance characteristics. A number of multivariate models combining these factors and some incorporating additional clinical data[50-55] have been studied. Although results are promising, inconsistencies

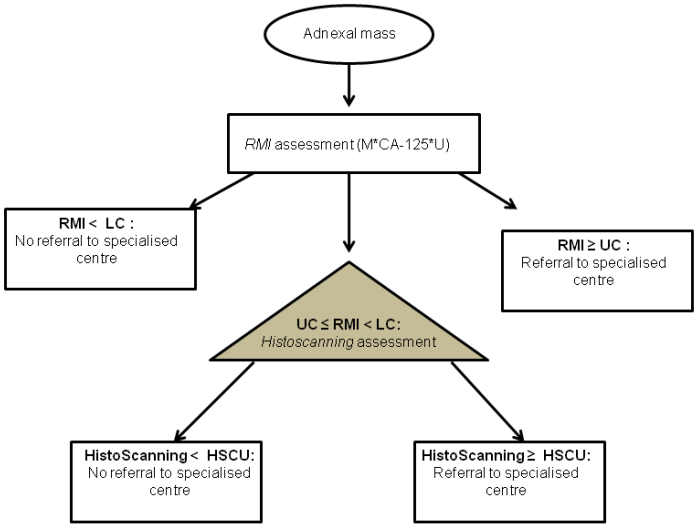
remain.[63,64] While such combinations outperform individual tests,[36] subjective evaluation of gray scale and Doppler images by experienced examiners outperforms most mathematical models.[68]

The Risk of Malignancy Index (RMI) is a standardized index commonly used for clinical evaluation of patients with an adnexal mass. In the United Kingdom it is routinely used to triage women with ovarian masses along appropriate referral pathways.[38-41] The RMI combines serum CA125 values, menopausal status and an US-morphology score and improves on the individual performance of each of these indices.[37] It compares well with most logistic regression models or artificial neural networks[57] and was advocated to be the prediction model of choice in preoperative assessment of adnexal masses.[58] The RMI has been validated in many European countries[38,40,59,60] with prospective studies reporting slightly lower sensitivity and specificity.

Ovarian HistoScanning is a new innovative US-based computerised tissue-characterization technique.[71] In its present form, Ovarian HistoScanning analyses US grayscale voxel data. This analysis is based on the fact that cancerous processes induce structural changes in the tissue texture which results in physical changes in the reflected ultrasound waves. Different tissue characterization algorithms were developed using mathematical methods that are optimized to quantify such changes in the backscattered waves.[70] The selection of the appropriate algorithms for a specific organ and their training is an empirical process. For Ovarian HistoScanning, three different characterization algorithms were selected and trained during clinical development studies in 2001-02 (unpublished data).

In this paper, we explore the possibility of using ovarian Histoscanning to improve the performance of RMI in the differential diagnosis of adnexal masses. We explore a strategy of using it as a sequential test in women with RMI values not associated with either very low or very high probability of cancer (figure 6) in order to improve clinical triage of patients with adnexal masses.

FIGURE 6: FLOW-CHART DESCRIBING THE STRATEGY FOR DIFFERENTIAL DIAGNOSIS OF ADNEXAL MASSES



1.1.3. PATIENTS AND METHODS

Patients

Data for this analysis was obtained from a cohort of patients, originally recruited to a prospective multicentric study[71] evaluating the accuracy of ovarian HistoScanning. A detailed description of the cohort and recruitment process has been described before.[71] Ovarian HistoScanning results were not used for patient management. Basic information collected included: age, menopausal status, detailed histopathology of surgical specimens, and stage of the cancers.

At the time of prospective data collection, clinicians were asked to classify the ovary as positive (malignant) or negative (benign) based on TVS-findings. RMI was not calculated at the time. Details of ovarian morphology at the time of initial TVS were not retained on the database, however, images were archived and stored. This makes it possible to retrospectively calculate RMI values for patients for whom CA125 data was available. Patients from the cohort for whom serum CA125 levels were available (199) were selected for this analysis.

Ultrasonography

A brief description of the Ultrasonography details that are important for this work are given underneath but more details are described in the earlier report.[71] At the time of the original study eligible women underwent TVS shortly before surgery by an expert sonographer with more than 10 years scanning experience. The Voluson 730 Expert scanner from General Electric equipped with a RIC5-9, 5-9 MHz Real Time 4D broadband endocavity transducer (Zipf, Austria) was used at all centres.

The examination involved conventional 2D-TVS for identifying ovaries and/or adnexal masses followed by a standardised 3D-TVS procedure to capture the 3D image of all structures. When needed, several 3D acquisitions were made to cover a mass. Ovaries and lesions were annotated and marked as regions of interest on the 2D-scan by the individual operators. All acquired data were anonymised, saved on a CD-ROM and send to the study coordinator. An independent expert matched the 2D annotated images to the structure depicted by the 3D-TVS data to determine the extent of the volume(s) of interest (VOI) to be analyzed by HistoScanning.

Ovarian HistoScanning™

HistoScanning analysis was applied to the raw (gray-scale values) voxel data that had been saved on the CD-ROM's. HistoScanning successively analyzed the VOI that were indicated on the 3D-scan by the operator. The characterization algorithms calculated a score value for each VOI in a scan. This was a continuous value ranging between 0 and 125. In this work, the final HistoScanning result for one individual patient was the maximum of this score over all available VOIs for that patient. This analysis was blinded with regard to the ultrasonographers' interpretation, CA125-value and the pathology results.

RMI-calculations

All archived 2D- and 3D-ultrasound images were reviewed at University College London by two independent experienced ultrasonographers (Mrs. G.Fletcher and Mrs. K.Ford). The ultrasonographers were blinded to the HistoScanning, CA125 and pathology results and assessed each US image for the presence of a number of key morphological features (multi-locularity of any cyst, presence of solid areas,

evidence of metastases, presence of ascites and bilaterality of lesions) necessary for the RMI.[37]

The RMI index is defined as the product of menopausal status (M), CA125 level and US-score (U). The US-score is based on the features described above each of which score one point. Currently there are 3 versions of RMI, each varying slightly in the method of scoring from the other. The original RMI described by Jacobs et al (RMI_1)[37] and 2 modifications (RMI_2 and RMI_3) by Tingulstad et al.[40,59] In RMI_1, U=0 for an US-score=0, U=1 for US-score=1 and U=3 for US-score≥2; M=1 for premenopausal status and M=3 for postmenopausal status. In RMI_2: M=4 for postmenopausal status; US-score≤1 leads to U=1 and for an US-score≥2, U=4. The changes in RMI_3 (compared to RMI_2) include, M=3 instead of 4 for postmenopausal status; and U=3 (instead of 4) when US-score≥2.

Analysis

Our aim was to develop a sequential strategy using RMI and ovarian HistoScanning that would enhance the accuracy of the RMI for differential diagnosis. The following scheme was adopted for this purpose (see also the pathway in figure 6): Patients with an RMI below a lower cut-off (LC) were considered to have a negative (non malignant) test outcome and patients with an RMI ≥ an upper cut-off (UC) were considered to have a positive (malignant) test outcome. The HistoScanning score was used to determine the test outcome for patients with an RMI value in between the upper and lower cut offs (LC≤RMI<UC), the outcome was positive when the HistoScanning score was ≥ a cut-off point (HSCU) and negative otherwise. Subsequently, we identified the optimal combination of the 3 cut-offs, LC, UC and HSCU that would together give the highest sensitivity and specificity for this sequential test (RMI followed by HistoScanning). This combination of cut-offs were then validated on an independent set of data not used for training.

A training set of 70% of the patients, selected randomly were used to identify these 3 cut-offs LC, UC and HSCU. This was done by varying the LC and UC for RMI and HSCU simultaneously in predefined ranges that made sense. The LC was varied between 5 and 265 with steps of 20, the UC between 500 and 2500 with steps of 50, and the HSCU was considered in its full range with steps of 5. Then the best combination was defined as the one that maximised the Youden Index (Y=sensitivity+specificity-1). This operating point was validated by calculating sensitivity and specificity in the remaining 30% of patients (testing set).

To test the variability that might exist in our data set, this random subsampling process was repeated 100 times so that 100 randomly selected training and testing sets were created.[90] At every iteration step the testing set didn't contribute to the optimization of the operating point and served as an independent assessment.

Additional note. *This implies that the following happened 100 times: a specific operating point was found to maximise the Youden Index in a resampled training set and subsequently this operating point was validated on the corresponding testing set. As a consequence, we had to decide which operating point to keep as a final result and how to report a 'summary accuracy' for this operating point. So it was decided to do this as follows: The number of resamplings (out of the 100) in which a certain operating point appeared to maximize the Youden index was reported. This was called the operating point frequency (OPF) and the operating point with the highest OPF was retained. This means that this operating point was validated in a number of testing sets equalling OPF and we then reported a mean sensitivity and specificity (together with 95% distribution intervals) at this operating point over this number of testing sets. It was a purposive choice to report the 'mean' accuracy over OPF testing sets and not over the total of 100 because the 100-OPF other testing sets correspond to other runs to find an operating point and including them might lead to overoptimistic results.*

Statistical analysis was performed using SAS version 9.1. and Matlab version R2008a. Means or medians of continuous variables were compared using ANOVA with F-test or the Kruskal-Wallis test. The Chi-square test was used to compare proportions. To measure the diagnostic accuracy, sensitivities and specificities were determined at fixed operating points. ROC curves were constructed and the area under the ROC curve (AUC) was calculated using the trapezoidal rule. The AUCs were compared using the nonparametric approach of DeLong et al.[81]

1.1.4. RESULTS

Patient characteristics

Of the 199 patients, 2 were excluded due to lack of agreement between the pathology report and the study CRF. In the remaining 197 patients, there was no significant age difference between women found to have malignant, benign and normal ovaries. However, there were more postmenopausal women in the normal

and benign when compared to the malignant group. Serum CA125 levels were significantly higher in women with malignant masses (table 6).

TABLE 6: PATIENT CHARACTERISTICS

Variable	**Malignant (N=98)**	**Benign (N=72)**	**Normal (N=27)**	p-value
Age (yrs) mean±SD	57 ± 12	58±10	62±13	0.095
CA 125 level (U/mL) Median (range)	2323 (7-32986)	12 (4-487)	17 (3-356)	<0.001
Menopausal-status %postmenopausal	66%	76%	89%	0.047

The 197 patients had 291 adnexal masses, 125 non malignant and 166 malignant. Details of ovarian histology are given in table 7. The majority (78%) of malignant tumours were primary invasive ovarian cancers and borderline ovarian cancers. 37% of all ovarian cancers and 94% of borderline cancers were early stage cancers (FIGO stages I and II). Overall, 69% of malignant tumours were of epithelial subtype.

TABLE 7: HISTOPATHOLOGY OF TUMOURS

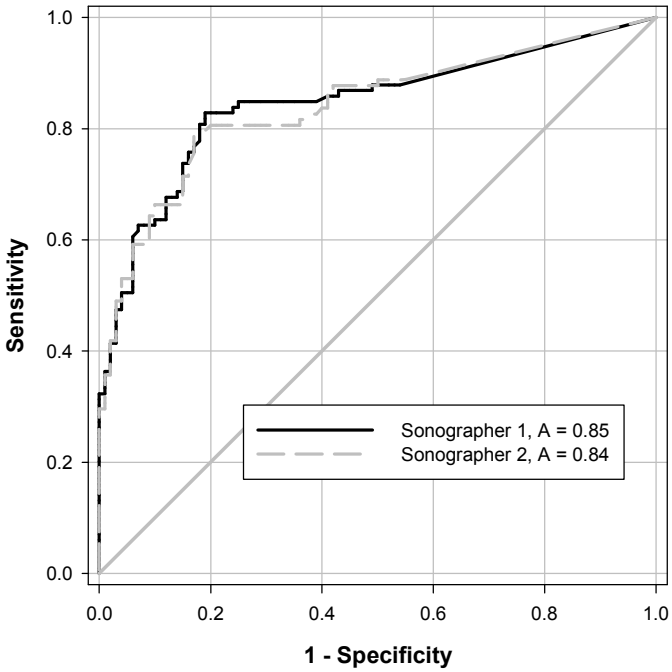
Type of tumour	N	%
Non Malignant (n=125)		
Ovarian cyst	30	24
Cystadenoma	28	22
Teratoma	4	3
Cystadenofibroma	15	12
Corpus Luteum cyst	7	6
Fibroma	6	5
Brenner Tumor	6	5
Corpus albicans	14	11
Hemorrhagic adnexal cyst	1	1
Thecoma	2	2
Endometriosis	8	6
Unknown	4	3
Malignant tumours (n=166)		
Ovarian cancer (OC)	89	54
Borderline ovarian cancer (BOC)	25	15
Secondary ovarian metastasis from cancer of another organ	16	10
Peritoneal carcinomatosis	15	9
Cancer of an organ invading the pelvis	4	2
Unknown*	17	10

* Note that about 50% of these malignant tumours of unknown histological origin had missing data and where therefore not contributing to our analysis and the other 50% had very high RMI values (≥ UC) and were thus always directly correctly classified by RMI (the first line test)

RMI

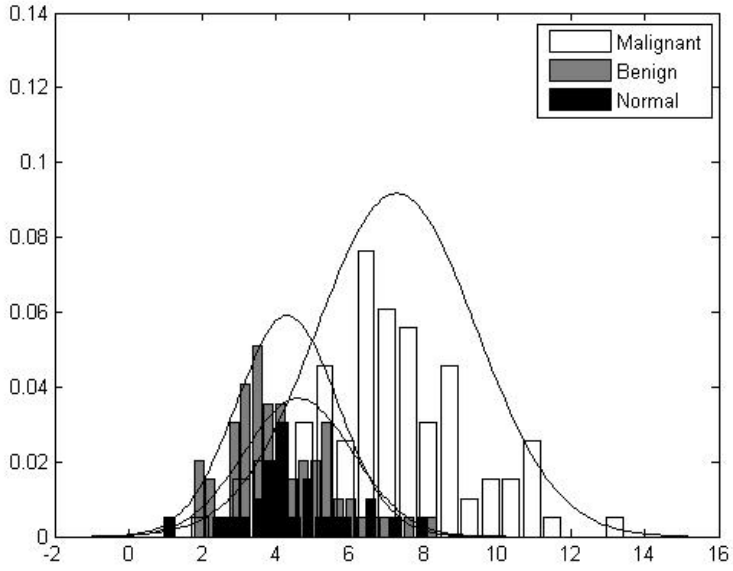
At University College London, two sets of US scores for each of the 197 patients were obtained independently from the two sonographers. The two RMI_1-scores that resulted from multiplying these scores with CA125 values and menopausal status showed very similar ROC-curves and AUC (see figure 7) on our patient population. Therefore it was decided to use the RMI based on the US-score of the 2^{nd} sonographer (random choice) for further analyses.

FIGURE 7: ROC CURVES OF RMI1 FOR BOTH SONOGRAPHERS



The ROC curves for the RMI_1, RMI_2 and RMI_3 were obtained and the respective AUC were estimated to be 0.85 (95%CI = [0.78;0.91]) for RMI_1, 0.87 (95%CI = [0.81;0.92]) for RMI_2 and 0.88 (95%CI = [0.83;0.93]) for RMI_3. The difference between the 3 AUCs was not statistically significant (p=0.99). Additionally, the ROC of the 3 RMI barely changed when the 27 normal patients were not included in the sample, thus confirming that the strength of the RMI is its ability to differentiate between malignant and non malignant adnexal masses. This can also be seen from the empirical histograms of the log RMI_2 score (figure 8) in which the normal and benign groups completely overlap.

FIGURE 8: EMPIRICAL HISTOGRAMS OF THE LOG RMI2 SCORE FOR THE DIFFERENT PATHOLOGY GROUPS



The sensitivity and specificity of the three RMIs on the ensemble of the 197 patients, was evaluated. Among the three RMIs, RMI_2 had the best sensitivity at the cut-off of 250, 82%(95%CI=[73;89]), and the specificity was 81%(95%CI=[72;88]). RMI_1 and RMI_3 at the cut-off of 250 had sensitivities of 68%(95%CI=[58;77]) and 74%(95%CI=[64;82]), respectively; and specificities of 89%(95%CI=[81;94]) and 86%(95%CI=[77;92]), respectively.

Sequential use of RMI and HistoScanning

Table 8 presents the results of the 100 resamplings carried out to find the combination of the 3 cut-offs (LC, UC and HSCU) that maximised the Youden Index. All combinations of RMI with HistoScanning showed improvements over RMI alone. The strategy combining RMI_3 and HistoScanning led to the highest mean accuracy in the iteratively selected testing sets (table 8). For RMI_3 with HistoScanning, the best combination of the 3 cut-offs (LC, UC, HSCU) was (105, 2100, 20) with mean sensitivity and specificity estimates of 88%(95%DI=[79;98]) and 95%(95%DI=[87;99]) respectively (testing).

TABLE 8: RESULTS OF A REPEATED RANDOM SUBSAMPLING PROCEDURE ON RMI1, RMI2 AND RMI3 SUPPLEMENTED WITH HISTOSCANNING TO FIND THE OPERATING POINT THAT MAXIMIZES THE YOUDEN INDEX.

	Training Set (70% of patients)		Testing set (30% of patients)	
Test	**Cut-offs[§] (LC,UC,HSCU)**	**OPF***	**Sens† [95%DI]**	**Spec† [95%DI]**
RMI_1 followed by HistoScanning	(25,2100,25)	33%	87% [78;97]	84% [73;94]
RMI_2 followed by HistoScanning	(165, 2050,20)	50%	84% [76;91]	93% [86;99]
RMI_3 followed by HistoScanning	(105,2100,20)	44%	88% [79;98]	95% [87;99]

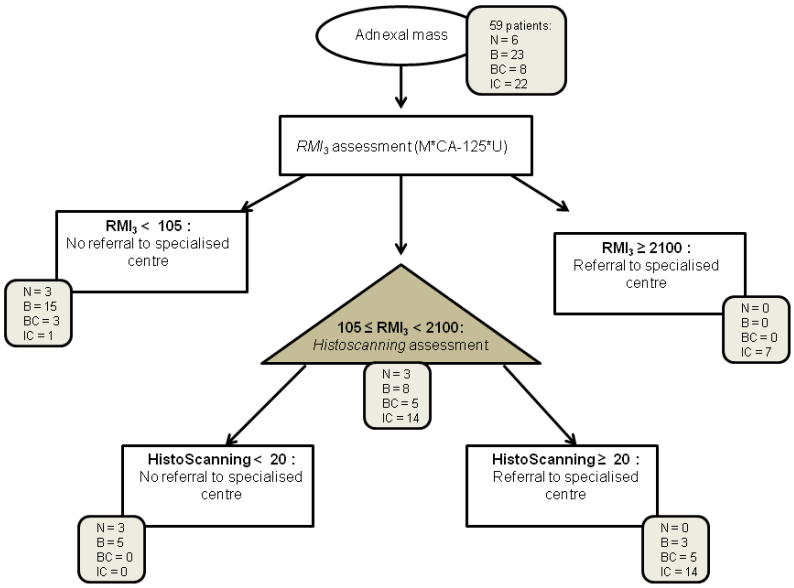
§ Combination of the 3 cut-off points; the first 2 are the lower (LC) and upper (UC) cut-off points for RMI and the third is the HistoScanning cut-off point (HSCU). HistoScanning is performed when LC ≤ RMI < UC.
*** Operating point frequency = Percentage of resamplings in which the 3 cut-off points maximize the Youden index.**
† Mean sensitivity/specificity with 95% distribution intervals over the number of resamplings in which the operating point maximized the Youden Index on the training data (=OPF). This served as a validation of the retained operating point.

Working example illustration

Since the sequential use of RMI_3 with HistoScanning achieved the best performance at optimised cut-offs of (105, 2100, 20), we applied this sequential strategy on 59 patients of one independent testing set (1 out of the 100 splits) (figure 9). Application of the RMI_3 would directly find 22 women with low (RMI_3<**105**) and 7 women with high (RMI_3≥**2100**) probability of adnexal cancer. The invasive carcinoma found among patients with RMI_3<105 was a 44 year old woman with CA125 = 36 U/ml and RMI_3 = 36, diagnosed with ovarian metastases from breast cancer in both ovaries. For the 30 women in the RMI_3 'grey-zone' (**105**≤RMI_3<**2100**), HistoScanning could further find out 22 women that should be referred to a specialised centre (women with HistoScanning score ≥ **20**). These 22 included 19 women with cancer and three with benign tumours. The 8 other women in the RMI_3 'grey zone' had HistoScanning values < **20** and could be

followed up by their doctor. The pathology of these women was the following: 3 had normal ovaries and 5 had benign tumours.

FIGURE 9: STRATEGY USING SEQUENTIALLY THE RMI_3 AND OVARIAN HISTOSCANNING FOR THE DIFFERENTIAL DIAGNOSIS OF ADNEXAL MASSES WITH RESULTS FROM ONE TESTING SET.



Abbreviations: N = normal, B = benign, BC = borderline cancer, IC = invasive cancer
In the small light gray square-boxes underneath, pathology results of patients in one testing set (=59 patients) are given as an example of how they are divided along the proposed pathway.

1.1.5. DISCUSSION AND CONCLUSIONS

This is the first study to explore the use of a novel US-based technique, Ovarian HistoScanning, to supplement the RMI for the differential diagnosis of adnexal masses. The study was able to identify a RMI 'grey-zone' in which HistoScanning could be used as a second line test to improve performance characteristics. The highest accuracy (mean sensitivity and specificity of 88% and 95%, respectively) was achieved using a sequential strategy of RMI_3 with HistoScanning as a second line test for RMI_3 cut-off values of ≥105 and <2100.

In this study, there was no significant difference in the performance characteristics of RMI_1, RMI_2 and RMI_3. This is consistent with earlier reports.[91,92] Only one study has reported improved accuracy of RMI_2 over RMI_1.[60]

External validation of the developed requires an independent testing set but this was impossible at this point in time. A possible option would have been to do one single split of the data in training and testing set (internal validation). However, given the high proportion of malignant cases and the moderate sample size, a single randomly chosen testing set would have been subject to chance (accuracy results in favour of the new method or in disfavour of the new method). In order to minimise this selection bias, we used a repeated subsampling approach that permits understanding the variability that might arise from data representative of the population under study. The cut-off determined using this method minimises selection bias, is more robust, likely to represent a true finding and be reproducible. However, the down side is that the results presented on one single testing set based on one single random data split (as in figure 9) should be considered with care.

Although we are cautiously optimistic regarding the findings using internal validation, further prospective validation on a totally independent data set is awaited.

While the data used in this work, were prospectively collected, they were obtained from a clinical study designed to study the accuracy of HistoScanning alone, without any attention to RMI. RMI was analysed retrospectively for the aim of our study. Two experienced sonographers reviewed archived 2D- and 3D-TVS images to derive morphological scores. This could be considered suboptimal. However, the images were always acquired by experts with at least 10 years experience and assessing such archived US-images has been noted to improve observer accuracy.[66] This is confirmed by the finding that the range of RMI scores for these patients are in line with other RMI validation studies.[40,59,61]

The quality of subjective assessment of adnexal masses on US-images varies widely, depending on operator expertise.[66,93] This variability may come from 2 sources: firstly, the definition of the region of interest and secondly, the interpretation of the morphological characteristics of the selected region. Ovarian HistoScanning being an automated computerised technique has the potential to significantly reduce this observer variability for image interpretation. Even so, Ovarian HistoScanning software is not completely without user dependency, as it requires the identification of the VOI, which falls under the first source of variability.

Unfortunately, within the context of the design of the clinical trial, it was impossible to assess inter-user variability related to the definition of the VOI. For a future application of our proposed strategy, the assessment of masses on the US-image as well as the initiation of the HistoScanning analysis on this image should ideally be done by the same operator during the same examination. Addressing the user/expertise dependency in such a set-up would thus be important.

Lucidarme et al[71] showed TVS sensitivity increased from 66% to 92% when TVS and HistoScanning were combined in parallel. However, RMI remains the current standard for differential diagnosis in many centers. Our study is thus novel as it for the first time shows that Histoscanning can be used as a sequential second line test to improve performance of RMI and more accurately triage women with adnexal masses.

Van Trappen et al[61] reported the use of a sequential strategy of two-staged second line tests comprising expert US and if needed subsequent MRI for patients with RMI_1 cut-off values of ≥25 and <1000 which achieved a sensitivity of 94% and specificity of 90%. The major advantage of the current strategy, applying HistoScanning for patients with RMI_3 cut-off values of ≥105 and <2100, is that similar accuracy is achieved without the need for patients to undergo a further expertise dependent and time consuming second line test (e.g. MRI or CT) as HistoScanning only involves computerised analysis of data already acquired at the time of the initial ultrasound. There are other second line tests with proven efficacy that are relatively easy to perform like color Doppler.[94] For this study we had no Doppler data available. For future investigations, it could be interesting to study a sequential approach with RMI and HistoScanning also containing Doppler information.

In conclusion, this paper is the first evidence of a potential role of Ovarian HistoScanning in improving the performance of RMI for the differential diagnosis of ovarian masses. The proposed strategy increases the proportion of patients appropriately referred to cancer centres /managed by general gynaecologists and has the potential to reduce costs by avoiding additional expensive second line tests, such as MRI/CT. Future research should include comparisons with other models in independent data sets and multicentre prospective clinical trials in a general population which incorporate cost-effectiveness/benefit analysis.

1.1.6. ADDITIONAL RESULTS AND DISCUSSION

1) At University College London, two sets of US scores for each of the 197 patients were obtained independently from the two sonographers. The stratified kappa and weighted kappa coefficients were calculated to measure the inter-observer agreement for this US-score and the Mc Nemar test was used to assess symmetry among the discordant pairs.

Table 9 reports Cohen's kappa and weighted kappa coefficients for the different pathology strata. In addition, the overall stratified kappa and weighted kappa were 0.38 (95%CI = [0.27;0.49]) and 0.45 (95%CI = [0.34;0.56]), respectively. The kappa reflects a slight agreement between the readers but the weighted kappa shows that the agreement improves to a moderate one when the close disagreements (i.e. 0 versus 1 and 1 versus ≥2) are considered less heavily then the distant disagreements (i.e. 0 versus ≥2). Symmetry was detected between the discordant pairs and only 10 distant disagreements exist with a 5/5 ratio (table 9). Additionally, the two RMI_1-scores that resulted from multiplying these scores with CA125 values and menopausal status showed very similar ROC-curves and AUC (see figure 7) on our patient population. Consequently, the agreement between the 2 sonographers is deemed acceptable and the US-score of the 2nd sonographer was chosen at random for further analyses.

TABLE 9: CROSS TABULAR FREQUENCIES AND CONCORDANCE MEASURES IN THE DIFFERENT PATHOLOGY STRATA OF THE US-SCORE ATTRIBUTED BY 2 SONOGRAPHERS

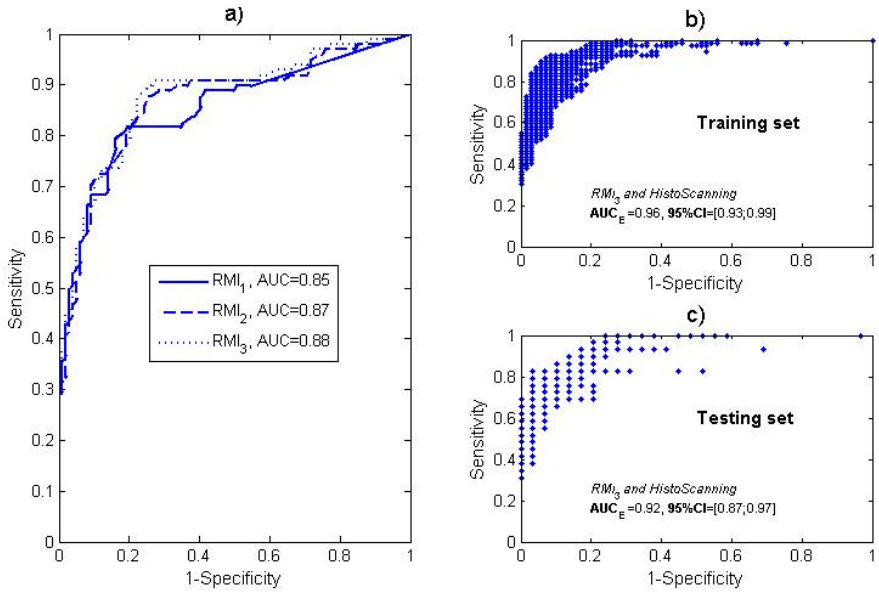
	US Score of S_1			
US Score of S_2	**0**	**1**	**≥2**	
0	40	10	5	55
1	12	22	17	51
≥2	5	22	64	91
Total	57	54	86	197

	Concordance	
Pathology	**$K^{\S}$ [95%CI]**	**K_w^{*} [95%CI]**
Malignant	0.32 [0.15;0.49]	0.38 [0.20;0.56]
Benign	0.36 [0.20;0.53]	0.45 [0.28;0.62]
Normal	0.58 [0.31;0.85]	0.64 [0.37;0.91]

Abbreviations: US=Ultrasound, S_1=Sonographer 1, S_2=Sonographer 2
§Kappa coefficient per pathology stratum
*weighted Kappa coefficient per pathology stratum

2) In order to identify the best combination of the three cut-off points (LC, UC, HSCU) for the sequential test RMI followed by HistoScanning, we have applied a random subsampling process. In addition, this process was repeated 100 times. In each of these 100 runs a training set of randomly selected patients (70% of sample) was selected to find LC, UC and HSCU by varying them simultaneously in predefined ranges that made sense. Every variation of one of the three cut-offs leads to a different sensitivity and specificity. These pairs of sensitivity and specificity obtained empirically at every variation of LC, UC or HSCU, can also be visualized by an empirical ROC curve by simply plotting the empirical sensitivities against 1-specificties. The resulting ROC curve will not consist of 1 jagged line (of points) as we are used to but is rather a collection of points. This is because we are handling 3 cut-offs simultaneously so we can find points that have a lower 'sensitivity' or lower '1-specificty' than their predecessors. An example of such an ROC curve that visualizes the search for the best combination of the three cut-offs is given in figure 10.

FIGURE 10: ROC CURVES OF RMI_1, RMI_2, RMI_3 AND ROC POINTS OF RMI_3 WITH HISTOSCANNING



Abbreviations: RMI=Risk of Malignancy index, AUC=Area under the curve

Figure 10.a) ROC curves and AUC of the 3 successive RMI

Figure 10.b) ROC points for RMI_3 with HistoScanning in one training set (=138 patients). The AUC is calculated with respect to the envelope of these ROC points.

Figure 10.c) ROC points for RMI_3 with HistoScanning in one testing set (=59 patients).

1.2.

MATHEMATICAL MODELS TO DISCRIMINATE BETWEEN BENIGN AND MALIGNANT ADNEXAL MASSES: POTENTIAL DIAGNOSTIC IMPROVEMENT USING OVARIAN HISTOSCANNING

Evelien Vaes[1] (MSc), Ranjit Manchanda[2] (MRCOG), Rina Nir[3] (MSc), Dror Nir[3] (PhD), Harry Bleiberg[4] (PhD), Philippe Autier[5] (MD), Usha Menon[2] (FRCOG) and Annie Robert[1] (PhD)

From [1]Université catholique de Louvain, Institut de Recherche Experimentale et Clinique, Epidemiology and Biostatistics Unit, Brussels, Belgium; [2]Gyneacoligical Oncology, UCL EGA Institute for Women's Health, London W1T 7DN, UK; [3]Advanced Medical Diagnostics, Waterloo, Belgium; [4]Jules Bordet Institute, Brussels, Belgium; [5]International Prevention Research Institute, Lyon, France

International Journal of Gynecological Cancer. In press

1.2.1. ABSTRACT

Purpose: Accurate preoperative clinical assessment of adnexal masses can optimize outcomes by ensuring appropriate and timely surgery. This paper addresses whether a new technology, ovarian HistoScanning, has an additional diagnostic value in mathematical models developed for the differential diagnosis of adnexal masses.

Patients and Methods: TVS-based morphological variables were obtained through blinded analysis of archived images in 199 women enrolled in a prospective study to assess the performance of ovarian HistoScanning. Logistic regression (LR) and neural network (NN) models including these variables, clinical- and patient-data along with the HistoScanning score (HSS, range 0-125, based on mathematical algorithms) were developed in a learning set (60% patients). The remaining 40% patients (evaluation set) were used to assess model performance.

Results: Of all morphological and clinical variables tested, serum CA125, presence of a solid component and HSS were most significant and used to develop the LR-model. The NN model included all variables. The novel variable, HSS, offered significant improvement in the LR- and NN-models' performance. The LR- and NN-models in an independent evaluation set were found to have AUC: 0.97

(95%CI=[94;99]) and 0.93 (95%CI=[88;98]), sensitivities: 83% (95%CI=[71;91]) and 80% (95%CI=[67;89]) and specificities: 98% (95%CI=[89;99]) and 86% (95%CI=[72;95]), respectively. In addition, these models showed an improved performance when compared to 3 other existing models (p values all <0.05)

Conclusion: This initial report shows a clear benefit of including ovarian HistoScanning into mathematical models used for discriminating benign from malignant ovarian masses. These models may be specifically helpful to the less experienced examiner. Future research should assess performance of these models in prospective clinical trials in different populations.

key words: differential diagnosis, ovarian cancer, multivariate mathematical models, HistoScanning, ultrasound

1.2.2. INTRODUCTION

Surgery by consultant gynaecological oncologists has been shown to lead to improved outcomes in women with ovarian cancer.[26] Accurate preoperative assessment of adnexal masses will help correctly triage women at high risk of malignancy for treatment by gynaecological oncologists. Furthermore this permits the larger proportion of patients with benign masses to be managed conservatively or by other gynaecologists.

The use of TVS to characterize adnexal masses is well established and over the years a number of scoring systems based on ultrasound determined morphological characteristics have been developed.[32,35] However, there is still no standardized index[5]. The addition of second line tests such as serum CA125 levels, Doppler features and clinical data to TVS morphological indices have been shown to outperform individual tests.[36] The Risk-of-malignancy index (RMI)[37] is such an index which is routinely used in the United Kingdom.[38] Other researchers have used multivariate mathematical models to determine the most important variables for differential diagnosis.[42,51-53,95] Performance of some models has been confirmed at validation[96] but for most of them inconsistencies remained at validation studies.[5,63] Moreover, most mathematical models do not offer much advantage over the RMI[57,65] nor do they outperform subjective evaluation by an experienced examiner.[64,68]

We explored whether a new computerized tissue characterization technology, ovarian HistoScanning,[71] adds an additional diagnostic value to prediction models.

Cancerous processes induce structural changes in the tissue texture which results in physical changes in the reflected ultrasound waves. HistoScannings' tissue characterization algorithms were optimized to quantify such changes in the backscattered waves. For Ovarian HistoScanning, three algorithms were selected and trained during clinical development studies in 2001-02 (unpublished data)

1.2.3. Patients and Methods

Patients

This report involves patients recruited to a prospective study evaluating ovarian HistoScanning. The original study protocol was approved by the Ethical Committees of all participating centres and was managed according to ICH-GCP (http://ichgcp.net/). Eligibility criteria were: women ≥ 18 years, scheduled for the surgical removal of one or two ovaries due to (a) a suspicious adnexal mass or (b) prophylactic oophorectomy or as part of hysterectomy for a uterine pathology. Exclusion criteria included: past history of ovarian cancer, past history of chemotherapy or radiotherapy and previous pelvic surgery. The latter helped avoiding alterations to ultrasound-data due to changes in the tissue morphology. All participants gave written informed consent. Baseline information included: age, menopausal status, histo-pathology of surgical specimens and FIGO staging of cancers. Only patients with CA125 measurements (199 out of 383) were included in this analysis.

Ultrasonography

TVS was performed on all eligible women shortly before surgery by an expert with more than 10 years of scanning experience. The Voluson 730 Expert scanner from General Electric equipped with a RIC5-9, 5-9 MHz Real Time 4D broadband endocavity transducer (Zipf, Austria) was used in all centres.

The examination consisted of conventional 2D-TVS for identifying ovaries and/or adnexal masses followed by a standardized 3D-TVS procedure to capture the 3D image of all structures present. Ovaries and lesions were annotated and marked as regions of interest on the 2D-scan by the individual operators. All acquired data were anonymized, saved on a CD-ROM and sent to the study coordinator. An

independent blinded expert matched the 2D annotated images to the structure depicted by the 3D-TVS to determine the volume(s) of interest (VOI) to be analyzed by HistoScanning.

Ovarian HistoScanning results

HistoScanning analysis was applied to the 3D raw voxel (gray-scale) data, saved on the CD-ROM's.

HistoScanning's algorithms successively analyzed the VOI, indicated on the 3D-scan by the operator, and calculated a value (ranging between 0 and 125) for each VOI in a scan. The HistoScanning-score of an individual patient was defined as the maximum value over all her available VOIs. This analysis was blinded with regard to the TVS-interpretation, CA125-value and the pathology results.

Retrospective assessment of TVS-morphological variables

The 2D- and 3D-ultrasound images of the 199 eligible patients were reviewed at University College London by an independent experienced ultrasonographer. The ultrasonographer was blinded to HistoScanning, CA125 and pathology results. She assessed each TVS-image for the morphological features (multi-locularity of any cyst, presence of solid areas, evidence of metastases, presence of ascites and bilaterality of lesions) that form the RMI.[37] In addition, a number of TVS-morphological features were also assessed for each ovary/adnexal lesion detected: cyst dimensions, cyst wall structure (smooth or irregular), echogenicity of cystic fluid, cyst structures (septae and/or papillations), septae thickness, size of papillation and overall impression of the lesion (as defined by the International Ovarian Tumour Analysis Criteria).[62] These features are part of the list of features analyzed in the UKCTOCS screening program (www.ukctocs.org.uk).[97]

Analysis

Statistical analysis was performed using SAS 9.1, SAS Enterprise Miner release 4.3 and SigmaPlot 10.0. Continuous variables are reported as medians with ranges and discrete variables as proportions. For continuous variables, pathology groups were compared using Kruskal-Wallis' test or Mann-Whitney's U-test (with Bonferroni-

correction) when the sample size in the normal pathology group was too small (n≤3) to be used in the comparison. The Chi-square test was used for discrete variables. Spearman's correlation coefficient was used to assess correlation between variables.

The data set was split into learning set (LS) and evaluation set (ES) by randomly selecting 60% of patients into the LS and the remaining 40% into the ES. Multivariate models were developed in the LS only. Models were subsequently tested in the ES and also compared to models reported in the literature.

Variables that were significant in univariate analysis or that had proven clinical interest were submitted to a multivariate logistic regression with pathology (malignant vs. non-malignant) as response variable. After checking for interactions and multicollinearity, a backward elimination procedure based on Akaike's information corrected criterion (AICC) was used to select the variables with a diagnostic value. The final model is referred to as the LR-model. Finally, 100 bootstrap-resampled data sets were created from the LS in which AICC selection was repeated to perform an internal validation of the input selection. As a result, inclusion frequencies of the variables entered in the full model as well as the most frequent final model among the 100 runs are reported.

Additionally, a neural network (NN) model[88] was developed. A feed-forward multilayer perceptron model was used, consisting of an input layer of variables connected to layer(s) of hidden neurons which is again connected to an output neuron. The neurons take a weighted sum of their inputs and put this value trough an activation function. Weights of the inputs of each neuron are estimated on the LS through a parametric error-minimization procedure. We used the logit function as activation function for the output neurons.The model complexity (e.g. number of hidden layers/neurons) was set by trial and error using different NN-models in our LS by splitting the LS into a training (70% patients, randomly chosen) and a test set (remaining 30% patients). The model with the lowest misclassification error on the test set was then selected.

To assess the added predictive ability of HistoScanning, the integrated discrimination improvement (IDI) with corresponding p-value, as recently proposed by Pencina et al,[87] was calculated for the LR- and NN-models.

The measured variables in our data allowed us to validate 3 other models for the differentiation of ovarian masses: the RMI,[37] the Logistic regression model, LR2, and the neural network model, NN2, developed by Timmerman et al.[53] The RMI is

defined as the product of menopausal status (M), CA125 and US-score (U) in which M=1 for premenopausal status and M=3 for postmenopausal status; U=0 for US-score=0, U=1 for US-score=1 and U=3 for US-score≥2. The probability of malignancy predicted by the LR2- and the NN2-model was taken from the mathematical formulas in Timmerman et al:[53]

$$\text{LR2: } z = -1.0889 + 0.5948(meno) + 0.0205(CA125) + 0.5446(ascites) - 0.7620(unilocular) - 1.1606(smooth) + 1.5049(papill) + 0.7633(bilateral)$$

and

$$\text{NN2: } z = -3.8616 + 2.9753\left(\frac{1}{1+\exp(-z_1)}\right) + 4.1980\left(\frac{1}{1+\exp(-z_2)}\right)$$

in which

$$z_1 = -0.5485 - 1.0792(meno) + 1.9383(CA125) + 0.7124(ascites) - 1.2664(unilocular) - 1.3741(smooth) + 0.8298(papill) + 1.5316(bilateral)$$

$$z_2 = -1.8129 + 1.0766(meno) + 0.1376(CA125) + 1.0112(ascites) - 0.8320(unilocular) - 1.6941(smooth) + 2.9541(papill) + 1.4654(bilateral)$$

ROC curves were constructed for our LR- and NN-models as well as for the earlier published RMI, LR2- and NN2-models. The area under the ROC curve (AUC) was calculated using the trapezoidal rule. The AUCs were compared using the nonparametric approach of DeLong et al.[81] A cut-off performance was also presented with corresponding sensitivity and specificity.

Our main analysis focused on the level of the ovary (full LS/ES), but modelling was also done on the level of the patient (reduced LS/ES) by taking the most severe pathology and maximum value over both ovaries for other variables as 1 data entry.

1.2.4. RESULTS

Two patients were excluded because of disagreements between the pathology report and the study's case report form. The remaining 197 patients had 291

adnexal masses (125 benign, 166 malignant). Description of tumour histology is given in table 10. A detailed histologic classification of the 166 malignant ovarian tumours is provided in table 11. Of all ovarian and borderline cancers, 37% and 94% were early stage cancers (FIGO stages I/II), respectively.

TABLE 10: HISTOPATHOLOGY OF THE 291 MASSES IDENTIFIED IN 197 WOMEN

Type of tumour	**n**	**%**
Non malignant Tumours (n=125)		
Ovarian cyst	30	24.0
Cystadenoma	28	22.4
Teratoma	4	3.2
Cystadenofibroma	15	12.0
Corpus Lutheum cyst	7	5.6
Fibroma	6	4.8
Brenner Tumor	6	4.8
Corpus albicans	14	11.2
Hemorrhagic adnexal cyst	1	0.8
Thecoma	2	1.6
Endometriosis	8	6.4
Unknown	4	3.2
Malignant tumours (n=166)		
Ovarian cancer (OC)	89	53.6
Borderline ovarian cancer (BOC)	25	15.1
Secondary ovarian metastasis from cancer of another organ	16	9.6
Peritoneal carcinomatosis	15	9.0
Cancer of an organ invading the pelvis	4	2.4
Unknown*	17	10.2

* Note that the majority of these malignant tumours of unknown histological origin had missing data for at least one of the variables used in the multivariate mathematical modelling and where therefore not contributing to our analysis.

TABLE 11: WHO ICD-O HISTOLOGIC CLASSIFICATION OF THE 166 MALIGNANT OVARIAN TUMOURS

Malignant tumours (n=166)	**n**	**%**
Ovarian cancer (n=89)		
Surface epithelial tumours	80	89.9
Sex-cord stromal tumours	2	2.2
Unknown	7	7.9
Borderline ovarian cancer (n=25)		
Surface epithelial tumours	24	96.0
Germ-cell tumours	1	4.0
Secondary ovarian metastasis from cancer of another organ (n=16)		
Surface epithelial tumours	2	12.5
Secondary (metastatic) tumour	12	75.0
Unknown	2	12.5
Peritoneal carcinomatosis (n=15)		
Surface epithelial tumours	8	53.3
Secondary (metastatic) tumour	1	6.7
Unknown	6	40.0
Cancer of an organ invading the pelvis (n=4)		
Secondary (metastatic) tumour	1	25.0
Unknown	3	75.0

Malignancy was found to be associated with significantly higher CA125 levels and US-score≥2 (table 12).

TABLE 12: CLINICAL AND MORPHOLOGICAL CHARACTERISTICS OF THE 197 PATIENTS

Variable		**Normal†** **(N=27)**	**Benign†** **(N=72)**	**Malignant†** **(N=98)**	p-value
Age (yrs) mean±SD		62±13	58±10	57 ± 12	0.095
CA 125 level (U/mL) Median (range)		17 (3-356)	12 (4-487)	2323 (7-32986)	<0.001
Menopausal-status %postmenopausal		89%	76%	66%	0.047
US-Score n (%)	0 1 ≥2	14 (51.9%) 6 (22.2%) 7 (25.9%)	31 (43.1%) 21 (29.2%) 20 (27.8%)	10 (10.2%) 24 (47.1%) 64 (70.8%)	<0.001

† A patient has been classified in a pathology group that corresponds to the most severe pathology group of her 2 ovaries

Malignant masses were significantly associated with US features of a solid component, irregular cyst walls, cystic fluid of random echogenicity, presence of papillations and septae, higher septae thickness and a higher HistoScanning-score (table 13).

TABLE 13: UNIVARIATE ANALYSIS OF CATEGORICAL AND CONTINUOUS VARIABLES FOR 365 OVARIES WITH PATHOLOGY INFORMATION IDENTIFIED IN 197 WOMEN

		Normal (n=74)		Benign (n=125)		Malignant (n=166)	
Variable		n	%	n	%	n	%
Locularity***	Unilocular cyst	6	8.6	35	28.5	16	10.0
	Multilocular cyst	0	0.0	27	21.9	10	6.3
	Unilocular Solid	0	0.0	4	3.3	17	10.6
	Multilocular Solid	3	4.3	8	6.5	36	22.5
	Solid	1	1.4	2	1.6	37	23.1
	None of the above	60	85.7	47	38.2	44	27.5
	Missing	4	--	2	--	6	--
Structure of cyst wall***	Smooth	6	8.6	61	49.6	50	30.3
	Irregular	4	5.7	15	12.2	61	38.1
	None of the above	60	85.7	47	38.2	49	30.6
	Missing	4	--	2	--	6	--
Echogenicity of cystic fluid***	Anechoic	3	4.3	55	44.7	39	24.4
	Uniform	1	1.4	6	4.9	7	4.4
	Random	5	7.2	14	11.4	57	35.6
	None of the above	61	87.1	48	39.0	57	35.6
	Missing	4	--	2	--	6	--
Structures in cyst***	Septae (S)	2	2.9	30	24.4	21	13.1
	Papillations (P)	0	0.0	7	5.7	15	9.4
	Both S and P	2	2.9	5	4.1	32	20.0
	None of the above	66	94.2	81	65.8	92	57.5
	Missing	4	--	2	--	6	--

	Normal		Benign		Malignant	
Variable	N	Median (Range)	n	Median (Range)	n	Median (Range)
Cystic volume (cc)	7	227 (25 – 2165)	67	105 (0.6 – 2597)	102	215 (8 – 3616)
Max size of papillation (mm) £	*2*	*Values: 22, 57*	12	14 (4 – 64)	45	19 (5 – 86)
Papillation volume (cc) £	*1*	*Value: 18*	7	5 (1 – 72)	30	5 (0.2 – 226)
Max septae thickness † (mm)	*3*	*Values: 1, 6, 9*	33	3 (2 – 10)	43	5 (2 – 39)
HistoScanning Score ††† (0-125)	5 3	2 (0 – 53)	85	9 (0 – 125)	131	101 (0 – 125)

*** p-value <0.001 by the Chi^2-test

†Adjusted p-value <0.05 for Mann-Whitney test (Benign versus malignant ovaries only)

††† p-value <0.001 by the Kruskall-Wallis test

£Adjusted p-value >0.05 (non-significant) for Mann-Whitney test (Benign versus malignant ovaries only)

Based on these results, the logistic regression was carried out in the LS on the following variables: Menopausal status (Meno: pre=0/post=1), US-Score (USS: 0/1/2), CA125 level (CA125), Ovarian HistoScanning-score (HSS), absence/presence of a solid component (Sol: 0/1), absence/presence of an irregular cyst wall (Wall: 0/1), absence/presence of random echogenic fluid (Flu: 0/1), absence/presence of papillations larger than 3mm (Pap: 0/1) and absence/presence of Septae larger than 4mm (Sept: 0/1). Table 14 shows that HSS was not highly correlated with any of the 8 others as Spearman's correlation coefficient is never >0.5 or <-0.5.

TABLE 14: SPEARMAN'S CORRELATION COEFFICIENT BETWEEN HISTOSCANNING AND ALL OTHER VARIABLES

	CA125	USS	Meno	Sol	Wall	Flu	Pap	Sept
HSS	0.52 <.0001 (269)	0.42 <.0001 (269)	-0.21 0.0005 (269)	0.50 <.0001 (264)	0.39 <.0001 (264)	0.32 <.0001 (264)	0.23 0.0002 (256)	0.21 0.0007 (251)

Abbreviations: HSS = Ovarian HistoScanning score, CA125 = CA125 serum level, USS = US-Score, Meno = Menopausal status, Sol = absence/presence of a solid component, Wall = absence/presence of an irregular cyst wall, Flu = absence/presence of random echogenic fluid, Pap = absence/presence of papillations larger than 3mm and Sept = absence/presence of Septae larger than 4mm

Spearman's correlation coefficients with underneath p-value for H_0: Rho=0 and (n)

Logistic regression

After the AICC-selection procedure, CA125, HSS and Sol were retained and the probability of malignancy predicted by the logistic regression model was given by:

$$\Pr(\textit{being malignant} \mid \textit{CA}125, \textit{HSS}, \textit{Sol}) = \frac{1}{1+\exp(-z)}$$

where

$$z = -4.45 + 0.0099(\textit{CA}125) + 0.034(\textit{HSS}) + 2.16(\textit{Sol})$$

The Analysis of the 100 bootstrap resampled LS showed the following inclusion frequencies for the 9 variables included in the multivariate model: HSS 100%, CA125 100%, Sol 74%, Flu 48%, Meno 41%, Pap 30%, USS 23%, Sept 21% and Wall 14%. The most frequent model (out of the 100 bootstrap-samples) was our final model (24 %).

Neural network

The NN-model with the lowest misclassification error on the LS had 1 hidden layer containing 3 hidden units with the Elliott function $\frac{z}{1+|z|}$ as activation function for hidden units. The estimated equation of this model is given by:

$$\Pr(\textit{being malignant} \mid \textit{CA}125, \textit{HSS}, \textit{USS}, \textit{Flu}, \textit{Wall}, \textit{Sol}, \textit{Pap}, \textit{Sept}, \textit{Meno}) = \frac{1}{1+\exp(-z)}$$

with $z = 2.22 + 15.13(h_1) - 24.15(h_2) - 14.12(h_3)$, in which

$$h_1 = \frac{1}{1+|z_1|} \text{ with } z_1 = -13.95 + 0.0084(CA125) + 0.025(HSS) - 0.92(USS) - 6.64(Flu) - 1.59(Wall) + 7.38(Sol) + 2.76(Pap) + 1.17(Sept) + 11.41(Meno),$$

$$h_2 = \frac{1}{1+|z_2|} \text{ with } z_2 = 1.32 - 0.0022(CA125) - 0.017(HSS) - 0.95(USS) - 0.78(Flu) - 0.33(Wall) + 1.18(Sol) + 0.83(Pap) - 0.60(Sept) + 1.38(Meno),$$

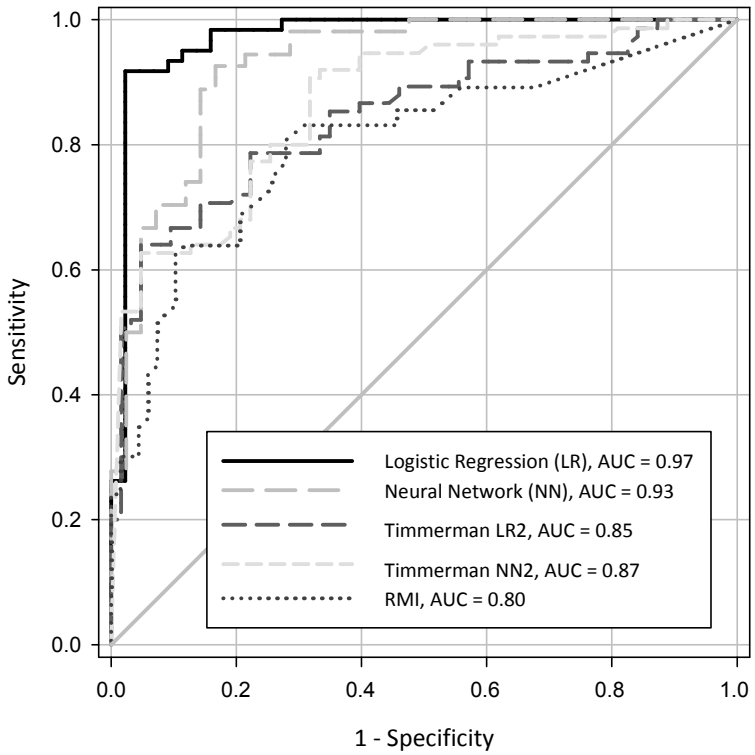
$$h_3 = \frac{1}{1+|z_3|} \text{ with } z_3 = -9.93 - 0.00081(CA125) + 0.058(HSS) + 1.72(USS) - 0.79(Flu) - 0.59(Wall) - 1.35(Sol) + 1.91(Pap) + 2.92(Sept) + 0.55(Meno).$$

Evaluation of model performance

In the ES, the IDI of the LR-model with HistoScanning compared to the LR-model without HistoScanning (so only containing CA125 and Sol) was 0.18 ($p<0.0001$). Similarly, for the NN-model with HistoScanning versus the NN-model without HistoScanning, IDI was 0.15 ($p=0.014$). Thus adding HistoScanning significantly improves model performance.

Figure 11 shows the ROC and AUC of the LR- and NN-models in the ES. The DeLong-test shows improved AUC performance for LR compared to RMI, LR2 and NN2 (p-values all <.0001). For NN compared to RMI, LR2 and NN2, the p-values were 0.00017, 0.0048 and 0.016, respectively.

FIGURE 11: ROC CURVES AND AUC ON FULL EVALUATION SET: UNIT OF ANALYSIS IS OVARY



At a probability of malignancy ≥54%, the LR-model reached an optimum sensitivity of 88.1% (95%CI=[77.8;94.7]) and a specificity of 95.7% (95%CI = [89.2;98.8]) on the LS. For the NN-model, the optimum performance was achieved at a probability of malignancy ≥23% with a sensitivity and specificity of 96.7% (95%CI=[88.7;99.6]) and 95.6% (95%CI=[89.1;98.8]), respectively. When these cut-offs were cross-validated on the ES, the good results for the LR-model were confirmed but the sensitivity and specificity of the NN-model were slightly lower than in the LS (table 15). Cut-off performance on our data of the RMI, LR2- and NN2-models can also be found in table 15.

TABLE 15: SENSITIVITY AND SPECIFICITY FOR ALL VALIDATED MODELS (USING THE ORIGINALLY REPORTED CUTOFFS) AND FOR THE NEW MODELS CROSS-VALIDATED ON THE EVALUATION SET

	Original report					Our study data (Evaluation set only)						
		test set				unit of analysis = ovary				unit of analysis = patient*		
Method	total pop.	n_{pat}	sens(%)	spec(%)	cutoff	n_{pat} n_{ov}	sens(%) [95%CI]	spec(%) [95%CI]	cutoff	n_{pat} n_{ov}	sens(%) [95%CI]	spec(%) [95%CI]
RMI[37]	143	-	85	97	200	79, 148	69 [58;78]	79 [68;88]	200	79, 79	69 [53;81]	84 [67;95]
LR2 [53]	173	57	94	83	60%	79, 135	79 [68;87]	70 [57; 81]	60%	79, 79	91 [79;98]	69 [49;85]
NN2[53]	173	57	94	95	60%	79, 135	>99 [95;100[	10 [4;20]	60%	79, 79	>99 [92;100[	7 [1;23]
LR	-	-	-	-	54%£	79, 103	83 [71;91]	98 [89;99]	75%£	79, 79	68 [52;81]	88 [71;97]
NN	-	-	-	-	23%£	79, 94	80 [67;89]	86 [72;95]	63%£	79, 79	65 [50;78]	90 [74;98]

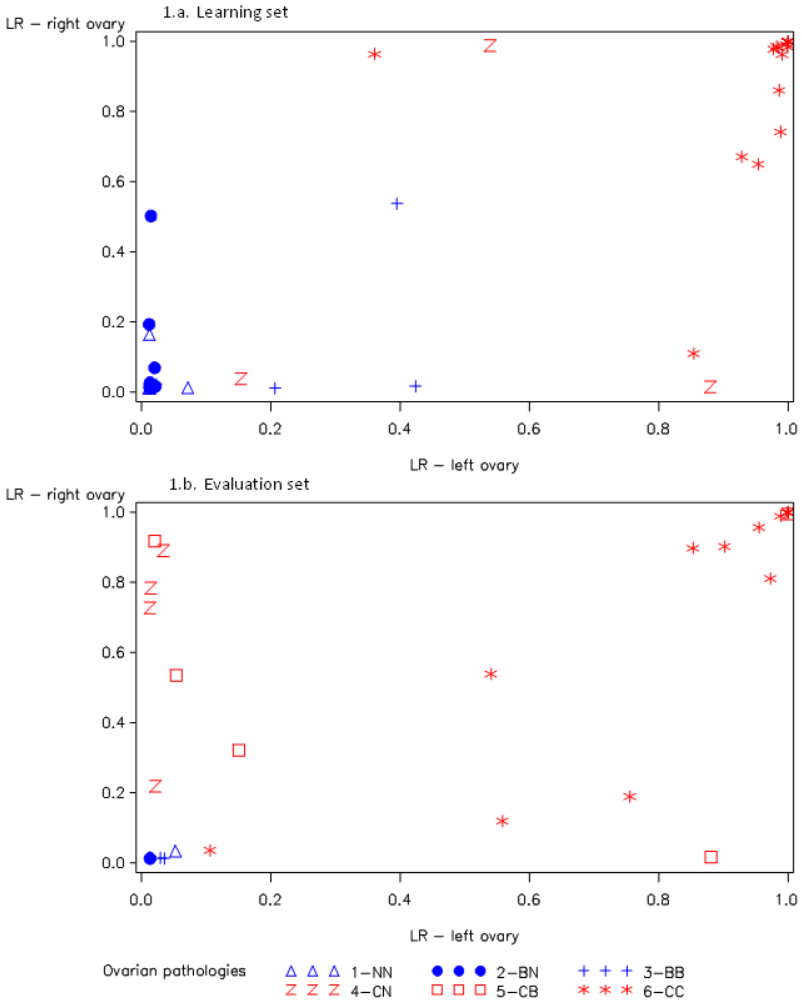
Abbreviations: sens = sensitivity, spec = specificity, n_{pat} = number of patients, n_{ov} = number of ovaries

* For every patient, pathology was selected to be (1) the pathology of the single ovary that was present, (2) the most severe pathology of the 2 ovaries in our data base

£ These are the cutoffs that optimized the Youden index on the learning set

Figure 12 shows that for all pathology groups, the values predicted by the LR-model for the left ovary are not correlated with the values for the right ovary. This implies that a malignant left ovary didn't necessarily force the LR-equation to be elevated for the non-malignant right ovary of the same patient (true in LS and ES).

FIGURE 12: PROBABILITY OF MALIGNANCY (PREDICTED BY THE LOGISTIC REGRESSION) OF RIGHT VERSUS LEFT OVARY



Abbreviations: LR = Logistic Regression

Figure 12.a and 12.b.: This graph shows the probability of malignancy of right versus left ovary. Patients were attributed to 6 groups as follows: of the 83 women with a single adnexal mass, 29 women underwent conservative surgery (not in this graph) while the contra-lateral normal ovary was removed in 63 women. For 9 out of these 63 women, both ovaries were found to be normal at histology

('NN'). Of the remaining 54 patients with unilateral disease, 31 had one benign and one normal ovary ('BN'), while 23 had one malignant and one normal ovary ('CN'). Of patients with bilateral abnormal ovaries: 35 were bilateral benign ('BB'), 60 were bilateral malignant ('CC'), while 10 had one benign and one malignant ovary ('CB')

Models on reduced data set

The LR- and NN-models were also developed and validated in a reduced LS and ES containing only 1 data entry per patient (using patients as units of analysis rather than ovaries). The AICC-selection on the reduced LS also retained CA125, HSS, and Sol and the LR-model equation estimated to:

$$z = -4.65 + 0.0095(CA125) + 0.032(HSS) + 2.10(Sol)$$

The IDI of this model compared to the same model without HSS was 0.34 (p<0.001) so HSS adds predictive ability. When validated on the reduced ES this model reached an AUC of 0.90 (95%CI=[0.82;0.97]) which was no longer significantly higher than the AUC of the RMI (AUC=0.84, 95%CI=[0.75;0.93]), the LR2-model (AUC=0.92, 95%CI=[0.85;0.98]) and the NN2-model (AUC=0.92, 95%CI=[0.86;0.98]). The best NN-model on the reduced LS, was a model with the hyperbolic tangent activation function $1-\frac{2}{1+\exp(2z)}$ for hidden layers, 1 hidden layer and 2 hidden units with the estimated equation as follows:

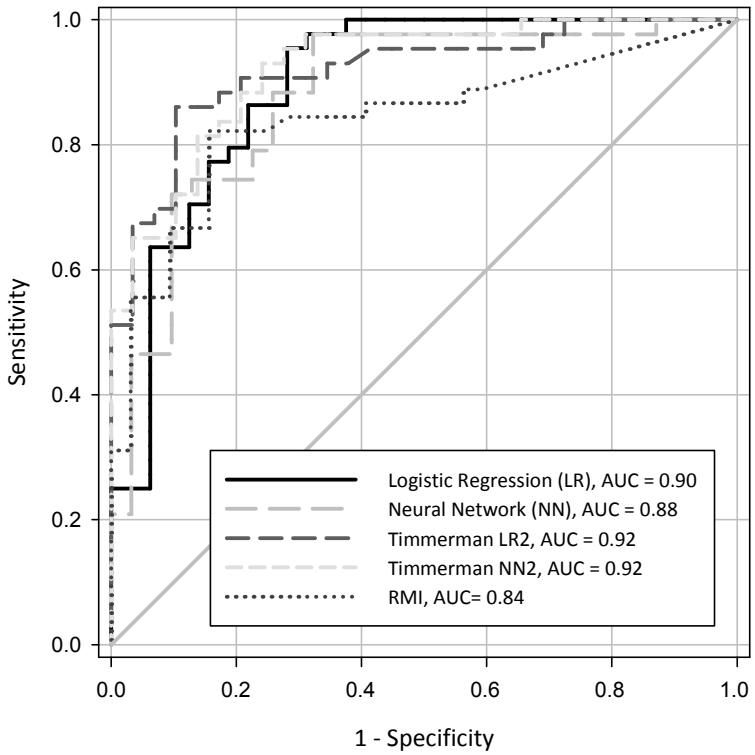
$$z = -5.28 - 7.57(h_1) - 10.61(h_2), \text{ in which}$$

$$h_1 = 1-\frac{2}{1+\exp(2z_1)} \text{ with } z_1 = 6.28 - 0.0025(CA125) - 0.067(HSS) - 0.88(USS) + 2.87(Flu) - 2.15(Wall) - 4.32(Sol) + 2.01(Pap) + 5.12(Sept) - 2.65(Meno),$$

$$h_2 = 1-\frac{2}{1+\exp(2z_2)} \text{ with } z_2 = -0.49 - 0.0037(CA125) + 0.0029(HSS) + 0.31(USS) - 0.14(Flu) + 0.18(Wall) - 0.24(Sol) + 0.086(Pap) - 0.35(Sept) - 0.31(Meno).$$

The hierarchical comparison of the NN-model with and without HSS showed an IDI of 0.11 that was no longer significant (p=0.17). Also, the NN-model had an AUC of 0.88 (95%CI=[0.79;0.96]) which was no longer significantly different from the AUC of the RMI, the LR2- and NN2-models (table 15 and figure 13).

FIGURE 13: ROC CURVES AND AUC ON REDUCED EVALUATION SET: UNIT OF ANALYSIS IS PATIENT



1.2.5. DISCUSSION AND CONCLUSIONS

This paper provides the first evidence of the potential advantage of including HistoScanning in multivariate mathematical models developed for the differential diagnosis of adnexal masses. We have developed and validated 2 multivariate models (LR and NN) using a number of morphological and patient related variables along with a novel variable: HistoScanning-score. In the validation stage, both the LR- and NN-model showed high AUC values (0.97, 0.93 respectively) and performed significantly better than RMI[37] and 2 other previously reported models: LR2 and NN2.[53] In addition, the IDI assessment on the ES showed that adding HistoScanning to our LR and NN-models significantly improves model performance.

Strengths of this work include prospectively collected data, blinded analysis and subsequent validation of the models using a separate ES.

The data used in this work, though prospectively collected, were obtained from a study designed to test the accuracy of HistoScanning and therefore the protocol did not include assessment of all TVS-morphological features. Consequently, derivation of the TVS-morphological parameters was done based on archived 2D- and 3D-TVS images. This could be considered suboptimal.[98] However, the data were acquired by experts with at least 10 years experience and assessing archived TVS-images acquired by experts has been found to improve on the initial observer accuracy.[66]

As CRF requirements of the original study did not include storage of Doppler examinations, this was not included in our model development. Although adding colour Doppler to morphological features improves model performance,[52,64] models incorporating Doppler variables can be difficult to reproduce.[57] Furthermore, a recent meta-analysis found that addition of 3D-Power Doppler did not improve performance of LR-models.[5]

Differential diagnosis is done at the patient level and developing models at patient level would thus be a natural choice. However, given that two thirds of variables used here characterize the ovary (table 13), our rationale and preference was to use the ovary as unit of analysis. This still allows diagnosis at patient level by considering results of both ovaries. However, one single patient may contribute 2 different masses. This might lead to dependence in the data which could potentially bias the results. We have evaluated and tried to exclude such a 'dependence bias' by showing that no correlation exists between the modelled results of left and right ovaries of one patient (figure 12). We also refitted the models on a reduced data set in which the unit of analysis was the patient. On this reduced data set, the performance (AUC (figure 13) and sensitivity/specificity (table 15)) of the LR- and NN-models at validation-stage was less than that achieved using the full data set (per ovary analysis). Our models perform well at the level of each ovary, despite of the presence of systemic variables such as CA125, even for patients with unilateral disease. The improved performance of LR2- and NN2-models on per patient compared to per ovary analysis is probably related to their developmental methodology which was at the patient level.

An AICC-based selection for the LR-model found CA125, HistoScannig-score and presence of a solid component to be statistically significant. The strength of our LR-model based on these 3 variables was further confirmed by bootstrap analysis, showing that they had the highest inclusion frequencies. This LR-model, using just 3

variables, has the advantage of being much simpler and easier to reproduce compared to the NN-, LR2- and NN2-models. The fall in performance of the NN-model in the ES may be explained by our moderate sample size, which may be too low for the high number of parameters that need to be estimated. Also there might be issues with overfitting and generalizability.

The other models that were reproduced on our data: RMI and the Timmerman models LR2 and NN2[53] have been validated by other researchers. An external validation study[57] has reported sensitivities/specificities and AUC values for LR2 and NN2 to be similar to what we found. The performance of RMI on our data are very much in line with other reports.[58]

Appropriate triage of women with adnexal masses seen in gynaecology clinics is necessary so that those at high risk for malignancy can be referred to gynaecological oncologists for further management. 'Subjective' ultrasound pattern recognition by experienced examiners is a very good method for discriminating malignant from benign masses[64,99] with sensitivities and specificities up to 96% and 90%, respectively.[66] Complicated multivariable models using >10 variables also achieve high sensitivities and specificities.[99] However, the sensitivity and specificity achieved by a less experienced examiner is lower: 86% and 80% respectively.[66] Thus, there is need for tests that improve discriminatory ability of less experienced examiners. Although use of a 3D-probe and delineation of 'volume of interest' is required with HistoScanning, this is increasingly within the remit of most sonographers who routinely undertake 2D-scanning. We feel that our LR-model using just 3 variables can lead to significantly improved performance in the hands of a less experienced examiner.

In conclusion, this report is the first to show the potential benefit of including HistoScanning into mathematical models for discriminating between benign and malignant ovarian masses. The LR-model developed in this report achieved high performance and might be of benefit to less experienced operators. Although the internal validation of our models showed encouraging results, future research should assess these models in prospective clinical trials on different populations.

1.2.6. Additional results and discussion

1) In the original study, details about the volume or size of the ovarian tumours found during pathology were not retained on the database. The largest diameter of the ovary itself, measured at pathology, was saved in the database. Note however that the study started in 2004 and pathology practise has evolved a lot ever since. Therefore, for illustrative purposes only, the median sizes of the ovaries for the

three different pathology groups (normal, benign and malignant) are given in table 16.

TABLE 16: LARGEST DIAMETER OF THE OVARY (AT PATHOLOGY) FOR THE THREE DIFFERENT PATHOLOGY CLASSES

Variable	Normal (n=74)		Benign (n=125)		Malignant (n=166)	
	n	Median (IQR*)	n	Median (IQR*)	n	Median (IQR*)
Largest diameter of ovary (mm)	26	27 (20 – 30)	73	42 (35 –100)	92	60 (40 –100)
*IQR = Interquartile Range						

2) The LR- and NN-models were developed on the learning set (LS) and then subsequently validated on the evaluation set (ES) that was set aside. Both LS and ES are random subsets of the same original data set, though, which is why such a validation is often referred to as 'internal validation'. The earlier developed models, RMI[37] and also LR2 and NN2 from the publication by Timmerman et al[53] were also validated on our ES but since they have been developed on a completely different study population, this latter validation is referred to as an 'external' validation. The comparison of our models to the existing models on our ES might therefore be optimistic; this is why external validation of our models on independent data is awaited, as pointed out in the discussion in section 1.2.5.

CHAPTER 5: DISCUSSION, CONCLUSION AND PERSPECTIVES

1. GENERAL DISCUSSION AND PERSPECTIVES

1.1. PROSPECTIVELY COLLECTED DATA, RETROSPECTIVELY ASSESSED VARIABLES

Our data concerned patients coming from a prospective study designed to study the accuracy of HistoScanning alone, without any attention to RMI or other variables that characterize the US-morphology. The RMI and US-morphology variables were therefore assessed retrospectively for the aims of our work. Since HistoScanning is not used in clinical practise today and given that the RMI and US-variables were assessed blindly on high quality US-data, we believe that the retrospective nature of the latter two did not influence or bias the models build in this work.

1.2. APPLICATION OF OUR RESULTS ON DIFFERENT POPULATIONS

All patients of our study population (a clinical series) were patients scheduled for surgery. This means that they were mostly symptomatic patients or at least patients with clinical indications for which surgery was deemed necessary. This enabled us to have a gold-standard (histopathology results) to compare the outcome results of our tests to. Since the interest of our work was to develop novel tests and mathematical models for the **differential** diagnosis of adnexal masses, we believe that this specific patient population is very suitable. However, this implies that extending our results to asymptomatic patients is not straightforward. Indeed, prevalence of malignancy varies across different populations, making a comparison between different populations difficult, because separation cut-offs might be affected.[74] On an asymptomatic population, the sequential test (RMI_3 and HistoScanning) might require other cut-offs than the ones defined in chapter 4. The same is true for the developed mathematical models, LR and NN. In Addition, one might be willing to achieve a higher sensitivity for similar specificity (or the other way round) on a different population which means that one is looking for a

different trade-off between FP and FN results and therefore different cut-offs need to be chosen.

Validation of our models on independent data is awaited and if needed adjustment of the cut-offs of the sequential test and of the cut-offs of the mathematical prediction models to other populations is important.

1.3. SAMPLE SIZE

In this work, all novel tests and mathematical models have been developed on a sample size of 197 patients (total of 199 minus 2 patients with disagreements between pathology and the study CRFs). Despite of this moderate size, we feel cautiously optimistic about the performance of our novel test (RMI followed by HistoScanning) and mathematical models (LR and NN), because the internal validations that were performed showed reassuring results.

1.4. PATHOLOGY RESULTS

The patients used in our work came from a multicentric prospective study[71] and we have used histopathology results as the 'gold standard'. So it is important to note that at all centres, pathology was performed according to local standards and may therefore differ slightly from one centre to the other. This might have introduced a bias in our differential diagnostic tests and models. Although, since we used only the outcome 'malignant' versus 'non-malignant' in model development, which is a simplification of a detailed pathology report, we don't believe that this potential difference between pathology standards has biased our models a lot.

The development as well as the validation of a tissue characterization tool like HistoScanning depends very much on pathology. The HistoScanning algorithms locally analyse tissue volumes of 0.6 to 1.6 mm^3 (20 by 20 voxels of US-data) in a VOI, resulting in the final score for a VOI (ranging from 0 to 125) that was used in this work. The spatial distribution of the red circles (the 20 by 20 voxels regions of US-data that were likely to present a cancerous region) on the 2D or 3D grid of the VOI was not studied thoroughly in this work as detailed pathology results giving an indication of the spatial emplacement of the tumours are difficult to produce for

the case of ovarian cancer. The final score for each VOI, ranging between 0 and 125, reflects however a computerised analysis of consecutive volume units and therefore it was decided to use these continuous scores for each VOI in this work.

1.5. ALTERNATIVE MODELS FOR PAIRED ORGAN DATA

In chapter 4, multivariate mathematical models have been developed for the differential diagnosis of ovarian masses. It has been discussed in chapter 4 that on our data, the ovary is to be preferred as unit of analysis to the patient; a fortiori because of the fact that model outcomes of left ovaries didn't correlate to outcomes of right ovaries. In the statistical literature, special logistic regression models for paired organ data exist.[100] Some of those models predict probabilities of malignancies for one organ considering patient related covariates and the status of the fellow organ[101] at the same time. Others choose not only to consider the status of the fellow organ but also the organ specific covariates of the fellow organ for making predictions.[100] In either way, information of both organs is used while modelling the predicted probabilities for a patient.

Regarding the mathematical models for the differential diagnosis of ovarian masses, most researchers have been modelling on the patient level. In doing so, the worst outcomes of both ovaries were considered for the patient and covariates such as 'bilaterality of lesions' to consider the paired organ status were included. Since we saw no correlation between predicted probabilities of right and left ovaries indicating no loss of information by ignoring the paired status in our data, we have decided to model on the ovarian level which would increase our sample size and make our models more accurate in predicting malignancy for an ovary. This appeared to be a good choice, despite of the presence of systemic variables such as CA125. Nevertheless, it would be interesting to consider the paired organ models[100] for the case of ovarian cancer in order to study whether they would improve model accuracy and diagnosis.

1.6. VOI SELECTION FOR HISTOSCANNING

The quality of subjective assessment (visual pattern recognition by the operator) of adnexal masses on US-images varies widely, depending on examiner expertise.[66,93] Very experienced examiners achieve sensitivities and specificities up to 96% and 90%, respectively[66] while the sensitivity and specificity achieved by a less experienced examiner is lower: 86% and 80% respectively.[66] The 2 sources introducing this variability may be:

1. The definition of the region of interest
2. The interpretation of the morphological characteristics of the selected region.

Ovarian HistoScanning being an automated computerised technique has the potential to significantly reduce this observer variability for image interpretation. Even so, Ovarian HistoScanning software is not completely without user dependency, as it requires the identification of the VOI, which falls under the first source of variability. Unfortunately, within the context of the design of the study, it was impossible to assess inter-user variability related to the definition of the VOI.

We believe that ideally, the assessment of masses on the US-image as well as the initiation of the HistoScanning analysis on this image should be done by the same operator during the same examination. Indeed, it is a major advantage that HistoScanning requires nothing more than the US-data and can therefore run in parallel while the operator finishes the examination. For the patient, this would mean no additional discomfort while especially the less experienced operator will have an extra parameter to base his/her diagnosis upon. Nevertheless, addressing the user/expertise dependency with respect to the VOI selection in such a set-up would be important.

Although use of a 3D-probe is required with HistoScanning, this is increasingly within the remit of most sonographers who routinely undertake 2D-scanning.

1.7. SCREENING

More accurate staging operations, more effective chemotherapy and more radical surgery have improved duration of survival for patients with ovarian cancer over the years.[102] Unfortunately, however, there has only been little change in the overall ovarian cancer mortality rates.
(http://info.cancerresearchuk.org/cancerstats/types/ovary/mortality/)

As a result, researchers have been making efforts to develop effective strategies for early detection in high risk as well as in asymptomatic populations. For a test to be interesting as a screening modality, it needs to have very high specificity and PPV. A one time use of TVS or CA125 as screening modalities was found to lack the necessary specificity and PPV for screening.[103,104] Combining CA125 and US did show encouraging results, especially for postmenopausal women.[105] Screening strategies comprising out of multimodal approaches have been more refined, achieved higher specificities and PPVs[106,107] and trial results were encouraging.[108,109] All these studies have shown that CA125 is to be used as a first level screen, followed by US. Lately, the Risk of Ovarian Cancer algorithm based on age and a CA125 profile rather than CA125 at a fixed cut-off, followed by TVS with CA125 as level 2 screen has shown its feasibility to screen for ovarian cancer.[110]

The future perspective of studying the potential role of ovarian HistoScanning in the screening of high risk and asymptomatic populations is interesting. Based on the high specificities that were achieved for the sequential test (RMI_3 followed by HistoScanning), it seems plausible to consider adding HistoScanning as a level 3 screen after TVS. Health economics and physical and psychological morbidities should then also be studied.

1.8. A COMMENT ON COST-EFFECTIVENESS

In this section, the objective was not to discuss the results of a cost-effectiveness analysis regarding the diagnostic models developed in this thesis but rather to set the scene for a small discussion on this topic. To illustrate this discussion, we will start with presenting some cost figures.

1.8.1. SOME COST FIGURES

Cost figures for medical imaging procedures were collected from the administration of the St. Luc hospital (Brussels, Belgium): Computer Tomography (CT), Magnetic resonance imaging (MRI), and TVS cost 119.53€, 117.59€ and 21.48€ (in 2007 €'s), respectively. When considering any of these medical imaging procedures, it is important to note that actually the fees of the medical staff performing the imaging have to be counted too.

Costs for surgery were obtained from the CMDS[B] and FMDS[B] databases for the year 2005. In total 18 different operating room procedures that could possibly be performed on patients with benign or malignant neoplasms, were taken into account. These include a wide range of possible procedures (from diagnostic laparoscopy to debulking surgery) performed depending on the origin and severity of the disease. For the year 2005, a total of 81 hospital stays for malignant patients and of 417 hospital stays for benign patients (the reason for admitting these patients being the surgery) were identified. For malignant patients, the median cost of a surgery was 2867€ (range: 891€ - 16964€) and for benign patients this was 1621€ (range: 845€ - 7134€) (both expressed in 2005 €'s). These figures include costs for laboratory tests, pharmaceutical products, the surgical intervention in question, synthetic materials, doctor and staff fees, hospital and room charges, and extra charges admitted during the hospital stay.

All of the above mentioned costs are full charges, i.e. a sum of the part paid by the social security system and the part paid by the patient.

1.8.2. THE UNITED KINGDOM

The RMI, which has been extensively discussed in this work, is used as a triage tool for women presenting with adnexal masses in the United Kingdom. All over the United Kingdom, this triage is performed according to well defined guidelines.[39,41] In the United Kingdom, the US-scan is acquired by sonographers (not by medical doctors) and the US-score, needed for the RMI calculation, is also assessed by sonographers. If one would advocate the use of RMI_3 supplemented with HistoScanning for the differential diagnosis (or triage) of ovarian masses instead of RMI alone for triage, this would imply of course the additional cost of the ovarian

[B]The Clinical Minimum Data Set (CMDS) contains relevant clinical and other information, such as primary and secondary diagnosis, age, sex, length of stay and more. The Financial Minimum Data Set (FMDS) is based on the billing data for hospitalized patients. More information can be found in the book published by Dirk Corens[111], which is available at:
http://www.euro.who.int/__data/assets/pdf_file/0007/96442/E90059.pdf.

HistoScanning analysis. Moreover, given the higher effectiveness (sensitivity) of the combined approach (at least achieved on our data), this would mean that more malignant patients are correctly identified and referred to specialized centres where they would undergo more expensive surgical procedures by very experienced surgeons. Most importantly, however, is that a FN (lower sensitivity) case is presumably 'lost' if treatment is deferred until too late. In other words, the question is how much more cost can HistoScanning accrue per additionally gained 'life year'? Then, on the other hand, higher specificity will imply less wrong (FP) referrals to specialized centres and will in this way avoid more expensive surgical procedures.

These arguments merely launch a debate about cost-effectiveness and, as mentioned before, a cost-effectiveness analysis considering them and other arguments is awaited.

1.8.3. BELGIUM

In Belgium, the use of RMI is not adopted in clinical practice for the differential diagnosis of women present with adnexal masses, nor for triaging patients along appropriate referral pathways. Ultrasonic imaging, on the other hand, is often used to detect the presence of an adnexal mass in the first step of the work-up of the patient. Then, for women with an equivocal image, various evaluation modalities can be used alone or together for distinguishing benign from malignant masses: MRI, CT, US-morphology assessment, Colour Doppler, etc. The sensitivity and specificity of MRI for detecting malignancy when a patient presents with an equivocal US-image is of 89% and 86%, respectively.[5,112] This is very much comparable to the accuracy results we have achieved here both with combining RMI with HistoScanning as with the LR model that was developed. The major advantage of our approaches (including HistoScanning) compared to MRI is the fact that HistoScanning only involves computerised analysis of data already acquired at the time of the initial US, implying the following positive attributes:

- HistoScanning avoids additional medical staff fees as the MRI has to be performed by a specialist (radiologist) which will induce additional fees
- The patient doesn't have to wait for the MRI to be performed (which will presumably be performed during another appointed at a later data) so days or weeks of patient anxiety can thus be avoided and important time may be saved

Similarly as for the previous section, these points are here for the purpose of a discussion and a complete cost-effectiveness analysis is awaited.

1.9. RECOMMENDATIONS

Based on the items that were previously discussed, we would recommend the following items for the future: A validation on independent data (external validation) of RMI followed by HistoScanning and of the LR and NN models is important before adopting them in clinical practise. This data set should preferentially be coming from a prospective, multicentric study to assess diagnostic accuracy. The information that is to be collected for every patient should be thoroughly discussed by a multidisciplinary team and this information should definitely contain: a large list of US-morphology based variables, Us-Doppler measurements, a large list of patient related information and very detailed pathology reports (e.g. lesion dimensions etc.). This study should also consider the important issues of inter-user variability with respect to the VOI selection for HistoScanning as well as the inter-user variability with respect to the interpretation of US-based morphology scores and variables. Regarding the inter-user variability, a panel of 3 experienced sonographers or examiners could be a solution to avoid disagreements between 2 of them. Ideally the HistoScanning analysis and the US-interpretation should be done by the same examiner.

2. CONCLUSIONS

The main goal of our work was to define the scope that exists in clinical practise for a new computerised technology in the diagnosis and work-up of ovarian cancer. In other words, defining how and where this technology could improve accuracy, detection and patient care. This main objective has been broken down into smaller objectives (see chapter 2):

A. Would applying ovarian HistoScanning for the patients in an RMI grey-zone improve triage of women with ovarian masses?

B. Could adding ovarian HistoScanning to multivariate mathematical prediction models for ovarian masses improve model performance?

Ovarian HistoScanning is an US-based computerised tissue characterization technique. The HistoScanning analysis is based on the fact that cancerous processes induce structural changes in the tissue texture which result in physical changes in the reflected US waves. Moreover, the characterization algorithms are fine-tuned to find such changes on a small scale. In other words, in its current form, the technique is very useful to analyse tissue regions of an adnexal mass that are difficult to interpret visually, i.e. before symptoms on the cellular level (irregular borders, irregular solid areas) become visible on the image or to analyse these cellular level symptoms in case of hesitation (benign versus malignant), etc. Ovarian HistoScanning serves thus as liable assistant (computerised so less subjectivity) to the operator (that has to select the VOI for HistoScanning) in diagnosing adnexal masses. For his/her diagnosis, the clinician will also rely on CA125, US-morphology, colour Doppler parameters, patient history and other patient characteristics. Multivariate mathematical models translating a high number of such variables into one 'probability of malignancy' have been proposed in order to try to predict the ovarian histology. For less experienced examiners whom achieve much lower accuracy in predicting malignancy than experienced examiners,[66] such models are very useful. In this work, the added value of HistoScanning as a new variable in such models was investigated and it was shown that HistoScanning can significantly improve model performance (aim A).

The factors used in the RMI score calculation also result from a multivariate logistic regression model.[37,40,59] In case of the RMI, the mathematical model development lead to a simple and robust tool that was advocated to be the prediction model of choice.[58] As far as we know, the RMI is the only model that is widely adopted in some countries for defining the patient referral path. Therefore, we also focussed on the added value of HistoScanning for the patients with intermediate RMI values (those for which RMI has difficulties deciding). A sequential test, applying HistoScanning only for patients in an RMI 'grey-zone', was developed and accuracy results of this test were very encouraging. The use of HistoScanning as a second level test for patients in the RMI grey-zone can thus be envisaged in clinical practise (aim B).

In conclusion, differentiating between malignant and benign ovarian masses has been studied intensively by researchers over the past decades. A large number of different variables and parameters characterizing ovarian masses have been

investigated and combined. We believe that the ovarian HistoScanning algorithms bring different information about the ovarian masses that can assist and improve the existing models for differential diagnosis.

PART II

OVARIAN SONOHISTOLOGY:

STATISTICAL MODELS FOR ULTRASOUND DATA AND THEIR ABILITY TO DIFFERENTIATE TISSUE TYPES

OUTLINE

Ovarian cancer is one of the leading causes of death from gynaecological malignancies.[3] Early diagnosis of ovarian cancer is the key for improving outcomes for the disease.

Medical imaging techniques have revolutionised medicine during the last decades. Ultrasound (US) in particular gives access to vital data in a non-invasive way and is effective for imaging soft tissues of the body. Compared to other medical imaging modalities, US has the following positive attributes:

- US is a real-time, easy operation medical imaging technique
- US has a non-invasive and radiation free nature
- US is relatively low-priced,

Hence US has become widely used as a diagnostic technique in general clinical practice.

In gynaecology, US is one of the most important and primary diagnostic tools. Its use continues to increase and it is now an essential part of the diagnostic procedure in examining the female pelvis. Indeed, the use of US morphology to characterize adnexal masses and thus diagnose ovarian cancer is well established.[4]

When an US image is assessed on a macroscopic scale, large entities can be distinguished in soft tissues. On a smaller scale, it is seen that each such entity has a very detailed granular structure, random in nature and variously described by the terms texture or speckle. This is explained by the fact that an US image is the result of a rather complicated set of physical phenomena, namely, the insonification and resulting absorption, reflection, and coherent scattering from a tissue medium of ultrasonic pressure waves, and the electronic detection of the backscattered or echo pulses for display as an image.

The macroscopic entities as well as their speckle patterns are important when designing methods for automatically analysing and deriving information from US-data.

We will describe the use of statistical models that characterize US-data from speckle areas. In this work, we apply these statistical models for classifying ovarian tissues and more specifically for detecting ovarian cancer.

Part II is organized as follows: Chapter 1 gives a more detailed introduction to US imaging and the various statistical models that will be analysed in this manuscript. In chapter 2, detailed research aims are described. Chapter 3 concerns the description of the data, the patients and the methods used in this work. In chapter 4 the results of our work are presented and finally chapter 5 concludes and discusses the results and some future perspectives are introduced.

CHAPTER 1: BACKGROUND

The necessary definitions and notations for this research work are described in this chapter: Section 1 concerns the US device and the basic (physical) principals of ultrasonic imaging. In section 2 different notations used for dealing with US signals are introduced. Finally section 3 discusses the use of statistics for US data and introduces different statistical models derived for US data.

1. ULTRASONIC IMAGING

1.1. THE ULTRASOUND DEVICE

Basically, a medical US device consists of three main parts:

1) The transducer or probe that emits ultrasound waves into the medium and also receives backscattered waves.

2) The scanner, to which the probe is connected, that handles and processes all the data from the probe.

3) The screen where the processed data is displayed as an image, to be analyzed by the physician.

FIGURE 14: EXAMPLES OF ULTRASOUND DEVICES

1.2. THE RADIOFREQUENCY SIGNAL

In this section the basic functionality of the transducer is introduced as well as the definition of radiofrequency signals.

1.2.1. PIEZOELECTRIC CRYSTALS IN THE TRANSDUCER

A piezoelectric crystal is a crystal that generates a mechanical force resulting from an applied electrical field or, the other way round, that generates an electrical charge in response to an applied mechanical force. These piezoelectric crystals form the basis of the emitting and receiving of US-waves in the transducer of an US device. A transducer presents with an array of crystals that can be linear or curved but always contains of a certain number of individual crystals.

1.2.2. THE RADIOFREQUENCY SIGNAL DESCRIPTION

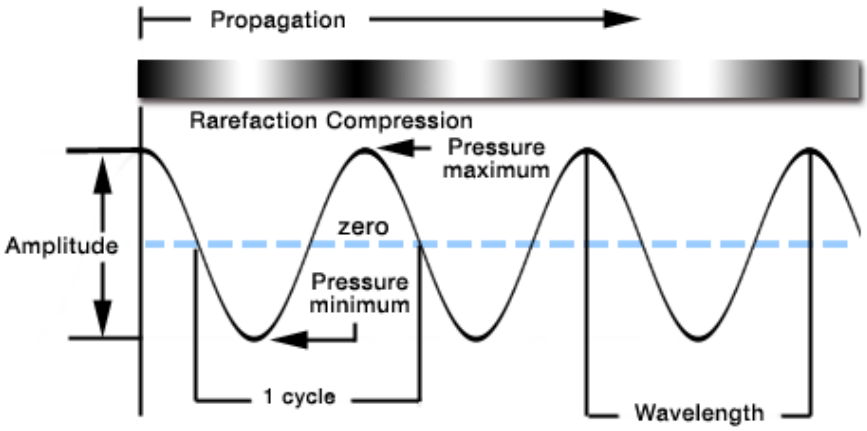
In US, the radiofrequency (RF) signal is the 'raw', unprocessed signal that is produced by the US transducer. In theory it is the electrical signal produced by the piezoelectric crystals in response to the US pulses. In practise however, the RF signal is the signal that is received after (what manufacturers refer to as) the beam former, which combines the signals of the individual crystals in a meaningful way. A thorough explanation about beamforming and about the acquisition and processing of RF data by the US device is beyond the scope of this work, however, more details can be found in other manuscripts.[113]

1.3. PHYSICAL BASICS OF ULTRASOUND IMAGING

1.3.1. ULTRASOUND WAVES

US waves are longitudinal pressure waves that propagate through plastic media like e.g. tissues. The medium/tissue particles are compressed (compression areas) and released (rarefaction areas) in an alternating way and this particle movement is characterized by a sinusoidal wave as shown in figure 15. The propagation speed (c) depends on the medium. The wave frequency is given by the number of cycles in one unit of time and the wavelength (λ) is the minimal distance between 2 maxima with the same pressure (see also figure 15).

FIGURE 15: SOUND WAVE



1.3.2. ULTRASOUND REFLECTION AND IMAGING

The tissue resists to the US-wave propagation and this resistance can be characterized by the acoustic impedance ($Z = \rho c$, where c is the speed of sound and ρ the medium density). When a difference in acoustic impedance is encountered (interface), part of the energy of the US-wave will be reflected and part will be transmitted. These reflections or backscattering form the basis of US imaging.

Traditionally, the reflections are split into 3 categories:

1) Specular reflections: originate from interfaces between large (compared to the wavelength) homogeneous structures

2) Diffuse scattering (shortly scattering): The dimensions of the interface are smaller than the wavelength. The interface is called a **scatterer** and it will reflect the incoming waves in all directions.

3) Diffractive scattering: Occurs when the interface size is in between the 2 extremes above

In general, soft tissue is modelled as a collection of small scatterers. The signal coming from this tissue is therefore seen as the sum of many small reflections.

2. TERMINOLOGY AND REPRESENTATION OF SIGNALS

A signal is a physical representation of information that is often a function of time and can be represented by a known model, e.g. the sinusoidal model.

This section lists different notational representations of sinusoidal signals in general. Then notations are given for signals received from scatterer areas. Finally, the notion of envelope of an US signal is introduced. The envelope of the signal is the random variable for which the statistical models were developed.

2.1. PHASOR NOTATION

A sinusoidal signal is represented by the following formula: $s(t) = A\cos(2\pi\omega t + \phi)$, where A, ω and ϕ are the amplitude, the frequency and the phase of the signal, respectively.

Often, it is preferred to consider the sinusoidal signal alternatively as the projection of the following complex signal:

$$A\cos(2\pi\omega t + \phi) + iA\sin(2\pi\omega t + \phi)$$

Euler's notation is then used and we write: $s(t) = Ae^{i(2\pi\omega t + \phi)}$ without loss of generality.

Similarly, the phasor notation[114] of the sinusoidal signal is used, which is noted by: $Ae^{i\phi}$, enabling us to add and multiply (for addition and rotation) in an easy way signals of this kind by using complex number arithmetic.

2.2. SCATTERER SIGNAL AS PHASOR SUM

2.2.1. RESOLUTION CELL

The area being imaged by an US scanning system can be divided into small volumes or cells. If two structures are so small that they will fit into the same cell, they will not be resolved by the US scanning system. These cells are resolution cells, with sometimes complex shapes but always described by three dimensions: an axial length, l, a lateral width, x, and a slice thickness, t (figure 16). The scatterers for example, are details that are small enough to fit entirely in a resolution cell.

FIGURE 16: RESOLUTION CELL

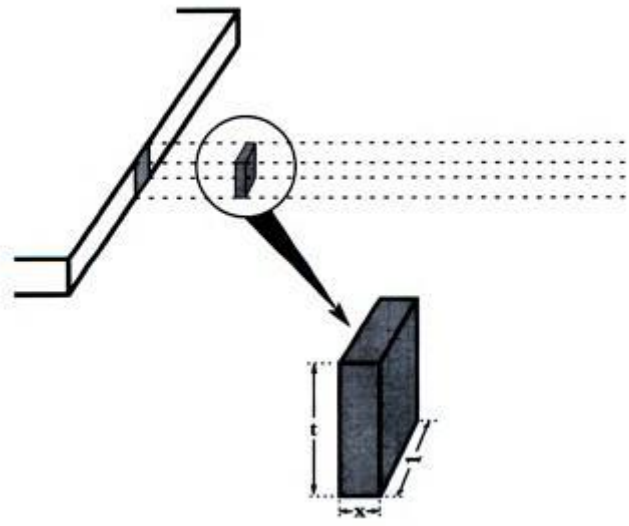


Image taken from the handbook (pp. 2) by Bates et al.[115]

The shaded area represents one single resolution cell for the scanning system. Note that the 3 dimensions of the resolution cell are values in each direction at the position of the specific cell. Elsewhere, these values may be different.

2.2.2. PHASOR SUM

In scattering conditions, all individual scatterers retransmit the US energy into all directions. The resulting reflection is determined by the sum over all individual scatterers:

$$S = \sum_{n=1}^{N} \alpha_n e^{(i\phi_n)} = \mathrm{R}\, e^{i\Phi} = I + iQ$$

Where N is the number of scatterers in the resolution cell, α_n is the amplitude of the n[th] scatterer and ϕ_n the phase of the n[th] scatterer.

2.3. ENVELOPE DETECTION

Envelope detection or also envelope extraction is performed to find the more global behaviour of the US pulse.[113] The envelope of an RF signal is obtained from the complex representation of the real signal by taking the original signal as real part and its Hilbert transform[116] as imaginary part. The real and imaginary components of these complex numbers are often referred to as I (in phase) and Q (quadrature) components. The magnitude of the resulting complex signal is then used as detected signal for imaging (envelope).

In the line of the notation stated above, the envelope equals:

$$R = \sqrt{I^2 + Q^2}$$

Similarly to the envelope, the instantaneous phase of the signal is given by: $Ph = \arctan\left(\frac{Q}{I}\right)$.

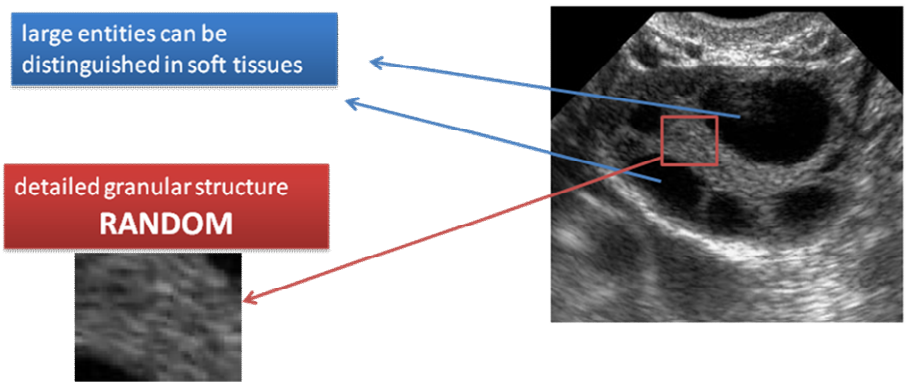
Additional low pass filtering with decimation or interpolation may be carried out on processed signal to bring it in line with display properties. By log compressing the envelope signal a linear display range is obtained. The resulting data are also called gray level data which take integer values between 0 and 255.

3. STATISTICAL MODELS FOR THE ULTRASONIC BACKSCATTER

3.1. RESEARCH MOTIVATION

The scatterers present in tissue will reflect the signals in a multidirectional way and this leads to interference (constructive or destructive) between the different backscattered waves. In the image, the impact of this shows up as alternating regions of high and low intensity (i.e. high and low envelope values). This effect is called speckle. Figure 17 shows the typical granular structure of the speckle.

FIGURE 17: 2D-FRAME OF AN US-IMAGE OF AN OVARY



Moreover, the spatial emplacement of the scatters inside one resolution cell is in general and for normal circumstances completely random. The resulting signal can thus be seen as a random variable. For this reason it is interesting to study the statistical distributions of US signals and investigate whether different distributions will characterize different tissue types. With respect to figure 17, this implies that statistical distributions were derived for patches of envelope data coming for an US-image as the one given by the red square and its zoom that shows clearly the speckle pattern.

Over the years, researchers have investigated probability distribution models that characterize the envelope of the backscattered echo. The earlier work by Wagner et al.[117] firstly presented a general description of the statistics of speckle patterns in ultrasound backscatter. Other researchers that were focussed on tissue characterization took interest in these results because the parameters of these probability distribution models could be a source of information regarding the underlying characteristics of the tissue.[114,118-121] Applications of the usage of these parameters for tissue characterization can e.g. be found in breast cancer diagnosis[119,122,123] and prostate cancer diagnosis.[74]

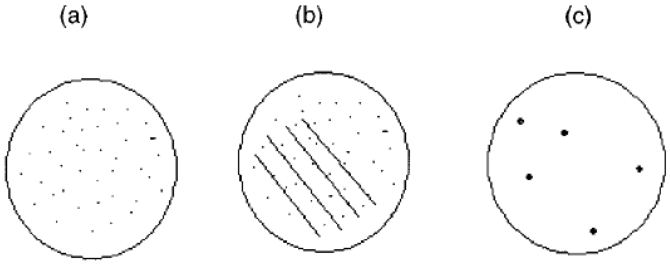
In this work, we would like to describe the use of those parameter estimations for characterizing ovarian tissue. Additionally, the 'closeness' of the empirical data to these theoretical distributions was also investigated by means of a goodness-of-fit statistic.

3.2. POSSIBLE SCATTERING CONDITIONS FOR THE RESOLUTION CELL

In general, three different scattering conditions (see also figure 18) for soft tissues are distinguished:[124]

a) A large number of scatterers randomly aligned within the range cell and no correlation between the scatterers (Rayleigh situation)
b) The scattering region contains subresolvable periodic alignment of scatterers in addition to a collection of randomly located scatterers (Post-Rayleigh situation)
c) Very few scatterers within in the range cell (Pre-Rayleigh Situation)

FIGURE 18: SCATTERING CONDITIONS (IMAGE TAKEN FROM SHANKAR ET AL.[124])



As introduced, several statistical distribution models for the envelope of the speckle have been developed in the literature for the case of US data from soft tissues. They were derived in a theoretical way and sometimes specially adapted to the different scattering conditions stated above. In the following sections these distribution models will be introduced in detail.

3.3. STATISTICAL MODELS FOR THE ENVELOPE

In the following sections, five different statistical models are fully described and the Signal to Noise Ratio (SNR) of the envelope is also linked to the scattering conditions in figure 18.

3.3.1. THE RAYLEIGH MODEL

When the real, I, and imaginary, Q, components of the signal follow both a Gaussian distribution with zero mean and equal variances ($\sim N(0,\sigma)$), the envelope (R) will be Rayleigh distributed.[114,125] The Rayleigh model holds when the following conditions are satisfied:[125]

- A large number of scatterers in the resolution cell (generally more than 10) in order to make use of the Central Limit Theorem
- Randomly located scatterers ($\phi_i's$ are uniformly distributed between $-\pi$ and π)

- Small scatterer spacings compared to the wavelength

The Rayleigh model has been defined for the 'ideal' scatterer condition, presented in figure 18 (a), also referred to as the Rayleigh situation. The probability distribution function (pdf) for the Rayleigh model equals:

$$f(r) = \frac{re^{\left(\frac{-r^2}{2\sigma^2}\right)}}{\sigma^2} \quad \text{for } r \geq 0; \sigma > 0$$

Where r stands for the envelope data and σ for the scale parameter of the Rayleigh distribution.

3.3.2 THE RICIAN MODEL

If, in addition to the conditions leading to the Rayleigh situation, the tissue contains a non-random alignment of periodically spaced scatterers (see figure 18 (b)), a coherent component (strong signal) is added to the signal. Under this additional condition the signal becomes:

$$S = \sum_{n=1}^{N_1} \alpha_n e^{(i\phi_n)} + \sum_{m=1}^{N_2} \bar{\alpha}_m e^{(i\bar{\phi}_m)} = \left(I + \bar{I}\right) + i\left(Q + \bar{Q}\right),$$

Where N_2 represents the number of scatterers that give rise to the coherent component. It can be shown[124] that if the spacing between the scatterers is $\lambda/2^k$, for $k = 0,1,2,\ldots$, then $\bar{\phi}_m$ will equal π/ε, where $\varepsilon = 2^k/4$.

In the particular case of $k = 0,1$, or 2, the envelope of the speckle noise will be Rician distributed with the following pdf:

$$f(r) = \frac{r}{\sigma^2} e^{\left(\frac{-r^2+v^2}{2\sigma^2}\right)} I_0\left(\frac{vr}{\sigma^2}\right) \text{ for } r \geq 0; v \geq 0, \sigma > 0 ,$$

where $I_0(.)$ is the modified Bessel function of the first kind of order 0. In the Rician or Rice pdf, v represents the amplitude or strength of the coherent component. It can be seen that the smaller v, the more the Rician pdf approaches the Rayleigh distribution.

The 'Rician-ness' will depend on the spacing of the periodically located scatterers. If, as above, this spacing equals $\lambda/2$ or $\lambda/4$ the Rician pdf is applicable. On the other hand, for spacings of $\lambda/2^k$ (with $k \geq 3$), the real and imaginary parts of the signals will have different variances but equal means leading to the generalized Rician situation.[124] A common aspect to both the Rician and the generalized Rician situation is the fact that the phase of the backscattered echo will no longer be independent of the envelope and will not be uniform between $-\pi$ and π.

3.3.3. The Signal to Noise ratio

The experimental signal to noise ratio (SNR) of the envelope was linked[125,126] to the tissue conditions presented in figure 18. The SNR of the envelope is given by:

$$SNR(R) = \frac{\mathrm{E}[R]}{\sqrt{\mathrm{Var}[R]}} .$$

The models of Rayleigh and Rice require a large number of scatterers for the central limit theorem to hold. It has been shown[125,126] that the experimental SNR of the envelope signal decreases when the scatterer density falls below 10 scatterers per resolution cell. Hence, the tissue conditions of figure 18 can also be classified with respect to the SNR:

a) Pre-Rayleigh situation when 0 < SNR < 1.91
b) Rayleigh situation when SNR $\cong$ 1.91
c) Post-Rayleigh situation when SNR > 1.91

3.3.4. THE K MODEL

The K-model has been introduced by Jakeman and Pusey[127] and it adapts well to the situation of a small 'effective' number of scatterers case (figure 18 (c)). It arises when the number of scatterers follows a negative binomial distribution and it is also applicable for the envelope signal when this number is low (pre-Rayleigh situation)[114]. The pdf of the K-distribution is given by:

$$f(r) = 2\left(\frac{r}{2}\right)^{a} \frac{b^{a+1}}{\Gamma(a)} K_{a-1}(br) \text{ for } r \geq 0; a, b > 0,$$

where $K_{\omega}(.)$ is the modified Bessel function of second kind and order ω and a and b are the shape and scale parameters respectively. Moreover, the shape parameter a should be seen as the 'effective' number of scatterers (a percentage of the real number) contributing to the signal in the range cell[114] through the following relation: $a = N(1+\varepsilon)$, where $\varepsilon \in]-1, 0[$ stands for the global variation in scatterers' cross-sections and N is the total number of scatterers.

When the different scatterers' cross-sections are of the same magnitude ($\varepsilon \to 0$) or when the number of scatterers becomes large ($N \to \infty$), the K distribution approaches the Rayleigh distribution. On the other hand when there is a great difference in scatterers' cross-sections ($\varepsilon \to -1$) the K distribution becomes more and more pre-Rayleigh. The K distribution therefore encompasses the Rayleigh as well as the pre-Rayleigh situation and could therefore be seen as an improvement of the Rayleigh distribution. However, the estimation of its parameters can be very complex and becomes less stable when the underlying scatterers' condition approaches the Rayleigh regime.[114]

3.3.5. THE WEIBULL MODEL

There are only few arguments in the literature for using this distribution in modelling the envelope signal of US data.[128] But the relation between the SNR and a parameter of the Weibull distribution shows that it can model a wide range of tissue conditions. The pdf of the Weibull distribution is:

$$f(r)=\frac{a}{b}\left(\frac{r}{b}\right)^{a-1} e^{-\left(\frac{r}{b}\right)^{a}} \text{ for } r\geq 0; a,b>0,$$

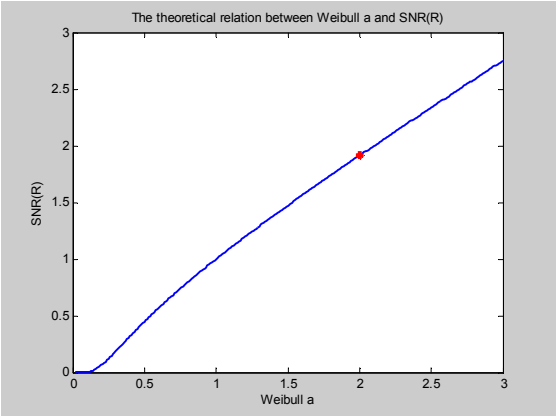
where a is the shape and b the scale parameter.

When assessing the Weibull distribution for the envelope, it can be shown that the SNR increases with the parameter a through[128]

$$SNR=\frac{\Gamma\left(1+\frac{1}{a}\right)}{\sqrt{\Gamma\left(1+\frac{2}{a}\right)-\Gamma^{2}\left(1+\frac{1}{a}\right)}}$$

Figure 19 depicts this relation for $0<a<3$: When $0<a<2$, the SNR will be smaller than 1.91 (Pre-Rayleigh). Also, $a=2$ corresponds to a SNR of 1.91 (the red dot on figure 19) representing the Rayleigh situation, and $a>2$ yields an SNR > 1.91 (post-Rayleigh). This relation evidences of an indirect link to the 3 scattering conditions. That is why we were also interested in testing the Weibull distribution on our data.

FIGURE 19: THE THEORETICAL RELATION BETWEEN WEIBULL a AND THE SNR



3.3.6. THE NAKAGAMI MODEL

The idea of using the Nakagami distribution for modelling the US backscatter was introduced by Shankar et al.[122,124] Modelling the envelope with a Nakagami distribution requires to estimate two parameters m and Ω (the shape and scale parameter respectively) in the following pdf:

$$f(r) = \frac{2m^m}{\Gamma(m)\Omega^m} r^{(2m-1)} e^{\left(-\frac{mr^2}{\Omega}\right)} \text{ for } r \geq 0; m, \Omega > 0$$

The parameter m conveys information about the envelope statistics, as does the parameter a of the Weibull distribution. Values of m in $]0,1[$ correspond to the pre-Rayleigh regime, $m = 1$ corresponds to the Rayleigh distribution, and $m > 1$ leads to the post-Rayleigh regime. The latter statement has also been refined by Shankar et al[124]: When there is a periodically located structure with $\lambda/2$ or $\lambda/4$ spacings, m will be greater than 1 and the Rician distribution will be applicable. However, if the spacing is $\lambda/2^k$ (with $k \geq 3$), the generalized Rician will fit, and m will be in $]0.5;1[$, corresponding also to a post-Rayleigh situation. To allow a proper separation between the latter situation and the pre-Rayleigh situation because m is in $]0.5;1[$ for both, the phase data are used, keeping in mind that in the post-Rayleigh situation, the phases are not uniformly distributed.

CHAPTER 2: AIMS AND OBJECTIVES

The Rayleigh distribution characterizes ultrasound scattering in a 'normal' situation (see also figure 18 (a)). We have seen that this implies a large number of independently distributed scatterers in a resolution cell with random locations. Normal ovarian tissue could therefore be Rayleigh distributed. Nevertheless, if the 'real' number of scatterers is too low (as in figure 18 (c)), the theorem leading to Rayleigh statistics breaks down.[129] Under these circumstances, the K-distribution might be an alternative, meaning that normal ovarian tissue (with a low number of scatterers) can also be K-distributed without constraints on values for the parameters a and b.

In case of tumour growth in the tissue the situation of figure 18 (b) seems plausible. Thus the Rician distribution could potentially be characteristic of tumour growth. But can its parameters help us distinguishing benign from malignant tumours?

On the other hand, when abnormalities arise in specific regions of the tissue, it is more likely that these regions will have widely varying scatterers' cross-sections (this means that $\varepsilon \to -1$). This would induce a decrease in the number of 'effective' scatterers ($N(1+\varepsilon)$) and so the K-model holds, albeit for abnormal tissue in this case.

Based on these findings, we could state the following **research hypotheses**:

h1) Normal ovarian tissue follows the Rayleigh distribution
h2) Normal and abnormal ovarian tissue can both be K-distributed
h3) Abnormal ovarian tissue is more likely to be Rician distributed

Also the following **research questions** could be stated:

q1) Are the parameters of the K distribution different for normal and abnormal ovarian tissues?
q2) Can the parameters of the Rician distribution differentiate between benign and malignant regions?

The Weibull and Nakagami models were shown to be applicable to the 3 situations (Rayleigh, post- and pre-Rayleigh) of scattering conditions. Therefore, both normal

and abnormal tissue types can be Weibull as well as Nakagami distributed. Using the previously described relations between the other 3 distributions and the Weibull and Nakagami parameters, respectively, the abovementioned hypotheses and questions could be translated into the statements that can be found in table 17 and table 18.

TABLE 17: RESEARCH HYPOTHESES REPHRASED FOR THE WEIBULL AND NAKAGAMI MODELS

Hyp.		Weibull	Nakagami
h1	Normal ovarian tissue	a ~2	m ~1
h2	Normal and abnormal ovarian tissue	a <2	m <1
h3	Abnormal ovarian tissue	a >2	m >1 **or** in]0.5;1[AND non-uniformly distributed phases

TABLE 18: RESEARCH QUESTIONS REPHRASED FOR THE WEIBULL AND NAKAGAMI MODELS

Ques.	Weibull	Nakagami
q1	Are the a 's that are < 2 different for normal and abnormal ovarian tissue?	Are the m 's that are <1 different for normal and abnormal ovarian tissue?
q2	Can a 's that are >2 differentiate between benign and malignant?	Can m 's that are >1, **or** in]0.5;1[together with phase information, differentiate between benign and malignant?

CHAPTER 3: PATIENTS AND METHODS

1. PATIENTS, ULTRASONOGRAPHY AND RF-DATA

1.1. CLINICAL SETTING AND PATIENTS

We have used data from an ongoing translational-like study that has the purpose of developing Ovarian HistoScanning on RF-data. The study takes place in Oslo, Norway, at the Norwegian Radium Hospital. Eligibility criteria are: women ≥ 18 years with Ultrasound scans performed according to validated guidelines and scheduled for the surgical removal of one or two ovaries due to (a) a suspicion of cancer (b) prophylactic oophorectomy either alone or as part of hysterectomy for a uterine pathology. Exclusion criteria are: unsuitable or unwilling to enter the study or to proceed to surgical investigation. The study is managed according to the guidelines provided in the declaration of Helsinki and the EU Directive on Good Clinical Practice. The study protocol was approved by the local ethical committee. An informed consent was obtained from all patients that have participated so far.

The main objective of this ongoing study is for the Advanced Medical Diagnostics (AMD) research and development team to develop algorithms and classifiers for the purpose of detecting malignant ovarian tissue areas on the RF-data. This study has started in July 2008 and will end as soon as 90 eligible patients with good quality data are included.

1.2. ULTRASONOGRAPHY

All eligible women underwent Transvaginal sonography (TVS) shortly before surgery by an expert having more than 10 years of experience. A Logiq9 expert scanner from General Electric (GE) equipped with a 4DE7C, 3.3-10 MHz Real Time 3D/4D intracavity convex-volume probe was used. For the purpose of this study, the GE Logiq9 scanner used at the study centre was enabled to allow the user to

save the RF voxel data, as well as the usual gray-scale voxel data. These RF-data were used for this work.

The examination involved conventional 2D-TVS for identifying ovaries and/or adnexal masses followed by a standardised 3D-TVS procedure to capture the 3D image of all structures and to save the RF-data. Often, several 3D acquisitions had to be made to cover a mass because the RF-data were too heavy for the machine to store in one acquisition. Every 3D RF-acquisition was done according to validated guidelines and by fixing the US machine's preset to a predefined one. All these US-data was taken to the AMD-offices at Waterloo together with the completed CRFs.

1.3. RF-DATA WORK-UP

In this work, Matlab version R2008a was used (The Math Works Inc., Natick, MA) and PRSD studio version 2.2.4 (PR Sys Design, Delft, The Netherlands).

For 17 patients for whom we had RF-data, completed pathology reports and completed CRFs in the (still growing) database, a total of 63 scans were taken (sometimes up to 9 scans were taken for one ovary for the reasons mentioned above). It was decided to use one scan per ovary for this work (a total of N=18 scans). The pathology of these 18 ovaries (one patient had left and right ovary in the database) was as follows: 5 normal, 5 benign, 3 borderline and 5 malignant ovaries.

Subsequently, RF-data of these 18 scans have been loaded into the HistoScanning software application. This application depicts 3 different views of the full 3D-volume and thus allows marking a volume of interest (VOI) on the 3D-data 'box'. For each scan (ovary) in our data base, VOIs that typically characterize malignant, benign and normal ovaries were marked together with an independent expert gynaecologist. Then, the points (voxels) that belong to these VOI were exported to a file that Matlab could read for further analysis. In (every subsequent frame of) these VOI, we selected areas of 11 by 11 pixels, also referred to as units. Sixteen features were estimated for each of these units: the Rayleigh- σ, the Rician- v and σ, the K- a and b, the Weibull- a and b, the Nakagami m and Ω, 5 χ^2-statistics assessing the fit to these 5 distributions (envelope data) and 1 χ^2 assessing the fit to the uniform distribution (phase data) and finally the SNR.

In the next section, more computational details on all these 16 estimations (also called features) are given.

2. CALCULATION OF THE DIFFERENT FEATURES

2.1. ESTIMATION OF THE PARAMETERS OF THE DIFFERENT STATISTICAL DISTRIBUTIONS

The parameters of the Rayleigh and Weibull distributions have been estimated by the normal Maximum Likelihood (MLE) methods ('raylfit' and 'weibfit' respectively in Matlab) available in the Matlab stats toolbox (The Math Works Inc., Natick, MA). For the other distributions, the MLE procedure could be implemented as an optimization problem using the Nelder-Mead simplex method available in the Matlab optimization toolbox. This method requires an initial guess of the solution and then it starts looking for a maximum of the log-likelihood function. On our data, this optimization method often ran into problems and significantly slowed down the calculation speed; it was therefore not done and only the initial guesses were used. In order to obtain an initial guess for the different parameters of the Rician, K and Nakagami distributions, the moments estimators based on the second and fourth order moments of the data have been used. The initial guesses are given by:[128]

$s_{ini} = \sqrt[4]{2\left[E\left(r^2\right)\right]^2 - E\left(r^4\right)}$ and $\sigma_{ini} = \sqrt{\frac{E(r^2) - s_{ini}^2}{2}}$ for the Rician distribution

$a_{ini} = \frac{2}{\frac{E\left(r^4\right)}{\left[E\left(r^2\right)\right]^2} - 2}$ and $b_{ini} = \sqrt{\frac{4a_{ini}}{E\left(r^2\right)}}$ for the K-distribution

$m_{ini} = \frac{\left[E\left(r^2\right)\right]^2}{E\left(r^4\right) - \left[E\left(r^2\right)\right]^2}$ and $\Omega_{ini} = E\left(r^2\right)$ for the Nakagami distribution.

Intuitively it can easily be seen that the Rician and K distribution encounter for two different scattering conditions. So when the data are pre-Rayleigh distributed (K-distributed), the estimation of s_{ini} is a complex number. On the other hand, if the data are post-Rayleigh (Rician), a_{ini} is negative so it never occurs that both of them are calculated on the same empirical data.

2.2. GOODNESS OF FIT STATISTICS

In the previous section, estimation methods for the parameters of Rayleigh, Rician, K, Weibull, and Nakagami distributions were described. It has been shown how these different statistical distribution models apply to different tissue conditions (see figure 18). The goodness-of-fit of these probability models on our (empirical) US-data was assessed using the standard χ^2-statistic[130] for goodness-of-fit. Additionally, as explained above, the uniformity of the phase characterizes whether the data adheres to the post-Rayleigh situation. So, the χ^2-statistic was also used to assess the goodness-of-fit of the uniform distribution $\sim U(]-\pi,\pi[)$ on the phase data.

We have implemented a Matlab algorithm that calculates the χ^2-statistics using 20 bins for the 5 above-mentioned distributions (envelope data) and for the uniform distribution (phase data).

CHAPTER 4: RESULTS

1. PARAMETER ESTIMATIONS AND GOODNESS OF FIT STATISTICS

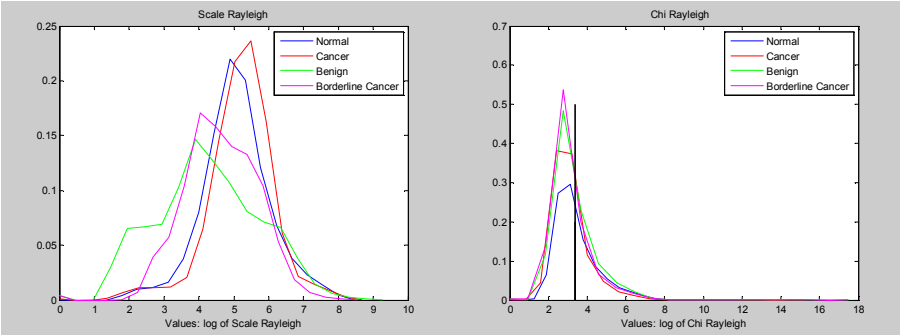
1.1. RAYLEIGH

TABLE 19: MEAN AND VARIANCE OF THE ESTIMATIONS OF RAYLEIGH σ AND OF THE χ^2_{ra} OF FIT TO THE RAYLEIGH DISTRIBUTION

σ	n_u*	Mean	Variance
Normal	48261	5.07	0.96
Benign	167469	4.27	2.22
Borderline Cancer	251578	4.63	1.10
Cancer	207901	5.21	0.92
χ^2_{ra}	n_u*	Mean	Variance
Normal	48261	3.40	1.17
Benign	167469	3.26	1.13
Borderline Cancer	251578	3.10	0.96
Cancer	207901	3.10	0.81

***n_u** = number of units

FIGURE 20: DISTRIBUTION OF THE ESTIMATIONS OF RAYLEIGH σ AND OF THE χ^2_{ra} GOODNESS OF FIT STATISTIC



Right panel: The black line represents the critical value of a χ^2-distribution with 18 degrees of freedom (20 bins – 1 parameter -1). This value corresponds to about the 75th percentile for all 4 pathology groups.

Findings:

- The Rayleigh σ estimations were on average the highest for units coming from cancerous tissue (see also table 19 and figure 20)
- The units from benign tissue had on average the lowest values for σ
- For about 75% of all units a Rayleigh distribution could not be rejected, regardless of the pathology group they belong to.

1.2. RICIAN

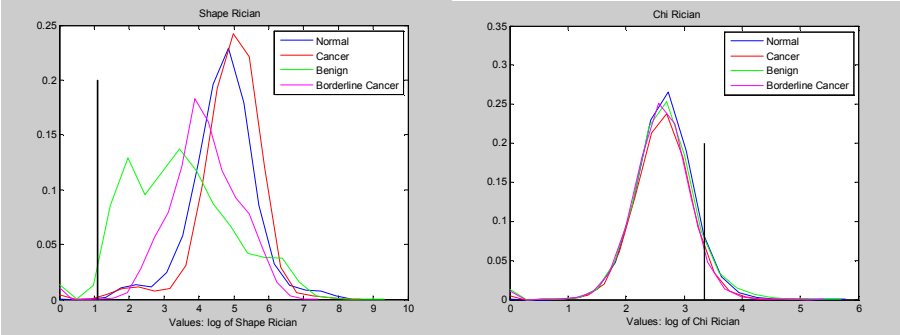
For a total of 33% of all units the Rician parameters were calculated (the other 67% had K parameters). The per-pathology percentages can be found in table 20.

TABLE 20: MEAN AND VARIANCE OF THE ESTIMATIONS OF RICIAN ν AND OF THE χ^2_{ri} OF FIT TO THE RICIAN DISTRIBUTION

ν	n_u* (n%)	Mean	Variance
Normal	48261 (25%)	4.70	0.87
Benign	167469 (30%)	3.45	2.19
Borderline Cancer	251578 (37%)	4.14	1.03
Cancer	207901 (33%)	4.91	0.86
χ^2_{ri}	n_u* (n%)	Mean	Variance
Normal	48261 (25%)	2.67	0.22
Benign	167469 (30%)	2.64	0.33
Borderline Cancer	251578 (37%)	2.59	0.27
Cancer	207901 (33%)	2.63	0.24

*n_u = number of units

FIGURE 21: DISTRIBUTION OF THE ESTIMATIONS OF RICIAN ν AND OF THE χ^2_{ri} GOODNESS OF FIT STATISTIC



Left panel: The black line represents the value of ν that corresponds to 2 on a regular scale. When ν is larger than 2 the Rician distribution starts to deviate from the Rayleigh a lot

Right panel: The black line represents the critical value of a χ^2 -distribution with 17 degrees of freedom (20 bins – 2 parameter -1). This value corresponds to about the 95th percentile for all 4 pathology groups.

Findings:

- The Rician v parameter of the envelopes had on average higher values for units coming from cancerous tissue (see also table 20 and figure 21)
- The units representing benign tissue, on the other hand, presented with on average low values for v
- For the majority of all units with parameter estimations for the Rician distribution (33%), the fit of their data to the Rician distribution couldn't be rejected (χ^2_{ri}), regardless of the pathology they belong to (figure 21).
- These 33% of units showed similar values for χ^2_{ra} as they did for χ^2_{ri}. This implies that for the majority of these units, a Rayleigh distribution could not be rejected either. Whether these units are thus really of a 'post-Rayleigh' regime is not evident.

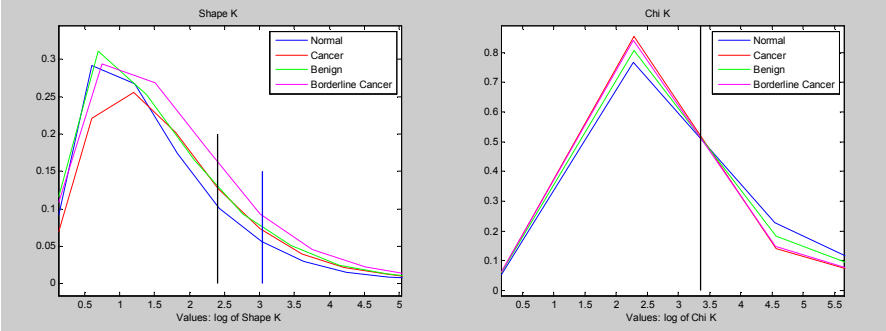
1.3. K

TABLE 21: MEAN AND VARIANCE OF THE ESTIMATIONS OF K a AND OF THE χ^2_K OF FIT TO THE K DISTRIBUTION

a	**n_u* (n%)**	**Mean**	**Variance**
Normal	48261 (75%)	1.52	1.16
Benign	167469 (70%)	1.59	1.34
Borderline Cancer	251578 (63%)	1.71	1.40
Cancer	207901 (67%)	1.72	1.29
χ^2_K	**n_u* (n%)**	**Mean**	**Variance**
Normal	48261 (75%)	3.02	0.98
Benign	167469 (70%)	2.94	1.47
Borderline Cancer	251578 (63%)	2.86	1.41
Cancer	207901 (67%)	2.87	1.27

***n_u** = number of units

FIGURE 22: DISTRIBUTION OF THE ESTIMATIONS OF K α AND OF THE χ^2_K GOODNESS OF FIT STATISTIC



Left panel: The black/blue line represents the value of α that corresponds to 10/20 on a regular scale. The 88th percentile corresponds to that blue line (α =20).

Right panel: The black line represents the critical value of a χ^2 -distribution with 17 degrees of freedom (20 bins – 2 parameter -1). This value corresponds to about the 80th percentile for all 4 pathology groups.

Findings:

- The K-α parameter had on average higher values for units coming from borderline and cancerous tissue (see also table 21 and figure 22) compared to the benign and normal units
- The effective number of scatterers, as represented by α, was indeed low (≤20) for the majority of these 67% of units with K parameters.
- For the majority of units of all 4 pathologies, a fit to the K-distribution could not be rejected.
- Nearly half of these 67% of all units, had values of χ^2_{ra} that were higher than the critical value meaning that for them, a Rayleigh distribution could be rejected. The pre-Rayleigh regime is thus applicable for these units and K is a good model for them.

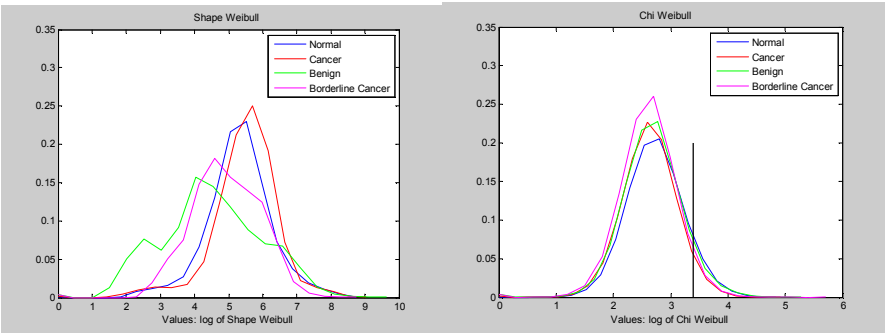
1.4. WEIBULL

TABLE 22: MEAN AND VARIANCE OF THE ESTIMATIONS OF WEIBULL a AND OF THE χ^2_{Wei} OF FIT TO THE WEIBULL DISTRIBUTION

a	n_u*	Mean	Variance
Normal	48261	5.31	0.90
Benign	167469	4.51	2.12
Borderline Cancer	251578	4.91	1.02
Cancer	207901	5.49	0.92
χ^2_{Wei}	n_u*	Mean	Variance
Normal	48261	2.73	0.24
Benign	167469	2.68	0.27
Borderline Cancer	251578	2.61	0.24
Cancer	207901	2.63	0.22

***n_u** = number of units

FIGURE 23: DISTRIBUTION OF THE ESTIMATIONS OF WEIBULL a AND OF THE χ^2_{Wei} GOODNESS OF FIT STATISTIC



Right panel: The black line represents the critical value of a χ^2-distribution with 17 degrees of freedom (20 bins – 2 parameter -1). This value corresponds to about the 95th percentile for all 4 pathology groups.

Findings :

- The Weibull a estimations behave very similar to the estimations of the Rayleigh σ in terms of mean values (table 22) as well as for the shape of their distributions (figure 23)

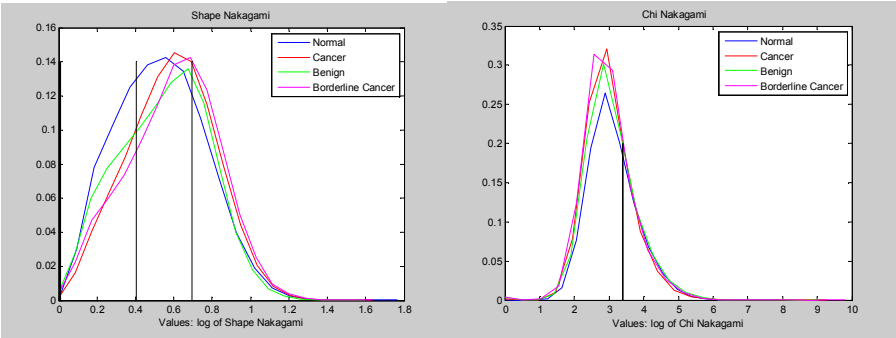
1.5. NAKAGAMI

TABLE 23: MEAN AND VARIANCE OF THE ESTIMATIONS OF NAKAGAMI m AND OF THE χ^2_{Nak} OF FIT TO THE NAKAGAMI DISTRIBUTION

m	n_u*	Mean	Variance
Normal	48261	0.53	0.054
Benign	167469	0.55	0.055
Borderline Cancer	251578	0.59	0.057
Cancer	207901	0.59	0.052
χ^2_{Nak}	n_u*	Mean	Variance
Normal	48261	3.11	0.49
Benign	167469	3.04	0.54
Borderline Cancer	251578	2.97	0.50
Cancer	207901	2.98	0.45

***n_u** = number of units

FIGURE 24: DISTRIBUTION OF THE ESTIMATIONS OF NAKAGAMI m AND OF THE χ^2_{Nak} GOODNESS OF FIT STATISTIC



Left panel: In between the 2 black lines lie all the m values in]0.5; 1[on a regular scale.

Right panel: The black line represents the critical value of a χ^2-distribution with 17 degrees of freedom (20 bins – 2 parameter -1). This value corresponds to about the 75th percentile for all 4 pathology groups.

Findings :

- The Nakagami m parameter had on average higher values for borderline and cancerous tissue (figure 24 and table 23) and lower values for benign and normal.
- All units for which the Nakagami m parameter is < 1 have estimations for K and not for Rician and the reverse is true for units with $m \geq 1$
- For the units with m in]0.5;1[(a total of about 168000), the goodness of fit of the phase data to the uniform distribution was assessed and for nearly half of them the uniform distribution was rejected for the phase data. However, this χ^2 of goodness of fit for the phase data did not show any differentiating capacities.

CHAPTER 5: CONCLUSIONS AND DISCUSSION

1. CONCLUSIONS

Statistical distribution models were derived for the case of US data from scatterer-areas. Researchers have tried to give those models a sensible meaning according to the possible scattering conditions (see also figure 18) for soft tissues and have distinguished 3 regimes: Rayleigh, post and pre-Rayleigh. That the 5 distributions, introduced in this work, apply to one of the regimes for US data has been shown by computer simulations and by using phantom data.[118,124,131] In this work we have investigated whether they apply to US data (RF) coming from ovarian scans. In addition, the usefulness of these distributions to differentiate between different ovarian tissue types was also assessed. We have formulated a number of research hypothesis and questions in chapter 2 and have performed a number of analyses in chapter 4 to gain insight into the situation of US data of ovarian tissues and to investigate the posed hypotheses:

h1) The majority (about 75%) of normal ovarian units indeed follow the Rayleigh distribution. However, the same is true for the 3 other pathologies. Moreover, the χ^2 - statistic that characterizes the goodness-of-fit to the Rayleigh distribution is not distributed differently for the normal and abnormal pathologies (figure 20). The Rayleigh σ parameter, however, is on average the lowest for units representing benign tissue.

h2) The K-distribution was applicable on 67% of all units. It was indeed a representative model for normal as well as abnormal ovarian tissue.

h3) On our data, we did not see a tendency of finding the Rician distribution more often for abnormal ovarian tissue.

The following questions were also answered:

q1) Units representing abnormal ovarian tissue have on average higher values for the parameter a of the K-distribution. This is not true though for benign units and in general the distributions of a for the 4 pathology groups highly overlap.

q2) On these data, the Rician v clearly shows the capacity to differentiate between benign and malignant regions.

Alternatively, these hypotheses and questions were rephrased for the more general Weibull and Nakagami models. The results in chapter 4 show the following:

h1) Weibull a values are not concentrated around 2, regardless of the pathology. For the Nakagami m, values are concentrated around 1 and normal units are slightly more likely to take values around 1 compared to abnormal ones.

h2) Only very few units seem to take Weibull a's < 2. All units with Nakagami m's < 1 were indeed also K-distributed thus confirming the link between those 2 models. Also, one could find all 4 pathologies taking m values < 1.

h3) On our data, we did not see a tendency of finding more abnormal ovarian regions with values for Weibull a > 2 but we did see a higher tendency of Nakagami m > 1 in the abnormal regions. For the units that had m in]0.5;1[the distribution of the phase data didn't help in differentiating normal from abnormal.

q1) Only very few units seem to take Weibull a's < 2. Units representing abnormal ovarian tissue have on average higher values for Nakagami m's < 1.

q2) On these data, the Weibull a's > 2 clearly show the capacity to differentiate between benign and malignant regions but Nakagami m's that are >1 don't. Analysing the distribution of the phase data for units with m in]0.5;1[doesn't add differentiating value.

From this work, it results that normal ovarian regions are not Rayleigh-distributed and abnormal ovarian regions are not more likely to follow Rician-statistics.

Nevertheless, there seems to be potential to use the parameter of the Rician distribution to distinguish between certain benign and malignant regions. The Weibull distribution fits ovarian data but it doesn't 'obey' to the rules about its parameters as stated by Raju et al.[128]

In conclusion, this study supports the hypothesis that pre-Rayleigh statistics are present in the case of ovarian tissues (at least for the imaging system used for this work). Indeed, the K-model was representative of normal as well as abnormal data, wherever it was applicable (here this was on 67% of the data). Moreover, the parameters of the K-distribution might show capacity to find certain malignant ovarian regions. Also, the more general Nakagami that is applicable in the Rayleigh, post- and pre-Rayleigh regimes, is a suitable model for ovarian data. In a certain region the m -parameter has some potential to do tissue classification.

2. DISCUSSION

Raju and Srinivasan[128] have studied the Statistics of the envelope of Ultrasonic backscatter from human skin tissue. The K and Weibull distributions fitted their skin data best and they have shown that the parameters of the statistical distributions could differentiate between skin tissue types. To our knowledge, studies involving other tissues have rarely assessed and compared the use of multiple distributions to their data. But still the parameters of these distributions have been widely used to do tissue classification.[74,119,122,123] This work evaluated which statistical regime was present in the case of backscatter from ovarian tissues and which distributions modelled those data best. Additionally, it was shown how parameters of these distributions could characterize certain ovarian tissue types.

The main difference between the results presented here and other tissue characterization studies,[74,119,122,123] is the unit of analysis that was used to estimate the parameters. In most other studies, a full region that represented the lesion in question was used. In other words, for each scan, there was 1 estimated value for each parameter. In our case, units of only 11 by 11 pixels were used, and so for each scan there could be tens, hundreds or even thousands of units and just as much parameter estimations. This was a deliberate choice, because our objective was to classify tissues on a small scale.

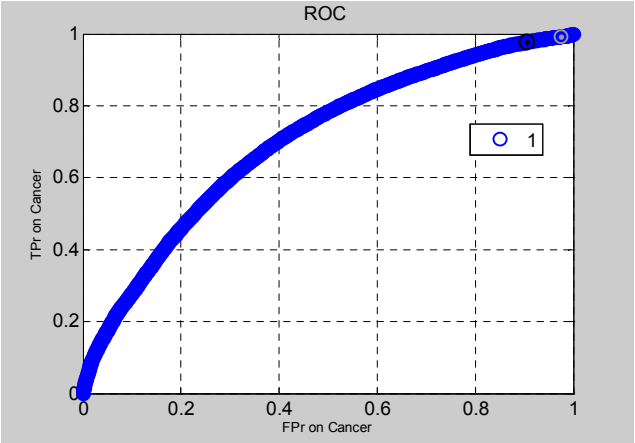
In part I of this manuscript, it was already discussed how difficult it is to mark a volume of interest (VOI) in ovarian scans. Some of the units that belong to an abnormal pathology class might in reality only lie on the border of the lesion or even fully outside the lesion. This might explain the huge overlap in the histograms of the different parameters and of the goodness-of-fit statistics.

In terms of the number of different patients, the sample size of this study was rather moderate. But given the fact that scans were analysed at the level of small units, we hope to be able to cover a wider range of possible morphological situations present in ovarian tissue.

This work presents an evaluation of how statistical models can allow a differentiation between ovarian tissue types. The next step would be to develop a so-called tissue classifier. This classifier would be the result of a huge classification exercise that would consist of finding the best statistical classification model based on the best (sub)set of these parameters and/or other features that are of different origin. Similar multi-parameter approaches have been used to do tissue classification in breast[123,132] and prostate cancer,[74] often combining very large sets of features of different origin. The overlap in the different histograms of the statistical features analysed here (figure 20 to figure 24) is too important to consider them alone.

We will give an illustrative example of how such a classifier would work: a simple classifier has been developed based on the following 7 features: Rician v and σ, the a and b of the K-distribution, Nakagami m and Ω and one extra feature of another origin. The classifier was a simple mixture of 2 Gaussian models, trained to distinguish between 'non-cancer' (normal and benign scans) and 'cancer' (borderline cancer and cancer scans) on 70% of our data. The remaining 30% of data was used to assess the performance of this classifier. The ROC-curve on this 30% of data for the Gaussian mixture classifier is shown in figure 25. The operating point that maximised the Youden Index presented a sensitivity of 68% and a specificity of 62% or similarly, 32% error on the cancer class and 38% error on the non-cancer class. Given that this classifier classifies **units** of a cancer or a non-cancer region (and not an entire cancer versus non-cancer scan) and that a huge overlap is therefore to be expected for the reasons explained above, this first attempt is already very encouraging. For instance, an additional analysis that extends this units-based classifier to a decision for each scan could improve sensitivities and specificities a lot.

FIGURE 25: ROC-CURVE FOR THE MIXTURE CLASSIFIER FOR CANCER/NON CANCER



This figure was made in PRSD studio version 2.2.4 (PR Sys Design, Delft, The Netherlands). The ROC curve is an empirical curve as described in Part I, chapter 3 and every cut-off point for which a sensitivity and specificity is calculated is represented by a blue circle. As the continuous results of the classifier is very dense, the resulting ROC curve is based on a high number of cut-off points which may give it a smooth appearance.

Future research should focus on building a cancer versus non-cancer classifier for RF-data of ovarian tissues. The Rician, K and Nakagami parameters are to be used together with additional features of a different origin. Secondarily, a classifier that attempts to distinguish between the 4 tissue types (normal, benign, borderline and cancer) could also be envisaged. Finally, all developed classifiers should be tested on independent patients for correct validation.

For the research presented in part II of this thesis, it was decided to focus on a rather broad notion of pathology (normal, benign, borderline and cancer) and also to use an entire region on the US-image (represented by a VOI) to be characteristic of this notion. This was done because it was the most interesting choice with respect to the aim of our work. One may wonder, however, whether these statistical models and/or their parameters can differentiate between different tissue types from a histological point of view? For example, classifying 'units' representing serous or mucinous histological tissue types. Would the 5 statistical models that were introduced here, be more representative of one or the other?

Would their parameters assist us in differentiating between such tissue types? These questions represent interesting future perspectives.

Similarly, when a radiologist (or any other examiner) analyses an ovarian mass on an US-image, he or she might be looking for specific features like the edges of the mass, as having smooth or ruptured edges may mean the difference between being a borderline cancer or a malignant cancer.[62] Picking up blood vessels can also be an indicative feature of malignancy. The disorganisation of the tissue that is locally caused by rupturing borders or vessel formation might be specifically modelled by the Rician or K distribution? And this might only happen on the edges and not inside the mass? As we are using units of analysis that are very small (11 by 11 pixels only), we might be picking such typical 'edge' phenomena up by these statistical models or their parameters. To asses whether this happens in particular around edges, the local analysis (units) would have to be extended to the consecutive units by means of a 2D (or even 3D) surface analysis.

These latter future perspectives are nevertheless only feasible for ovarian cancer if one can define a more 'quantitative' and detailed gold standard notion on the level of the VOI. This means that instead of just defining a VOI representative of malignant or benign tissue in general, a VOI would contain a more detailed spatial information (e.g. where are the borders and how do they look like?), a more detailed histological information (implying a very detailed pathology report), correct measurements of the lesion and a sound way to map such details from histopathology to the image. In such an ideal scenario, training a huge classification scheme with as starting point a local analysis (level of the units) including these statistical models and their parameters seems interesting.

REFERENCES

1. Ferlay J, Parkin DM, Steliarova-Foucher E. Estimates of cancer incidence and mortality in Europe in 2008. *Eur J Cancer* 2010;46:765--781.

2. Quinn MJ, d'Onofrio A, Moller B, et al. Cancer mortality trends in the EU and acceding countries up to 2015. *Ann Oncol* 2003;14:1148--1152.

3. Sankaranarayanan R, Ferlay J. Worldwide burden of gynaecological cancer: the size of the problem. *Best Pract Res Clin Obstet Gynaecol* 2006;20:207--225.

4. Granberg S, Wikland M, Jansson I. Macroscopic characterization of ovarian tumors and the relation to the histological diagnosis: criteria to be used for ultrasound evaluation. *Gynecol Oncol* 1989;35:139--144.

5. Myers ER, Bastian LA, Havrilesky LJ et al: Management of Adnexal Mass: Evidence Report/Technology Assessment No. 130. Rockville, MD, Agency for healthcare Research and Quality, AHRQ Publication No. 06-E004, 2006

6. Tavassoli FA, Devilee P. World Health Organisation Classification of Tumours. Pathology & Genetics of Tumours of the Breast and Female Genital Organs. IARC press: Lyon, 3rd edition, 2003

7. Benedet JL, Bender H, Jones H, III, et al. FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. FIGO Committee on Gynecologic Oncology. *Int J Gynaecol Obstet* 2000;70:209--262.

8. Ferlay J, Shin HR, Bray F, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;

9. Board RE, Bruijns CT, Pronk AE, et al. Stage- and CA125-related survival in patients with epithelial ovarian cancer treated at a cancer center. *Int J Gynecol Cancer* 2006;16:18--24.

10. Parkin DM, Bray F, Ferlay J, et al. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74--108.

11. Kristensen GB, Trope C. Epithelial ovarian carcinoma. *Lancet* 1997;349:113--117.

12. Godard B, Foulkes WD, Provencher D, et al. Risk factors for familial and sporadic ovarian cancer among French Canadians: a case-control study. *Am J Obstet Gynecol* 1998;179:403--410.

13. Whittemore AS, Harris R, Itnyre J. Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case-control studies. II. Invasive epithelial ovarian cancers in white women. Collaborative Ovarian Cancer Group. *Am J Epidemiol* 1992;136:1184--1203.

14. Riman T, Dickman PW, Nilsson S, et al. Hormone replacement therapy and the risk of invasive epithelial ovarian cancer in Swedish women. *J Natl Cancer Inst* 2002;94:497--504.

15. Riman T, Dickman PW, Nilsson S, et al. Risk factors for invasive epithelial ovarian cancer: results from a Swedish case-control study. *Am J Epidemiol* 2002;156:363--373.

16. Hankinson SE, Hunter DJ, Colditz GA, et al. Tubal ligation, hysterectomy, and risk of ovarian cancer. A prospective study. *JAMA* 1993;270:2813--2818.

17. Harlow BL, Hartge PA. A review of perineal talc exposure and risk of ovarian cancer. *Regul Toxicol Pharmacol* 1995;21:254--260.

18. Whittemore AS, Wu ML, Paffenbarger RS Jr, et al. Personal and environmental characteristics related to epithelial ovarian cancer. II. Exposures to talcum powder, tobacco, alcohol, and coffee. *Am J Epidemiol* 1988;128:1228--1240.

19. Le T, Giede C, Salem S, et al. Initial evaluation and referral guidelines for management of pelvic/ovarian masses. *J Obstet Gynaecol Can* 2009;31:668--680.

20. Goff BA, Mandel LS, Melancon CH, et al. Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. *JAMA* 2004;291:2705--2712.

21. Ghezzi F, Cromi A, Siesto G, et al. Laparoscopy staging of early ovarian cancer: our experience and review of the literature. *Int J Gynecol Cancer* 2009;19 Suppl 2:S7--S13.

22. Ghezzi F, Cromi A, Uccella S, et al. Laparoscopy versus laparotomy for the surgical management of apparent early stage ovarian cancer. *Gynecol Oncol* 2007;105:409--413.

23. Le T, Krepart GV, Lotocki RJ, et al. Does debulking surgery improve survival in biologically aggressive ovarian carcinoma? *Gynecol Oncol* 1997;67:208--214.

24. Rubin SC, Randall TC, Armstrong KA, et al. Ten-year follow-up of ovarian cancer patients after second-look laparotomy with negative findings. *Obstet Gynecol* 1999;93:21--24.

25. Vergote I, Trope CG, Amant F, et al. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *N Engl J Med* 2010;363:943--953.

26. Chan JK, Kapp DS, Shin JY, et al. Influence of the gynecologic oncologist on the survival of ovarian cancer patients. *Obstet Gynecol* 2007;109:1342--1350.

27. Engelen MJ, Kos HE, Willemse PH, et al. Surgery by consultant gynecologic oncologists improves survival in patients with ovarian carcinoma. *Cancer* 2006;106:589--598.

28. Giede KC, Kieser K, Dodge J, et al. Who should operate on patients with ovarian cancer? An evidence-based review. *Gynecol Oncol* 2005;99:447--461.

29. Earle CC, Schrag D, Neville BA, et al. Effect of surgeon specialty on processes of care and outcomes for ovarian cancer patients. *J Natl Cancer Inst* 2006;98:172--180.

30. Paulsen T, Kjaerheim K, Kaern J, et al. Improved short-term survival for advanced ovarian, tubal, and peritoneal cancer patients operated at teaching hospitals. *Int J Gynecol Cancer* 2006;16 Suppl 1:11--17.

31. Eisenkop SM, Spirtos NM, Montag TW, et al. The impact of subspecialty training on the management of advanced ovarian cancer. *Gynecol Oncol* 1992;47:203--209.

32. Sassone AM, Timor-Tritsch IE, Artner A, et al. Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy. *Obstet Gynecol* 1991;78:70--76.

33. DePriest PD, Varner E, Powell J, et al. The efficacy of a sonographic morphology index in identifying ovarian cancer: a multi-institutional investigation. *Gynecol Oncol* 1994;55:174--178.

34. Lerner JP, Timor-Tritsch IE, Federman A, et al. Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system. *Am J Obstet Gynecol* 1994;170:81--85.

35. Ueland FR, DePriest PD, Pavlik EJ, et al. Preoperative differentiation of malignant from benign ovarian tumors: the efficacy of morphology indexing and Doppler flow sonography. *Gynecol Oncol* 2003;91:46--50.

36. Kinkel K, Hricak H, Lu Y, et al. US characterization of ovarian masses: a meta-analysis. *Radiology* 2000;217:803--811.

37. Jacobs I, Oram D, Fairbanks J, et al. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *Br J Obstet Gynaecol* 1990;97:922--929.

38. Bailey J, Tailor A, Naik R, et al. Risk of malignancy index for referral of ovarian cancer cases to a tertiary center: does it identify the correct cases? *Int J Gynecol Cancer* 2006;16 Suppl 1:30--34.

39. Royal College of Obstetricians and Gynaecologists. Ovarian cysts in postmenopausal women, Evidence-Based Clinical Guideline No. 34. Available at:http://www.rcog.org.uk/womens-health/clinical-guidance/ovarian-cysts-postmenopausal-women-green-top-34. Accessed 9 February, 2009.

40. Tingulstad S, Hagen B, Skjeldestad FE, et al. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals. *Obstet Gynecol* 1999;93:448--452.

41. Bullinaria J. Introduction to Neural Computation. Available at: http://www.cs.bham.ac.uk/~jxb/inc.html. Accessed 24 December, 2010.

42. Timmerman D, Testa AC, Bourne T, et al. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group. *J Clin Oncol* 2005;23:8794--8801.

43. Tailor A, Jurkovic D, Bourne TH, et al. Sonographic prediction of malignancy in adnexal masses using multivariate logistic regression analysis. *Ultrasound Obstet Gynecol* 1997;10:41--47.

44. Yamashita Y, Hatanaka Y, Torashima M, et al. Characterization of sonographically indeterminate ovarian tumors with MR imaging. A logistic regression analysis. *Acta Radiol* 1997;38:572--577.

45. Alcazar JL, Errasti T, Laparte C, et al. Assessment of a new logistic model in the preoperative evaluation of adnexal masses. *J Ultrasound Med* 2001;20:841--848.

46. Prompeler HJ, Madjar H, Sauerbrei W, et al. Diagnostic formula for the differentiation of adnexal tumors by transvaginal sonography. *Obstet Gynecol* 1997;89:428--433.

47. Tailor A, Jurkovic D, Bourne TH, et al. Sonographic prediction of malignancy in adnexal masses using an artificial neural network. *Br J Obstet Gynaecol* 1999;106:21--30.

48. Van Calster B, Timmerman D, Lu C, et al. Preoperative diagnosis of ovarian tumors using Bayesian kernel-based methods. *Ultrasound Obstet Gynecol* 2007;29:496--504.

49. Smolen A, Czekierdowski A, Stachowicz N, et al. [Use of multilayer perception artificial neutral networks for the prediction of the probability of malignancy in adnexal tumors]. *Ginekol Pol* 2003;74:855--862.

50. Szpurek D, Moszynski R, Smolen A, et al. Using logistic regression analysis in preliminary differential diagnosis of adnexal masses. *Int J Gynecol Cancer* 2005;15:817--823.

51. Moszynski R, Szpurek D, Smolen A, et al. Comparison of diagnostic usefulness of predictive models in preliminary differentiation of adnexal masses. *Int J Gynecol Cancer* 2006;16:45--51.

52. Mousavi AS, Borna S, Moeinoddini S. Estimation of probability of malignancy using a logistic model combining color Doppler ultrasonography, serum CA125 level in women with a pelvic mass. *Int J Gynecol Cancer* 2006;16 Suppl 1:92--98.

53. Timmerman D, Verrelst H, Bourne TH, et al. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. *Ultrasound Obstet Gynecol* 1999;13:17--25.

54. Alcazar JL, Jurado M. Using a logistic model to predict malignancy of adnexal masses based on menopausal status, ultrasound morphology, and color Doppler findings. *Gynecol Oncol* 1998;69:146--150.

55. Schutter EM, Sohn C, Kristen P, et al. Estimation of probability of malignancy using a logistic model combining physical examination, ultrasound, serum CA 125, and serum CA 72-4 in postmenopausal women with a pelvic mass: an international multicenter study. *Gynecol Oncol* 1998;69:56--63.

56. Lu C, Van Gestel T, Suykens JA, et al. Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. *Artif Intell Med* 2003;28:281--306.

57. Van Holsbeke C, Van Calster B, Valentin L, et al. External validation of mathematical models to distinguish between benign and malignant adnexal tumors: a multicenter study by the International Ovarian Tumor Analysis Group. *Clin Cancer Res* 2007;13:4440--4447.

58. Geomini P, Kruitwagen R, Bremer GL, et al. The accuracy of risk scores in predicting ovarian malignancy: a systematic review. *Obstet Gynecol* 2009;113:384--394.

59. Tingulstad S, Hagen B, Skjeldestad FE, et al. Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses. *Br J Obstet Gynaecol* 1996;103:826--831.

60. Morgante G, la Marca A, Ditto A, et al. Comparison of two malignancy risk indices based on serum CA125, ultrasound score and menopausal status in the diagnosis of ovarian masses. *Br J Obstet Gynaecol* 1999;106:524--527.

61. van Trappen PO, Rufford BD, Mills TD, et al. Differential diagnosis of adnexal masses: risk of malignancy index, ultrasonography, magnetic resonance imaging, and radioimmunoscintigraphy. *Int J Gynecol Cancer* 2007;17:61--67.

62. Timmerman D, Valentin L, Bourne TH, et al. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group. *Ultrasound Obstet Gynecol* 2000;16:500--505.

63. Aslam N, Banerjee S, Carr JV, et al. Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer. *Obstet Gynecol* 2000;96:75--80.

64. Timmerman D. The use of mathematical models to evaluate pelvic masses; can they beat an expert operator? *Best Pract Res Clin Obstet Gynaecol* 2004;18:91--104.

65. Timmerman D, Van Calster B, Jurkovic D, et al. Inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumors. *J Clin Oncol* 2007;25:4194--4200.

66. Timmerman D, Schwarzler P, Collins WP, et al. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. *Ultrasound Obstet Gynecol* 1999;13:11--16.

67. Valentin L. High-quality gynecological ultrasound can be highly beneficial, but poor-quality gynecological ultrasound can do harm. *Ultrasound Obstet Gynecol* 1999;13:1--7.

68. Valentin L, Hagen B, Tingulstad S, et al. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. *Ultrasound Obstet Gynecol* 2001;18:357--365.

69. Braeckman J, Autier P, Soviany C, et al. The accuracy of transrectal ultrasonography supplemented with computer-aided ultrasonography for detecting small prostate cancers. *BJU Int* 2008;102:1560--1565.

70. Braeckman J, Autier P, Garbar C, et al. Computer-aided ultrasonography (HistoScanning): a novel technology for locating and characterizing prostate cancer. *BJU Int* 2008;101:293--298.

71. Lucidarme O, Akakpo JP, Granberg S, et al. A new computer-aided diagnostic tool for non-invasive characterisation of malignant ovarian masses: results of a multicentre validation study. *Eur Radiol* 2010;20:1822--1830.

72. Tingulstad S, Skjeldestad FE, Halvorsen TB, et al. Survival and prognostic factors in patients with ovarian cancer. *Obstet Gynecol* 2003;101:885--891.

73. Weide R, Arndt M, Pandorf A, et al. Ovarian cancer treatment reality in northern Rheinland-Pfalz (Germany). Suboptimal surgical treatment as a possible cause for inferior survival. *Onkologie* 2007;30:611--617.

74. Scheipers U. Sonohistology - Methods and Systems for Ultrasonic Tissue Characterization based on a Multifeature Approach and Fuzzy Inference Systems (Thesis dissertation). Available at: http://www.scheipers.org/ Accessed 9 February, 2009.

75. Grenier B. Evaluation de la décision médicale - Introduction à l'analyse médico-économique. Masson S.A., Saint-Germain, Paris, 3rd edition, 1999

76. YOUDEN WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32--35.

77. Zou KH, Hall WJ, Shapiro DE. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Stat Med* 1997;16:2143--2156.

78. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561--577.

79. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29--36.

80. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839--843.

81. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837--845.

82. Metz CE. ROC analysis in medical imaging: a tutorial review of the literature. *Radiol Phys Technol* 2008;1:2--12.

83. Gur D, Bandos AI, Rockette HE. Comparing areas under receiver operating characteristic curves: potential impact of the "Last" experimentally measured operating point. *Radiology* 2008;247:12--15.

84. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. *Radiology* 2003;229:3--8.

85. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* 2006;163:670--675.

86. Wang TJ, Gona P, Larson MG, et al. Multiple biomarkers for the prediction of first major cardiovascular events and death. *N Engl J Med* 2006;355:2631--2639.

87. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157--172.

88. Bishop CM. Neural Networks for Pattern Recognition. Oxford University Press Inc., New York, USA, 3rd edition, 1995

89. Dorum A, Blom GP, Ekerhovd E, et al. Prevalence and histologic diagnosis of adnexal cysts in postmenopausal women: an autopsy study. *Am J Obstet Gynecol* 2005;192:48--54.

90. Han Jiawei and Kamber Micheline. Classification and Prediction, Data Mining - Concepts and Techniques. Elsevier, San Francisco, CA, 2003; pp 285-383

91. Aslam N, Tailor A, Lawton F, et al. Prospective evaluation of three different models for the pre-operative diagnosis of ovarian cancer. *BJOG* 2000;107:1347--1353.

92. Manjunath AP, Pratapkumar, Sujatha K, et al. Comparison of three risk of malignancy indices in evaluation of pelvic masses. *Gynecol Oncol* 2001;81:225--229.

93. Guerriero S, Alcazar JL, Pascual MA, et al. Intraobserver and interobserver agreement of grayscale typical ultrasonographic patterns for the diagnosis of ovarian cancer. *Ultrasound Med Biol* 2008;34:1711--1716.

94. Medeiros LR, Rosa DD, da Rosa MI, et al. Accuracy of ultrasonography with color Doppler in ovarian tumor: a systematic quantitative review. *Int J Gynecol Cancer* 2009;19:1214--1220.

95. Szpurek D, Moszynski R, Smolen A, et al. Artificial neural network computer prediction of ovarian malignancy in women with adnexal masses. *Int J Gynaecol Obstet* 2005;89:108--113.

96. Timmerman D, Van Calster B, Testa AC, et al. Ovarian cancer prediction in adnexal masses using ultrasound-based logistic regression models: a temporal and external validation study by the IOTA group. *Ultrasound Obstet Gynecol* 2010;

97. Hamilton W, Menon U. Ovarian cancer. *BMJ* 2009;339:b4650--b4650.

98. Van Holsbeke C, Yazbek J, Holland TK, et al. Real-time ultrasound vs. evaluation of static images in the preoperative assessment of adnexal masses. *Ultrasound Obstet Gynecol* 2008;32:828--831.

99. Van Holsbeke C, Van Calster B, Testa AC, et al. Prospective internal validation of mathematical models to predict malignancy in adnexal masses: results from the international ovarian tumor analysis study. *Clin Cancer Res* 2009;15:684--691.

100. Martus P. Statistical methods for the evaluation of diagnostic measurements concerning paired organs. *Stat Med* 2000;19:525--540.

101. Rosner B. Multivariate methods in ophthalmology with application to other paired-data situations. *Biometrics* 1984;40:1025--1035.

102. Engel J, Eckel R, Schubert-Fritschle G, et al. Moderate progress for ovarian cancer in the last 20 years: prolongation of survival, but no improvement in the cure rate. *Eur J Cancer* 2002;38:2435--2445.

103. Schapira MM, Matchar DB, Young MJ. The effectiveness of ovarian cancer screening. A decision analysis model. *Ann Intern Med* 1993;118:838--843.

104. Buys SS, Partridge E, Greene MH, et al. Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: findings from the initial screen of a randomized trial. *Am J Obstet Gynecol* 2005;193:1630--1639.

105. Menon U, Talaat A, Jeyarajah AR, et al. Ultrasound assessment of ovarian cancer risk in postmenopausal women with CA125 elevation. *Br J Cancer* 1999;80:1644--1647.

106. Jacobs I, Davies AP, Bridges J, et al. Prevalence screening for ovarian cancer in postmenopausal women by CA 125 measurement and ultrasonography. *BMJ* 1993;306:1030--1034.

107. Jacobs I, Stabile I, Bridges J, et al. Multimodal approach to screening for ovarian cancer. *Lancet* 1988;1:268--271.

108. Jacobs IJ, Skates SJ, MacDonald N, et al. Screening for ovarian cancer: a pilot randomised controlled trial. *Lancet* 1999;353:1207--1210.

109. Menon U, Jacobs I. Screening for ovarian cancer. *Best Pract Res Clin Obstet Gynaecol* 2002;16:469--482.

110. Menon U, Skates SJ, Lewis S, et al. Prospective study using the risk of ovarian cancer algorithm to screen for ovarian cancer. *J Clin Oncol* 2005;23:7919--7926.

111. Corens D. Health system review: Belgium. *Health Systems in Transition* 2007;9:1--172.

112. Liu J, Xu Y, Wang J. Ultrasonography, computed tomography and magnetic resonance imaging for diagnosis of ovarian carcinoma. *Eur J Radiol* 2007;62:328--334.

113. Bijnens B. Exploiting radiofrequency information in echocardiography. Acquisition, processing and illustrated applications. *Acta Biomedica Lovaniensia* 1997;

114. Dutt V. Statistical Analysis Of Ultrasound Echo Envelope (Thesis dissertation). Available at: http://mayoresearch.mayo.edu/mayo/research/ultrasound/upload/dutt.pdf Accessed 9 February, 2009.

115. Bates J, NetLibrary I. Practical gynaecological ultrasound. Greenwich Medical Media, 2nd edition, 1997

116. Johansson M. The Hilbert transform. Available at: http://w3.msi.vxu.se/exarb/mj_ex.pdf Accessed 9 February, 2009.

117. Wagner RF, Smith SW, Sandrik JM, et al. Statistics of Speckle in Ultrasound B-Scans. *IEEE Trans Son Ultrason* 1983;30:156--163.

118. Georgiou G, Cohen FS. Statistical Characterization of Diffuse Scattering in Ultrasound Images. *IEEE Trans Ultrason Ferroelectr Freq Control* 1998;45:57--64.

119. Shankar PM, Dumane VA, Reid JM, et al. Use of the K-distribution for classification of breast masses. *Ultrasound Med Biol* 2000;26:1503--1510.

120. Shankar PM. Ultrasonic tissue characterization using a generalized Nakagami model. *IEEE Trans Ultrason Ferroelectr Freq Control* 2001;48:1716--1720.

121. Dumane VA, Shankar PM. Use of frequency diversity and Nakagami statistics in ultrasonic tissue characterization. *IEEE Trans Ultrason Ferroelectr Freq Control* 2001;48:1139--1146.

122. Shankar PM, Dumane VA, Reid JM, et al. Classification of ultrasonic B-mode images of breast masses using Nakagami distribution. *IEEE Trans Ultrason Ferroelectr Freq Control* 2001;48:569--580.

123. Shankar PM, Dumane VA, Piccoli CW, et al. Computer-aided classification of breast masses in ultrasonic B-scans using a multiparameter approach. *IEEE Trans Ultrason Ferroelectr Freq Control* 2003;50:1002--1009.

124. Shankar PM. A General Statistical Model for Ultrasonic Backscattering from Tissues. *IEEE Trans Ultrason Ferroelectr Freq Control* 2000;47:727--736.

125. Tuthill TA, Sperry RH, Parker KJ. Deviations from Rayleigh statistics in ultrasonic speckle. *Ultrason Imaging* 1988;10:81--89.

126. Verhoeven JT, Thijssen JM, Theeuwes AG. Improvement of lesion detection by echographic image processing: signal-to-noise-ratio imaging. *Ultrason Imaging* 1991;13:238--251.

127. Jakeman E, Pusey PN. A model for non-Rayleigh sea echo. *IEEE Transactions on Antennas and Propagation* 1976;24:806--814.

128. Raju BI, Srinivasan MA. Statistics of Envelope of High-Frequency Ultrasonic Backscatter from Human Skin In Vivo. *IEEE Trans Ultrason Ferroelectr Freq Control* 2002;49:871--882.

129. Shankar PM, Reid JM, Ortega H, et al. Use of Non-Rayleigh Statistics for the Identification of Tumors in Ultrasonic B-Scans of the Breast. *IEEE Transactions on Medical Imaging* 1993;12:687--692.

130. Rosner B. Fundamentals of Biostatistics. Brooks/Cole, Duxbury, CA, 5th edition, 2000; pp 403-406

131. Narayanan VM, Shankar PM, Reid JM. Non-Rayleigh Statistics of Ultrasonic Backscattered Signals. *IEEE Trans Ultrason Ferroelectr Freq Control* 1994;41:845--852.

132. Gefen S, Tretiak OJ, Piccoli CW, et al. ROC analysis of ultrasound tissue characterization classifiers for breast cancer diagnosis. *IEEE Trans Med Imaging* 2003;22:170--177.

Scientific Curriculum Vitae

Evelien Vaes received her BSc and MSc degrees in mathematics from the Vrije Universiteit Brussel, Belgium and her MSc degree in epidemiology and biostatistics from the Université catholique de Louvain, Louvain-la-Neuve, Belgium. From 2006 to 2010, she worked as a PhD student at the Université catholique de Louvain, Brussels, Belgium and as a researcher at Advanced Medical Diagnostics, Waterloo, Belgium.

PUBLICATIONS

- Francart J, Vaes E, Henrard S, Legrand C, Baas P, Gaafar R, Van Meerbeeck JP, Sylvester R, Robert A. *A Prognostic index for progression free survival in malignant mesothelioma with application to the design of phase II trails: a combined analysis of 10 EORTC trials.* **European Journal of Cancer**, 45 (13):2304-2311, 2009

- Vaes E, Manchanda R, Nir R, Nir D, Bleiberg H, Autier P, Menon U, Robert A. *Mathematical models to discriminate between benign and malignant adnexal masses: potential diagnostic improvement using Ovarian HistoScanning.* **International Journal of Gynaecological Cancer, in press.**

- Vaes E, Manchanda R, Nir R, Nir D, Bleiberg H, Autier P, Robert A, Menon U. *Differential diagnosis of adnexal masses: A sequential use of the Risk of Malignancy Index and Ovarian HistoScanning.* **Submitted**

SCIENTIFIC COMMUNICATIONS

- Vaes E, Nir D, Nir R, Soviany C and Robert A. *Statistical Models Towards Ultrasonic Characterization of Ovarian Tissues*. **Poster presentation** at the 14th Annual meeting of the Belgian Statistical Society, October 2006, Houffalize, Belgium.

- Francart F, Vaes E, Sylvester R, Van Glabbeke M, van Meerbeeck JP, Robert A. *Prognostic factors in malignant mesothelioma of progression-free-survival: a better approach for phase II mesothelioma trials.* **Oral presentation** at the 14th

Annual meeting of the Belgian Statistical Society, October 2006, Houffalize, Belgium.

- Vaes E, Nir D, Nir R and Robert A. *Decision analytical models to define the cost-effectiveness of ovarian cancer detection techniques.* **Poster presentation** at the 15th Annual meeting of the Belgian Statistical Society, October 2007, Antwerp, Belgium.

- Vaes E, Nir D, Nir R and Robert A. *Statistical models for Ultrasound data that allow a differentiation between tissue types.* **Oral presentation** at the 16th Annual meeting of the Belgian Statistical Society, October 2008, Namur, Belgium.

- Francart F, Vaes E, Sylvester R, and Robert A. *Adapting the cancer phase II design for a better assessment of therapeutic activity: application to malignant mesothelioma.* **Poster presentation** at the 16th Annual meeting of the Belgian Statistical Society, October 2008, Namur, Belgium.

- Vaes E, Manchanda R, Nir R, Nir D, Bleiberg H, Robert A and Menon U. *A sequential use of the Risk of Malignancy Index and Ovarian HistoScanning for the differential diagnosis of adnexal masses.* **Poster presentation** at the Joint 15th ECCO/34stESMO conference, September 2009, Berlin, Germany.

NATIONAL AND INTERNATIONAL TRAINING

- 5-day Course in Statistical Pattern recognition, October 2006, Technische Universiteit Delft, Delft, The Netherlands

- 2-day training in Digital Signal Processing' by Dr. Chris Bore, February 2007, Waterloo, Belgium.

- 1-day Workshop on Longitudinal data analysis subject of missingness, October 2007, Antwerp, Belgium

- 2-day Course on Introduction to Statistical Genetics, November 2007, Louvain-la-neuve, Belgium

- 3-day Course on Survival Analysis and Competing Risks, September 2008, Brussels, Belgium
- 4-day Course on Mixed Model Analysis of Medical Data Using SAS, September 2009, Louvain-la-Neuve, Belgium