MD Thesis

Title

Novel Approaches to Ultrasound Based Evaluation and Management of Adnexal Masses

By

Natalie Paloma NUNES

Gynaecology Diagnostic Outpatient Treatment Unit

University College Hospital, London

Thesis submitted for the Degree of Doctor of Medicine (Research), University College London

Declaration Statement of Originality and Personal Contribution to Work

I, Natalie Nunes, confirm that all the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis and permissions have been sought and granted. I was personally involved in the design of all studies. I completed all applications for ethical approval, wrote all of protocols and patient information leaflets as well as arranged recruitment and follow up of all patients. I performed all level II ultrasound examinations except where indicated, collected all the data, performed all the data entry, updated the ethics committee yearly and ensured the correct running of all studies. I did all statistical analyses in conjunction with UCL statistician Dr Gareth Ambler apart from study 8, the randomized controlled trial, where the sample size and the randomisation list were calculated and generated by Mr Paul Bassett, another UCL statistician.

_____

Natalie Nunes

Abstract

Hypothesis and Aim:

This thesis aims to assess current ultrasound tools and protocols for the diagnosis and management of ovarian tumours. I hypothesized that the IOTA (International Ovarian Tumour Analysis) tools would have less diagnostic accuracy when performed by a less experienced operator and that the IOTA LR models would therefore not be useful as a triage tool but that while the 'simple rules' (SR) model may be useful as a triage tool as it is most similar to pattern recognition. I also hypothesized that the SR modified protocol would offer fewer operations to women with benign disease as compared with Royal College of Obstetricians and Gynaecologists [RCOG] protocol using the risk of malignancy index [RMI].

Objectives:

The aim of this thesis was to investigate and explore recently developed and currently utilised ultrasound based tools as methods of evaluation and means of determining the management of adnexal tumours.

The novel approach was to investigate the use of these tools when performed by an average (level II) ultrasound operator as compared with the experts who developed them and to determine the management for asymptomatic postmenopausal women with incidentally detected adnexal tumours.

Studies: Determining the accuracy of LR1/LR2 and 'Simple Rules' (SR) by an average operator compared with pattern recognition and evaluating the performance

of LR1/LR2 and SR using two reference standards, (histology and follow-up ultrasound scans) as a triage tool. Performing a meta-analysis of the SR model. Comparing the performance of two management protocols (Royal College of Obstetricians and Gynaecologists [RCOG] using the risk of malignancy index [RMI] and an SR-based protocol) for the likely intervention rates in asymptomatic postmenopausal women, and conducting a randomised controlled trial to compare the performance of those two management protocols (the RMI/RCOG guidance and the SR-based protocol) for the actual intervention rates in asymptomatic postmenopausal women.

Findings:

LR2 (average operator) had a similar sensitivity but the specificity was significantly lower, LR1/LR2 showed higher sensitivities and significantly lower specificities compared to the experts and pattern recognition and LR1/LR2 can be used as a triage tool once you accept the greater false positive rate. SR (average operator) showed similar sensitivity but a poorer specificity. When indeterminate tumours were assumed to be malignant there was a significantly increased sensitivity, decreased specificity and decreased diagnostic accuracy versus when pattern recognition was used, the sensitivity and diagnostic accuracy were maintained with increased specificity. SR performed well for the diagnosis of malignancy (meta-analysis) and SR works well as a triage tool, With the RMI/RCOG protocol, the likely intervention rate was 6 x more compared with SR and the RMI/RCOG protocol offered women surgery 9 x more often than SR and the actual surgical rate was 2.6 x more than SR (RCT).

Conclusions:

The IOTA LR1/LR2 models performed fairly well when used pre-operatively or as a

triage test by a less experienced operator. Comparatively, the SR model worked very well when the rules were applicable but this leaves the problem of deciding what to do with the 25% of women for whose ovarian tumours, the rules are not applicable. This thesis confirmed that the SRMP protocol had fewer indeterminately defined tumours and therefore offered surgery to fewer women with benign tumours when compared with the RCOG/RMI protocol.

Table of Contents

List of Figures

List of Tables

List of Equations

Abbreviations

α-FP	Alpha-fetoprotein
ANN	Artificial neural networks
AUC	Area under the curve
β-hCG	Beta-Human chorionic gonadotropin
BOT	Borderline ovarian tumour
BSO	Bilateral salpingo-oopherectomy
CA125	Cancer Antigen 125
CA19-9	Cancer Antigen 19-9
CEA	Carcinoembryonic antigen
CT	Computed Tomography
EOC	Epithelial ovarian carcinoma
FIGO	International Federation of Gynecology and Obstetrics
GDOTU	Gynaecology Diagnostic Outpatient Treatment Unit
GCT	Granulosa Cell Tumours
GP	General practitioner
hCG	Human chorionic gonadotropin
HE4	Human Epididymis Protein 4 (HE4)
HRT	Hormone replacement therapy
IOM	The Institute of Medicine
IOTA	International Ovarian Tumour Analysis
LDH	Lactate dehydrogenase
LR+ve	Positive Likelihood ratio
LR-ve	Negative Likelihood ratio

LR	Logistic Regression
LR1	IOTA Logistic Regression Model 1
LR2	IOTA Logistic Regression Model 2
MOOSE	Meta-analysis of Observational Studies in Epidemiology
MRI	Magnetic resonance imaging
NHS	National Health Service
PID	Pelvic inflammatory disease
PMP	Pseudomyxoma peritonei
PR	Pattern Recognition
PPC	Primary peritoneal cancer
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies 2
QUOROM	Quality of Reporting of Meta-analyses
RCOG	Royal College of Obstetricians and Gynaecologists
RMI	Risk of Malignancy Index
ROC	Receiver operator characteristics
ROMA	Risk of malignancy algorithm
SE	Standard error
SR	Simple Rules
SR+MA	Simple Rules and malignancy assumed when the simple rules were not applicable.
SR+PR	Simple Rules and use of pattern recognition when the simple rules were not applicable.
STARD	Standards for the Reporting of Diagnostic accuracy studies
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology

TAS	Transabdominal scan
TPA	Tissue polypeptide antigen
TVS	Transvaginal scan
UCH	University College Hospital
UKCTOCS	United Kingdom collaborative trial of ovarian cancer screening
USS	Ultrasound scan

Acknowledgements

I would like to thank my teacher and supervisor, Mr Davor Jurkovic (Consultant Gynaecologist, University College Hospital, London) for his unending guidance, advice, encouragement and support during my years as a research fellow. The training he gave me in gynaecological ultrasound was world-class and has forever changed my career path and my practice of Obstetrics and Gynaecology. His thought provoking discussions have deepened my analytical processes to do the work and the writing of this thesis.

I am grateful to my supervisor Professor Widschwendter (Consultant Gynaecological Oncologist, Department of Women's Cancer, Institute for Women's Health, University College Hospital, London) for his supervision and in particular his expertise in the field of ovarian cancer.

I am forever indebted to my colleagues Dr Xulin Foo, Dr Joel Naftalin, Dr Will Hoo, Dr Amna Malik, Dr Michelle Swer and Dr Kate Pateman-Graham for recognising patients suitable for my studies and identifying them to me, for being a great team to learn with during the day and to pull data with during the evenings. I would also like to thank the other clinic-supervising consultants Miss Naaila Aslam and Mr Reham Salim for their support of my studies and our clinic sonographer Sara-Louise Pointer who was always ready to offer me suitable potential recruits.

To the nurses who were supportive of my studies, I wish to say thank you. Sister Jacinta O'Neil, Ronke and Sarah Ekladios.

I am extremely thankful to the administrative staff who tirelessly printed clinic lists, booked patients, rebooked patients, responded to my many emails and patiently dealt with my numerous requests. These studies would not have been completed without you. Particular thanks go to Odechi Anodu, Hakeem Kuku, Lisa Hardy and Funmi Obende.

Finally I want to thank my McNeil family and all my friends who supported me and cheered me along the way. Particularly I want to thank my brother-in-law and my sister Kevern and Caryn Jameson, along with my darling nieces Kayla-Riis, Kaitlyn-Rose and Kristen-Rae who allowed me repeatedly into that paradise they call home, where I was able to retreat, refresh, regroup and then return. Thank you for allowing me into your space, your home and your hearts. To my parents Carl Nunes and Cheryl (RIP) Nunes, who were my first teachers, I thank you for the discipline and perseverance you instilled in me. To my dear friend Crispin Asiedu who departed this earth just before we each submitted our own theses, may you share in my joy of completion (from the cloud of witnesses), as I have shared in the joy of your completion of life's work and to the rest of my Congress world, I thank you for your support during my process.

Part I – Introduction

Adnexal tumours can arise from the ovary, fallopian tube or para-ovarian / para-tubal structures. There are many tools used to assess these adnexal tumours.

Pavlik et al found the prevalence of ovarian cysts in premenopausal woman to be higher than in postmenopausal women (35% versus 17% respectively). The incidence of new cyst formation was also significantly higher for premenopausal women (15.3% premenopausal; 8.2% postmenopausal, $P<0.001$). (Pavlik, 2013) Nezhat found amongst 1101 women, only 0.4% of the tumours in women under the age of 55 years were malignant while 80% were functional or endometriomas. (Nezhat, 1992) Another study with women at or below the age of 45 found a 1.4% incidence of malignancy. (Mecke, 1992)

In a cohort study of 31,834 women in whom 6,807 had an ultrasound scan abnormality, 63% of these had complete resolution of their cysts. (Pavlik, 2013) In a cohort of 15,106 asymptomatic postmenopausal women, a similar percentage (69%) was found to resolve spontaneously. (Modesitt, 2003)

When a woman is found to have an ovarian cyst or an adnexal mass, the nature of this mass and the symptoms she experiences determine the management of this tumour. The most important determination is whether the tumour is benign or malignant. Benign lesions can be managed expectantly if a woman is asymptomatic. Alternatively benign lesions can be managed by conservative surgery with laparoscopic excision. Minimally invasive surgery requires shorter hospitalization, affords the patient a faster

recovery and allows a speedier return to work. Laparoscopic surgery is also more aesthetically pleasing to women with smaller incision scars. Malignant lesions on the other hand will need to be managed by an oncology team in a tertiary unit. This has been found to provide the best outcome for these women. These lesions often require a staging laparotomy and debulking surgery. Women managed and operated on by a gynaecological oncologist have better operative staging, more complete debulking with less residual tissue and better survival rates and should therefore be referred pre-operatively. (Elattar, 2011)

Non-invasive diagnosis has therefore become even more relevant and important. It is imperative that all efforts are made to determine the nature of a tumour before embarking on empirical surgical management.

The tool with the most widespread use in the United Kingdom is the risk of malignancy index (RMI). (Jacobs, 1990) It combines features of the tumour on ultrasound scan with the absolute value of a tumour marker CA 125 (U/ml). The tumour marker CA 125 though, is not specific for the diagnosis of ovarian cancer. It can be raised in many benign conditions or non-ovarian malignancies while it may be normal in borderline, early stage epithelial and non-epithelial ovarian malignancies.

A number of newer models have been developed since then. Artificial neural networks are computer-based learning statistical algorithms, which are modelled after the brain. (Timmerman, 1999a; Tailor, 1999) They have not though shown to work consistently. There are logistic regression models such as Tailor's, the IOTA LR1 and LR2, which require complex calculations. (Tailor, 1997; Timmerman, 2005)

There are ultrasound only based models such as the simple rules method, which is simple indeed but requires a greater degree of skill and assessment than the RMI. There is a recent model called the risk of malignancy algorithm (ROMA), which involves a score for the patient's menopausal status and the laboratory value of two tumour markers (CA 125 and HE4). Moore et al found that combining the two tumour markers gave a sensitivity of 76.4% and a specificity of 95%. (Moore, 2008) Moore et al went on to describe the ROMA algorithm, which classified 93.8% of epithelial ovarian carcinomas as high risk. (Moore, 2009)

Part II – Ovarian / Adnexal Tumours

2.1 Ultrasound Scan Appearance and Histology of the Normal Ovary and Adnexa

2.1.1 Normal Ovary

2.2.1.1. The Normal Ovarian Embryology

The ovaries begin to develop from the 3rd week after fertilization. Primordial germ cells migrate from endoderm located at the hindgut-yolk sac interface and migrate to the gonadal ridge. The indifferent gonad develops at a crown-rump length of 5.5mm and becomes ovarian around day 42 at a crown-rump length of 17mm. Pre-granulosa cells (some of which are derived from the surface epithelium) surround the germ cells, which form the primordial follicles. The germ cells proliferate during early pregnancy to reach 6 million in the 6-month-old fetus. At birth the ovary is primarily made up of follicles with scant stroma and most stromal development occurs during the first year of life. (Robboy, 2008)

2.2.1.2. The Normal Ovarian Anatomy and Histology

A single layer of cells called the serosa or the surface epithelium, which is white on

examination, covers the surface of the ovary. Beneath the surface epithelium is the cortex, which contains primordial and developing follicles with the interstitial lutein and theca cells. Beneath this layer is the medulla where the blood vessels run from the ovarian hilum. This layer is pauci-cellular. The cortex and medulla appear pinkish grey.

2.2.1.3. The Normal Ovary on Ultrasound Scan

The ovaries in a normal woman are located in the pelvis and can vary in size and scan appearance. In the pre-pubertal female the ovaries are small and smooth. In the post-pubertal female they are larger (30 x 20 x 10 mm on average) and they vary in appearance throughout the hormonal cycle with follicular development, ovulation and the development and then regression of the corpus luteum. During the early follicular phase of the menstrual cycle the normal ovary contains 3-11 follicles of ≤10mm, which appear as anechoic structures within the ovary. One follicle usually becomes dominant and enlarges to 18-20mm before ovulation occurs around day 14 (of a 28 day cycle). The dominant follicle becomes a corpus luteum after ovulation, which can appear cystic or solid on ultrasound scan with a circumferential increased vascular pattern. In the post-menopausal female the ovary are often smaller and contain no follicles.

2.1.2 Normal Fallopian Tube and other Adnexal Organs

2.2.2.1. Embryology

Fallopian tubes are derivatives of the unfused cephalic aspect of the Müllerian

(paramesonephric) ducts. This occurs during the 5th to 8th week after fertilisation. Caudal to this is the aspect of the Müllerian ducts, which fuse and cannulate to form the halves of the upper uterine corpus. Medial to the Müllerian ducts are the Wolffian (mesonephric) ducts, which regress, in female embryos.

2.2.2.2. Anatomy and Histology

The 9-12 cm long fallopian tubes have 4 segments, which, from medial to lateral, are the interstitial portion, the isthmus, the ampulla and the infundibulum. (Robboy, 2008) The wall of the interstitial aspect of the tube is the most muscular and is continuous with the myometrium. Lateral to this is the narrow but also muscular isthmus, which is 2-3 cm in length. The ampulla is the longest segment of the tube (half its length) and runs a meandering course. The most lateral aspect of the tube is the funnel-shaped infundibulum, which is 1cm in length and diameter and has a number of irregular, tongue-like extensions, the fimbriae.

The tubes have 3 layers, which are the mucosa, the muscle and the outer serosa. The mucosa contains 4 cell types, which are the ciliated cells, secretory cells, intercalary (or peg) cells, and reserve basal (or undifferentiated) cells. The ciliated cells have tongue like cilia, which protrude into the lumen and waft the contents along. This is important for the function of the tubes to fertilization and movement of the fertilized embryo to the uterine cavity.

Remnants of the Wolffian ducts may persist, forming cysts from the fimbria, along the mesosalpinx and in the broad ligament.

2.2.2.3. Ultrasound Scan

Apart from the interstitial portions, the fallopian tubes are not easily visible on ultrasound scan when no pathology or surrounding fluid is present. In transverse view, when sweeping the ultrasound probe towards the fundus, the interstitial portions are seen. From here, if the probe is moved laterally, the tube can be followed to the fimbria. If there is a trace of retrograde menstruation within the tube or surrounding free fluid in the pelvis, this makes visualisation more apparent. Wolffian duct remnants, which form cysts or masses, will be seen in the adnexa but will be separate from the ovary and uterus.

2.2 Ovarian Tumours: Ultrasound Scan Appearance, Histology, Demographics and Treatment

All tumours within this thesis were assigned a tumour histological type according to the World Health Organisation (WHO) tumour classification and borderline and primary invasive malignancies were further grouped according to the International Federation of Gynecology and Obstetrics' staging. (WHO, 1973; FIGO, 1971; Shepherd, 1989)

2.2.1 Physiological Cysts

2.3.1.1. Follicular Cysts

Follicular cysts are thin walled unilocular cysts, which are due to persistence of the normal Graafian follicle. Unluteinised follicular cysts produce estradiol and result from overstimulation of the ovary by follicle-stimulating hormone (FSH) or ovulation inducing drugs. Granulosa lutein cysts result from failure of rupture of the dominant follicle for release of the ovum. A layer of luteinised granulosa cells lines them with uniform round or oval nuclei and little cytoplasm. The fluid within is usually serous but it may also be haemorrhagic. They produce progesterone. Theca-lutein cysts predominantly produce androstendione and due to prolonged ovarian exposure to luteinizing hormone (LH) or beta-human chorionic gonadotropin (β-hCG). They contain luteinised theca interna cells.

On ultrasound, these cysts appear to be simple thin-walled, unilocular cysts with anechoic or haemorrhagic (web-like) fluid. They are predominantly 2-5cm in size but they can be larger and they often resolve on their own. They can be multiple or bilateral and they are poorly vascular. These are often found incidentally as they are usually asymptomatic unless haemorrhage has occurred. A repeat ultrasound scan can be done to confirm resolution in the early follicular phase of the woman's next menstrual cycle.

2.3.1.2. Corpus Luteal Cysts

A corpus luteum develops after ovulation. A cystic corpus luteum contains the same layers as the follicular cysts. The luteinized granulosa cells are intermixed with blood and macrophages laden with hemosiderin. Unlike those from follicular cysts, these cells are large, polyhedral in shape with a low nuclear: cytoplasmic ratio. There is often central haemorrhage

Corpus luteal cysts often appear as thick-walled haemorrhagic cysts with the classical "ring of fire" vascularity on Doppler interrogation.

2.2.2 Epithelial Tumours

Tumours that arise from the surface epithelium of the ovary comprise 60-75% of all ovarian tumours and 90% of primary ovarian cancers.

2.3.2.1. Serous Tumours

Benign serous cystadenomas are thin walled cysts with watery or thin mucinous fluid within. The inner wall can be smooth or can have polypoid projections, which may be firm if fibrous or soft if edematous. The external wall can also be smooth or it can be exophytic with external projections. They average 10 cms and are bilateral in 10-20% cases. A single layer of ciliated or non-ciliated epithelium usually lines the inner cyst wall and the papillae usually consist mostly of stroma. These cells secrete mucin and

mitoses are unusual.

2.3.2.2. Mucinous Tumours

These tumours are usually thick-walled cysts with a smooth capsule and with thick mucinous fluid within them. They are usually up to 50 cm in size though much larger ones have been documented. They are usually multilocular while mucinous cystadenofibromas are rare. Histologically, the walls may be collagenous and paucicellular with calcification present and they are be lined with columnar cells with high mucin content and small basal nuclei. Daughter cysts may arise from the cyst wall. Mitoses are unusual. On ultrasound these cysts are uni- or multilocular with smooth walls and low-level echogenic fluid with positive acoustic streaming.

2.3.2.3. Endometrioid Tumours

These tumours may arise de novo, on the surface epithelium, without the presence of endometriosis, while 20-30% of cases have endometriosis on the same ovary or elsewhere. This is a rare tumour and occurs primarily as a cystadenofibroma in women with an average age of 57years. Ultrasound appearance is believed to be the same as other benign cystadenomas/cystadenofibromas.

2.3.2.4. Müllerian Mesenchymal (stromal) Tumours

These occur uncommonly in the ovary but more often occur in the uterus. They

contain stromal of mesenchymal cells with varying degrees of proliferation.

2.3.2.5. Clear Cell Tumours

Benign clear cell cystadenomas/cystadenofibromas are rare tumours occurring at the mean age of 45 years. Microscopically they contain large polyhedral clear cells thus their name.

2.3.2.6. Transitional Cell (Brenner) Tumours

These are autochthonous tumours that arise from metaplasia of the surface epithelium into urothelial-like cells. Polyhedral or elongated cells are seen with grooved long 'coffee-bean' nuclei. Alternatively cells can be mucinous, endocervical or pseudostratified and ciliated. Brenner tumours are usually benign and account for 2% of primary ovarian tumours. They are usually unilateral and occur in women in their 4^{th} to 8^{th} decades. These are primary solid tumours < 2cm with a bosselated surface.

2.2.3 Germ Cell Tumours

These tumours account for 20% of all ovarian neoplasms.

2.3.3.1. Dysgerminoma

These tumours originate from pluripotential premeiotic oocytes. They account for 1% of all primary ovarian malignant tumours, about 1-2% of all ovarian germ cell tumours, and something over 25% of malignant ovarian germ cell tumours. (Robboy, 2008) These tumours are more commonly seen in adolescence and early adult life while it is rare in pre-pubertal children and postmenopausal women. Women with a Y chromosome and abnormal gonads due to gonadal dysgenesis and androgen insensitivity syndrome have a high risk of developing these tumours in the form of a gonadoblastoma.

Bilateral tumours are found in 15-20% cases and in half of those cases the second tumour is microscopic. Tumours are solid and highly vascular and are made up of sheets of pale uniform cells with a large central nucleus. Focal calcification is often seen in gonadoblastomas. Elevated serum lactate dehydrogenase (LDH) is a feature of these tumours.

2.3.3.2. Yolk Sac Tumours (Endodermal Sinus Tumours)

Pure yolk sac tumours are rare and are half as common as dysgerminomas. Most occur in women under the age of 30 years. Patients often present with abdominal distension and pain. Serum alpha-fetoprotein (α-FP) is often raised.

2.3.3.3. Teratomas

Teratomas are composed of mature or immature tissues deriving from the three pluripotential germ cell layers found in an embryo: ectoderm (skin or epidermal derivates and neural tissue), mesoderm (fat, bone, cartilage, connective tissue, muscle) and endoderm (epithelium of the gastrointestinal tract and the bronchial tree, thyroid tissue). (Saba, 2009; Robboy, 2008)

Mature Teratoma

Mature cystic teratomas of the ovary are more commonly seen in premenopausal women, predominantly in the third to forth decades though they are seen in all age groups. (Westhoff, 1988) They are the most common type of germ cell tumour seen and account for approximately 12% of all ovarian tumours. One series found that 42% of their tumours were dermoids. (Ayhan, 1991) They are bilateral in 10-14% cases. (Ayhan, 1991; Papadias, 2005) Complications that may occur include torsion (8%), rupture (4%), infection 2% and malignant transformation which is reported in 0.3-5% of cases. (Ayhan, 1991; Papadias, 2005)

Mature ovarian cystic teratomas are often called dermoids because of their preponderance to contain epidermal tissue. They are generally unilocular but can be multilocular. They usually contain sebaceous fluid with sebum and keratin with dispersed hair and other tissue such as adipose tissue, thyroid tissue, bone or cartilage. Fat is found in 93% of dermoids. (Saba, 2009; Caspi, 1996) This contributes to its pathognomonic appearance on ultrasound and MRI scan.

On ultrasound the following features are seen in dermoid cysts: 1) predominantly cystic (18%) 2) Uniformly or predominantly dense echogenicity (33%) which represents adipose tissue or floating fat or ball of hair 3) Presence of echogenic thin band-like echoes which represent the dispersed hair 4) presence of a discrete highly echogenic focus with posterior acoustic shadowing (Rokitansky protuberance) 5) Presence of a fat fluid level (Caspi, 1996; Cohen, 1993; Hutton, 1979)

Dermoids cysts have been found to grow by 1.7-1.8mm/year. (Hoo, 2010; Caspi, 1997) Hoo et al found 5 factors, which increased the risk of surgical intervention. These were lower age, greater parity (≥ 2), past history of an ovarian cyst, bilateral cysts and larger cyst size. (Hoo, 2010) Alternatively expectant management has been found to be successful in up to 77% patients after 1 year. (Hoo, 2010)

Struma Ovarii

Struma ovarii is a teratoma where thyroid tissue is the only or major component (>50%). These primarily unilateral tumours account for 5% of mature teratomas but only 50-60% of strumas occur within a mature teratoma. The remainder are pure or are associated with a carcinoid tumour or a mucinous adenocarcinoma. These customarily occur in the 2nd to 5th decades and are incidental findings or present with vague symptoms. Less than 5% demonstrate hyperthyroidism while 33% may have ascites. Some patients have a pleural effusion and present with Meigs syndrome. These predominantly unilateral tumours are mostly benign and are solid and vascular with a smooth surface when pure. Cystic tumours also occur. Histologically it is identical to thyroid tissue and extraovarian spread is needed to confirm malignancy. Treatment

is oopherectomy when benign and pelvic clearance for malignant tumours.

Carcinoid and Strumal Carcinoid

Primary carcinoid tumours (< 5% of ovarian teratomas) are four times as prevalent as carcinoids metastatic to the ovaries. Types include insular (50% cases), trabecular (33%), strumal (17%) and mucinous. In the ovary, insular carcinoids are analogous to midgut carcinoids whereas trabecular carcinoids are analogous to foregut or hindgut carcinoids biochemically, histochemically and morphologically. Patients could be of any age but are usually peri- or post-menopausal. These tumours are usually unilateral but a 15% have a dermoid or a mucinous tumour in the contralateral ovary. Most primary ovarian carcinoids (60%) occur within dermoids or mucinous tumours. Metastatic carcinoids are usually bilateral and insular in type.

2.3.3.4. Mixed Germ Cell and Sex Chord-Stromal Tumours

These tumours have 2-3 cell type components.

Gonadoblastoma

An excess of 80% of these tumours occur in phenotypic females while the rest occur in males with cryptorchidism. The phenotypical females though have a high incidence (95%) of abnormal sex chromosomal combinations such as 46, XY (50%) and 45 X/46,XY (25%). Age at presentation is from birth up to the 4th decade. These

tumours are unilateral in 67% cases and are usually solid. These tumours can progress to be overgrown by a dysgerminoma, which can become malignant.

2.2.4 Sex cord-Stromal Tumours

This group of tumours account for 8% all ovarian tumours.

2.3.4.1. Granulosa Cell Tumours (GCT)

These tumours are classified by having ≥ 10% cells that resemble granulosa cells which are small cells with scanty cytoplasm and alongside theco-fibroma tumours they account for 70% of sex cord – stromal tumours. These predominently occur in peri-menopausal women (45-55 years of age) and present with vague abdominal symptoms (20%) or with endocrinological disruption.(50%) or both (30%). In 5% of cases there is a separate histological identity called jeuvenille GCT, which is more commonly seen in young girls and women. Meig's syndrome (benign ovarian tumour, ascites and pleural effusion) may occur. Raised serum estrogen is often found which is associated with irregular bleeding in premenopausal women or postmenopausal bleeding for those after the menopause. Half have associated endometrial hyperplasia and 10% with an endometrial cancer. Virilization is an alternative presentation and these tumours may occur in dysgenetic gonads. Serum inhibin is raised. These tumours are usually unilateral, solid-cystic and range in size from subclinical to the size of a term gestation. GCTs are treated as low-grade malignancies.

2.3.4.2. Thecoma-Fibroma Tumours

These tumours include the more frequently occuring fibromas (6% primary ovarian tumours) and the rarer thecomas as well as a range of tumours with features of a mix of them. These all arise from ovarian stroma and occur in a range of sizes. (Robboy, 2008)

Thecomas occur at any age but more commonly after the 4th decade with 75% in postmenopausal women. Thes tumour often present with estrogenic symptoms as with GCTs while androgenic thecomas are rare. Meigs syndrome an occur. These tumours are solid, well-circumscribed and usually unilateral. Other appearances have been described including calcification and posterior acoustic shadowing. (Athey,1987)

The theca cells are spindle shaped and have small pale round nuclei and mitoses are rare. These tumours are almost exclusively benign and may exibit cystic degeneration.

Thecafibromas are less likely to have estrogenic effects and look like fibromas on scan and macroscopically. Fibromas accound for 65% of the sex stromal tumours and over 90% occur in women over age 30. Ascites occurs in 10% of cases and in 40% of those with tumours>10cm. Also unilateral solid tumours, fibromas have a bosselated surface and often exibit cystic degeneration. Tumours may be edematous and may be confused with ovarian edema which tends to occur in polycystic ovaries. These tumours are hormonally inactive.

Sclerosing stromal tumours occur predominantly in the young (80% <30 years). They arise from peri-follicular myoid stromal cells and they are hormonally active secreting

estrogens, progestogens or androgens. These unilateral, solid, lobulated, well circumscribed tumours are 3-5cm in diameter and are all benign.

2.3.4.3. Sertoli-Leydig Cell Tumours (Androblastomas)

These tumours are rare (1% sex cord stromal and 0.1-0.5% primary ovarian tumours). Diagnosis is made due to the presence of Sertoli cells, Leydig cells or fibroblastic cells and are characterised by the finding of Reinke crystals. These are primarily found in the 2nd and 3rd decades and unilateral in 98% cases. These encapsulated and lobulated tumours are solid with small cystic areas and usually <10cms. They may be associated with raised alpha-fetoprotein levels but not as high as is seen in yolk sac tumours. They may be estrogen- or androgen- producing but are often hormonally inert. Well-differentiated Sertoli cell tumours are seen with testicular feminization syndrome (XY females).

2.2.5 Other

2.3.5.1. Endometriomas

Endometriosis occurs when ectopic endometrial glands and stroma are found outside of the endometrial cavity. There is an environment supporting growth due to the presence of growth factors. This endometriosis commonly occurs in the pelvis such as the pelvic sidewalls, the pouch of Douglas and any pelvic peritoneal surfaces. When this tissue is found on the ovaries, this is where the classic endometriomas develop. The

ovarian ectopic endometrium bleeds each month at the same time the normally located endometrium sheds (menses) causing the cyst to accumulate this old blood. There are 2 theories for the cause of endometriosis. The first is due to receptive pelvic peritoneum, which allows implantation of endometrium, which makes its way to the pelvic peritoneal cavity via retrograde menstruation. The second is thought to be due to metaplastic change in the tissue which is present premenstrually and which activates later in life.

On ultrasound scan endometriomas are typically described as thick walled well-circumscribed cysts. Within them homogenous low-level fluid is seen which is classically described as “ground glass” is this is due to the presence of old blood. Cysts are usually unilocular but up to 30% can be multilocular. (Van Holsbeke, 2010b; Kupfer, 1992) Papillary projections have been described in up to 20% of endometriomas. (Patel, 1999) Hyperechoic cyst wall foci are also described. (Kinkel, 2006; Patel, 1999) Acoustic streaming is absent in up to 91% of endometriomas. (Van Holsbeke, 2010a) A carpet like debris is often seen in the dependant part o the cyst. (Fried, 1993; Kupfer, 1992) Kupfer et al saw this in 82% of their endometriomas. These cysts are usually poorly vascular but vascularity can be seen around the cyst in the normal ovarian tissue near the ovarian hilum.

During pregnancy endometriomas and deposits of endometriosis often undergo decidualization including within cutaneous endometriosis. (Tazegül, 2013) There is 1 reported case of cutaneous decidualized endometriosis in a non-pregnant woman. (DeClerck, 210) The cysts can be misdiagnosed as malignancies at this time as the appearance changes. The fluid become anechoic or less echogenic, the papillary

projections become larger, more prominent or appear where there were none before, and with the typical pregnancy–induced increase in vascularity, these cysts become more vascular.

Management of endometriomas involves medical and surgical options (gold standard of laparoscopic excision or alternatively laparotomy for clearance). Treatment often depends on symptoms and fertility requirements rather that the sonographic appearance of the disease as these may not correlate to symptoms.

2.2.6 Borderline Ovarian Tumours

The International Federation of Gynecology and Obstetrics (FIGO) classify borderline tumours (BOT) (previously called carcinomas of low malignant potential) separately from primary invasive ovarian tumours since 1971. (FIGO, 1971) Both borderline and primary invasive tumours are classified into 1 of 4 stages with further classification into subsections. (Shepherd, 1989)

BOTs have varying ultrasound appearances. They tend to be tumours that get quite large despite having a slower growth rate compared to invasive malignancies. They are usually well-circumscribed tumours with no evidence of invasion or ascites and they are often multilocular with papillary projections. (Valentin, 2006) Normal surrounding ovarian tissue “crescent sign” is usually seen. (Yazbek, 2007a)

Borderline ovarian tumours (BOT) are tumours, which have proliferating cells and nuclear atypia with mitotic activity but with no infiltrative destructive growth or stromal invasion whereas primary invasive tumours do. BOTs account for 10-15% of all ovarian tumours. (Tinelli, 2009) BOTs are more commonly found in pre-menopausal women as compared to primary invasive tumours. (Gotlieb, 2005) BOTs peak in the 5th decade versus primary invasive tumours which peak in the 6th and 7th decades. (Heintz, 2006; Cancer Research UK, 2013) These women are more likely to be diagnosed in early stages (82% stage 1), have a lower relapse rate (7.8%) and have an overall better prognosis. (du Bois, 2013; Heintz, 2006). They have a better 5-year survival 87.3% versus 49.7% when compared to that of primary invasive tumours but it is suggested that this is heavily influenced by the fact that BOTs usually present at stage 1 whereas invasive tumours are often diagnosed at later stages. (Heintz, 2006)

Treatment of BOTs tends to be more conservative and minimally invasive with fertility sparing surgery for those who wish to maintain their reproductive function. (Maneo, 2004; Tinelli, 2009; Seracchioli, 2001) Surgery may be a laparoscopic salpingo-oopherectomy for unilateral disease or an ovarian cystectomy for 1 of the ovaries when bilateral disease exists. Fertility and pregnancy outcomes appear to be good and ultrasound surveillance is possible for recurrence. (Seracchioli, 2001) Further appropriate laparoscopic procedures include appendicectomy for mucinous tumours, total laparoscopic hysterectomy and bilateral salpingo-oopherectomy, peritoneal biopsies, omentectomy and pelvic and paraaortic lymphadenectomy.

2.3.6.1. Epithelial tumours

Serous BOTs (SBOTs) are bilateral in 35-40% cases. These tumours tend to have multiple papillary projections as do endocervical-type BOTs. (Yazbek, 2007a)

Mucinous BOTs (MBOTs) tend to be more multilocular when compared to their benign counterparts. Classic honeycomb nodules (multilocular nodules) are seen in MBOTs of gastrointestinal type on ultrasound scans. (Yazbek, 2007a)

Patients with endometriod BOTs have a mean age of 55-60 years and the tumour may be associated with synchronous or metachronous endometrial hyperplasia or an adenocarcinoma. (Robboy, 2008)

Transitional cell (Brenner) BOTs behave like benign tumours though they get much larger (up to 23cm) and primarily only have symptoms of mass effect. There are highly multilocular tumours and are rarely associated with a synchronous or metachronous bladder transitional cell carcinoma.

Stage 1 invasive tumours have a 5 year survival of 83-90% while serous BOTs stage Ia and Ib have a near 100% 5 year survival. (Silverberg, 2004) Serous BOTs are just over twice as common as mucinous BOTs while invasive serous tumours are 4 times as common as mucinous invasive tumours. (du Bois, 2013; Heintz, 2006). Of all the histological types at stage IV, clear cell was found to have the worse prognosis/ 5 year survival rate (12.9%) followed closely by mixed epithelial (13.2%) and mucinous (17.3%).

2.3.6.2. Germ cell Tumours

Germ cell tumours provide a number of diagnostic challenges to pathologists because of the varying cell line components. The literature does not describe a distinct subclass of borderline germ cell tumours.

2.3.6.3. Sex cord-Stromal Tumours

Granulosa cell tumours are treated as tumours of low malignant potential. Diagnosis is usually at an early stage and recurrence is often very late (30-50% at 20 years). (See above)

2.2.7 Primary Malignant Ovarian Tumours

2.3.7.1. Epithelial tumours

Epithelial primary invasive malignancies are usually multilocular-solid tumours which are highly vascular, fairly large with irregular borders and which begin to invade surrounding structures. For primary invasive ovarian malignancies, the principle of treatment is debulking followed by adjuvant chemotherapy.

Primary undifferentiated and serous carcinomas are commonly bilateral whereas

endometrioid and mucinous are often unilateral. Mucinous carcinomas are usually 15-30cm and multilocular or multilocular-solid and rarely they may be almost entirely solid with a soft mucoid surface.

Endometroid carcinoma is the second commonest ovarian epithelial malignancy and is associated with endometriosis in 30% of cases and with endometrial carcinoma in 25-30% of cases. They are bilateral in 28% cases and confined to the pelvis in 70% cases. Those that occur within endometriomas are often in younger women but generally these occur in the 4^{th} and 5^{th} decades of life. While up to 30% of cases have a synchronous endometrial cancer, in 15-25% it is histologically similar which raises the suspicion of metastases. These though are usually considered as a separate independent tumour. Primary tumours are more likely to be cystic-solid, unilateral and larger than the metastatic versions, which are usually multinodular solid, bilateral and smaller. The stage of the uterine tumour can also help distinguish primary from metastatic as the latter ones tend to have higher-grade tumours with deep myometrial invasion. Rarely this can also occur with a synchronous or metachronous endometriod tumour of the cervix, which is autochthonous. These tumours are usually cystic, 12-20cm in diameter and with multiple papillary projections and a smooth outer surface. The fluid may be blood stained. Alternatively it may be solid with haemorrhage and necrosis or a large solid papillary projection if within an endometrioma.

Primary ovarian sarcomas are rare and are usually solid tumours with tongue like projections into adjacent tissue.

Malignant clear cell tumours account for 5-10% primary ovarian malignancies and

occur predominantly in the 5th to 7th decades with 50-67% of the women being nulliparous. Paraneoplastic hypercalcaemia is a recognised association of these primarily unilateral tumours. Half are diagnosed at stage 1. Endometriosis is seen in 60-65% of these women though synchronous endometrial tumours are not often seen. These 15-20 cm large tumours are thick-walled unilocular cysts with papillary projections and bloodstained fluid.

Malignant Brenner tumours occur in the 6th decade and ranges from 10-30cm in diameter. These are usually unilateral cysts with soft multiple papillary projections.

2.3.7.2. Germ cell tumours

Dysgerminomas are fast growing tumours, but they metastases late. Two thirds of cases are FIGO Stage I at the time of diagnosis.

Choriocarcinomas are characterised by presence of trophoblastic malignant cells and they can be gestational or non-gestational. Gestational choriocarcinomas include ovarian metastasis from primary uterine choriocarcinoma tumours, which can be associated with or occur after a normal pregnancy, spontaneous miscarriage, complete hydatidiform mole, or partial mole, and primary gestational ovarian choriocarcinoma tumours (1 in 369 million pregnancies) that arise from an ovarian ectopic pregnancy. (Axe, 1985) The nongestational type which accounts for ≤0.6% of all ovarian tumours, is as a component of a mixed germ cell tumour and a pure ovarian choriocarcinoma is a very rare aggressively malignant tumour. (Lv, 2011) These tumours are associated with high serum β-hCG levels and immunohistochemical staining with β-

hCG, and placental lactogen are diagnostic. While they typically metastasize early via venous, lymphatic and transcoelomic routes, these tumours are highly chemosensitive.

Embryonal Carcinoma are the least differentiated of the germ cell tumours with pluripotent potential and are homologous to their counterparts in the testes. Pure tumours are rare with peak presentation in the 2nd and 3rd decade of life with a mass or hormonal expressions in 67% (precocious puberty or abnormal vaginal bleeding). Tumours are usually large, soft solid and unilateral but they are highly malignant invading locally as well as spreading intra-peritoneally and distantly. Tumour cells express β-hCG and α-FP. Treatment is surgical followed by chemotherapy in all patients even though 60% present at stage 1A. Surveillance is with serum α-FP and β-hCG alongside imaging.

2.3.7.3. Sex cord-Stromal Tumours

Thecomas, which are malignant, are rare and only a few examples are documented.

2.2.8 Primary Peritoneal Cancer (PPC)

PPC is a serous carcinoma of the peritoneal lining of the abdomen. Staging is the same for ovarian cancer but there only exists stages III and IV. Up to 15% patients diagnosed with advanced serous carcinomas are thought to have PPC instead. After a prophylactic BSO, women with BRA1/2 still carry a 1-5% risk of developing PPC after 7-27 years and low risk women have a 0.02% after up to 18 years. (Piver, 1993; Casey, 2005;

Gotlieb, 2006)

Ultrasonographically the peritoneum appears globally thickened and ascites almost always present. Tumour grade was the main reported significant prognostic factor and treatment is as for ovarian cancer. (Lee, 2013)

2.2.9 Pseudomyxoma peritonei (PMP)

This is a condition where there is the presence of free gelatinous mucin in the peritoneal cavity and this may or may not have a cellular component. This gelatinous ascites is thought to be due to spread from a mucinous ovarian or gastrointestinal tumour and rarely of a tubal or an extraabdominal tumour. The ovarian tumour tends to be bilateral (66%) and shows penetration of the stroma with tracks of mucin (pseudomyxoma ovarii) while pseudomyxoma ovarii can occur without pseudomyxoma peritonei. Appendiceal mucoceles have a similar process causing the leakage of the mucin. Prognosis is dependant on the primary tumour and amenability of the tumour to clearance or cytoreduction and is therefore excellent for stage 1 tumours despite the presence of this condition. This suggests that PMP is not malignant spread. Recurrence of PMP is common.

2.2.10 Metastatic Ovarian Tumours

The ovary is commonly a site of metastases from malignant tumours, most often seen from the gastrointestinal tract (39-42%), the breast (28-29%) and the endometrium

(20%). (Young, 1991; Antila, 2006; de Waal, 2009). Metastases account for 5-15% of all ovarian tumours. (Antila, 2006; de Waal, 2009) Ovarian metastatic tumours are bilateral in up to 69% cases. (de Waal, 2009) Tumours are often smaller, solid, well circumscribed tumours of high vascularity (stomach, breast, lymphoma and uterine) but they may also be larger, multilocular-solid or multilocular with irregular borders and less vascularity (colon, appendix, rectum and biliary tree). (Testa, 2007) The second group were often larger than the first (12cm versus 7cm) while the metastatic tumours as a group are often smaller than their primary counterparts (<10cm versus >15cm). (Young, 2006; Testa, 2007; Guerriero, 2012)

Krukenberg tumours are the most well known of the metastatic tumours though not the most common. (Young, 2006) This name is assigned to metastatic tumours with a >10% component of the tumour made up of mucin secreting signet-ring cells and where the stroma often demonstrates a sarcomatoid reaction. They primarily metastasize from gastric tumours and can be seen in up to 55% patients with a gastric tumour. (Young, 2006) These are typically bilateral bosselated surface solid tumours without necrosis unlike gastrointestinal type non-Krukenberg metastases, which tend to be unilateral, solid and cystic with necrosis. (Young, 2007)

2.3 Other Adnexal Tumours: Ultrasound Scan Appearance, Histology, Demographics and Treatment

2.3.1 Non-ovarian Adnexal Masses

2.4.1.1. Hydrosalpinx

A dilated fallopian tube has a characteristic appearance of a sausage-shaped cystic structure with anechoic or low level fluid and 'incomplete septa' (septa that do not reach the opposite wall of the cystic structure in all planes), which are due to protruding mucosal folds. The pathognomonic feature is described as a 'cog-wheel' appearance or as a 'beads-on-a-string' appearance for more longstanding cases. (Timor-Tritsch, 1998). Treatment is conservative unless a woman is to have in vitro fertilisation (IVF) treatment, as success rates are higher after removal.

2.4.1.2. Pyosalpinx / Hematosalpinx / Tubo-Ovarian abscess

When the above described structure contained echogenic fluid this is often a sign of a pyosalpinx (pus containing) or a hematosalpinx (blood containing). When there is a tubo-ovarian abscess present, the structure involves the ovary and is often multilocular-solid with thick septae and ground glass contents or it may be predominantly solid in appearance. (Valentin, 2004) It may demonstrate increased vascularity and is tender on palpation. Treatment of a pyosapinx or abscess is initially antibiotics with analgesia but

ultrasound guided transvaginal drainage or surgical drainage may be required if the patient does not respond to this. (Aboulghar, 1995) This may be misdiagnosed as a malignancy if important factors such as age, history, examination and inflammatory markers are not considered. (Hata, 1989; Varras, 2003)

2.4.1.3. Fimbrial, Para-tubal, Paraovarian and Broad Ligament cysts

These cysts are remnants from embryonic ducts of which most are paramesonephric (Müllerian) in origin and occurs anywhere between the ovary and the tube as well as lateral to the uterus within the broad ligament. (Genadry, 1977; Samaha, 1985; Grant, 1992) These cysts are lined with epithelium and usually contain anechoic fluid. Most cysts are unilocular but they may have a range of appearances up to multi-locular solid with papillary projections in cystadenomas and cystadenofibromas. (Athey, 1985; Korbin, 1998; Savelli, 2006) They are separate from the ovary but this may occasionally be difficult to demonstrate. (Kim, 1995)

Hydatid cysts of Morgagni are typically referred to as fimbrial cysts and are usually small, multiple, avascular, pedunculated cysts with anechoic fluid and which are attached to the fimbria. These are more commonly seen in women with unexplained infertility; 52% versus 26% of those with explained fertility. (Rasheed, 2011)

Management is usually conservative unless the cyst changes in appearance or becomes symptomatic which may occur if torsion occurs. (Said, 2008)

Broad ligament leiomyoms (fibroids) may also occur and are often softer that their

intrauterine counterparts. (Chmaj-Wierzchowska, 2012)

2.4.1.4. Peritoneal Pseudocysts

When fluid is entrapped between peritoneal adhesions and pelvic organs, peritoneal pseudocysts are formed. Adhesions are fibrous bands of tissue between the abdominal organs and abdominal walls that are usually formed in response to peritoneal injury such as after infections commonly pelvic inflammatory disease and appendicitis, as seen in endometriosis or after a surgical procedure. On ultrasound these cysts are rarely rounded but instead are commonly irregular superiorly and follow the contours of the pelvic sidewalls laterally. Up to 81% of peritoneal pseudocysts are describes as having blurred, undefined margins and bizarre morphological features. The fluid is usually anechoic and septae may be seen within. The septae may be incomplete and the septae may be mobile on palpation with the endovaginal ultrasound scan probe, which is known as the “flapping sail sign”. (Savelli, 2004) The ipsilateral ovary is seen in up to 84% of cases and may be seen adjacent or within the pseudocyst like a 'spider in a web'.

Treatment is often conservative as surgical treatment has a moderately high risk of complications such as visceral injury or infection and a high risk of recurrence.

2.4.1.5. Borderline and Malignant Non-Ovarian Adnexal Tumours

Tubal malignancies are uncommon but are being detected more often during prophylactic risk-reducing bilateral salpingo-oopherectomies done for BRCA1/BRCA2

carriers or those with a strong family history. (Manchanda, 2011) In this series of 308 high-risk women, 10 occult tubal cancers while only 4 occult ovarian cancers were detected. There are theories and some evidence that some primary ovarian epithelial adenocarcinomas specifically high-grade serous carcinomas and primary peritoneal cancers have a tubal origin. (Kurman, 2013; Nasser, 2014)

Most paraovarian tumours are benign but borderline tumours have been reported. (Savelli, 2006; Suzuki, 2013) The literature describes a 2% malignancy rate. (Stein, 1990)

2.4 Epidemiology and Demographics of Ovarian cancer

In the UK, ovarian cancer is the 5th most common cancer in women after breast, lung, bowel and uterus with 7,011 new diagnoses in 2010, 7,116 new diagnoses in 2011 and with 6,692 new diagnoses in 2012 (where it was the 6th most common after breast, lung, bowel, skin melanoma and uterine). (ISD Scotland online, 2013; Office for national statistics, 2012; Welsh cancer intelligence and surveillance unit, 2013; Northern Ireland cancer registry, 2013; IARC, 2012; Ferlay, 2010). The incidence rate is 17.1 new cases per 100,000 population and the mortality rate is 9.0 per 100,000 population. It was the 4th most common cause of cancer deaths in 2011 after lung, breast and bowel and therefore the most common gynaecological malignancy causing death. (Cancer research UK, 2013; IARC, 2012)

The lifetime risk of developing ovarian cancer is approximately 2% or 1 in 51 for women in England and Wales. This risk increases to 3% for women with one

affected first degree relative (up to 5% for when the affected relative was under the age of 50 years versus 2.5% when they were over 50 years). (Jervis, 2013) This risk increased even further to between 30 and 40% for women with two affected first-degree relatives. Ovarian cancer is predominantly a disease of post-menopausal women with almost 85% of cases being diagnosed in women over 50 years of age with the peak incidence rates seen in women aged 65. (Figure 2.4-1) Despite the fact that most cases of ovarian cancer occur in postmenopausal women, since the mid-1970s, the incidence in females aged 15-39 has increased by 56%. (Cancer research UK)

Over the last 30 years the incidence of ovarian cancer has increased in England and Wales but it has decreased in Northern Ireland. The incidence went from 15 per 100,000 women in 1975 to about 19 per 100,000 in the late 1990s and in 2011 it was 17 per 100,000. (Figure 2.4-2)While the incidence has increased in England, the mortality rate has halved with a 43.5% 5-year survival. Stage I disease has a 92% 5-year survival while stage IV has a 5.6% 5-year survival. (Table 2.4-1)Survival rates at 1 year have significantly improved while longer-term survival has improved less. (Figure 2.4-3)Survival has improved for women less than 65 years of age but it has worsened for those over the age of 70. (Figure 2.4-4)Worldwide the incidence of ovarian cancer was over 230,000 with a mortality of just over 150,000 in 2012. (IARC, 2012) It is the 7th most common cancer worldwide in women.

The most important prognostic factor in ovarian cancer remains the stage of the disease at presentation. Unfortunately, around 75% of affected women present with advanced stage disease, which is thought to be due to the vague or absent symptoms of early stage disease.

Lataifeh et al. performed a retrospective study of 100 women with early stage ovarian cancer (stage I) and 100 women with advanced stage ovarian cancer (stage III) looking at symptomatology at presentation. (Lataifeh, 2005) They found that 90% of women with early stage ovarian cancer and 100% of women with advanced ovarian cancer had at least one symptom. When comparing women with early stage disease versus those with late stage disease, 51% versus 44% reported abdominal pain and 32% versus 62% reported abdominal distension. Most women (70%) in both groups had symptoms of a short duration (less than 3 months). Smaller tumours (<5 cm diameter) were three times more likely to have advanced disease (p=0.02). The authors found that those with early disease were more likely to be younger and to have well-differentiated tumours and that advanced disease was not usually due to a delay in diagnosis. Women with a BOT had a significantly longer average symptom time interval than women with a Stage I or II invasive malignancy (8.0 +/- 7.7 versus 3.4 +/- 3.7 months, respectively (P = 0.03). (Eltabbakh, 1999)

Early ovarian cancer causes minimal, nonspecific, or no symptoms at all. Symptoms that may be experienced include bloating, abdominal distension, abdominal discomfort, fatigue, vaginal bleeding, constipation or sensation of a mass with or without pressure effects on the bladder and rectum. Women with borderline or early stage cancer are more likely to report a sensation of an abdominal mass or of urinary pressure symptoms due to the presence of large low-grade tumours. They are less likely to report gastrointestinal problems or general malaise as compared to those with advanced stage cancer. General malaise and other non-specific symptoms are least common in borderline disease.

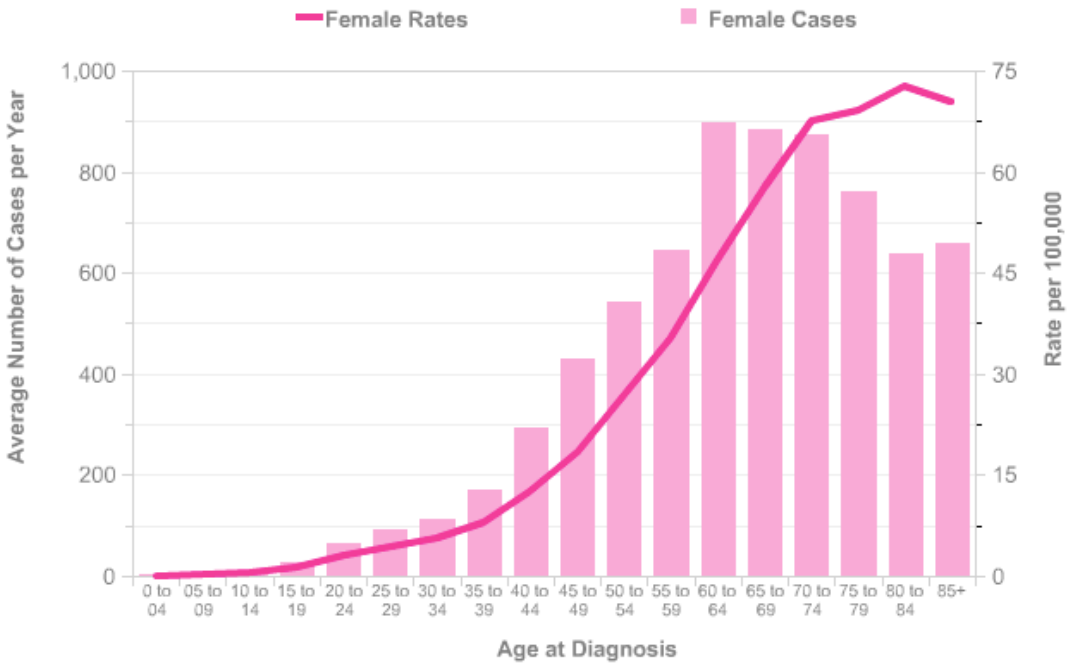
Hippisley-Cox and Coupland performed a cohort study over 10 years, 9 months at the general practice level to find the risk factors for ovarian cancer and develop an algorithm. They found the independent predictors of increasing age, abdominal distension (23-fold higher risk), family history of ovarian cancer (9.8-fold higher), abdominal pain (7-fold higher), postmenopausal bleeding (6.6-fold higher), loss of appetite (5.2-fold higher), anaemia (2.3-fold increased risk), bleeding per rectum (2-fold higher) and weight loss (2-fold higher). (Hippisley-Cox, 2011) Using their algorithm, the 10% of women predicted to have the highest risk had 63% of all the ovarian cancers diagnosed over the consecutive 2 years. Higher parity older women, and those with a family history of breast or ovarian cancer, are more likely to be diagnosed at an advanced stage of disease (Webb, 2004).

Risk factors associated with ovarian cancer include exposure to an increased number of ovulatory cycles seen in women with an early menarche, late menopause, nulliparity, infertility (36-46% risk increase) and recurrent fertility treatment. (Edmondson, 2001) Multiparous women have a 50% decreased risk compared to their nulliparous counterparts and those with multiple pregnancies have greater protection than those with singletons. Breastfeeding is also associated with reduced risk (34%) as well as the use of the combined oral contraceptive pill. (Table 2.4-2)

Carrier status of genetic predispositions like BRCA1 and BRCA2 gene mutations (Malander, 2004), and family history of ovarian, breast or colorectal cancer are also risk factors for ovarian cancer. The average cumulative risk of ovarian cancer by age 70 years was 39% (18%-54%) for BRCA1-mutation carriers and 11% (2.4%-19%) for

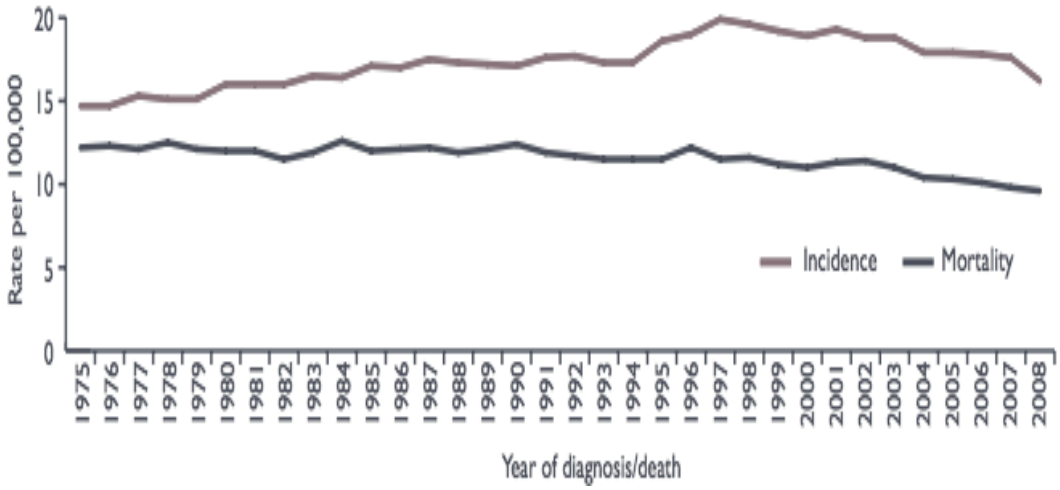
BRCA-2 carriers and of breast cancer of 65% (95% confidence interval 44%-78%) for BRCA-1 and of 45% (31%-56%) for BRCA-2 carriers. (Antoniou, 2003) There are other genetic mutations which have a higher risk of developing ovarian cancer such as the autosomal dominant Lynch II syndrome or HNPCC (hereditary nonpolyposis colorectal cancer) which is associated with a high risk for developing colorectal, endometrial, stomach, small bowel, breast, pancreas as well as ovarian cancers.

Figure 2.4-1: Ovarian Cancer (C56-C57), Average Number of New Cases per Year and Age-Specific Incidence Rates per 100,000 Population, Females, UK, 2009-2011 (Cancer Research UK)



http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer

Figure 2.4-2: Ovarian Cancer (C56-C57), European Age-Standardised Incidence and Mortality Rates, Great Britain, 1975-2008 (Cancer Research UK, 2011)



http://publications.cancerresearchuk.org/downloads/product/CS_CS_OVARY.pdf

March 2011.

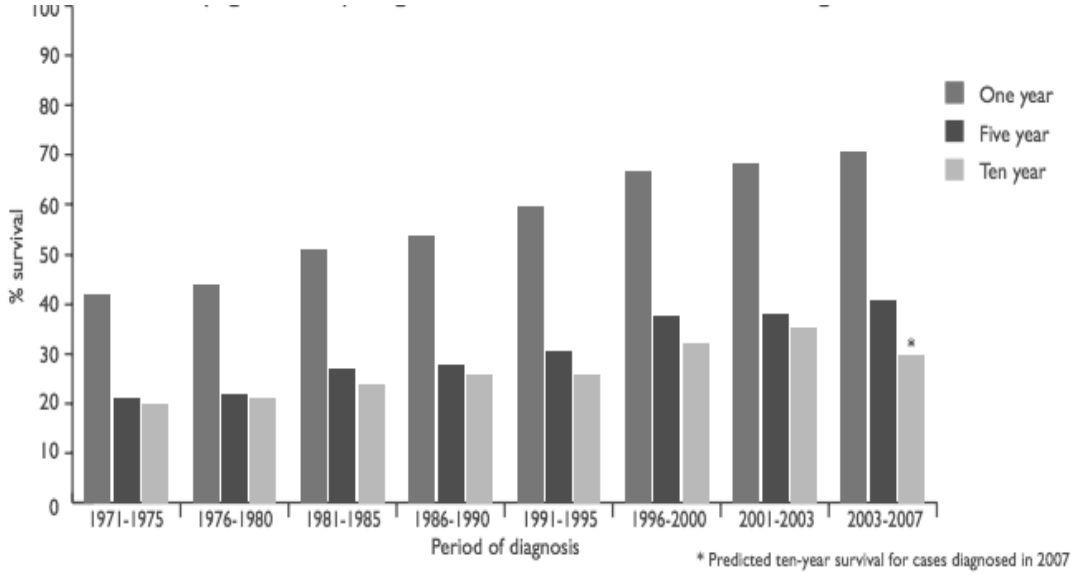
Table 2.4-1: Ovarian Cancer, Five-Year Stage-Specific Relative Survival Rates, Adults (Ages 15-99), Anglia Cancer Network, 2004-2008 (Cancer Research UK, 2011)

Stage at Diagnosis	Number of cases	% of all cases	5-year relative survival (%)	95% Confidence Interval
Stage I	424	29	92.0	(86.5-97.6)
Stage II	62	4	55.1	(36.8-73.5)
Stage III	652	45	21.9	(17.3-26.4)
Stage IV	216	15	5.6	(1.9-9.4)
Unstaged	89	6	27.6	(16.0-39.3)
All stages	1443	100	43.5	(39.9-47.0)

http://publications.cancerresearchuk.org/downloads/product/CS_CS_OVARY.pdf

March 2011.

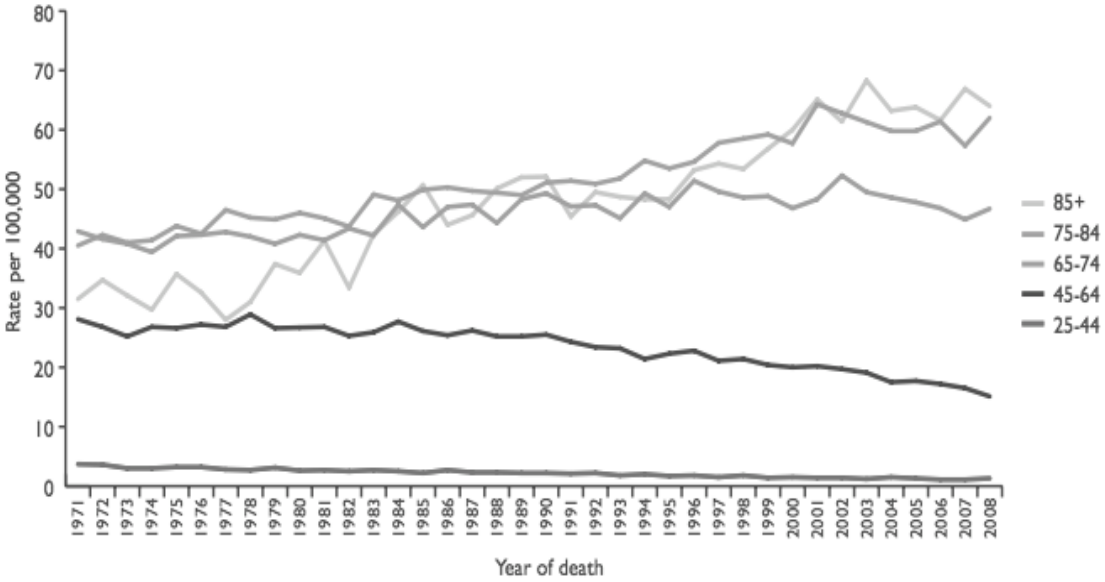
Figure 2.4-3: Ovarian Cancer, One-, Five- and Ten-Year Age-Standardised Relative Survival Rates, Adults (Ages 15-99), England and Wales, 1971-1995, and England, 1996-2007 (Cancer Research UK, 2011)



http://publications.cancerresearchuk.org/downloads/product/CS_CS_OVARY.pdf

March 2011.

Figure 2.4-4: Ovarian Cancer (C56-C57), Age-Specific Mortality Rates, UK, 1971-2008 (Cancer Research UK, 2011)



http://publications.cancerresearchuk.org/downloads/product/CS_CS_OVARY.pdf

March 2011.

Table 2.4-2: Ovarian Cancer, Relative Risk by Parity and Duration of Oral Contraceptive (OC) Use (Cancer Research UK, 2011)

Relative risk for ovarian cancer by parity	**Number of children**	**Relative risk (95% CI)**
	3+	1
	2	1.21 (1.10-1.32)
	1	1.60 (1.43-1.79)
	0	2.12 (1.81-2.48)
Relative risk for ovarian cancer by duration of oral contraceptive (OC) use (mean)	**OC use**	**Relative risk (99% CI)***
	Never	1.00 (0.96-1.04)
	Less than 1 year (0.4 years)	1.00 (0.91-1.10)
	1-4 years (2.4 years)	0.78 (0.73-0.83)
	5-9 years (6.8 years)	0.64 (0.59-0.69)
	10-14 years (11.6 years)	0.56 (0.50-0.62)
	15 years or more (18.3 years)	0.42 (0.36-0.49)
Risk reduction for ovarian cancer by time elapsed since cease of OC use (per five years of OC use)	**Time elapsed since cease**	**Proportional risk reduction**
	Less than 10 years	29%
	10-19 years	19%
	20-29 years	15%

* Relative risk stratified by study, age, parity and hysterectomy.

Data sources: Beral et al. Lancet 2008; 371, 303-14; Granstrom C et al. BJC 2008; 98 (1), 199-205

http://publications.cancerresearchuk.org/downloads/product/CS_CS_OVARY.pdf

March 2011.

2.5 Importance and Clinical Significance and Natural History of Ovarian Cancer

2.5.1 Clinical Significance

Ovarian cancer is the most common gynaecological malignancy causing death in the UK despite being only the second most commonly occurring malignancy after uterine. (Cancer research UK, 2011) Symptoms are vague and non-specific which contribute to diagnoses occurring at a late stage and therefore the reduced survival. Various models of tumourigenesis (as below) may also explain the reason for diagnoses occurring in the latter stages. No precursor markers on scan or in serum tests have been identified which thwarts plans to develop screening protocols. For these reasons, correct diagnosis and management of ovarian tumours is of paramount importance.

2.5.2 Natural History

A small study has suggested that approximately half of ovarian carcinomas develop from pre-existing, benign appearing cysts. This pathway is called adenoma-carcinoma sequence. The remaining half seem to develop from normal appearing ovaries and this pathway is called de novo carcinogenesis. (Horiuchi, 2003)

Shih and Kurman have suggested a different theoretical model of two main pathways of tumourigenesis. It suggests that surface epithelial tumours are divided into two broad categories designated type I and type II tumours. (Shih, 2004; Kurman, 2010)

Type I malignant ovarian tumours are comprised of low-grade serous, low-grade endometrioid, clear cell and mucinous carcinomas and Brenner (transitional cell) tumours. They are generally indolent, thought to develop from borderline tumours and generally present early in stage I. Type II tumours though are comprised of high-grade serous, high-grade endometrioid, malignant mixed mesodermal tumours (carcinosarcomas), and undifferentiated carcinomas. They are aggressive, are theorised to arise in normal ovaries and present in an advanced stage (III/IV). (Kurman, 2011)

Li et al suggested that ovarian serous cancers might actually have a tubal origin rather than originating from Müllerian metaplasia of ovarian surface epithelium. (Li, 2011) It has been proposed that these serous tumours arise from the implantation of tubal epithelium (benign or malignant). Manchanda et al found more occult invasive/in situ tubal cancers than ovarian ones (4 ovarian and 10 tubal) amongst 308 high-risk women (191 with unknown mutation status; 117 known BRCA1/BRCA2 carriers) who chose risk-reducing surgery. The overall occult invasive cancer rate was 5.1% (95% CI 1.9-10.83) in BRCA1/BRCA2 carriers and 1.05% (95% CI 0.13-3.73) in untested women.

Endometriosis, known to be associated with endometrioid and clear cell tumors is regarded as the precursor of these malignant tumors. It is presumed that endometrial tissue is implanted on the ovary via retrograde menstruation. New data now also suggests that mucinous and transitional (Brenner) tumors arise via metaplasia from transitional-type epithelial nests at the tubal-mesothelial junction. (Kurman, 2010)

2.6 Ultrasound Scan Modality for the Assessment of Ovarian and Adnexal Tumours

2.6.1 Lesion

A lesion is defined as an area of the ovary or of an adnexal mass, which is considered to be a non-physiological change. Assessing this requires knowledge of the physiology of the ovary and this is dependant on the age and menopausal status of the woman and the day of her menstrual cycle.

2.6.2 Solid versus Cystic

Solid tumours are typically hyperechoic but may be isoechoic. A solid tumour has been arbitrarily defined, as one in which greater than 80% of the tumour is solid. (Timmerman, 2000) There are few instances where confusion may arise as to whether a tumour is solid or not such as with cysts with thick fluid, blood clot within a cyst and fat within a dermoid which appears white. Checking for vascularity, movement on palpation or looking for acoustic streaming may help. If after employing those techniques it is still unclear, it is recommended that the tumour be assumed to be solid. Solid tumours are noted to have a higher risk of malignancy when compared to other types of cysts but in particular to unilocular cysts. (Granberg, 1989)

2.6.3 Locularity

A locule is a cyst wall cavity or compartment, which is separated by septae. A cyst may

be unilocular with a single cavity, bilocular with 2 cavities or multilocular with multiple cavities. Unilocular cysts can have incomplete septae within it. To be considered bilocular or multilocular there must be complete septae present. Tumours can be broadly classified as unilocular, unilocular-solid, multilocular, multilocular-solid or solid. Cysts judged to be unilocular on ultrasound scan have been demonstrated to have a very low risk of malignancy while multilocular cysts have a greater risk. (Meire, 1979; Granberg, 1989; Valentin, 2013)

2.6.4 Cyst Contents

The fluid within a cyst can vary in appearance. Thin fluid tends to be anechoic so serous cysts can be anechoic or contain low levels. Low-level echogenic fluid is seen as predominantly anechoic but with some echogenic specks within it. Thicker fluid tends to be more echogenic. Some cysts have a pathognomonic appearance. Haemorrhagic cysts can be described as having “cobweb” or “star-shaped” contents which may appear solid but which are in fact only clot within the cyst. Endometriomas are cysts that contain thick old blood and the fluid within them is described as having a “ground glass” appearance on ultrasound scan. Fat within cysts appear white or echogenic and hair has a classic appearance of streaks of bright white. Both of these are typically seen in dermoid cysts.

2.6.5 Cyst Wall

As a cyst is a fluid filled structure, it has an inner and outer surface called the cyst wall.

The cyst wall surface can be smooth or it can be irregular. Irregular cyst walls have papillary projections. Thicker cysts walls are associated with malignancy. (Granberg, 1989)

2.6.6 Septum

A septum is a thin solid structure that extends from one part of the internal cyst wall to another. If it extends to another aspect of the internal cyst wall in all scanning planes visualised, then it is a complete septum. If it extends into the cyst cavity without touching the cyst wall in all scanning planes then it is an incomplete septum. A septum may be vascular (demonstrating blood flow though it) or it may be avascular. Thick septae are defined as those greater than 3mm.

2.6.7 Papillary projections

Papillary projections are solid areas arising from the cyst wall and typically extend into the cyst cavity. Its perpendicular distance from the inner cyst wall must be ≥3mm. Papillary projections themselves can have a smooth or irregular surface and they may or may not demonstrate vascularity on Doppler examination. The risk of malignancy has been found to be higher the greater the number of papillary projections is present. (Granberg, 1990; Bailey, 1998) Hydrosalpinges have a classic “cogwheel sign” or “beads on a string” appearance and as such can often be identified easily but if describing the cyst and the inner surface has raised areas ≥3mm, these should be described as papillary projections. (Timor-Tritsch, 1998) While papillary projections

occur in benign cysts, they have been found to be characteristic of borderline ovarian tumours (BOT) and stage 1 primary invasive epithelial ovarian cancer. (Valentin, 2006b)

2.6.8 Acoustic Shadowing

Acoustic shadowing is the ultrasound phenomenon where a structure being examined reflects almost all the sound waves causing loss of the acoustic echo behind the structure and an anechoic "shadow" behind the structure. This commonly occurs with calcified structures such as pedunculated calcified fibroids, fibromas or dermoid cysts.

2.6.9 Ascites

Ascites is defined as free fluid in the peritoneal cavity, which is sufficient to be seen outside of the pouch of Douglas or to be seen anterior to the uterus. Free fluid in the peritoneal cavity is predominantly found in the pouch of Douglas due to gravity. A small amount of fluid, which is anechoic, or of low echogenicity is normal in a premenopausal woman. Ascites is a characteristic of intra-abdominal malignancies and is typical of ovarian cancers. The clinical detection of ascites to predict the presence of an ovarian malignancy has a positive predictive value (PPV) of 95% and a negative predictive value (NPV) of 64%. (Shen-Gunther, 2002) Only 17% of stage I and II cases have been found to have ascites and this is of smaller volume (all with <500mls) as opposed to 89% in stage III and IV malignancies (66% with >500mls). Fifty-two percent of borderline tumours had no clinical evidence of ascites. (Shen-Gunther, 2002)

2.6.10 Colour Doppler

Doppler is a mode of ultrasound scanning that allows you to detect movement of fluid such as blood. Use of Doppler allows you to detect the velocity and direction of the blood flow within a vessel or within a tumour. Calculations such as the resistance index or the pulsatility index can be calculated. More often a tumour is subjectively described as avascular, poorly vascular or highly vascular. Alternatively a subjective score is given from '1' to '4' where '1' is avascular or no colour seen in the lesion and '4' is highly vascular. Highly vascular tumours have a greater risk of malignancy. (Timor-Tritsch et al., 1998)

2.6.11 Acoustic Streaming

Acoustic streaming is use of Doppler power within a cyst to produce a steady current within the fluid which is driven by the absorption of the high amplitude acoustic oscillations (sound waves) from the probe. The fluid within the cyst is then seen to flow away from the source of the sound waves. Endometriomas, which contain thick fluid, may not demonstrate acoustic streaming.

2.6.12 Qualitative classification

Cysts or tumours can be classified as within 1 of the following 5 tumour groups.

1. Unilocular tumours have a single cyst cavity with no or incomplete septae

and no papillary projections. The risk of malignancy has been round to be 0.3-1% with a benign odds ratio of 12.6 (95% CI: 1.61-99.10), $P < 0.001$. (Valentin, 2013; Granberg, 1989; Gramellini, 2008)

2. Unilocular-solid tumours are cysts with a single cavity but also contain at least 1 papillary projection or a solid component. The risk of malignancy was found to be marginally higher at 2% by Granberg. (Granberg, 1989; Gramellini, 2008)

3. Multilocular tumours are cysts with more than a single cyst cavity but no papillary projections or solid components. The malignancy risk has been found to rise slightly to 8-12.6% for these tumours but the odds ratio of it being benign was still 7.9 (95% CI: 1.00-62.38), $P < 0.05$. (Granberg, 1989; Saunders, 2010)

4. Multilocular-solid tumours are those with multiple cyst compartments and also at least 1 papillary projection or a solid component. The risk of malignancy for these tumours has been found to be high at 36% with Gramellini quoting an odds ratio of malignancy 6.4 (95% CI: 1.81-22.70), $P < 0.001$. (Granberg, 1989; Gramellini, 2008)

5. Solid tumours, as classified by Timmerman, are tumours with more than 80% solid components. (Timmerman, 2000) The risk of malignancy has been found to be the highest (39%) when compared to the other tumour groups with Gramellini recording an odds ratio of 5.5 (95% CI: 1.48-20.92), $P < 0.05$. (Granberg, 1989; Gramellini, 2008)

2.6.13 Palpation and mobility

Assessing mobility of the pelvic organs is often the part forgotten during an ultrasound scan assessment of the pelvis. Important information can be garnered from the location of pelvic organs in relation to other organs and to the pelvic sidewall but also their mobility in relation to the adjacent organs.

In a normal pelvis the organs slide over each other due to their visceral and parietal peritoneal coverings. Loss of this demonstrable mobility is usually due to adhesions and occasionally it may be due to the weight of the tumour preventing easy movement without causing discomfort to the patient. (Yazbek, 2007b; Guerriero, 2009) These adhesions may be due to previous pelvic infection such as pelvic inflammatory disease (PID) or appendicitis, endometriosis, previous operations, inflammatory conditions such as Crohn's disease or due to a malignancy invading adjacent tissues. Knowledge of a patient's medical history can give help determine which of the above conditions may have caused adhesions in a particular case. The type of adhesions and the location of the ovaries can give further information. Previous PID tends to cause filmy thin mobile adhesions while chronic inflammatory conditions or previous surgery tend to cause dense firm adhesions. Endometriosis adhesions tend to cause obliteration of the pouch of Douglas and tend to cause pathognomonic "kissing ovaries" which are ovaries adherent to the postero-lateral aspect of the uterus and that touch in the midline. (Ghezzi, 2005) Malignancies that invade surround tissues are often primary epithelial adenocarcinomas as opposed to non-epithelial or metastatic tumours.

2.7 Current Modalities for Assessing Ovarian Cysts

Many models have been developed to increase the accuracy of pre-operative diagnosis because non-invasive assessment of ovarian cysts is difficult.

2.7.1 Main Tumour Markers – CA125, HE4

CA125 (cancer antigen 125, carcinoma antigen 125, or carbohydrate antigen 125) is a high molecular weight glycoprotein of the mucin family (also known as mucin 16), which is expressed on cell membrane surfaces and is also been found free in the serum.

Serum measurements are used primarily for the diagnosis of ovarian cancers as well as monitoring of progression, response to treatment and recurrence. Increased tissue expression and raised serum levels have been found in epithelial ovarian cancers though this is often not the case in early stage and borderline tumours.

Serum levels may also be raised in several benign conditions such as menstruation, endometriosis, recent ovulation, pelvic inflammatory disease, heart failure, liver cirrhosis, lung disease and recent surgery. (Miralles, 2003) It may also be raised in non-ovarian malignancies, especially those with serosal involvement or spread. (Topalak, 2002)

An early study (1989) gave CA 125 a sensitivity of 86% and a specificity of 78% in the pre-operative identification of malignant ovarian tumours. (Mogensen, 1989)
A systematic review of use of CA125 to diagnose ovarian malignancies including 17 studies with 2374 women showed a pooled sensitivity of 80% (95% CI: 76%-82%) and the specificity was 75% (95% CI: 73%-77%). (Medeiros, 2009)

Hirai et al looked at 146 women with stage IA ovarian cancers and found that women with normal CA125 levels (n=87, 60%) tended to have smaller tumours (<20 cm) with

less solid components (<50%) and had a slightly different histological type distribution when compared with those with elevated CA125 levels. (Hirai, 2011) Those with normal levels were more likely to have a clear cell adenocarcinoma with a unilocular cyst, which may have less than 50% solid components (n=12). Seventy percent of the clear cell tumours had normal CA 125 levels. Those with elevated CA125 levels tended to have unilocular cysts with 50% to 80% solid components (n=4, 100%) or an endometrioid adenocarcinoma in solid tumours (≥80% solid components). Mucinous adenocarcinoma tumours presenting as multilocular cysts with less than 50% solid components were found to have normal CA 125 levels (n=25) more often than elevated ones (n=19). Pattern recognition was found to be more accurate than CA 125 [93% (95% CI: 90.9% to 94.6%) versus 83% (95% CI = 80.3% to 85.6%)] (Van Calster, 2007)

Human epididymis protein 4 (HE4) is a four-disulfide core domain protein-2 glycoprotein. The gene, found of chromosome 20 encodes a protein of the WFDC domain family and it may be involved in sperm maturation.

The HE4 gene had highest expression in normal human trachea, salivary gland, lung, prostate, pituitary gland, thyroid, and kidney. (Galgano, 2006) It is also found in the glandular epithelium of the female genital tract though in low quantities and usually in tissue of Müllerian origin. (Hellström, 2002) Among tumours, gene expression was highest in ovarian serous and endometrioid carcinomas followed by lung adenocarcinomas, breast, renal, transitional cell, gastrointestinal, pancreatic carcinomas as well as mesotheliomas. (Galgano, 2006; Drapkin, 2005)

The median CA125 and HE4 serum concentrations were significantly higher in patients with epithelial ovarian cancers when compared with healthy females and those with benign tumours. (Montagnana, 2011; Sandri, 2013)

Clinically HE4 can be used for diagnosis, response to treatment, progression and for surveillance for recurrence. A systematic review in 2013 with 13 studies found that HE4 performed better when compared with CA125 for the diagnosis of ovarian cancer (ORs HE4 was 37.2 (95% CI: 19.0 to 72.7, adjusted for publication bias) versus CA 125 which was 15.4 (95% CI: 10.4 to 22.8)]. (Ferraro, 2013) A further systematic review of 45 studies with 10,671 women and 3946 ovarian malignancies, the pooled sensitivity for the diagnosis of malignant tumours (borderline and invasive) was 78% (95% CI: 77%-79%), and the specificity was 86% (95% CI: 85%-87%). (Macedo, 2014)

2.7.2 Other Tumour Markers – LDH, α-FP, β-hCG, CEA, CA19-9

Lactate dehydrogenase (LDH) is an enzyme involved in anaerobic glycolysis by converting pyruvate to lactate in low oxygen conditions. Serum levels are often raised in times of increased cell breakdown such as hemolysis but also in patients with a variety of cancer types. Very high levels are often found specifically in women with dysgerminomas. (Kawai, 1992; Konishi, 1986; Fujii, 1985)

Human chorionic gonadotropin (hCG) is a glycoprotein produced by the embryonic syncytiotrophoblast or placenta. Apart from in normal or molar pregnancies, elevated levels of the β-chains (β-hCG) can be seen in women with choriocarcinomas, germ cell tumours, teratomas with elements of choriocarcinoma, and islet cell tumours.

Alpha-fetoprotein (α-FP) is the most abundant fetal plasma protein and is produced by the fetal yolk sac and liver. The levels of his glycoprotein fall dramatically after birth until it reaches adult levels by 1 year of life. Raised levels are found specifically in women with yolk sac tumours. (Kawai, 1992; Konishi, 1986)

Kawai et al performed a study of 7 tumour markers (α-FP, CA125, CA19-9, tissue polypeptide antigen (TPA), carcinoembryonic antigen (CEA), human chorionic gonadotropin (hCG), and lactate dehydrogenase (LDH)) in 135 women with germ cell tumours. (Kawai, 1992) Dysgerminomas, particularly and yolk sac tumours, to a lesser extent had raised levels of LDH while hCG was raised in 11 of the 36 yolk sac tumours. The authors found raised α-FP in 100% of the yolk sac tumours (n=36), 61.9% of the 21 immature teratomas and in 11.8% of the 17 dysgerminomas. The α-FP was normal in all the mature cystic teratomas with malignant transformation (n=7) and the mature cystic teratoma (n=31). (Kawai, 1992)

Elevated serum levels of the antigens CA19-9 and CEA are predominantly found during the diagnosis of colon and biliopancreatic malignancies.

Raised CA19-9 levels are also seen in teratomas. This includes benign mature cystic teratomas as well as malignant immature teratomas and mature cystic teratomas with malignant transformation. It therefore is unlikely to aid in the determination of whether the tumour is benign or malignant. TPA and CEA were not considered useful tumour markers for ovarian germ cell tumours. (Kawai, 1992; Dede, 2006)

Elevated CEA levels are found in 40.6% of epithelial adenocarcinomas and these are mostly in women with a mucinous adenocarcinoma. It is also found in women with pseudomyxoma peritonei, or a Krukenberg tumour (Konishi, 1986)

2.7.3 Risk Malignancy Index (RMI)

The most commonly used model for the diagnosis of ovarian cancer is the RMI (Jacobs, 1990). This model calculates a score which is the product of the CA125 (U/ml) laboratory value, a score for menopausal status (1 if pre-menopausal and 3 if postmenopausal) and a greyscale ultrasound score of 0, 1 or 3 where 1 score is given each for bilaterality, ascites, multilocularity, solid areas and intra-abdominal metastases.

Equation 2.7-1: RMI

RMI = CA125 x M x U

M = Menopausal status score

U = Ultrasound score

In the original study by Jacobs a score of >200 gave a sensitivity of 85% and a specificity of 97% for the diagnosis of ovarian cancer. A recent systematic review of 16 studies assessing the diagnostic performance of RMI showed an overall sensitivity of 78% (95% CI 71– 85%) and a specificity of 87% (95% CI 83–91%). (Geomini, 2009)

The RMI is simple and widely used which is advantageous. The main disadvantage is that it includes a blood test for CA125, which adds to the cost, as well as to the

time taken to come to a conclusion. The other concern is that the diagnosis of an ovarian cancer may be delayed in up to 18-22% of cases due to normal CA125 values which occur particularly with early stage epithelial and as well as non-epithelial malignancies. (Bast, 1983; Rosen, 2005)

The RCOG guidance uses the RMI for diagnostic evaluation of adnexal tumours. The absolute value of the RMI and the serum level of CA125 are used to determine the management plan. For postmenopausal women with an RMI of < 25 are managed by a general gynaecologist. Those with a simple unilateral cyst of <5cm and who have a CA 125 of < 30u/ml can be managed conservatively. If the CA125 ≥ 30u/ml or the cyst is larger, has septations or solid areas then surgery is advised. For those with a RMI of 25 – 250, management should be in a cancer unit. Those with a RMI >250 are offered management by gynaecological oncologists in a tertiary cancer centre. (RCOG, 2003)

In premenopausal women, the RMI may still be used but its specificity falls significantly due to false positive results of the CA125 in this population. (Miralles, 2003)

2.7.4 The Risk of Malignancy Algorithm (ROMA)

ROMA is the calculation using a patient's menopausal status with the CA 125 (U/ml) and the HE4 laboratory values to estimate the risk of a patient having an ovarian malignancy. Ultrasound features do not influence this result, as it is purely laboratory based.

Moore et al found it to have a sensitivity of 94% and a specificity of 75% for the detection of epithelial ovarian malignancies. (Moore, 2011) The ROMA has shown better diagnostic performance in post-menopausal women (sensitivity of 82.5%) when compared to that in pre-menopausal women (sensitivity of 53.3%). (Sandri, 2013)

The AUC for ROMA was significantly higher in comparison to CA125 alone (93.3% versus 90.3%, p=0.0018) in postmenopausal patients. Sub-analysis considering only patients with endometrioid disease showed the highest accuracy of HE4. (Sandri, 2013)

The combination of HE4 and CA125 (ROMA) has not show better performance than HE4 alone. (Montagnana, 2011)

2.7.5 Ultrasound scan Logistic Regression Models

These are mathematical models using multivariate regression analyses of tumour characteristics, personal or family history or tumour markers to determine the risk of a tumour being malignant.

In 1997 a model was created where 3 of the 10 recorded variables (age, papillary projection score [present=1, absent=0] and the time averaged maximum velocity [TAMXV]) were found to be useful. (Tailor, 1997) The probability of malignancy was calculated using the following formula:

Equation 2.7-2: Tailor Logistic Regression Model

$1/(1 + e^{-z})$

Where z = 0.1273 x (Age) + 0.2794 x (TAMXV) + 4.4136 x (papillary projection score) – 14.2046. Using a cut-off value of 25%, they achieved the best sensitivity of 93.3% and sensitivity of 90.4%.

The IOTA Group (International Ovarian Tumour Analysis) is a European multicentre group that started in 2000 and which devised a set of terms to define adnexal lesions. (Timmerman, 2000) There were almost 3000 patients assessed at 19 centres in 8 countries.

In 2005 the IOTA Group devised a logistic regression model based on those ultrasound factors and patient characteristics to determine the risk of a lesion being malignant. (Timmerman, 2005) The 12 factors are 1) personal history of ovarian cancer (yes=1, no=0), 2) current use of hormonal therapy (yes=1, no=0), 3) age of the patient (years), 4) maximum diameter of lesion (mm), 5) presence of pain during the examination (yes=1, no=0), 6) presence of ascites (yes=1, no=0), 7) presence of blood flow within a solid papillary projection (yes=1, no=0), 8) purely solid tumour, 9) maximum diameter of the solid component (maximum of 50mm), 10) irregular internal cyst walls (yes=1, no=0), 11) presence of acoustic shadows (yes=1, no=0), 12) colour flow score (1-4 where 1 is no flow and 4 is maximum flow). (Timmerman, 2005)

LR1 used all 12 variables to determine the probability of malignancy. This was

calculated using the formula:

Equation 2.7-3: IOTA Logistic Regression Model LR1

$y = 1/ (1 + \exp[-z])$

where $z = – 6.7468 +1.5985 (1) – 0.9983 (2)+0.0326 (3) + 0.00841 (4) – 0.8577 (5) + 1.5513 (6) + 1.1737 (7) + 0.9281 (8) + 0.0496 (9) + 1.1421 (10) – 2.3550 (11) + 0.4916 (12)$ as described in the original IOTA study.The probability y is dichotomised at 0.1 to give a predictive diagnosis and this gave a sensitivity of 93% and a specificity of 76%. These tests were subsequently tested prospectively by the IOTA group. LR1 had a sensitivity of 92.2% and a specificity of 86.5% for external groups (997 women) and 92.7% and 80.6% respectively for the internal groups (941 women). (Timmerman, 2010a)

Six variables were used for the calculation of LR2: 3) age of the patient (in years), 6) presence of ascites (yes=1, no=0), 7) presence of blood flow within a solid papillary projection (yes=1, no=0), 9) maximal diameter of the solid component (expressed in millimetres, but with no increase > 50 mm), 10) irregular internal cyst walls (yes=1, no=0), 11) presence of acoustic shadows (yes=1, no=0).

The probability of malignancy was calculated using the formula:

Equation 2.7-4: IOTA Logistic Regression Model LR2

$y = 1/ (1 + \exp[-z])$

where $z = - 5.3718 + 0.0354 (3) + 1.6159 (6) + 1.1768 (7) + 0.0697 (9) + 0.9586 (10) - 2.9486 (11)$ as described in the original IOTA study.

The probability y is dichotomised at 0.1 to give a predictive diagnosis, which had a sensitivity of 92% and a specificity of 75% for the diagnosis of ovarian cancer. When tested prospectively by the IOTA group LR2 had a sensitivity of 91.8% and a specificity of 85.6% for external groups (997 women) and 89.2% and 79.8% respectively for the internal groups (941 women). (Timmerman, 2010a)

2.7.6 Other ultrasound scoring models

In 2008 the IOTA collaboration proposed use of 'Simple Rules' to assess tumours. (Timmerman, 2008) The 'Simple Rules' are based on morphological features assessed on ultrasound scan. The 'Simple Rules' represent a simplified form of "pattern recognition" (see below) and allows systematic examination of tumour morphology to be performed in a structured way by less experienced operators.

The ultrasound scan-based 'Simple Rules' use 10 rules to assess adnexal masses. There are five rules that predict malignancy (M-rules): (1) irregular solid tumour; (2) ascites; (3) at least four papillary structures; (4) irregular multilocular–solid tumour with a largest diameter of at least 100 mm; and (5) very high colour content on colour

Doppler examination (score 4). There are five rules to suggest that the tumour is benign (B-rules): (1) unilocular cyst; (2) presence of solid components where the largest solid component has a largest diameter of <7 mm; (3) acoustic shadows; (4) smooth multilocular tumour less than 100 mm in largest diameter; and (5) no detectable blood flow on Doppler examination (score 1).

If one or more M-rules apply in the absence of a B-rule, the mass is classified as malignant. If one or more B-rules apply in the absence of an M-rule, the mass is classified as benign. If no rule applies or both M and B rules apply, the mass cannot be classified and an expert examiner is advised. (Table 2.7-1)

These ten rules were applicable to 76% (937/1233) of all tumours in the study, where they resulted in a sensitivity of 93%, specificity of 90%, positive likelihood ratio (LR+) of 9.45 and negative likelihood ratio (LR−) of 0.08. Among those cysts that were not classifiable 82% were benign in the original study. When prospectively tested the rules were applicable in 76% (386/507) of the tumours, where they had a sensitivity of 95% (106/112), a specificity of 91% (249/274), LR+ of 10.37, and LR− of 0.06. The ‘Simple Rules’ were subsequently tested prospectively. (Timmerman, 2010b) The rules were applicable in 77% of the 1938 women with a sensitivity of 92% (95% confidence interval 89% to 94%) and specificity of 96% (94% to 97%). The ‘Simple Rules’ cannot be applied to all adnexal tumours, which makes this method less effective than pattern recognition. However, in cases where ‘Simple Rules’ are applicable, their diagnostic accuracy is very high.

Table 2.7-1: Simple Rules for the classification of benign and malignant tumours (Timmerman, 2008)

Rules for predicting a malignant tumour (M-rules)		**Rules for predicting a benign tumour (B-rules)**	
M1	Irregular solid tumour	B1	Unilocular
M2	Presence of ascites	B2	Presence of solid components: largest solid component < 7 mm
M3	At least four papillary structures	B3	Presence of acoustic shadows
M4	Irregular multilocular solid tumour with largest diameter ≥ 100 mm	B4	Smooth multilocular tumour with largest diameter < 100 mm
M5	Very strong blood flow (colour score 4)	B5	No blood flow (colour score 1)

2.7.7 Artificial Neural Network Models

Artificial neural networks (ANNs) are computer-based learning statistical algorithms which are modelled after the human central nervous system and which investigate non-linear interactions between variables. ANNs consist of interconnected processing elements that respond simultaneously to various inputted data. (Shepherd, 1990; Haykin, 1994; Dayhoff, 2001) The ANN, through a learning process called training

acquires knowledge and interneuron connection strengths (known as synaptic weights) are used to store that knowledge. Information is exchanged between neurons in the form of numerical values via synaptic interconnections. ANNs have the ability to perform multifactorial analyses and to generalise.

In a prospective study on 173 consecutive patients who were scheduled to undergo surgery for adnexal tumours, an ANN model using simple clinical and ultrasound-derived criteria was tested to determine if the tumour was malignant. The prevalence of malignancy was 28.3% and the best ANN gave a sensitivity of 95.9%, specificity of 93.5%, LR+ of 14.75 and LR- of 0.04. (Timmerman, 1999a) Another study tested ANN models on 67 women with adnexal masses, of which 22.4% were malignant. With an optimum cut-off value of 0.45, they reached a sensitivity of 100%, a specificity of 98.1%, a LR+ of 52.6 and a LR- of 0. (Tailor, 1999) When prospectively tested, Tailor's model did not perform similarly and only achieved a sensitivity of 43% (95% CI, 23 – 64), specificity of 92% (95% CI, 84 – 100), LR+ of 5.38 and LR- of 0.62. (Aslam, 2000b)

2.7.8 Pattern Recognition

While the differentiation between benign and malignant adnexal tumours is often difficult, he most accurate method thus far to assess the nature of an adnexal tumour is called 'pattern recognition'. (Valentin, 1999a; Timmerman 1999b) This technique is based on systematic examination of the tumour's morphological features on ultrasound scan. The ultrasound operator determines if the tumour is benign or malignant based on this examination finding and their experience.

Van Calster found pattern recognition to have an accuracy of [93% (95% CI: 90.9% to 94.6%)]. (Van Calster, 2007) In a multicentre trial with 1066 women, pattern recognition was shown to predict malignancy in 97% (95% CI, 93–99%) (140/144) of the primary invasive malignant tumours. This included 93% of Stage 1 primary invasive tumours, 99% of Stage 2–4 primary invasive tumours, 93% (95% CI, 80–97%) (39/42) of the metastatic tumours and 60% (95% CI, 47–72%) (33/55) of the borderline tumours. (Sokalska, 2009) Yazbek et al found a similar accuracy with 68.6% (24/35) of the women with BOTs receiving the correct pre-operative diagnosis. (Yazbek, 2007a)

The accuracy of pattern recognition has been shown to be affected by the degree of certainty or diagnostic confidence of the ultrasound operator with an accuracy varying between 75% and 98% for 'certain' or sure diagnoses. (Yazbek, 2010) Among 300 patients with tumours, using the static ultrasound images, the greatest accuracy (91.7 %, 95% CI 88 – 94.3) and the highest level of interobserver agreement was seen amongst the ultrasound operators with the greatest experience. (Timmerman 1999b) Its usefulness is therefore related to the experience and confidence of the operator.

While this method is the most accurate, it can only be used by highly skilled and experienced ultrasound operators, which are not widely available.

2.7.9 Magnetic Resonance Imaging (MRI) / Computed Tomography (CT) Scans

When basic grey scale ultrasounds scans are unable to determine the risk of

malignancy of a tumour, second line imaging tests are utilised. A meta-analysis of 9 studies evaluated the value of a second radiological test for an indeterminate adnexal mass detected on basic gray-scale ultrasound. (Kinkel, 2005)

The prevalence of ovarian cancer was 8.75% in the premenopausal population and 32.40% the postmenopausal population with an ovarian mass. After the initial gray-scale ultrasound scan assessment, the positive predictive value (PPV) in pre-menopausal women changed from 8.75% to 25% and in postmenopausal women from 32.4% to 63% for indeterminate results. Alternatively it fell to 2% (premenopausal women) and 7% (postmenopausal women) for benign results.

After performing the second test for those indeterminate tumours, the PPV increased more after MRI with IV contrast administration (premenopausal women, 80%; postmenopausal women, 95%) or with non-enhanced MRI (premenopausal women, 70%; postmenopausal women, 92%) than it did after combined gray-scale and Doppler US (premenopausal women, 30%; postmenopausal women, 69%) or after CT scan (premenopausal women, 38%; postmenopausal women, 76%) ($P < 0.001$). It therefore determined that contrast-enhanced MRI provided the highest positive predictive value for the diagnosis of ovarian cancer when compared with CT, ultrasound scan with Doppler, or MRI without contrast administration.

Both Doppler ultrasound and MRI are highly sensitive for identifying malignant lesions (ultrasound 100%, MRI 96.6%), while the specificity of both tests vary (ultrasound 39.5%, and MRI 83.7% in 1 prospective study). (Sohaib, 2005; Iyer, 2010)

MRI criteria for the diagnosis of malignancy are a large solid component, wall thickness greater than 3 mm, septums greater than 3 mm thick with or without nodularity (papillary projections), and necrosis. Additional features include evidence of invasion into other pelvic organs or the sidewalls, evidence of peritoneal or omental disease, presence of ascites and lymphadenopathy. Borderline tumours as with ultrasound are more difficult to recognise and have no specific criteria.

Amongst the benign tumours, MRIs are quite good at recognising the old blood in endometriomas and the fat in dermoid cysts. MRI also has some advantages when ultrasound scans may reach their limits such as being able to examine all parts of a cyst even when very large and even when the patient's habitus or pathology such as fibroids impedes the ultrasound waves.

CT scans find similar features of ovarian cancers as MRIs do and the tumours often demonstrate contrast uptake.

Despite its lesser performance in the diagnosis of ovarian cancer, CT scans are currently the preferred technique for the evaluation and staging of tumours pre-treatment. CT scans are also used to predict the likelihood of successful surgical clearance. Diaphragmatic and the large bowel involvement are noted to be the most predictable determinant of suboptimal cytoreduction. Other features are involvement of the sidewall with hydroureter / hydronephrosis, omental disease involving the spleen, stomach or lesser sac and suprarenal paraaortic lymphadenopathy. (Iyer, 2010)

2.8 Current Management Options for Ovarian Cysts

2.8.1 Conservative management with or without ultrasound surveillance

The likelihood of an asymptomatic benign cysts requiring surgical intervention is very low. In a study of 134 postmenopausal women with asymptomatic adnexal cysts judged to be benign on ultrasound scan, only 9% required surgery. (Valentin, 2002) Histological diagnosis confirmed benign tumours in all surgical cases. (Valentin, 2002) There was cyst regression in 54% of women and the remaining cysts were unchanged during follow up. In large cohort studies of 31,834 - 15,106 women in whom 2,763-6,807 had cysts, 63-69% of these had complete resolution of their cysts. (Modesitt, 2003; Pavlik, 2013) Simple unilocular cysts had a larger resolution rate (69%) when compared with more 'complex' appearing cysts (55%). (Modesitt, 2003; Bailey, 1998) Ovarian cysts can therefore resolve spontaneously and not all women need surgical management. Overall this data suggests that expectant management of adnexal cysts, which appear benign on ultrasound scan, is safe and the risk of subsequent surgical intervention is low.

Despite this information, the knowledge of the presence of a cyst still produces anxiety and prompts further investigations and blood tests and subsequent removal even when asymptomatic.

The RCOG recommends conservative management for all simple, anechoic, unilocular cysts, which are <50mm in diameter in postmenopausal women. Cysts that are less that 20mm do not need follow up with ultrasound scan surveillance. (RCOG, 2003) In premenopausal women 'simple' cysts < 50mm do not require any follow up according to the RCOG, as these are usually physiological and resolve in the subsequent menstrual cycles. (RCOG, 2011)

2.8.2 Laparoscopic surgery

For women who do not meet the RCOG criteria for conservative management are offered surgical management. When women are offered surgical management, it is recommended that this should be via a laparoscopic route where possible due to shorter hospital stays, lower morbidity and faster recovery. (Hilger, 2006; Mais, 1995) This applies to premenopausal women with cysts between 50mm and 70mm and postmenopausal women with intermediate risk. These are women for whom the risk of malignancy is considered to be moderate, the disease is considered to be early stage or for whom fertility sparing surgery is being considered.

2.8.3 Fertility sparing surgery

Younger women who have not completed their families and particularly those with early stage (stage 1) or low-grade tumours may wish to have fertility sparing surgery. The rate of tumour recurrence following fertility sparing surgery and radical surgery in early-stage and low-grade tumours are 9% and 11.6%, respectively. (Ayhan, 2003)

Overall survival rates and disease-free intervals are not dissimilar.

2.8.4 Staging Laparotomy

The best survival rates for women with ovarian cancer are achieved when treatment is organised and carried out by gynaecological oncologists working in cancer centres because the prognosis is better when they have optimal debulking. (Im, 2005; Junor, 1999; Curtin, 1997)

Favourable independent prognostic factors for 5-year overall survival for ovarian cancer include younger age, good performance status, histological types other than mucinous or clear cell, well-differentiated histological tumor grade, early stage at diagnosis, smaller tumor volume before surgical debulking, no residual tumor after debulking, and absence of ascites. (Pomel, 2007; Pereira, 2016)

Table 2.8-1: International Federation of Gynecology and Obstetrics (FIGO) staging of primary ovarian carcinoma (Shepherd, 1989)

Stage	Tumour characteristics
I	**Growth limited to the ovaries**
Ia	Growth limited to one ovary, no ascites present containing malignant cells, no tumour on the external surface, capsule intact
Ib	Growth limited to both ovaries, no ascites present containing malignant cells, no tumour on the external surfaces, capsules intact
Ic	Tumour either stage Ia or Ib but tumour found on surface of one or both ovaries or capsule(s) ruptured or ascites present containing malignant cells or peritoneal washings positive
II	**Growth involving one or both ovaries with pelvic extension**
IIa	Extension or metastases to the uterus or fallopian tubes
IIb	Extension to other pelvic tissues
IIc	Tumour either stage IIa or IIb but tumour found on surface of one or both ovaries or capsule(s) ruptured or ascites present containing malignant cells or peritoneal washings positive
III	**Tumour involving one or both ovaries with histologically confirmed peritoneal implants outside the pelvis and / or positive retroperitoneal or inguinal nodes; superficial liver metastases; tumour limited to the true pelvis but with histologically confirmed malignant extension to small bowel or omentum**
IIIa	Tumour grossly limited to the true pelvis with negative nodes but with histologically confirmed microscopic seeding of abdominal peritoneal surfaces or extension to small bowel or mesentery

IIIb	Tumour involving one or both ovaries with histologically confirmed implants of metastasis to abdominal peritoneal surfaces (lesions ≤2 cm), negative nodes
IIIc	Tumour involving one or both ovaries with histologically confirmed extrapelvic peritoneal metastasis (lesions >2 cm) or positive retroperitoneal or inguinal nodes
IV	**Growth involving one or both ovaries with distant metastases, positive cytologic findings if pleural effusion is present; parenchymal liver metastases**

Part III – Performing a Meta-Analysis

A meta-analysis is a systematic quantitative approach to identify, appraise, synthesize and if suitable to combine and summate results of previous research to make conclusions about the entire body of research on a particular topic or to answer a particular question. It is a statistical process bringing together individual sections of evidence, which is similar enough to be combined. There are a number of pitfalls to be avoided to prevent the results of the meta-analysis from being biased and causing publication of misinformation. There are a number of important steps to be followed, which will be detailed below.

The first step is to define the research question to be answered. A protocol should be written with objectives, patient characteristics, interventions, outcomes, index test investigated, reference test used, inclusion and exclusion criteria, study types, time flow and length of follow up. (Egger, 1997)

The next step is to identify potential studies. Useful online databases are MEDLINE, which has published articles, EMBASE, which also includes conference abstracts and Cochrane. Cochrane includes the Cochrane Database of Systematic Reviews (CDSR), the Cochrane Central Register of Controlled Trials (CENTRAL), the Cochrane Methodology register (CMR) and the Database of Abstracts of Effects (DARE). Other strategies for finding useful studies include searching through bibliographies of articles and especially of review articles as well as using online databases such as Cancer Research UK or trial registration websites to find studies with an International Standard Randomised Controlled Trial Number (ISRCTN). Including unpublished studies helps

reduce publication bias. Publication bias is the effect caused by the tendency for researchers to publish only positive results. This causes the overall effect to appear positive.

Once the potential studies have been identified, those of most use must be chosen. These are usually those of highest quality and those that actually answer the question. Well-conducted double-blinded randomized controlled trials are often those of highest quality but these may not be available depending on the research question. Depending on the type of studies to be included, there are several tools to aid in the process and to specifically assess the quality of the studies.

The Quality of Reporting of Meta-analyses (QUOROM) guidelines gives guidance on standards to improve the quality of reporting of meta-analyses of clinical randomized controlled trials (RCTs). (Moher, 1999) This includes 21 headings and subheadings for the abstract, introduction, methods, results and discussion on what should be reported on objectives, data searches, study selection criteria and characteristics, assessment of validity, data extraction, and quantitative data synthesis.

CONSORT statement includes a 25-item checklist and a flow diagram and it provides instructions for reporting all randomized controlled trials. It is mainly for individually randomized, two group and parallel trials whereas other types require further information. (Shultz, 2010)

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is a 27-point checklist for conducting and reporting a systematic review or a meta-

analysis with a 4-step flow diagram for the selection of studies (identification, screening, eligibility and included). (Moher, 2009; Liberati, 2009) It is used to maintain the standards for reporting of RCTs and other types of studies such as intervention evaluation and has mostly superseded the QUOROM statement.

The Institute of Medicine (IOM) has also issued guidance on recommended standards for systematic reviews of the comparative effectiveness of medical or surgical interventions. (Eden, 2011) They aim to ensure the systematic reviews are objective, transparent and produce high quality, scientifically valid results which can be used.

Newcastle-Ottawa Scale is reported to be an easy, convenient tool for quality assessment of non-randomized studies such as case-controlled and cohort studies. (Wells, 2000; Wells, 2003; Wells, 2013; Hartling, 2012)

Meta-analysis of Observational Studies in Epidemiology (MOOSE) describes recommendations for the conduct and reporting of meta-analyses in observational studies. (Stroup, 2000) A checklist has been created to aid with every step of the meta-analysis process. This is the checklist that was used for our meta-analysis, as it was the checklist most appropriate for the type of studies done on our research question.

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement are a list of recommendations developed on what should be included in an accurate and complete report of an observational study including cohort, case-control, and cross-sectional studies. (von Elm, 2007; V, 2007) The aim is for clarity, completeness, and transparency of the reporting, which will allow readers to appreciate

the true value and applicability of the results.

Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) is a modified assessment tool to individually assess each potential study for a meta-analysis of diagnostic accuracy. (Schueler, 2012; Whiting, 2011) This tool has 4 main domains on patient selection, the index test, the reference standard and flow of patients through the study alongside timing of the index test and reference standard.

Diagnostic studies should have Standards for the Reporting of Diagnostic accuracy studies statement (STARD). (Bossuyt, 2003) Included studies should ideally have many if not all the standards and descriptions within the publication.

When choosing the studies to include, those with high risk of bias should be noted as they can affect the final result. Some more common types of biases to scrutinise studies for are selection, exclusion, publication and verification bias. Selection bias occurs statistically when there is an error in choosing participants for the study. Sample selection bias is due to non-random or non-consecutive sampling of the population. This may be due to volunteering or self-selection when patients choose to be in the study or choose which group they wish to be in. (Hernán, 2004) Informative censoring occurs when there is a differential loss to follow up. This must be noted and accounted for when doing the statistical tests. Exclusion bias occurs when different criteria are used to select cases and controls.

Publication bias is seen when studies are evaluating a medical or surgical treatment, as negative results are less likely to be published and would then affect the final meta-

analysis. Funnel plots of reported sensitivities and specificities against study size can be used to evaluate this is such cases. It is dependant on the assumption that the large studies will be near the average and the small studies will be fairly equally dispersed around the larger studies. Therefore a symmetrical plot would be reassuring, whereas an asymmetric plot would suggest the presence of publication bias. (Sterne, 2011) Sourcing unpublished studies would help in reducing this type of bias. Verification bias is seen in a study when a new test is being investigated and the result of this new test determines if the reference test is performed (the gold standard test). Such verification bias can be adjusted for by using an alternative second reference test though it is likely to be of inferior quality and it may define the outcome differently to the original reference test. (de Groot, 2011) These 2 tests must not be considered equivalent though and this must also be adjusted for to avoid differential-verification bias.

It is best if 2 independent authors do data extraction, as this helps to avoid errors. A standardised proforma is useful to ensure the correct standards are checked and that the all the necessary information is documented. Usual data extracted include sensitivity, specificity, prevalence of the disease in the study population, patient characteristics, confidence intervals, odds ratios, relative risks, standard errors and data for creation of the 2 x 2 contingency tables.

Study methods must be scrutinised to ensure the study was carried out in a way that is suitable to allow it to be included in the meta-analysis and appropriate attempts were made to reduce bias. Outcome measures must be standardised if not already consistent between studies.

A forest plot is a graphical display of all the studies included in a meta-analysis or

systematic review. It demonstrates the relative strength of each study, the range of its confidence intervals and its relation to the average or the line of no effect. For accuracy studies, the line would represent the pooled average sensitivity or specificity and each study will lie on either side or on the line itself. This range of difference can be calculated and this value indicates statistically how heterogeneous the studies' results are. For studies assessing a treatment the vertical is called the line of no effect and any study whose line crosses this line of no effect suggests that at that confidence interval, the effect is no different from no effect. If the pooled data lies on the line then there is no evidence of effect but if it lies completely to one side or the other, it would demonstrate either a positive or a negative effect. (Sedgwick, 2012b)

Statistical homogeneity exists when the relative risks of the studies are similar in size and when the variation between them is no more than expected when taking samples from the same population—that is, there was minimal variation between them. If this is not the case then there is statistical heterogeneity present. (Sedgwick, 2012a) Cochran's Q Test is a statistical test of heterogeneity and any variation is assumed to be due to differences between studies when sampling from the same population, or alternatively minor differences in the study methodologies. Higgins I^2 statistic is an alternative test and the lower the I^2 test result, the greater the homogeneity of the study results. If >50%, significant heterogeneity is considered to be present. (Sedgwick, 2012a) If the studies are heterogeneous then efforts must be made to examine and assess why that is the case.

Overall, when a meta-analysis is conducted well with use of the appropriate checklist and the included studies are of equally high quality the resulting data should be

usefulness to authors, reviewers, editors, readers, clinicians and decision- and policy-makers.

Part IV – Performing a Randomized Controlled Trial

A randomized controlled trial (RCT) is a type of clinical trial, which is considered the way to find the best evidence for relationship between cause and effect and to determine the cost effectiveness of the treatment.

There are many important features of an RCT that improve its validity compared to other types of trials. Randomization of trial participants aims to ensure that there are no systematic differences whether known or unknown, between the intervention group and the control group (or the group with an alternative intervention) that can affect the outcome. When participants are randomized between 2 groups, it is assumed that the 2 groups are similar and therefore any difference in outcome is due to the intervention. Randomization also avoids selection bias, which may be due to self-selection of selection based on symptoms as this can affect outcomes in the groups.

Blinding is a trial method when either the participants or the trial operator is unaware of which group the participant is in. It is called single blinding when only the participant is unaware and double blinding when both participant and the trial operator is unaware. The benefit of blinding the participant is the avoidance of any impact on reporting of beneficial or side effects because of the knowledge of group allocation and pre-determined perceived views of the groups. Meanwhile the benefit of blinding the trial operator is avoiding the affect that knowledge of the participant's allocation can have on reporting of complications and outcomes. Blinding avoids bias but it is not always feasible or appropriate.

During the conduct of the trial, all intervention groups should be treated identically except for the experimental treatment, as this will ensure that any differences in outcome are due to the experimental treatment and not due to any other differences.

Patients are normally analysed on an intention to treat basis. This is analysis based on allocation and not based on final intervention received, irrespective of whether they experienced the intended or the alternative intervention.

The analysis is focused on estimating the size of the difference in predefined outcomes between the intervention groups.

Part V – Principles of Ethical Assessment and Approval

The main purpose of an ethical committee's review and approval is to safeguard researchers, participants and future research. Ethical committees base their assessment and decisions on the World Medical Association's Declaration of Helsinki. (World Medical Association, 2013) This declaration discusses the duties of a doctor to their patients and that the clinician's actions should always be in the best interest of the patients as stated in the World Medical Association's Declaration of Geneva and International code of medical ethics. (World Medical Association, 1948; World Medical Association, 1949) The declaration also states that medical progress needs research, which must include patients but that the pursuit of this knowledge and progress must never override the rights of the patients and the conduct of the study must always be morally justifiable.

A favourable ethical opinion allows researchers to conduct their study and publish the results without repeatedly having to defend the ethics of the study. During the ethical review process, questions may be raised that the researchers may not have considered and this can therefore facilitate, promote and champion ethical research. By obtaining ethical approval from an impartial committee that has robust systems in place, all ensure that any research conducted maintains a high ethical standard, is of sound integrity and is in accordance with good research governance and legal requirements.

Clinicians have a duty to consider the ethical aspects of their study design and to ensure the health, dignity, safety, rights and wellbeing of all participants are maintained. There are many aspects to this, which will be detailed below.

The public must also be reassured that their interests are being guarded as this will help avoid potential participants declining to take part in future studies due to fear of being exploited. This therefore protects future research.

Factors considered in an application for ethical approval are many.

How much any new treatment or test deviate from the current, normal, accepted, local or national guidance or clinical practice will be noted. RCTs have raised ethical issues about patients sacrificing themselves for the good of future patients and future society. The concern about this is reduced if the trial employs the principle of clinical equipoise, which is the belief that there must be no decisive evidence that either arm of the trial is superior to the next. This allows the trial to begin with a null hypothesis. (Edwards, 1998) This suggests that placebo-controlled trials are ethically wrong unless doing nothing is an equitable safe option.

The evidence behind the research question must be presented to ensure that such a study is needed and that the question has not already been adequately answered by previous research. The scientific merit or value of the study will be assessed. The study should be designed so that it can answer the research question and address the study aim.

The method of recruitment is examined to ensure there is no evidence of planned coercing. It must be clear how potential participants will be identified, approached and recruited and who are the persons performing this role. Any advertising or use of posters must be factually correct, restrained in tone and must not overemphasize any

inducements or payments for participating. Suitability of recruitment of vulnerable persons will be investigated to decide if they should be excluded. Non-English speakers should be offered professional translation services to aid communication rather than excluding them or using a relative.

The potential risks (likelihood of occurrence and severity) and the burden (for example blood tests, number of hospital visits) will be weighed against the potential benefits to the participant (new treatment or increased surveillance). The importance of the objective outweighs the potential risks and burden to the participant. Potential benefits to society, healthcare and future patients will be noted.

The planned procedure for gaining informed consent will be considered. This will scrutinise use of an information leaflet and the consent form itself. Information must be given using lay terms and making sure potential participants are aware that it is completely voluntary, does not affect their future medical care and that they can opt out at any time without jeopardising their treatment. Potential participants should be made aware of the name and contact details of the point person, the possible risks and the complaint procedure.

Patients must be assured of confidentiality in care of their data. Persons with access to participants' data should be kept to a minimum and their details should be anonymized and kept in a safe place such as a locked drawer or on password protected electronic storage. Participants' general practitioners must be informed of the patient's participation in a study, as they are often a patient's first port of call for ailments. This information would therefore better be able to help the GP treat or give guidance to their

patient.

All potential conflicts of interest should be declared as if a researcher stands to gain financially or otherwise, it could bias their conduct of the study. Source of funding should also be declared as if the funding source should also be unbiased.

Peer review of the study protocol outside of the research group is encouraged including laypersons or patients who could be potential participants as they can provide insight to enhance and improve the protocol. Insurance/indemnity arrangements must be made clear so that should any participant be harmed, it is clear what arrangements are in place for them. Authors must ensure honest reporting of results.

Part VI – Aim

Aims of this thesis and the studies are:

- To evaluate current ultrasound tools used in the screening or the diagnosis of ovarian tumours.
- To assess the accuracy of the IOTA (International Ovarian Tumour Analysis) Logistic Regression models LR1 and LR2 at diagnosing ovarian cancer in the hands of a level II ultrasound operator.
- To evaluate the use of the IOTA (International Ovarian Tumour Analysis) Logistic Regression models LR1 and LR2 in an outpatient clinical setting including those with expectant management.
- To assess the accuracy of the IOTA (International Ovarian Tumour Analysis) Simple Rules in the hands of a level II ultrasound operator to diagnose ovarian cancer.
- To perform a meta-analysis to evaluate the overall accuracy of the IOTA (International Ovarian Tumour Analysis) Simple Rules in the diagnosis of ovarian cancer from published studies and available data.
- To evaluate the use of the IOTA (International Ovarian Tumour Analysis) Simple Rules in an outpatient clinical setting including those with expectant management.
- To test the hypothesis that use of the simple rules can reduce operative rates for women with benign disease when compared with the current Royal College of Obstetricians and Gynaecologists (RCOG) guidance with the use of RMI with a randomized controlled trial.

Part VII – Studies – Materials and Methods

7.1 Setting and Timing

All studies were conducted in the Gynaecological Diagnostic Outpatient Treatment Unit (GDOTU), which is the general gynaecology clinic at University College Hospital in London. Studies were conducted between May 2009 and May 2014.

7.2 Patient population

The University College Hospital, which is situated in the north of central London, is a teaching hospital and tertiary referral centre which primarily provides care for the people of Camden and Islington. It forms part of the University College London Hospitals NHS Foundation Trust (UCLH). The hospital was founded as the 'North London Hospital' in 1834. The name changed to UCH in 1837 and after further collaborations and incorporations of other hospitals, the trust gained foundation status in 2004.

In 2012/13 the trust had an annual turnover of £840 million with contracts with over 70 clinical commissioning bodies. There were 870,000 outpatients, 120,000 A&E admissions and 150,000 inpatients that year. There are 665 in-patient beds and 12 operating theatres.

The department of Women's Health offers a range of specialist clinics including the GDOTU a gynaecological scanning clinic, a walk-in early pregnancy clinic, ambulatory gynaecology, urogynaecology, colposcopy, menopause, psychosexual health, gynaecological oncology, paediatric gynaecology, a specialist endometriosis centre, an assisted conception unit and a specialist clinic for African women.

Patients recruited to the studies in this thesis were all seen in the gynaecological scanning clinic and were referred by their general practitioner (GP) or came via A&E, other hospital clinics or they were tertiary referrals from other hospitals requesting a second opinion.

7.3 Ultrasound scans

Ultrasound scans were performed on a General Electric Healthcare Voluson E8 ultrasound scan machine. (© 2012 General Electric Company, Milwaukee, WI, USA) Transvaginal scans were done using a 4 – 9 MHz Wide Band Endocavity Volume Transducer with three dimensional capability while the transabdominal scans were done with the Wide Band Convex Transducer of 2 – 8 MHz. Transvaginal scans were performed in lithotomy using a standardised protocol and transabdominal scans were performed with the patient supine. Transabdominal scans were performed when the mass in question could not be seen in its entirety using the transvaginal probe. The adnexae were always examined after the uterus (if present) and in both the longitudinal and transverse planes.

During ultrasound of the adnexa, detailed systematic examination and

documentation of its morphological features was conducted including details relevant to the RMI, the IOTA LR1/LR2 and the simple rules. Examination by an expert level III ultrasound operator performing pattern recognition was executed subsequently.

7.4 Ultrasound Operators

Gynaecologists who had a special interest in gynaecological ultrasonography, who had more than 10 years' experience in gynaecological ultrasonography, who accepted tertiary referrals, who participated in multidisciplinary meeting, who audited their practice, who were recognised by the RCOG for teaching and training in gynaecological ultrasonography and who had been involved in research in this field, did level III ultrasonography in keeping with the guidelines of the European Federation of Societies for Ultrasonography in Medicine and Biology. (EFSUMB, 2006)

Ultrasound operators with at least one year of experience, who were trained in gynaecological ultrasonography and who routinely assessed patients with adnexal tumours, did level II ultrasonography. I did all primary ultrasound scans except for the 2nd half of the RCT.

7.5 Serum CA 125 measurements

Women who had a RMI assessment in any of the studies had a blood sample drawn for a CA 125 serum measurement.

The laboratory used the Roche Cobas Analyser Generation II (Cobas® CA 125 II, Roche Diagnostics GmbH, Sandhofer Strasse 116, D-68305 Mannheim) with a measuring range of 0.600-5000 kU/L (defined by the lower detection limit and the maximum of the master curve) with a further capability to dilute 5 x to measure up to a level of 25,000 kU/L.

7.6 Randomization

In the randomized controlled study, randomization took place in the GDOTU. Consecutively numbered, opaque, sealed envelopes were prepared at UCH by a nurse not involved in recruitment and the blocked randomization list was created using a computer-generated randomization sequence in Stata 12.1 (Stata Corp., College Station, Texas, USA) conducted by a hospital statistician and using varying block sizes. The envelopes were then kept securely in a box in a locked cabinet in a key-coded consultants' office. The nursing staff distributed the envelopes to the clinicians who were responsible for randomization and recruitment.

All eligible patients were given an information leaflet, in addition to, as much time they needed to make a decision. After questions were answered, those who wished to participate gave written informed consent. Once this was done, the next envelope was withdrawn and given to the researcher. Once opened, the patient was assessed according to the ultrasound assessment protocol within the envelope.

7.7 Histopathology and Staging

A trust histopathologist examined all tissue and fluid samples retrieved at surgery. Tumours were classified according to the World Health Organisation (WHO) guidelines and ovarian malignancies were staged according to the classification of the International Federation of Gynaecology and Obstetrics (FIGO). (Serov, 1973; Shepherd, 1989)

7.8 Statistical Analysis

Microsoft Excel (Redmond, Washington, USA) was used to establish the databases and for data entry. Statistical analyses were carried out using the software package Stata 11.1 ® to 14.0 ® (Stata Corp., College Station, TX, USA).

The required sample size was calculated before each study. A validation dataset should contain at least 100 events (borderline / malignant tumours) according to Harrell (2001) or it should satisfy the "rule of 10" according to Machin and Campbell (2005). These criteria were met for the validation studies. For the diagnostic accuracy studies, sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio (LR +ve), negative likelihood ratio (LR –ve) and the area under the receiver operating characteristic (ROC) curve (AUC) were calculated. The Chi-squared ($\chi 2$) method was used to compare difference in proportions between original data and current studies. A p value < 0.05 was considered statistically significant and exact 95% confidence intervals were calculated.

For the meta-analysis the sensitivity, specificity and accuracy of simple rules and pattern recognition were compared separately using McNemar's test for

paired binary outcomes. Random effects meta-analysis was used to calculate univariate pooled estimates of sensitivity and specificity for the IOTA simple rules tool. (DerSimonian, 1986)A bivariate approach was also investigated to calculate these values. (Harbord, 2007) Forest plots were constructed to summarise the results, and heterogeneity was quantified using the I^2 statistic. (Thompson, 2002a) Meta-regression was used to investigate any heterogeneity present in the results. (Thompson, 2002b)

For the randomized controlled trial the data was collected on a predesigned proforma, which is then filed. This is transferred to a Microsoft Excel database in preparation for analysis. A Chi-squared test (χ^2) test or Fisher's exact test was used to assess significant differences in the intervention rates between the two management protocols. Analysis was on an intention-to-treat basis of all patients who fit the inclusion criteria.

7.9 Ethical Committee Approval

As morphological analysis of the tumours using IOTA characteristics is performed routinely in the Department and no therapeutic decisions were based on the results of the IOTA and simple rules models, the UCH local Research and Development Department advised that the validation studies did not require formal ethical assessment and approval.

The North London Research Ethical Committee 2 ethics committee approved the randomized controlled trial (10/H0724/48) as well as the research and development committee at UCH and it was entered in the registry of randomized trials (ISRCTN89034131).

Part VIII – Studies – Introduction, Methods and Results

8.1 Study 1 - A prospective validation of the IOTA Logistic Regression Models (LR1 and LR2) in comparison to Subjective Pattern Recognition for the diagnosis of ovarian cancer when utilised by an average ultrasound operator assessing adnexal lesions

8.1.1 Introduction

Preoperative differentiation between benign and malignant ovarian tumours is difficult. A large number of scoring systems and diagnostic models have been developed in recent years in order to facilitate the detection of ovarian cancer on ultrasound. (Jacobs, 1990; Tailor, 1997; Alcazar, 1998) The majority of these tests have been developed based on retrospective analysis of small datasets and their accuracy tends to be poor when evaluated prospectively. (Aslam, 2000; Valentin, 2001) These findings indicate further work is required to improve pre-operative diagnosis of ovarian cancer by ultrasound.

The main strengths of the IOTA collaboration are a uniform approach to ultrasound assessment of adnexal lesions, a large dataset and the use of robust statistical methods. The main aim of the collaboration was to design models for the diagnosis of ovarian cancer, which could be used in routine clinical practice by non-expert operators of average skill and experience. Sonographers who are usually classified as Level II ultrasound operators carry out the majority of routine gynaecological ultrasound

examinations worldwide. (EFSUMB, 2006; EFSUMB, 2010; Salvesen, 2011) Level II examiners tend to describe morphological appearances of adenxal tumours in detail, but unlike the experts (Level III operators), they are not trained to differentiate subjectively between benign and malignant tumours on ultrasound. (Salvesen, 2011)

At present the subjective assessment of adnexal tumour morphology or "pattern recognition" method is the most accurate way to diagnose ovarian cancer on ultrasound scan. (Valentin, 2004) It has been hypothesised that a well-designed and accurate diagnostic model would help Level II sonographers and other examiners without particular expertise in gynaecological ultrasound to differentiate between benign and malignant adnexal tumours on ultrasound scan. Although, the initial published results with the IOTA models were promising, the models have been developed and subsequently tested by expert ultrasound operators (Level III), who were also able to use the "pattern recognition" method to determine the nature of adnexal lesions. (Timmerman, 2005; Yazbek, 2008; Van Holsbeke, 2010c) This could have positively contributed to the accuracy of the IOTA diagnostic models. It is therefore unknown whether the IOTA models would perform equally well if used by operators with less expertise in gynaecological ultrasound.

The aim of this study was to prospectively evaluate the diagnostic accuracy of the IOTA Logistic Regression Models (LR1 and LR2) on a representative sample of adnexal lesions when a non-expert ultrasound operator performed all ultrasound examinations.

8.1.2 Methods

This was a prospective single centre study, which was conducted initially over a 17-month period from May 2009 to September 2010 and then extended to January 2012 for a total of 33 months. All women attending the UCH Gynaecological Diagnostic Unit routinely undergo a detailed transvaginal and transabdominal scan, which included a systematic examination of the uterus, ovaries, adnexa and pouch of Douglas.

Women with ultrasound evidence of an adnexal tumour were assessed according to the IOTA protocol. The ultrasound examinations were all performed by myself, a single level II ultrasound operator (EFSUMB, 2006; EFSUMB, 2010; Salvesen, 2011), who received training in the systematic examination of ovarian tumours in accordance with the IOTA guidelines. (Timmerman, 2000) I did not receive training in tumour pattern recognition and I was discouraged from attempting to differentiate subjectively between benign and malignant tumours on ultrasound scan.

Demographic data including the patient's age, menopausal status as well as personal, medical and family history were all recorded as a part of routine assessment. Women ≥ 50 years of age who previously had a hysterectomy were defined as postmenopausal as is the standard in literature. In addition, morphological and Doppler characteristics of adnexal tumours according to the IOTA protocol were also recorded. (Timmerman, 2005) A family history that included the number of first-degree relatives with ovarian or breast cancer was taken from each patient routinely.

As with the original IOTA study (Timmerman, 2005), pregnant women, those unable to undergo a transvaginal scan and those who had surgery in excess of 120 days after the ultrasound scan, were excluded from the final data analysis. In women with bilateral

lesions, the lesion, which was more likely to be malignant according to the IOTA model, was included in the analysis.

The probability of an adnexal mass being malignant was estimated by using both the IOTA Logistic Regression Model (LR1 and LR2). Twelve variables were used for the LR1 calculation (1) personal history of ovarian cancer (yes=1, no=0): (2) current use of hormonal therapy (yes=1, no=0): (3) age of the patient (years): (4) maximum diameter of lesion (mm): (5) presence of pain during the examination (yes=1, no=0): (6) presence of ascites (yes=1, no=0): (7) presence of blood flow within a solid papillary projection (yes=1, no=0): (8) purely solid tumour: (9) maximum diameter of the solid component (expressed in millimetres, but with no increase > 50 mm): (10) irregular internal cyst walls (yes=1, no=0): (11) presence of acoustic shadows (yes=1, no=0): (12) colour flow score (1-4 where 1 is no flow and 4 is maximum flow). Six variables were used for LR2 [variable (3), (6), (7), (9), (10) and (11)]. The probability of malignancy was calculated using the formula described above in the section on logistic regressions models. The results of the calculations were stored in the research file and it was not made available to clinicians who were making decisions about the management.

A consultant expert operator then managed all the patients expectantly or surgically using the unit's standard clinical protocols and based on their assessment of the patient. The expert operator was blinded to the IOTA assessments and scores. Only women who underwent surgery were included in data analysis. Surgical options varied from laparoscopic ovarian cystectomy to primary ovarian debulking surgery. The accuracy of the IOTA models were assessed using histology as the gold standard. A histopathologist examined all specimens and was blinded to the IOTA scores. Tumours were classified

according to the World Health Organisation (WHO) guidelines and ovarian malignancies according to the criteria recommended by the International Federation of Gynecology and Obstetrics. (Serov, 1973; Benedet, 2000)

As morphological analysis of the tumours using IOTA protocol is performed routinely in the Department and no therapeutic decisions were based on the results of the IOTA models, the UCH local Research and Development Department advised that the study did not require formal ethical assessment and approval.

8.1.3 Statistical Analysis and Results

8.1.3.1. Statistical Techniques

The 17 month LR2 statistical analysis was performed using the software package Stata 11.1 ® (Stata Corp., College Station, TX, USA) while Stata 12.1 ® was used for the 33 month analysis. The diagnostic accuracy of the IOTA Logistic Regression Model (LR2) was assessed by calculating its sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio (LR +ve), negative likelihood ratio (LR –ve) and the area under the receiver operating characteristic (ROC) curve (AUC). The Chi-squared ($\chi 2$) method was used to compare difference in proportions between the results of original IOTA and current study.

The sample size was calculated prior to commencement. Harrell (2001) suggests that the validation dataset should contain at least 100 events (borderline / malignant

tumours), whereas Machin and Campbell (2005) suggest that the validation dataset should satisfy the "rule of 10" guideline, which means that in this study, the dataset should have at least 120 events as the LR1 model contains 12 predictors and at least 60 events for LR2. This validation dataset had 132 events, therefore satisfying both requirements.

8.1.3.2. Results

8.1.3.2.1. Results at 17 months

A total of 332 women were diagnosed with adnexal tumours during the initial 17-month study period. Of these, 141/332 (42.5%) women had surgery or a biopsy; whilst the remaining 191 of 332 (57.5%) were managed conservatively. 17/141 (12.1%) women were excluded. Three were pregnant and 14 did not have surgery within 120 days of ultrasound examination. A total of 124 women were included in the final analysis. (Figure 8.1-1) The mean age of the study population was 53.2 years (range 20 – 91 years). There were 63 (50.8%) premenopausal and 61 (49.2%) postmenopausal women. There were a range of referral sources; 46 (37.1%) were referred by their general practitioner, 68 (54.8%) were tertiary referrals from another unit, 8 (6.5%) attended as an emergency and 2 (1.6%) were referred via other routes. There were 58 (46.8%, 95% CI 38.2-55.5) benign, 9 (7.3%, 95% CI 3.9-13.2) borderline, 42 (33.9%, 95% CI 26.1-42.6) primary invasive malignant adnexal tumours and 15 (12.0%, 95% CI 7.5-19) metastatic tumours to the ovary. Of the 42 primary invasive malignant lesions there were 14 (33.3 %) Stage I, 3 (7.1%) Stage II, 13 (31%) Stage III and 12 (28.6%) Stage

IV tumours. Primary invasive epithelial, borderline and metastatic tumours were grouped together for the purpose of data analysis as in the original study so there were 66 (53.2%) malignant lesions in total.

The LR2 IOTA model had a sensitivity of 97.0% (95% CI, 89.5 to 99.6) and a specificity of 69.0% (95% CI, 55.5 to 80.5). The positive predictive value was 78.0% (95% CI 67.5 to 86.4) and the negative predictive value was 95.2% (95% CI 83.8 to 99.4). The positive likelihood ratio was 3.125 (95% CI, 2.124 – 4.597) and the negative likelihood ratio was 0.044 (95% CI, 0.011 - 0.174). The area under the ROC curve (AUC) was 0.93, standard error 0.022 (95% CI, 0.89 – 0.97).

There were 18 false positive and 2 false negative cases (Table 8.1-1). Benign ovarian cystadenomas and mature cystic teratomas were responsible for 55.6% (10/18) of all false positive diagnoses of ovarian cancer. The first false negative case was a premenopausal woman with large bilateral ovarian tumours whose histology showed a borderline mucinous tumour and an incidental finding of an appendix Goblet cell carcinoid tumour. The second patient was a postmenopausal woman, also with bilateral ovarian tumours, which were metastatic from a gastrointestinal primary. The tumours were smooth and multilocular with no solid areas and no ascites. There were tumour deposits seen though in the Pouch of Douglas but that factor is not tested in the IOTA models.

In order to test for the possibility of bias due to increased operator experience during the study period, the dataset was divided in two halves. The performance of the IOTA Logistic Regression Model (LR2) in the first 62 consecutive patients was not

significantly different form the results in the 62 subsequently recruited women [sensitivity 94% versus 100% (p=0.152); specificity 61% versus 78% (p=0.063)].

The AUC were not significantly different between this current study, the original IOTA report and the prospective IOTA validation study. (Table 8.1-2) (Timmerman, 2005; Timmerman, 2010) Sensitivity in this study was significantly better in comparison to the original study ($\chi 2=6.162$, p=0.013), but it was not statistically different to the sensitivities in the validation study. Specificities in the IOTA validation study were significantly higher compared to this study (external $\chi 2=21.1$, p=0.001; temporal $\chi 2=6.96$,p=0.008). The specificities in this and the original study, however, were not significantly different. Alternative cut-off points were investigated for this study's data. A cut-off of 6.4% instead of 10% would give us 100% sensitivity but the specificity would fall to 56.9%. (Figure 8.1-2)

Figure 8.1-1: Flow Diagram of eligibility for LR2. N = 124. (Nunes, 2012a)

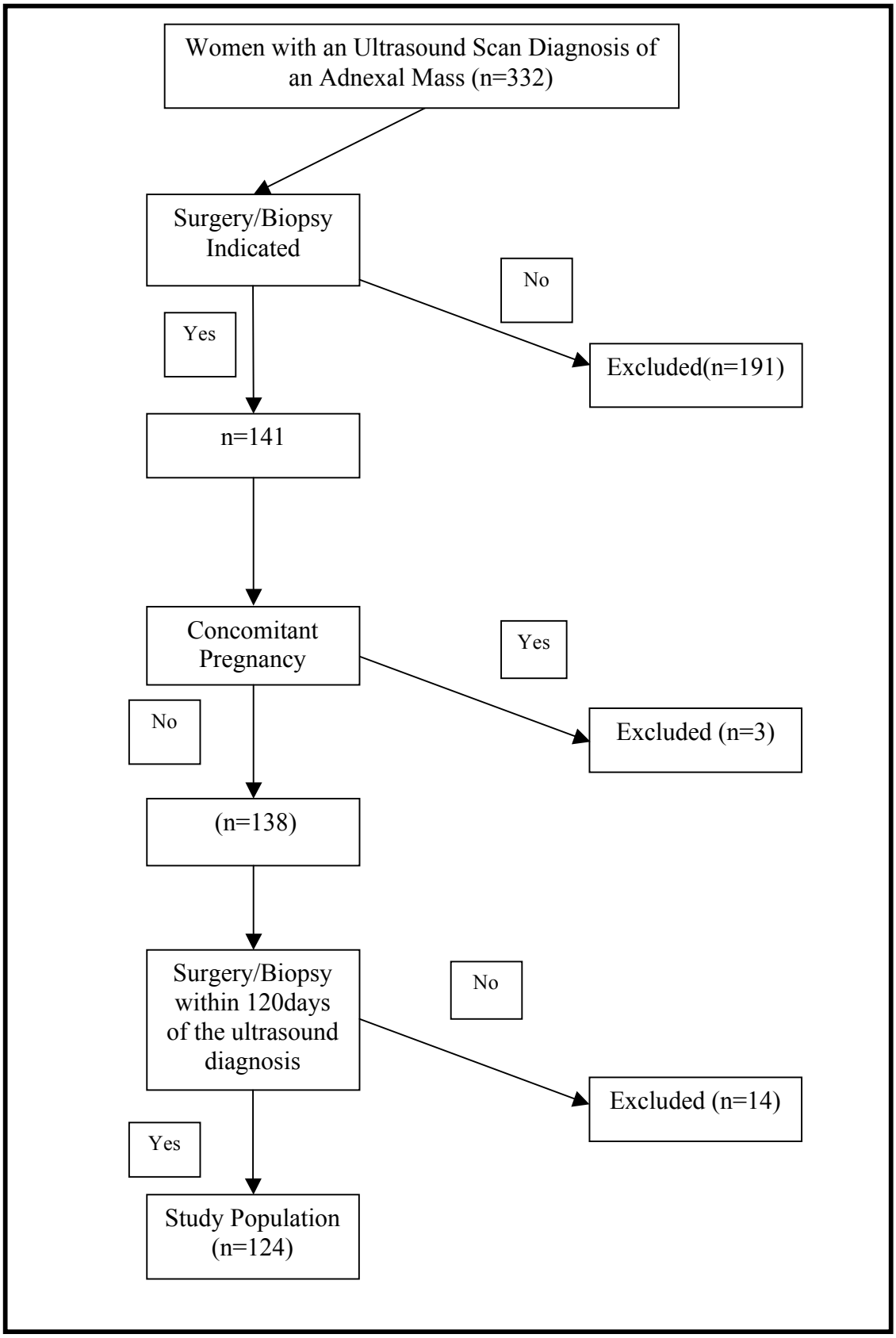


Table 8.1-1: Histological findings in women with false positive diagnoses of ovarian cancer using the IOTA logistic regression model LR2. (N = 18 of total 124) (Nunes, 2012a)

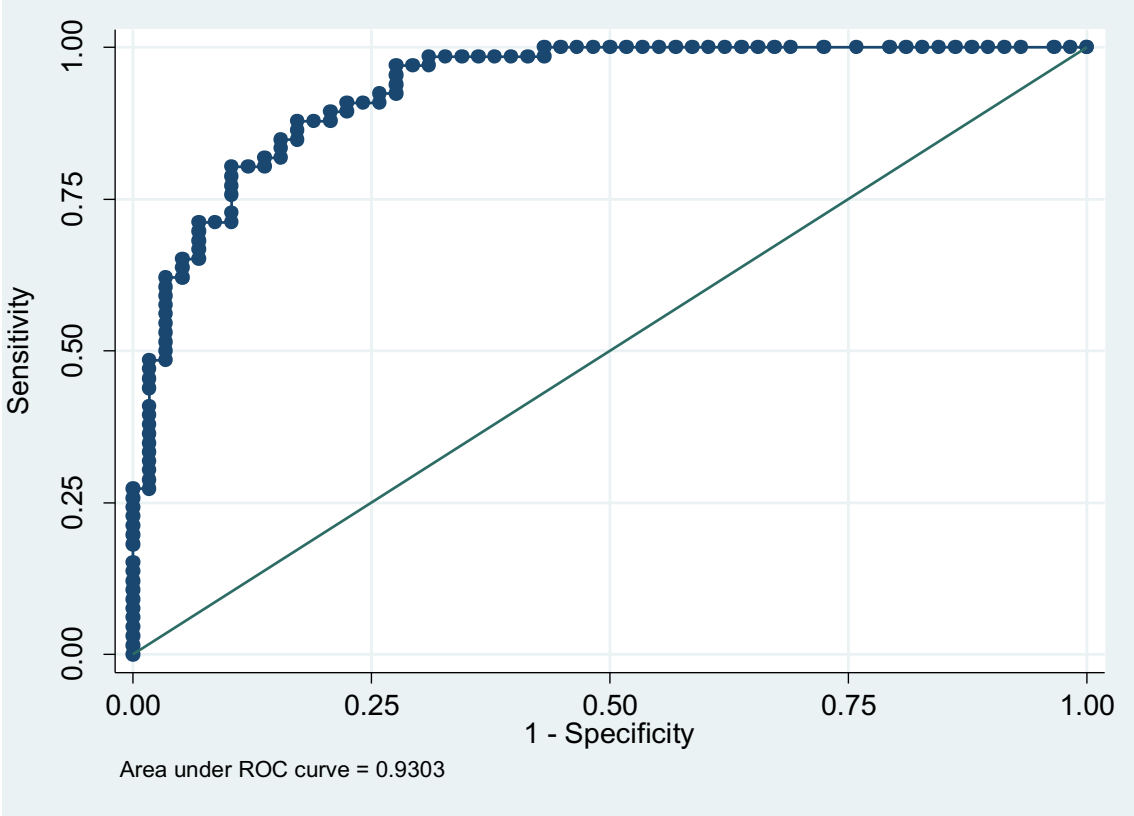
Histological diagnosis	N (%)
Cystadenoma/cystadenofibroma	5 (27.8)
Mature cystic teratoma	4 (22.2)
Dermoid with cystadenoma	1 (5.6)
Pedunculated leiomyoma	1 (5.6)
Benign aspirate/Pseudocyst	2 (11.1)
Torsion of benign cyst	2 (11.1)
Brenner tumour	1 (5.6)
Endometriosis	1 (5.6)
Fat necrosis and inflammation - suspected actinomycosis	1 (5.6)
Total	18 (100)

Table 8.1-2: Accuracy of the IOTA Logistic Regression Model (LR2): comparison of findings in the original (Timmerman, 2005) and prospective validation (Timmerman, 2010) IOTA studies with the current study. N = 124. (Nunes, 2012a)

	AUC (95% CI)	SE	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ve (95% CI)	LR-ve (95% CI)
Original IOTA study (Timmerman, 2005) (test set) N=312	0.92 -	0.018	89 -	73 -	3.3 -	0.15 -
Prospective validation IOTA study (Timmerman, 2010) External N=997	0.95 (0.93 – 0.96)	-	91.8 -	85.6 -	6.36 (5.33 – 7.63)	0.10 (0.06 – 0.14)
Temporal N=941	0.92 (0.90 – 0.94)	-	89.2 -	79.8 -	4.42 (3.78 – 5.19)	0.14 (0.1 – 0.19)

Current study N=124	0.93 (0.89 – 0.97)	0.022	97 (92.3 – 98.9)	69 (60.4 – 76.5)	3.12 (2.12 – 4.6)	0.044 (0.01 – 0.17)

Figure 8.1-2: The receiver operating characteristic curves of the IOTA Logistic Regression Model (LR2). N = 124. (Nunes, 2012a)



8.1.3.2.2. Results at 33 months

There were 555 women with adnexal tumours who underwent pelvic ultrasound examination during the 33-month study period. Of those 555, 317 (57.1%) women had surgical intervention and the remaining 238 (42.9%) were managed conservatively. There were 25 women who were excluded from the study: 21 women had surgery >120 days after ultrasound assessment and four women were pregnant.

There were therefore 292 women included in the final analysis. (Figure 8.1-3) The mean age of the 292 women included in the study was 51 years (range of 16 to 91). Of those 292 women, 155 (53.1%) were premenopausal and 137 (46.9%) were

postmenopausal. One hundred and sixty (54.8%) of the tumours were benign, 17 (5.8%) were borderline and 115 (39.4%) were invasive malignant tumours. The mean age of those with benign tumours was 43 years, which was significantly less than the mean age of 59 in those with malignant tumours ($P < 0.0001$). Of the 115 invasive malignant tumours, there were 25 (21.7%) Stage I, 5 (4.3%) Stage II, 34 (29.6%) Stage III, 22 (19.1%) Stage IV, 25 (21.7%) metastatic and 4 (3.5%) recurrent ovarian tumours. Borderline, primary invasive and metastatic tumours were analysed together in keeping with the original article and they accounted for a total of 132/292 (45.2%) of all tumours.

Table 8.1-3 shows the diagnostic performances of LR1, LR2 and PR in this and previously published studies. There were no significant differences in AUC between the previous studies and this study. Despite this, the sensitivity of LR1 was significantly higher and the specificity was significantly lower in our hands compared to the external and temporal validation studies carried out by the IOTA group. (Table 8.1-4) There were also significant differences in the sensitivity and specificity of LR2 between this and previous validation studies. (Table 8.1-5) We compared the performance of the models and the pattern recognition in the first half of the study with that in the second and we found no significant differences in the AUC. (Table 8.1-6)

There were 47 false positives and 3 false negatives with the LR1 model and 43 false positives and 5 false negatives with the LR2 model. (Table 8.1-7) Pattern recognition had only 11 false positives and 6 false negatives giving it the highest diagnostic accuracy of the three tests. The accuracy of PR was 0.942 (95%CI, 0.908 to 0.966) compared to that of LR1, which was 0.829 (95%CI, 0.781 to 0.870) and to that of LR2, which was 0.836 (95%CI, 0.788 to 0.872). This was statistically significant (LR1 versus

PR: $P < 0.001$ and LR2 versus PR: $P < 0.001$).

Cystadenomas and cystadenofibromas were the most commonly found tumours which were false positives in this study (15/47 (32%) LR1, 14/47 (30%) LR2, 6/47 (13%) PR). Mature cystic teratomas (Dermoid cysts) were the next single most common cause of false positives [5 of 24 (LR1), 6 of 24 (LR2), 1 of 24 (PR)]. These were though also the most common benign tumours found. The rare benign tumours, such as actinomycosis, myolipoma, Brenner tumour and struma ovarii were all misdiagnosed as cancers by both LR1 and LR2. Pedunculated leiomyomas (3 cases) were all wrongly diagnosed as cancers using LR1 and two out of three were also misdiagnosed by LR2. Four out of five cases of torsion of benign cysts were diagnosed as cancers by both models although the diagnosis was correct in all cases when pattern recognition was used.

The false negative results most commonly occurred in cases of borderline ovarian tumours. Two borderline tumours, which were both misclassified by LR2 and PR presented with large unilocular cyst, one of which contained a 13mm solitary papillary projection. All three diagnostic tests gave false negative results in a case of mucinous adenocarcinoma and a dermoid, which contained a 1.5mm focus of immature tissue.

Figure 8.1-3: Flow Diagram of eligibility LR1 and LR2. N = 292. (Nunes, 2013)

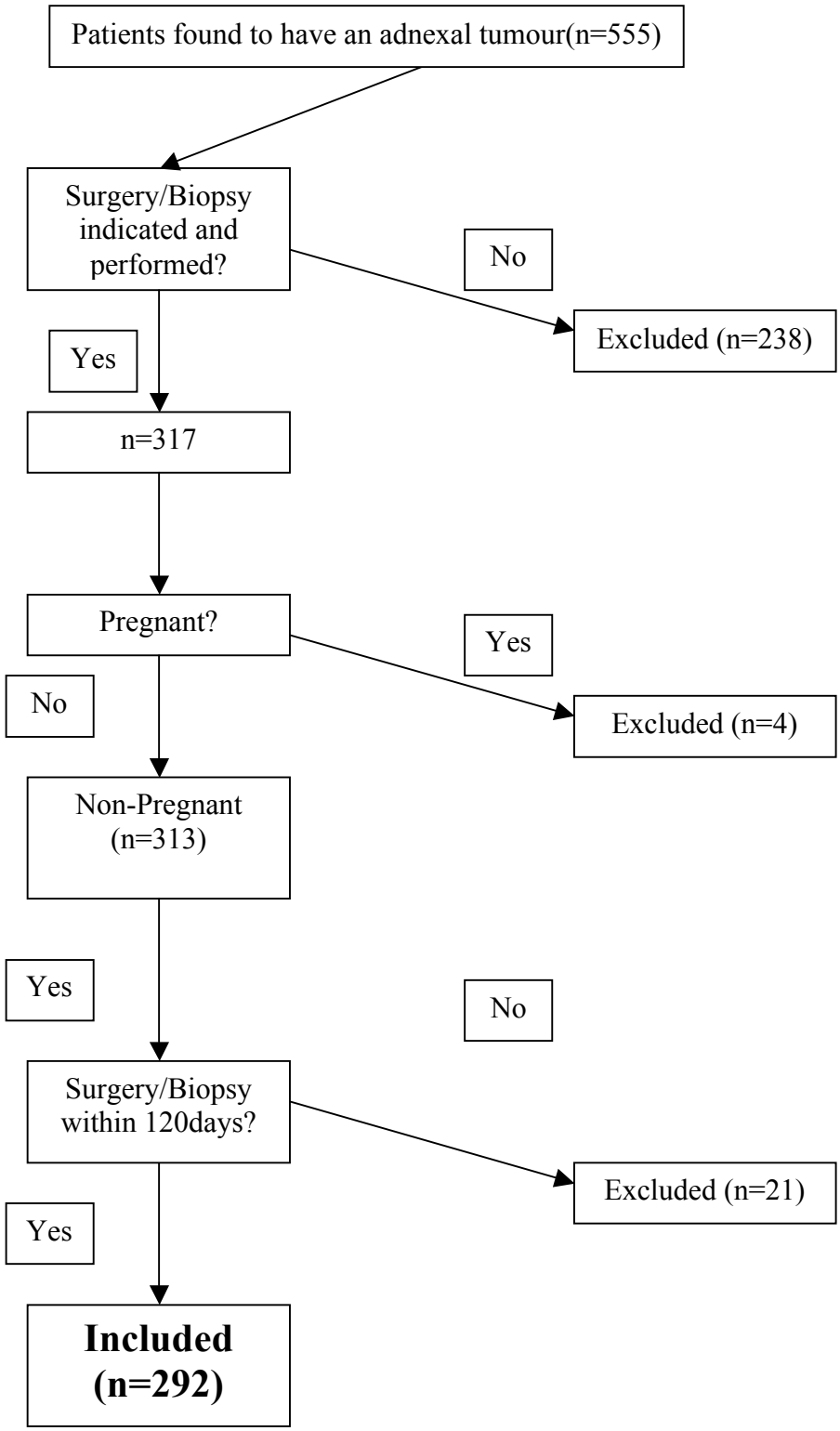


Table 8.1-3: Accuracy of the IOTA logistic regression models (LR1 and LR2): comparison of findings in the original and prospective validation IOTA studies with the current dataset Part A. N = 292. (Nunes, 2013)

LR1							
	AUC (95% CI)	SE	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ve (95%CI)	LR-ve (95%CI)	Accuracy (95%CI)
Original IOTA study (test set) N=312	0.94	0.017	93	76	3.81	0.09	

Prospective validation IOTA study External n=997	0.96 (0.94 – 0.97)	-	92.2 -	86.5 -	6.84 (5.69 – 8.25)	0.09 (0.06 – 0.14)	
Temporal n=941	0.95 (0.93 – 0.96)	-	92.7	80.6	4.77 (4.04 – 5.61)	0.09 (0.06 – 0.14)	
Current Study N=292	0.94 (0.92 – 0.97)	0.013	97.7 (93.6 – 99.4)	70.6 (67.2 – 72.0)	3.33 (2.85 – 3.55)	0.03 (0.01 – 0.10)	0.829 (0.781 – 0.870)

Accuracy of the IOTA logistic regression models (LR1 and LR2): comparison of findings in the original and prospective validation IOTA studies with the current dataset Part B. N = 292. (Nunes, 2013)

LR2							
	AUC (95% CI)	SE	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ve (95%CI)	LR-ve (95%CI)	Accuracy (95%CI)
Original IOTA study (test set) N=312	0.92 -	0.018	89 -	73 -	3.3 -	0.15 -	- -

Prospective validation							
IOTA study	0.95	-	91.8	85.6	6.36	0.10	-
External n=997	(0.93 – 0.96)		-	-	(5.33 – 7.63)	(0.06 – 0.14)	
Temporal n=941	0.92	-	89.2	79.8	4.42	0.14	-
	(0.90 – 0.94)		-	-	(3.78 – 5.19)	(0.1 – 0.19)	
Current Study	0.93	0.015	96.2	73.1	3.58	0.052	0.836
N=292	(0.90 – 0.96)		(91.4 – 98.8)	(65.6 – 79.8)	(2.77 – 4.63)	(0.022 – 0.123)	(0.788 – 0.872)

Accuracy of the IOTA logistic regression models (LR1 and LR2): comparison of findings in the original and prospective validation IOTA studies with the current dataset Part C. N = 292. (Nunes, 2013)

Pattern Recognition							
	AUC (95% CI)	SE	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ve (95%CI)	LR-ve (95%CI)	Accuracy (95%CI)

Prospective validation IOTA study External n=997	0.949 (0.930 – 0.964)	-	87.5	92.1	11.0 (8.6 – 14.1)	0.14 (0.10 – 0.19)	-
Temporal n=941	0.959 (0.944 – 0.973)	-	93.0	93.4	14.1 (10.6 – 19.0)	0.07 (0.05 – 0.11)	-
Current Study N=292	-	-	95.5 (90.4 – 98.3)	93.1 (88.0 – 96.5)	13.9 (7.84 – 24.6)	0.049 (0.022 – 0.107)	0.942 (0.908 – 0.966)

AUC – Area under the Curve; SE – Standard Error; LR+ve – Positive Likelihood Ratio; LR-ve – Negative Likelihood Ratio; CI – Confidence interval

Table 8.1-4: Comparison of sensitivity and specificity between the current dataset and previous datasets for LR1. N = 292. (Nunes, 2013)

Dataset	Sensitivity	Test**	Specificity	Test**
Original (n=312)	70/75	0.14*	179/237	0.28
External (n=997)	235/255	0.028	642/742	< 0.001
Temporal (n=941)	266/287	0.039	527/654	0.006
Current Study (n=292)	129/132	-	113/160	-

* Fisher's exact test (otherwise Chi-squared ($\chi 2$) test)

** With Current Study

Table 8.1-5: Comparison of sensitivity and specificity between the current dataset and previous datasets for LR2. N = 292. (Nunes, 2013)

Dataset	Sensitivity	Test**	Specificity	Test**
Original (n=312)	67/75	0.05	173/237 *	0.98
External (n=997)	234/255	0.098	635/742	< 0.001
Temporal (n=941)	256/287	0.017	522/654	0.065
Current Study (n=292)	127/132	-	117/160	

* Estimate (numerator between 172 – 174)

** With Current Study

Table 8.1-6: Comparison of the two halves of this study (LR1 and LR2). N = 292. (Nunes, 2013)

LR1 Model	*First half n=146*	*Second half n=146*	*P value*
LR1			
Sensitivity	98.6%	96.7%	
Specificity	62.7%	77.6%	0.038
AUC	0.93	0.96	0.34
LR2			
Sensitivity	95.8%	96.7%	
Specificity	68%	77.6%	0.17
AUC	0.91	0.95	0.28
PR			
Sensitivity	94.4%	96.7%	
Specificity	92%	94.1%	0.6

Table 8.1-7: Histological findings in women with false positive diagnoses of ovarian cancer using the IOTA logistic regression models (LR1 and LR2) and Pattern Recognition. N = 292. (Nunes, 2013)

Histological diagnosis	LR1 (N)	LR2 (N)	Pattern Recognition (N)
Cystadenoma	13	11	6
Cystadenofibroma	2	3	
Mature cystic teratoma	5	6	1
Benign simple cyst/Pseudocyst	7	5	1
Pedunculated leiomyoma	3	2	
Torsion of benign cyst	4	4	1
Endometriosis	3	3	1
Fat necrosis and inflammation - suspected actinomycosis	1	1	1
Fibroma	4	3	
Brenner tumor	1	1	
Myolipoma	1	1	
Struma ovarii	1	1	
Fibrothecoma	2	2	
Total	47	43	11

8.2 Study 2 – A prospective validation of the IOTA Simple Rules for the diagnosis of ovarian cancer

8.2.1 Introduction

Non-invasive pre-operative diagnosis of an adnexal tumour is the main goal of ultrasound assessments. Determination of the nature of the tumour is of paramount importance to guide management options. The "Simple Rules" were designed by the International Ovarian Tumour Analysis (IOTA) group collaboration using only ultrasound scan features to determine whether a tumour is benign or malignant.

There are 5 benign features and 5 malignant features, which are used to determine the nature of the tumour. Benign features were B1) Unilocular, B2) Presence of solid components where the largest solid component has a largest diameter < 7 mm, B3) Presence of acoustic shadows, B4) Smooth multilocular tumour with largest diameter < 100 mm and B5) No blood flow (colour score 1). Malignant features were M1) Irregular solid tumour, M2) Presence of ascites, M3) At least four papillary structures, M4) Irregular, multilocular solid tumour with largest diameter ≥ 100 mm and M5) Very strong blood flow (colour score 4). If one or more benign features exist in the absence of any malignant features being present then the tumour is defined as benign. If one or more malignant features exist in the absence of any benign features being present, then the tumour is defined as malignant. If no features exist or if both types of features are present then the tumour is unclassified.

The aim of this study is to assess the diagnostic accuracy of the simple rules protocol when performed by a level II operator to diagnose ovarian cancer. This is an external validation study.

8.2.2 Methods

This prospective single centre study was performed over a 33-month period from May 2009 to January 2012. This is the same population of women as in study 2 and the conduct of the study and the exclusion criteria are as described above for studies 1 and 2. In this case though, all consecutive patients diagnosed with an adnexal tumour were assessed according to the simple rules protocol. Once again the expert was blinded to the simple rules category to which the patient was assigned. This was of particular importance for the patients in whom the rules were not applicable. The level II ultrasound operator was instructed only to describe the various features of the tumour and assign it M or B features according to the simple rules criteria and not to assign a particular diagnosis.

In women with bilateral tumours, both tumours were assessed and categorised. If one was assessed as benign or indeterminate while the other assessed as malignant, then the malignant tumour was used in the final analysis. If one tumour was assessed as benign but the other as indeterminate, then her tumours were categorised as indeterminate and if both tumours fell into the same category then the larger was used.

Surgical intervention options and histological assessment and analysis were as

previously described. The histopathologist was blinded to the result of the simple rules assessment and was only aware of the surgical opinion of the surgeon.

The UCH local Research and Development Department determined that formal ethical assessment and approval was not required as ultrasound scan examination of all adnexal masses was routine practice including morphological analysis of the tumours using the IOTA terms (Timmerman, 2000) and also because therapeutic decisions were not based on the simple rules results.

Guidance on the STARD initiative was followed in the conduct, analysis and reporting of this study. (Bossuyt, 2003)

8.2.3 Statistical Analysis and Results

8.2.3.1. Statistical Techniques

Analysis was performed using software package Stata 11.1 ® (Stata Corp., College Station, TX, USA). A p value < 0.05 was considered statistically significant and exact 95% confidence intervals were calculated. The required sample size was calculated prior to commencement of the study. A validation dataset should contain at least 100 events (borderline / malignant tumours) according to Harrell (2001) or it should satisfy the "rule of 10" according to Machin and Campbell (2005) requiring the same 100 events in this study. These criteria were met. The sensitivity, specificity and accuracy of simple rules, pattern recognition, and simple rules followed by pattern recognition were compared separately using the Chi- squared ($\chi 2$) test for independent binary

outcomes. The sensitivity, specificity and accuracy of simple rules and pattern recognition were compared separately using McNemar's test for paired binary outcomes.

8.2.3.2. Results

There were 555 women who were initially assessed and diagnosed with adnexal tumours. Of those 335 (60.4%) went on to have surgery. There were 4 pregnant women in the surgical group who were excluded as per the original study as well as a further 28 who were excluded because of not having had surgery within the stipulated 120 days from last ultrasound scan. (Figure 8.2-1) 141 (46.5%) of the women were postmenopausal and the average age was 50 (range 16-91). Most of the referral routes (53.1%) were via the general practitioner, accident and emergency or an internal referral. The remainder were referred from other hospitals. There were 11 women with bilateral tumours and one of them had one of her tumours classified as malignant and the other as indeterminate. Her tumour that was classified as malignant was included in the final analysis. Of the histological confirmations, there were 168 (55.4%) benign tumours, 19 (6.3%) borderline tumours, 86 (28.4%) primary invasive ovarian cancers, 1 (0.3%) primary tubal cancer, 22 (7.3%) metastatic tumours, 4 (1.3%) primary non-gynaecological tumours and 3 (1%) recurrent ovarian cancers. Among 86 women with primary invasive ovarian cancers, 27 (31.4%) had stage I disease, 5 (5.8%) stage II, 34 (39.5%) stage III and 20 (23.3%) stage IV.

The rules were applicable in 237 (78.2%) of the 303 tumours. This proportion is in keeping with the original and validation studies. For these women in whom the rules

were applicable, the simple rules had a sensitivity of 96.2% (95%CI: 90.5-99.0) and a specificity of 88.6% (95%CI:82.0-93.5). The accuracy was 92.0% (95%CI:87.8-95.1) (Table 8.2-1) For the same population of women the sensitivity of pattern recognition was 97.1% (95%CI: 91.9-99.4) and the specificity was 93.2% (95%CI:87.5-96.8). The accuracy of pattern recognition was 94.9% (95%CI:91.3-97.4). For those 66 tumours for which the rules were not applicable 36 (55%) were benign. (Figure 8.2-2) For these patients there was a fairly equal distribution of those for who neither M nor B rules applied (34 / 52%) or whether both types of rules applied (32 / 48%). Whether both types of rules or neither types of rules applied, gave no strong indication to whether the tumour was benign or malignant as 56% of the benign tumours and 47% of the malignant tumours classified as indeterminate had no rules applicable. While the prevalence of malignant lesions was similar in women in whom neither or both rules applied [41.2 % versus 50% ($P>0.05$)], these malignancies were more likely to be borderline when both rules applied as opposed to when neither type applied [43.8% versus 7.1% ($p<0.05$)].

The prevalence, as expected was much higher for post-menopausal patients [67.4% (95/141)] than pre-menopausal patients [24.7% (40/162). This discordance was greater in the populations of women in whom the rules were applicable [postmenopausal 71.0% (76/107); premenopausal 22.3% (29/130). Among the 162 premenopausal women the rules were applicable in 80.2% (130/162) with a sensitivity of 89.7% (95%CI:72.6-97.8), a specificity of 89.1% (95%CI:81.3-94.4) and an accuracy of 89.2% (95% CI:82.6-94.0). There were 141 postmenopausal women in our population and the rules were applicable for 75.9% (107/141). For these women, when compared to the premenopausal women, there was an increased sensitivity of 87.1% (95%CI:70.2-96.4), similar specificity of 98.7% (95%CI:92.9-100) and increased accuracy of 95.3%

(95%CI:89.4-98.5). (Table 8.2-1)

When pattern recognition was used on the 66 tumours which could not be classified by the simple rules, the sensitivity was lower though the specificity was much higher (86.7%(95%CI:69.3-96.2) and 94.4% (95%CI:81.3-99.3) respectively, when compared to the performance of the simple rules. Further analysis was done using pattern recognition as the second line test when the rules were not applicable. The final analysis of the entire 303 population had a sensitivity of 94.1% (95%CI:88.7-97.4), a specificity of 89.9% (95%CI:84.3-94), and an accuracy of 91.7% (95%CI:88.1-94.6). When pattern recognition alone was used for all 303 tumours, there was a sensitivity of 94.8% (95%CI:89.6-97.9), a specificity of 93.5% (95%CI:88.6-96.7) and an accuracy of 94.1% (95%CI:90.8-96.4).

If all tumours which were not classifiable by the simple rules were assumed to be malignant the sensitivity increased but the specificity and accuracy decreased: 97% (95%CI:92.6-99.2), 69.6% 95%CI:62.1-76.5), and 81.8% (95%CI:77.0–86.0) respectively.

There were 4 false negative diagnoses and these included a borderline tumour, a dermoid cyst with a 1-2mm area of immature cells, an invasive stage III adenocarcinoma and a primary peritoneal cancer. (Table 8.2-2) There were 15 false positive diagnoses amongst which were found a range of histological diagnoses, as shown in table form. (Table 8.2-3)

The sensitivity, specificity and accuracy of simple rules, pattern recognition, and simple rules followed by pattern recognition were compared separately using the Chi-squared

($\chi 2$) test for independent binary outcomes. (Table 8.2-4) The sensitivity, specificity and accuracy of simple rules and pattern recognition were compared separately using McNemar's test for paired binary outcomes. These analyses are based on the 237 patients for whom the simple rules were applicable. Of these, only 11 patients had differing diagnoses under simple rules and pattern recognition. The results suggest that the specificity of simple rules is inferior to that of pattern recognition. ($p=0.03$) (Table 8.2-5) Comparative analyses between our results and the original and validation IOTA studies showed no significant difference for most of the tests apart from the specificity of simple rules in the validation study ($p< 0.001$). (Timmerman, 2010) (Table 8.2-6)

Figure 8.2-1: Flow Diagram of Study Eligibility for the Simple Rules. N = 303. (Nunes, 2014)

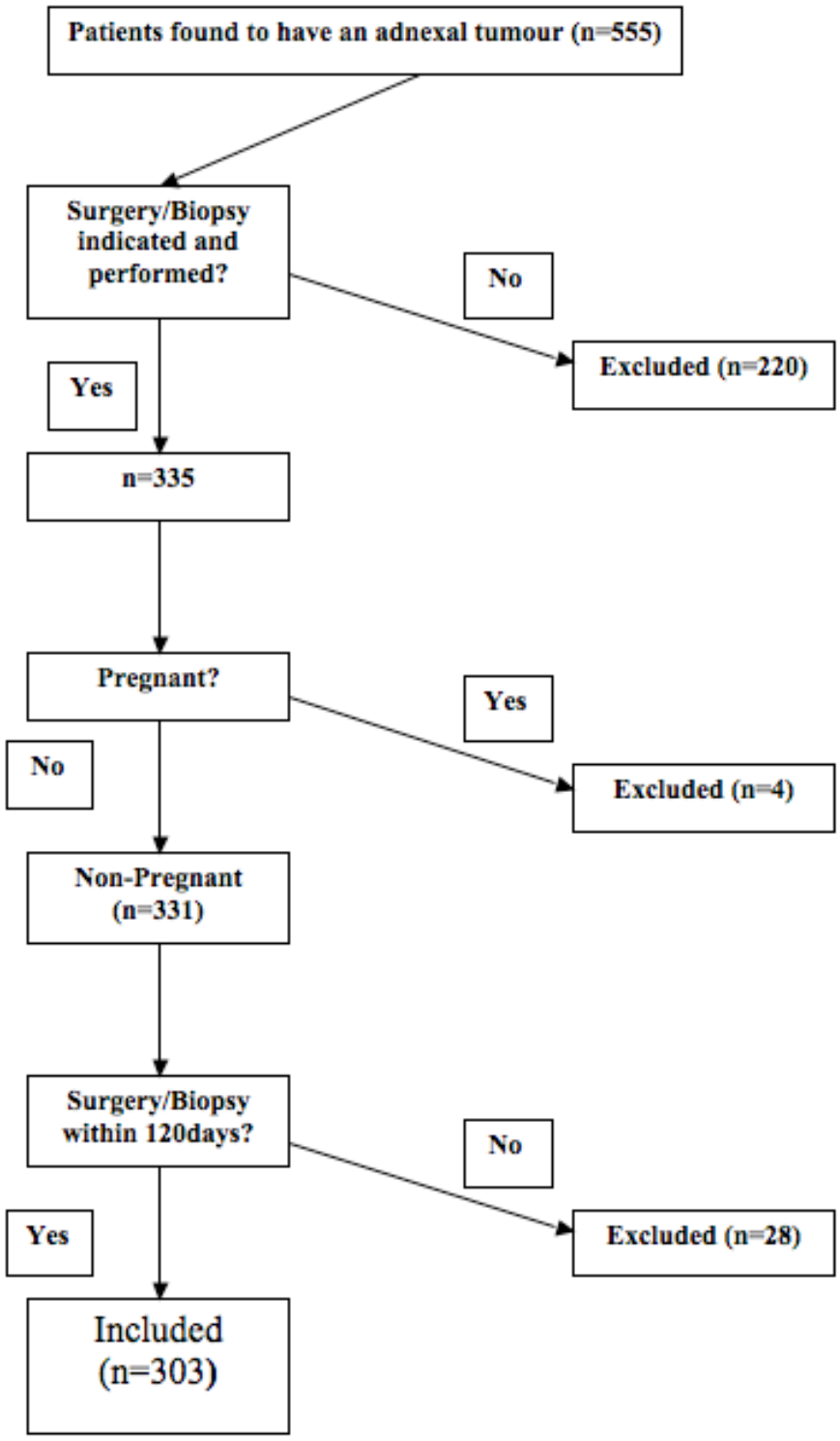


Figure 8.2-2: Flow Diagram of Results for the IOTA Simple Rules. (N = 237 of 303). (Nunes, 2014)

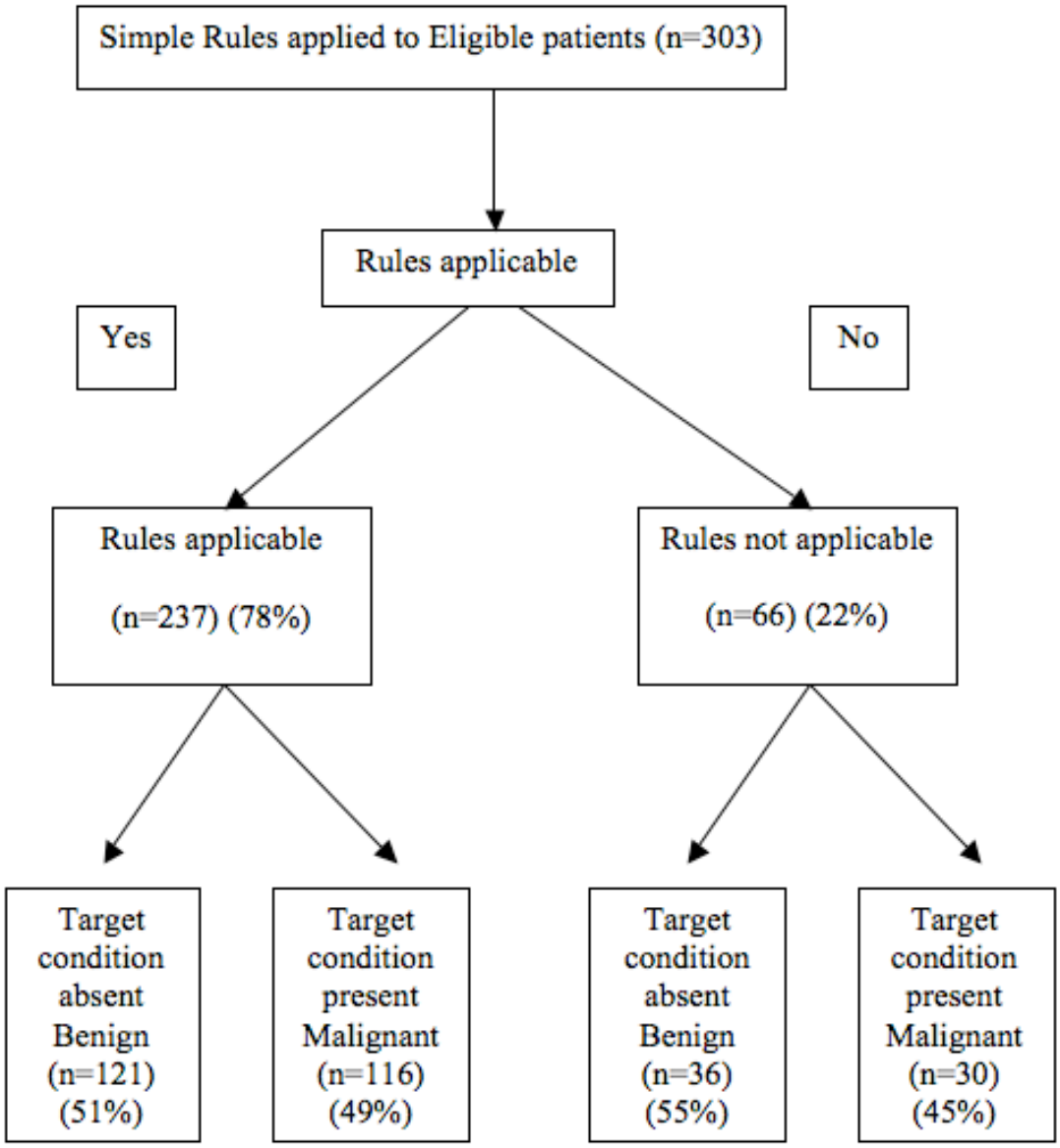


Table 8.2-1: Accuracy of the IOTA Simple Rules: comparison of findings in the original and prospective IOTA validation studies with the current study - Part A. N = 237. (Nunes, 2014)

Simple Rules								
	Rules applicable	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ve (95% CI)	LR-ve (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
Original SR study (test set) N=386	76% 386/ 507	94.6 (88.7-98.0)	90.9 (86.8-94.0)	10.4 (7.1-15.1)	0.059 (0.027-0.129)	80.9 (73.1-87.3)	97.6 (94.9-99.1)	92.0 (88.8-94.5)
Prospective validation SR study N=1501 External and Temporal	77% 1501/ 1938	92.1 (88.9-94.7)	95.7 (94.3-96.8)	21.3 (16.2-28.0)	0.082 (0.058-0.117)	87.4 (83.7-90.5)	97.4 (96.3-98.2)	94.8 (93.6-95.9)

Accuracy of the IOTA Simple Rules: comparison of findings in the original and prospective IOTA validation studies with the current study - Part B. N = 237. (Nunes, 2014)

Current Study								
	Rules applicable	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ve (95% CI)	LR-ve (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
Rules applicable population	78% 237/303	96.2 (90.5-99.0)	88.6 (82.0-93.5)	8.46 (5.3-13.7)	0.043 (0.016-0.113)	87.1 (79.6-92.6)	96.7 (91.8-99.1)	92.0 (87.8-95.1)
Pre-menopausal N=130	80% 130/162	89.7 (72.6 - 97.8)	89.1 (81.3 - 94.4)	8.23 (4.7 - 14.6)	0.116 (0.040 - 0.340)	70.3 (53 - 84.1)	96.8 (90.9 - 99.3)	89.2 (82.6 - 94.0)
Postmenopausal N=107	76% 107/141	98.7 (92.9 - 100)	87.1 (70.2 - 96.4)	7.65 (3.1 - 19.1)	0.015 (0.002 - 0.106)	94.9 (87.5 - 98.6)	96.4 (81.7 - 99.9)	95.3 (89.4 – 98.5)
SR+PR † N=303		94.1 (88.7 - 97.4)	89.9 (84.3 – 94)	9.3 (5.9 – 14.6)	0.066 (0.034 - 0.129)	88.2 (81.8 – 93)	95 (90.3 - 97.8)	91.7 (88.1 – 94.6)
PR ‡ N = 66		86.7 (69.3 - 96.2)	94.4 (81.3 - 99.3)	15.6 (4.0 - 60.4)	0.141 (0.057 - 0.353)	92.9 (76.5 - 99.1)	89.5 (75.2 - 97.1)	90.9 (81.3 – 96.6)

SR+MA § N=303		97 (92.6 - 99.2)	69.6 (62.1 - 76.5)	3.2 (2.5 - 4.0)	0.043 (0.016 - 0.112)	72 (64.9 - 78.4)	96.7 (91.8 - 99.1)	81.8 (77.0 – 86.0)

† SR+PR – Simple rules and use of pattern recognition for all those in whom the rules were not applicable (indeterminate tumours).

‡ PR – Pattern recognition used for those patients in whom the rules were not applicable (indeterminate tumours).

§ SR+MA – Simple rules and malignancy assumed for all those in whom the rules were not applicable.

Accuracy of the IOTA Simple Rules: comparison of findings in the original and prospective IOTA validation studies with the current study - Part C. N = 237. (Nunes, 2014)

Pattern Recognition							
Same population where rules applicable	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ve (95% CI)	LR-ve (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
Prospective validation SR study External and Temporal N=1501	91.1 (87.7-93.8)	95.7 (94.3-96.8)	21 (16.0-27.7)	0.094 (0.068-0.129)	87.3 (83.5-90.4)	97.0 (95.9-98.0)	94.5 (93.3-95.6)
Current Study N=237	97.1 (91.9-99.4)	93.2 (87.5-96.8)	14.2 (7.6-26.8)	0.031 (0.010-0.094)	91.9 (85.2-96.2)	97.6 (93.2-99.5)	94.9 91.3-97.4

Table 8.2-2: Histology Results for False Negative Simple Rules Analyses. (N = 4 of 237). (Nunes, 2014)

Histology	**Number**
Borderline	1
Dermoid cyst with 1-2mm foci of immature cells	1
Stage III Adenocarcinoma	1
Primary peritoneal cancer	1
Total	**4**

Table 8.2-3: Histology Results for False Positive Simple Rules Analyses. (N = 15 of 237). (Nunes, 2014)

Histology	**Number (%)**
Cystadenoma / Cystadenofibroma	6 (40)
Dermoid	2
Endometriosis	2
Leiomyoma	1
Actinomycosis	1
Torsion	1
Benign	1
Fibrothecoma	1
Total	**15**

Table 8.2-4: Accuracy of the IOTA Simple Rules: comparison of findings in the original and prospective validation studies with the current study for all patients. N = 303. (Nunes, 2014)

All Patients Simple Rules followed by pattern recognition							
	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ve (95% CI)	LR-ve (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
Prospective validation IOTA SR study External and Temporal N=1938	91.1 (88.4– 93.4)	92.7 (91.2– 94.0)	12.5 (10.3 – 15.1)	0.0955 (0.0729 – 0.125)	82.9 (79.6 – 85.8)	96.4 (95.3 – 97.4)	92.3 (91.0 – 93.4)
Current Study N=303	94.1 (88.7 – 97.4)	89.9 (84.3 – 94.0)	9.3 (5.91 – 14.6)	0.0659 (0.0336 – 0.129)	88.2 (81.8 – 93.0)	95.0 (90.3 – 97.8)	91.7 (88.1 – 94.6)

Pattern Recognition alone							
	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ve (95% CI)	LR-ve (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
Prospective validation IOTA SR study External and Temporal N=1938	90.4 (87.6 – 92.8)	92.7 (91.2 – 94.0)	12.4 (10.2 – 14.9)	0.104 (0.0799 – 0.134)	82.8 (79.6 – 85.7)	96.1 (95.0 – 97.1)	92.1 (90.8 – 93.2)
Current Study N=303	94.8 (89.6 – 97.9)	93.5 (88.6 – 96.7)	14.5 (8.17 – 25.7)	0.0555 (0.0269 – 0.114)	92.1 (86.3 – 96.0)	95.7 (91.4 – 98.3)	94.1 (90.8 - 96.4)

Table 8.2-5: P-values from comparisons of sensitivity, specificity and accuracy of simple rules and pattern recognition using Nunes (2014). N = 303. (Nunes, 2014)

Current Study	Sensitivity	Specificity	Accuracy
Simple vs. Pattern	0.56	0.03	0.13

Table 8.2-6: P-values from comparisons of sensitivity, specificity and accuracy of simple rules, pattern recognition and simple rules + pattern recognition (Nunes, 2014)

Current Study	Sensitivity	Specificity	Accuracy
SR vs. Original (2008)	0.59	0.48	1.00
SR vs. Validation (2010)	0.15	< 0.001	0.08
PR vs Validation. (2010) (SR patients only) (N=237)	0.04	0.20	0.80
SR+PR vs. Validation (2010)	0.27	0.19	0.76
PR vs. Validation (2010) (All patients) (N=303)	0.10	0.72	0.22

8.3 Study 3 – A prospective evaluation of the IOTA Logistic Regression Models (LR1 and LR2) in comparison to Subjective Pattern Recognition for the diagnosis of ovarian cancer in the outpatient setting using two reference standards

8.3.1 Introduction

The International Ovarian Tumour Analysis (IOTA) group developed two logistic regression models (LR1&LR2), aiming to improve the accuracy of pre-operative ultrasound diagnosis of ovarian cancer diagnosis by non-expert operators of average ability and experience.

Non-expert or Level II ultrasound operators tend to describe the morphology of adnexal tumours without attempting to state their nature, while expert operators (Level III) often use subjective pattern recognition to differentiate between benign and malignant lesions and to determine their likely histological type. (EFSUMB, 2006; EFSUMB, 2010)

A number of previous original and validation studies showed that LR1 and LR2 perform well in the hands of expert operators and can facilitate differentiation between benign and malignant tumours in women scheduled for surgical treatment of adnexal tumours. (Timmerman, 2005; Timmerman, 2010; Nunes, 2012; Nunes, 2013; Kaijser, 2014) Only a minority of women with an ultrasound diagnosis of an adnexal tumour however, require surgery and in the routine clinical practice the critical issue is not who will perform surgery but whether an intervention is required at all. There have been no studies so far, which assessed the suitability of IOTA models for the diagnosis of

ovarian cancer in outpatient setting and for triaging women with adnexal tumours for surgical interventions. The prevalence of malignancy in women attending outpatient clinics in likely to be lower, which could result in a larger number of interventions in women with benign disease even if the good diagnostic accuracy of IOTA models is maintained.

Another difficulty with using a test in outpatient setting is only a proportion of women would be selected for surgical treatment. Histological findings, which are traditionally used to assess as the reference standard for assessing the accuracy of models for the diagnosis of ovarian cancer are not applicable to population of women managed conservatively. In such circumstances delayed type cross sectional studies may provide best evidence that harvest the reference standard information from a prospectively planned and prolonged clinical follow up. (Knottnerus, 1997) It is unknown; however, what is the required length of follow up and appropriate frequency of visits to determine the nature of an adnexal tumour.

In this prospective study we assessed the accuracy of IOTA models for the diagnosis of ovarian cancer in outpatient settings using histology as the first reference standard and the tumour change at follow up ultrasound scans as a second reference test. We also assessed the potential impact on the intervention rates of a policy, which would replace pattern recognition with the IOTA models for triaging women with adnexal tumours for surgery

8.3.2 Methods

This was a single centre prospective observational study of consecutive women attending our gynaecological diagnostic unit for a variety of gynaecological complaints. The first phase of the study lasted for 33-months from May 2009 to January 2012. During that time women attended for their initial visits which included history taking, clinical examination and ultrasound scans. A woman's menopause is defined as her last menstrual period preceding 12 consecutive months of amenorrhoea with no physiological, pathological or medical cause other than the age-related loss of ovarian follicular activity. Women who were over the age of 50 years and had had a hysterectomy were also classified as being postmenopausal. All those with evidence of adnexal tumours were examined by a Level II operator (myself), who was trained in use of the IOTA protocol (Timmerman et al 2000), but not in tumour "pattern recognition". (Valentin, 1999b; Yazbek, 2007a) Attempting to designate a histological diagnosis or attempting to differentiate subjectively between benign and malignant tumours on ultrasound scan was discouraged. The probability of malignancy within an adnexal mass was estimated by using the IOTA Logistic Regression Model (LR1 and LR2) as per section 2.7.5.

All women had a detailed assessment including history as well as clinical and ultrasound examination.

The level II operator recorded the IOTA assessments in the research file and these assessments were not available to the clinicians who made the clinical decisions about the patients' plan of care. Pregnant women and those unable to undergo a transvaginal scan were excluded from the final data analysis. In cases of multiple lesions, the lesion, more likely to be malignant according to the IOTA model score, was included into the

analysis, as the diagnosis of malignancy in one lesion supersedes the diagnosis of any coexisting benign lesions. Following the examination by the level II operation, the women with adnexal tumours were re-examined independently by an expert ultrasound operator who used subjective pattern recognition to determine the nature of the adnexal tumour. Women with suspected ovarian cancer following the expert exam were referred to our gynaecological oncology team for further management. Women with presumed benign lesions were offered choice between conservative management and surgery taking into account their clinical symptoms and personal preferences.

Women who opted for conservative management were offered regular follow up ultrasound scans starting with 4 monthly intervals for a minimum of 12 months. Women who become symptomatic during follow up, those who requested intervention and those in whom ultrasound findings became suspicious of cancer were offered surgery. Only when the data collection was completed at the end of the study, were the IOTA LR1 and LR2 calculations of the risk of malignancy performed and included in the data collection sheets. Histopathology was the primary reference standard used. Tumours were classified according to the World Health Organisation (WHO) guidelines and malignancies were staged according to the International Federation of Gynecology and Obstetrics criteria. (Serov, 1973; Benedet, 2000) For the women in whom surgery was not required, an ultrasound scan at 12 months or more after the primary scan confirming the initial diagnosis of benign lesion was used as the second reference standard.

It was determined by the local Research and Development Department that formal ethical assessment and approval was not required as the steps in the conduct of the study were routine practice in the unit. This refers to ultrasound scan analysis including morphological analysis of the tumours using the IOTA protocol for the assessment of

adnexal masses. In addition to this, therapeutic decisions were not based on the IOTA model scores.

8.3.3 Statistical Analysis and Results

8.3.3.1. Statistical Techniques

The sample size for this study was determined using Harrell's recommendation that a validation dataset should contain at least 100 "events", that is borderline or malignant tumours (Harrell, 2001). Our validation dataset has 137 events, therefore satisfying this requirement.

Initially, the data was analysed after assuming that both reference tests had perfect (100%) sensitivity and specificity. The sensitivity, specificity and overall accuracy of the three diagnostic tests (LR1, LR2 and pattern recognition) were calculated under this assumption and presented with exact 95% confidence intervals. Formal comparisons across different index tests were made using McNemar's test; the exact version of this test was used when necessary. These analyses were performed in the software package Stata 14.0 © (Stata Corp., College Station, TX, USA).

A secondary analysis was performed using a Bayesian approach similar to that of de Groot (de Groot, 2011). In detail, we assumed that the choice of reference test (histology or ultrasound) was related to the underlying status of the patient (that is, patients with cancer were more likely to be assessed using histology). In addition, we assumed that the histology results were always correct whereas the ultrasound results

were imperfect. The prior distributions for ultrasound sensitivity and specificity were chosen to reflect our belief that both values were probably close to 90% and almost certainly within the range 80-100%. Low information priors were used for all other parameters. Two MCMC chains of 3000 samples were run in OpenBUGS (Lunn, 2012) with the first 1000 of each discarded as the burn-in period. All results are presented as medians with 95% credible intervals. To check the robustness of our results, additional analyses were run where the priors for ultrasound performance were changed to reflect worse performance.

8.3.3.2. Results

A total of 555 consecutive women attended for ultrasound assessment during the study period. 11 women were pregnant and they were excluded from the data analysis. A flow chart showing the management of 544 women included in the trial is shown in Fig. 1. (Figure. 8.3-1.) A total of 342/544 (62.9%) women had surgery while 147/544 (27.0%) were managed conservatively. (Table 8.3-1) 41/544 (7.5%) were lost to follow up, 13 women died soon after the diagnosis of adnexal tumour was made and one woman received chemotherapy as the primary treatment for presumed metastatic bowel cancer. Among the 13 women who died, 5 had a non-ovarian malignancy (oesophageal (2), pancreatic, cervical and endometrial), 4 had non-malignant medical conditions (amyloidosis, chronic renal failure and bronchiectasis, dementia with urinary sepsis and alcoholic induced liver failure) and 4 had suspected ovarian malignancies. The patients ranged from 16 to 91 years of age with an average age of 50 years. 237/544 (43.6%) of women were post-menopausal.

After excluding the 55/544 (10.1%) women who had neither reference test, the final diagnosis of a malignant tumour was made in 137/489 (28.0%) women and benign tumours were diagnosed in the remaining 352 (72.0%) women. Histological diagnosis was available in 342 women and they are shown in Tables 8.3-2, 8.3-3 and 8.3-4. The most common indications for surgery were suspected ovarian cancer and pelvic pain. (Table 8.3-5) Only 8 tumours increased in size during follow up whilst the remaining 139 (94.6%) were the same size, smaller or completely resolved. (Table 8.3-6) In these 8 women the cyst increased in size between 21% and 94% but none of the women had surgery for that reason as many of the cysts were still less than 5cm in diameter and the tumours were morphologically unchanged. In women managed expectantly the average follow-up was 27 months for all who had an ultrasound scan with a range from 12 to 68 months.

Using pattern recognition as the primary diagnostic test, 147 women were diagnosed with cancer. The management of women in respect of the predicted nature of pelvic tumour is shown in Fig 2. (Figure 8.3-2) Over 95% of women diagnosed with a malignant lesion had surgery compared to 50% of women with benign lesions. The predicted diagnoses of the nature of adnexal tumour at the initial visit using PR, LR1 and LR2 are shown in Table 8.8-7. There were significant difference in the proportion of women diagnosed with malignancy between PR and LR1/LR2. ($P < 0.0001$) (Table 8.8-8) Assuming that the rate of intervention would be identical in women diagnosed with cancer and benign disease regardless of the diagnostic method used, there would also be a significant difference in the number of women having surgery for presumed malignancy. This is because the false positive rate for LR1/LR2 was significantly higher that of pattern recognition. The overall intervention rates including both benign and malignant lesions for PR, LR 1 and LR2, however would not be significantly different.

(Table 8.3-7)

There were 10 patients who had surgery after completing 12 months follow up which enabled a comparison of the two tests performed. In one of them there were discordant results for histology and the follow-up ultrasound scan. These results suggest a good level of agreement between histology and the follow up visit strategy. The woman with the discordant results had surgery because her tumour had changed morphologically. Histology result however was benign polypoid endometriosis.

The primary analysis was performed assuming that both reference tests (histology and follow up ultrasound scans) had the same perfect accuracy. (Table 8.3-9) This showed LR1, LR2 and pattern recognition all demonstrated high sensitivities, but there were significant differences in the specificities for the diagnosis of ovarian cancer between PR and the IOTA models when tests were used at the initial visit. When assuming differing accuracy for histology (100%) and follow up ultrasound scans (90%), the results were similar to when both reference tests were assumed to be 100% accurate. (Table 8.3-10)

There were eight women with cancer (borderline [5] and invasive [3]) who were misdiagnosed as benign disease using pattern recognition (8/132; 6.1% [95% CI: 2.0% to 10.1%]). This caused delayed interventions in one woman who was eventually diagnosed with a borderline tumour. Among the invasive malignancies, all were stage 1 with two dermoid tumours with early malignant transformation within and the one seromucinous adenocarcinoma. Using LR1 there would have been four false negative cancer diagnoses (borderline [2] and invasive [2]) (3.0% [95%CI: 0.1% to 6.0%]) (P =

0.238 when compared to PR) compared to seven (borderline [3] and invasive [4]) (5.3% [95%CI: 1.5% to 9.1%]) using LR2 (P=0.787 when compared to PR) neither of which was statistically significant. The tumours missed by LR1/LR2 were those missed by PR except for 1 metastatic tumour detected by PR and missed by LR1.

Figure 8.3-1: Flowchart of Study Eligibility IOTA Logistic Regression Models LR1 and LR2 - IOTA LR1 and LR2 Follow-Up Study. N = 544

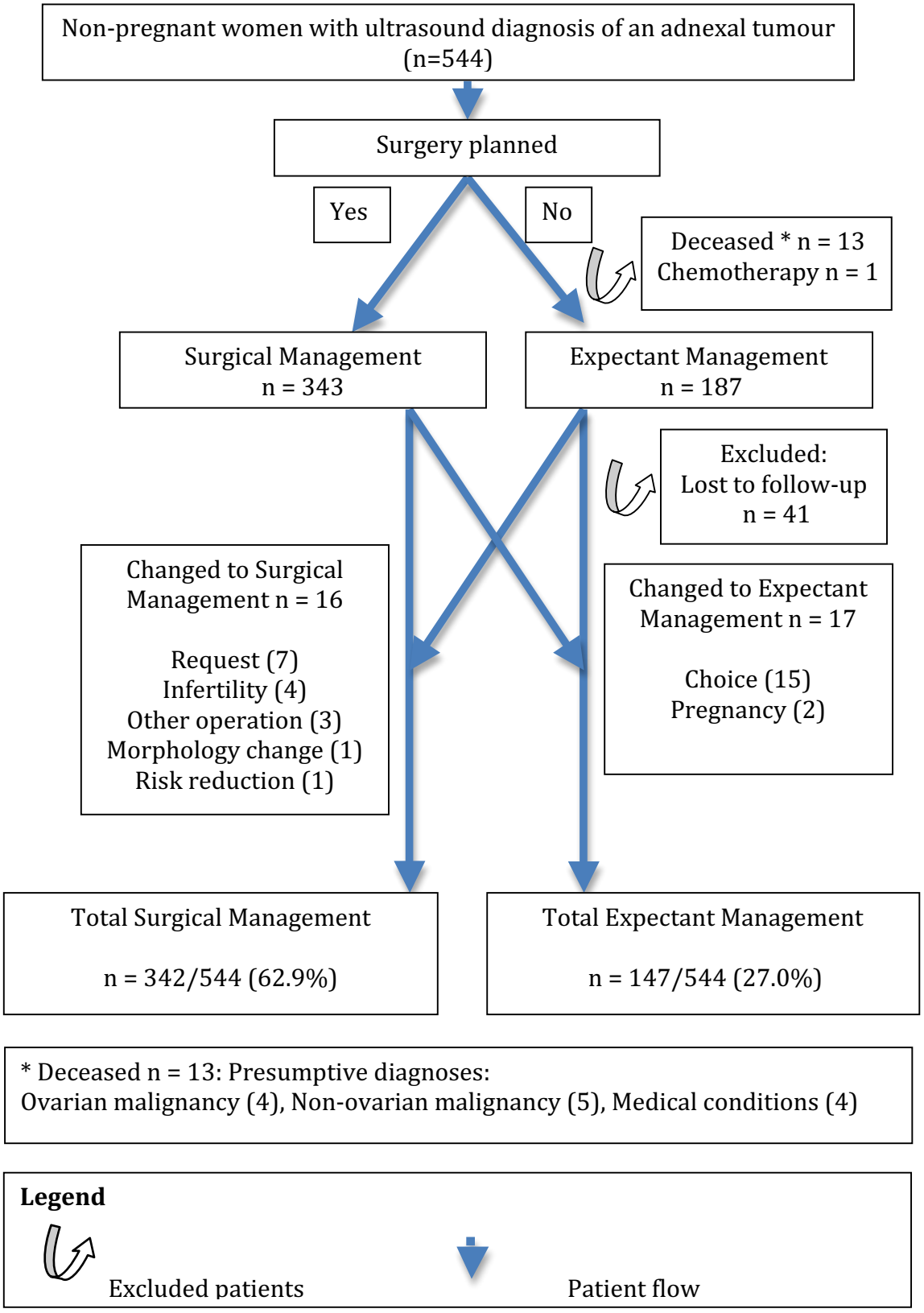


Table 8.3-1: Final Patient Reference Standard Allocation - IOTA LR1 and LR2 Follow-Up Study. N = 544.

	Minimum 12 month Follow up Number of Patients (%)
Histology Only	332 (61.0)
Follow-up USS Only	147 (27.0)
Both	10 (1.8)
Neither	55 (10.1)
Total	544 (100)

Table 8.3-2: Benign Histology - IOTA LR1 and LR2 Follow-Up Study. Histology N=342 Benign Tumours N=205.

Histology	**Number (%)**
Cystadenoma (n=43) /Cystadenofibroma (n=14)	57 (27.8)
Endometrioma	47 (22.9)
Mature cystic teratoma (Dermoid)	31 (15.1)
Benign functional (haemorrhagic cyst, follicular cyst)	29 (14.1)
Fibroma	10 (4.9)
Normal adnexae at surgery	5 (2.4)
Torsion (3 x Benign, 1x Cystadenofibroma, 1 x Fibroma)	5 (2.4)
Pedunculated Leiomyoma	4 (2.0)
Pelvic Inflammatory Disease	4 (2.0)
Fibrothecoma	3 (1.5)
Hydrosalpinx	3 (1.5)
Pseudocyst	2 (1.0)
Brenner	2 (1.0)
Actinomycosis	1 (0.5)
Struma Ovarii	1 (0.5)
Myolipoma	1 (0.5)
Total	**205**

Table 8.3-3: Malignant Histology - IOTA LR1 and LR2 Follow-Up Study. Histology N = 342. Malignancies N = 137. Outcome of Follow-up Ultrasound Scans - IOTA LR1 and LR2 Follow-Up Study. N = 171. N = 189.

Malignant Tumours	**Number (%)**
Borderline	21 (15.3% all Malignancies) (6.1% of all Histology)
Invasive:	116 (84.7% all Malignancies) (33.9% of all Histology)
Total	137

Table 8.3-4: Malignant Histology Tumour Type and Stage - IOTA LR1 and LR2 Follow-Up Study. Histology N = 342. Malignancies N = 137.

Tumour Type and Stage		**N (%)**
Borderline		**21 (15.3% all Malignancies)** **(6.1% of all Histology)**
Borderline n=21	Serous	7
	Mucinous	6
	Serous papillary	4
	Endometrioid	1
	GCT	1
	Undefined	2
Stage 1		21
Invasive:		**116 (84.7% all Malignancies)** **(33.9% of all Histology)**
Invasive Malignancies n=116	Epithelial	84
	Germ cell	3
	Sex chord Stromal	1
	Metastases, Recurrence, unknown	28
Invasive Malignancies n=116	Stage I	26/116 (22.4% of invasive malignancies)
	Stage II	5/116 (4.3% of invasive malignancies)
	Stage III	37/116 (31.9% of invasive malignancies)
	Stage IV	22/116 (10.0% of invasive malignancies)
	Metastases / Recurrence	26/116 (22.4% of invasive malignancies)

Table 8.3-5: Indication for Surgery - IOTA LR1 and LR2 Follow-Up Study. N = 342.

Suspected cancer	152
Pain	75
Women's Choice	62
Sub-fertility	30
Pressure symptoms	13
Other operation needed	5
UKCTOCS / Prophylactic	4
Abnormal Follow-up USS	1
TOTAL	**342**

Table 8.3-6: Morphological changes of expectantly managed adnexal tumours on follow up scans- IOTA LR1 and LR2 Follow-Up Study. N = 147.

	Follow-up USS Only (%)
Resolved	34 (23.1)
Smaller*	21 (14.3)
Unchanged	84 (57.1)
Larger*	8 (5.4)
Total	**147**

Figure 8.3-2: Flowchart Showing Management of Women with Adnexal Tumours with regard to ultrasound diagnosis of malignancy based on Pattern Recognition (n=544)

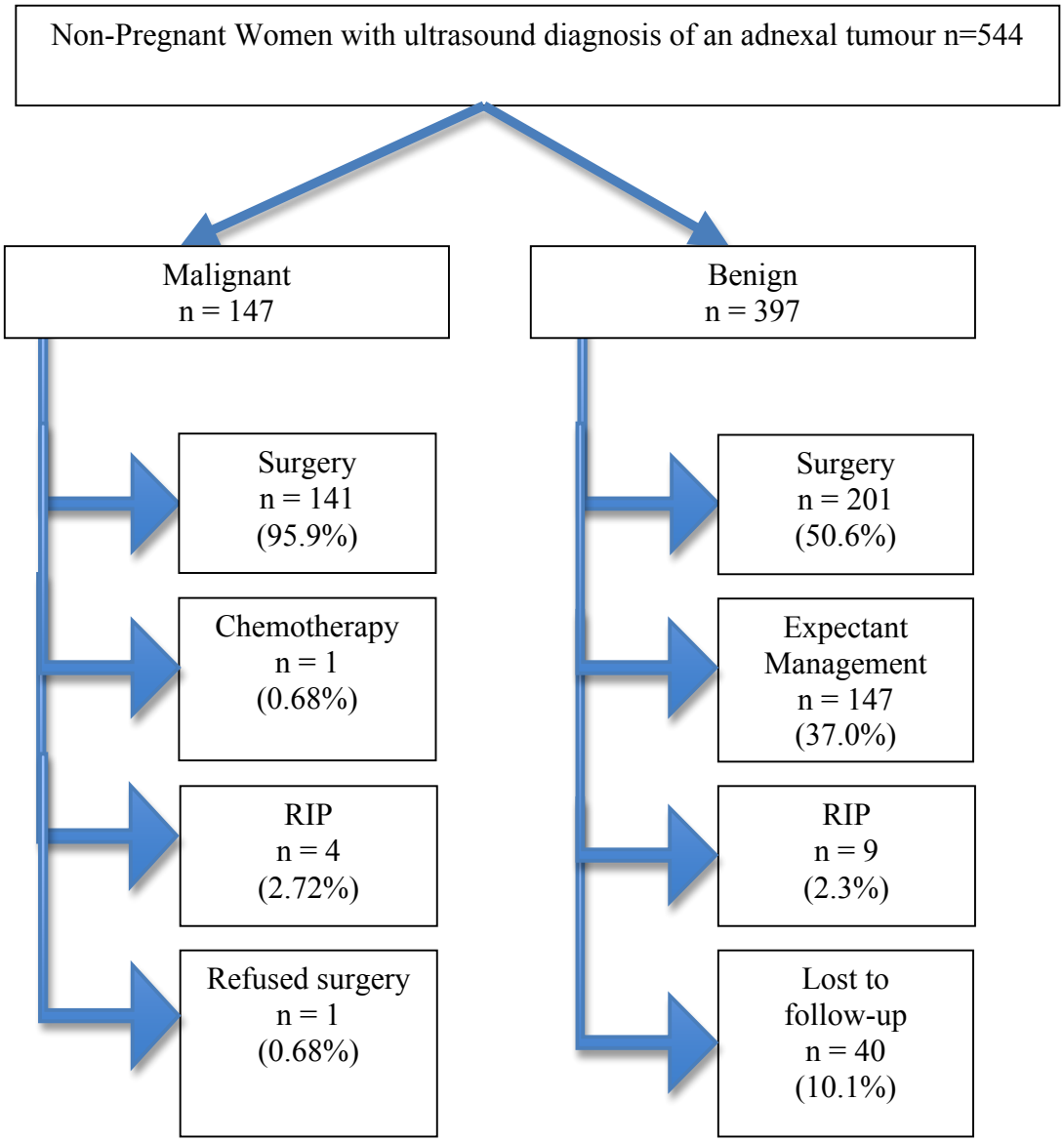


Table 8.3-7: Intervention based on initial ultrasound scan diagnosis by LR1, R2 and PR - IOTA LR1 and LR2 Follow-Up Study. N = 544.

	Malignant Diagnosis n	Intervention rates for Malignant Diagnoses n (%)	Intervention rates for All Diagnoses n (%)
Pattern Recognition	142	142/489 (29.0%)	342/489 (69.9%)
LR1	213	213/489 (43.6%)*	(76.1%)[#]
LR2	212	212/489 (43.4%)*	(76.0%)[#]

*Malignant diagnosis: Assumption that in keeping with PR (n=142), all women with a malignant diagnosis with LR1/2 would have had surgery.

[#]Benign Diagnosis: Assumption that in keeping with PR where 200/347 (57.6%) with benign diagnoses had surgery, 57.6% of women with benign diagnoses by LR1/2 would also have had surgery.

[#]LR1 rate = (213 + 57.6% of 276)/489 = 76.1%

[#]LR2 rate = (212 + 57.6% of 277)/489 = 76.0%

Table 8.3-8: LR1 versus Pattern Recognition - IOTA LR1 and LR2 Follow-Up Study. N = 489.

REF = Benign		PR	
		Benign	*Malign*
LR1	*Benign*	271	1
	Malign	68	12

P < 0.0001

REF = Malignant		PR	
		Benign	*Malign*
LR1	*Benign*	4	0
	Malign	4	129

P = 0.13 (Exact)

Table 8.3-9: Primary Analysis Both Reference Standards - Sensitivity, Specificity and Accuracy - IOTA LR1 and LR2 Follow-Up Study. All patients N = 480, 27.7% malignancy

Test	Number	Sensitivity	Specificity	Accuracy
PR	489	94.2 (88.8 to 97.4)	96.3 (93.8 to 98.0)	95.7 (93.5 to 97.3)
LR1	489	97.1 (92.7 to 99.2)	77.3 (72.5 to 81.5)	82.8 (79.2 to 86.1)
LR2	489	94.9 (89.8 to 97.9)	76.7 (71.9 to 81.0)	81.8 (78.1 to 85.1)

Table 8.3-10: Secondary Bayesian Analysis Assuming 'High' performance Follow up USS (90% accuracy) – Sensitivity and Specificity - IOTA LR1 and LR2 Follow-Up Study. All patients N = 489, 28.0% malignancy.

Test	Number	Sensitivity	Specificity
PR	489	93.8 (88.9 to 97.0)	96.4 (94.1 to 98.0)
LR1	489	96.7 (92.8 to 98.8)	77.5 (73.0 to 81.5)
LR2	489	94.5 (89.8 to 97.6)	76.8 (72.3 to 81.0)

Presented as medians and 95% credible intervals.

8.4 Study 4 - A prospective evaluation of the IOTA "Simple Rules" as a triage tool for the diagnosis of ovarian cancer in the outpatient setting using two reference standards

8.4.1 Introduction

Study 4 looked at the simple rules protocol and its accuracy for the diagnosis of ovarian cancer. In this study, the reference standard was histology in keeping with the literature. There were though, patients who were excluded because they did not have the preferred reference test (histology). These patients had an alternative reference test (follow up USS over at least 12 months) and the exclusion of these patients leads to some partial or incomplete verification bias. It should not be assumed that the alternative reference test has 100% accuracy and the difference in the accuracies of the 2 reference tests should ideally be considered during the analysis.

8.4.2 Methods

This was an extension of the prospective study (Study 2). Initial assessments were conducted over a 33-month period from May 2009 to January 2012 and follow up ultrasound scans were conducted until January 2014 as was done in Study 3. Patient assessment was as per the methods of study 2 but the analyses did not only include the women who had a histological diagnosis but also those who had expectant management and a follow up ultrasound scan of a minimum of 12 months, that showed complete resolution or which exceeded 6 months as per Study 3.

8.4.3 Statistical Analysis and Results

8.4.3.1. Statistical Techniques

As was the case in Study 3, the 2 reference standards were histology and follow-up ultrasound scans and both primary and secondary analyses were done. (See section **8.3.3 Statistical Analysis and Results** of Study 3 – Subsection **8.3.3.1 Statistical Techniques**).

8.4.3.2. Results

After excluding the 11 pregnant patients from the initial 555 consecutive women, there were 544 women included in the analysis. There were 342/544 (62.9%) women who had surgery while 147/544 (27.0%) women had expectant management with follow up ultrasound scans. (Figure 8.4-1) This is the same patient population with the same demographics as in Study 3. (See section **8.3.3 Statistical Analysis and Results** of Study 3 – Subsection **8.3.3.2 Results**). (Table 8.3-1, Table 8.3-2, Table 8.3-3, Table 8.3-4, Table 8.3-5, Table 8.3-6)

The simple rules were applicable in 80.3% (437/544) (95% CI: 76.7% to 83.6%) of the entire population while the rules were applicable in 77.2% (264/342) (95% CI: 72.8% to 81.7%) of women who had surgery and in 87.8% (129/147) (82.5% to 93.1%) of those who had expectant management only. (P = 0.0067)

When those who had surgery was analysed, the rules were applicable in 264/342

(77.2%) women with 146/342 (42.7%) classified as benign and 118/342 (34.5%) classified as malignant. Of those classified as benign, 142/146 (97%) were confirmed as benign and of those classified as malignant, 101/118 (85.6%) were confirmed as malignant. (Figure 8.4-2)

When the expectant group was analysed, the rules were applicable in 129/147 (87.8%) women. There were 128/147 (87.1%) assessed as benign and 1/147 (0.7%) as malignant and on follow up ultrasound scan, none of these showed any morphological changes to raise the suspicion of malignancy and the one assessed as malignant completely resolved on follow up ultrasound scan.

For the entire population who had either of the reference tests performed, the simple rules were applicable in 393/489 (80.4%) with 274/489 (56.0%) assessed as benign and 119/489 (24.4%) assessed as malignant. (Table 8.4-1) 270/274 (98.5%) of those assessed as benign were confirmed as benign and 101/119 (84.9%) of those assessed as malignant were confirmed as malignant.

Primary analysis assuming the same 100% accuracy for both reference tests showed a similar sensitivity and specificity for SR as for PR and for SR+PR. ($P>0.05$). As expected, SR+MA had a trend towards the highest sensitivity ($P>0.05$) and the lowest specificity ($P < 0.0001$) of all the tests. (Table 8.4-2)

Inclusion of the women who did not have surgery improved the specificity of all models, as the specificity of the models for those who did not have surgery was significantly greater than for those who did have surgery because of the greater likelihood that the women who had expectant management only, had benign disease.

(Tables 8.4-3 and 8.4-4)

There were eight women with cancer who were misdiagnosed as benign disease using pattern recognition (8/132; 6.1% [95% CI: 2.0% to 10.1%]). This caused a delay in surgery for one woman, who was eventually diagnosed with a borderline tumour. Using SR there would have been four false negative cancer diagnoses (3.0% [95%CI: 0.1% to 6.0%]), which was not statistically significant. (P = 0.238) When assuming differing accuracy for histology (100%) and a follow up ultrasound scan (90%) the median sensitivity and specificity with the 95% credible intervals for SR+PR, SR+MA and pattern recognition were similar to when both reference tests were assumed to be 100% accurate. (Table 8.4-5)

On comparing the results of PR, LR1, LR2 and SR, SR was seen to match the performance of PR when the rules were applicable (Specificity P = 0.04 [exact], Sensitivity P = 1.00 [exact]). (Table 8.4-6) SR had a better specificity when compared with LR1/LR2 (P = 0.0001) though sensitivity was similar (P = 0.50 [LR1]). (Table 8.4-7)

Figure 8.4-1: Flowchart of Study Eligibility Simple Rules - IOTA Simple Rules Follow-Up Study. N = 544.

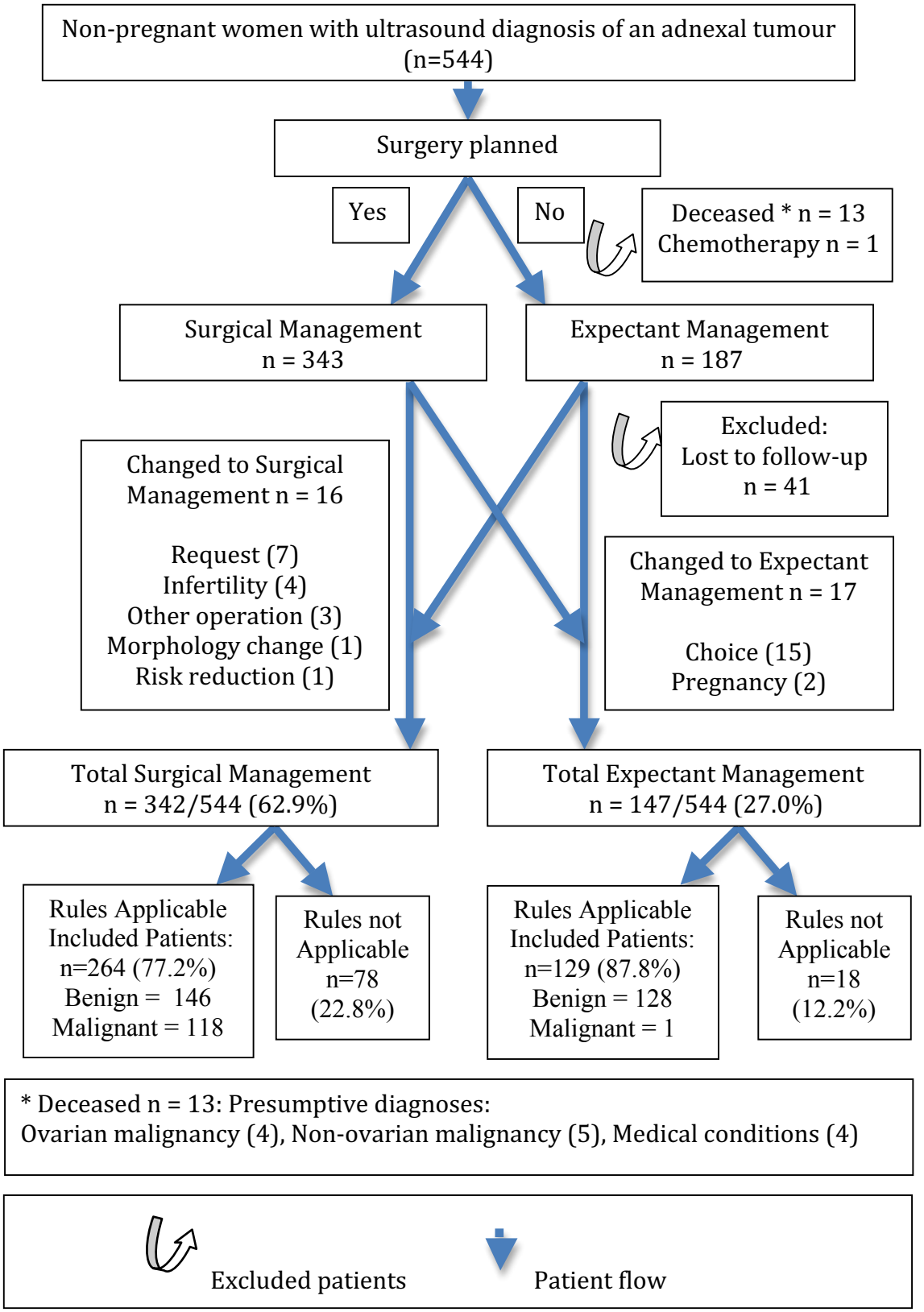


Figure 8.4-2: Flow Diagram of Study Results Stratified by Management Simple Rules - IOTA Simple Rules Follow-Up Study. N = 544.

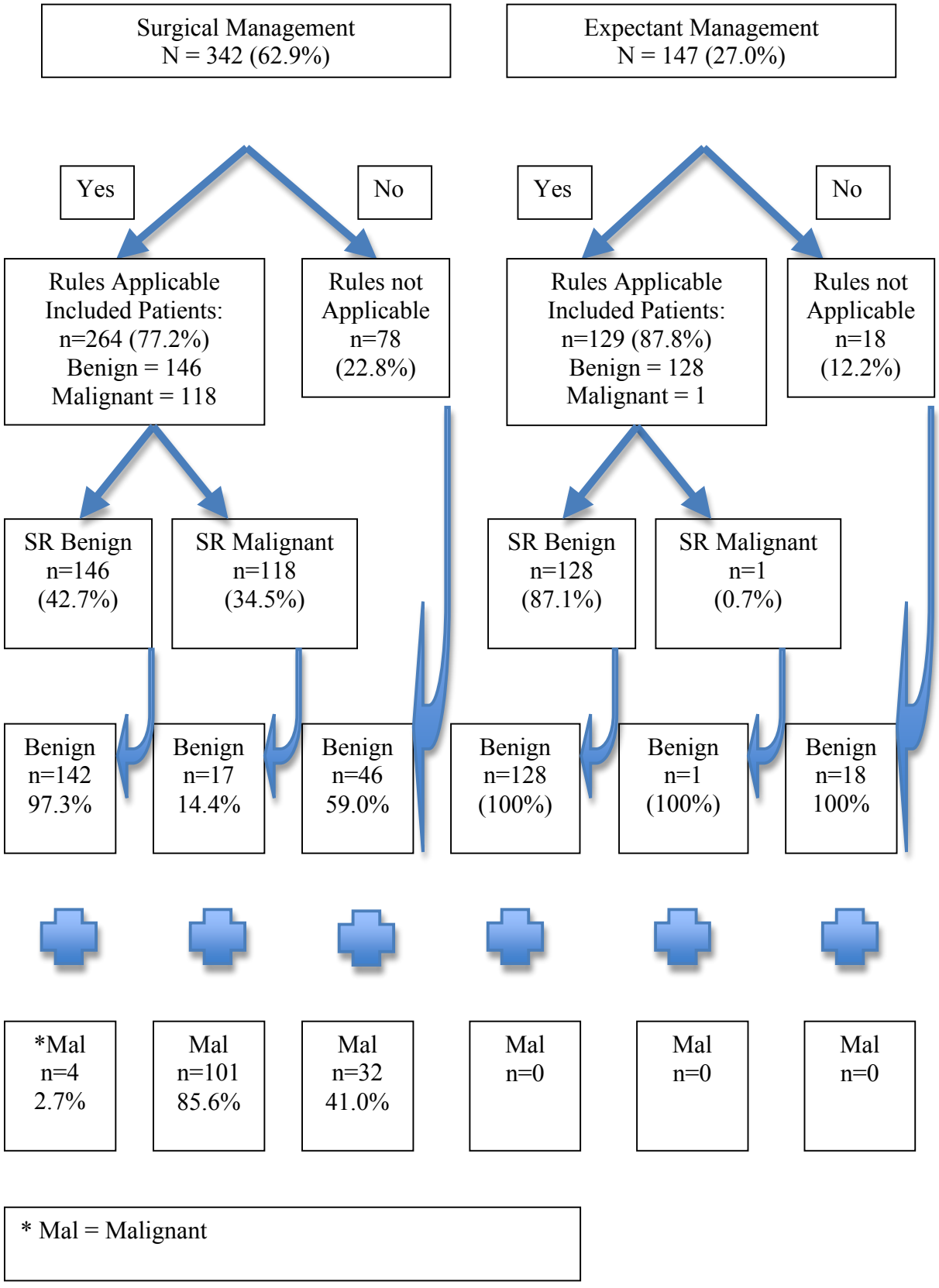


Table 8.4-1: Patient Outcomes after use of either of the 2 Reference Standards Simple Rules - IOTA Simple Rules Follow-Up Study. N = 489.

	Rules Applicable N = 393 (80.4%)				Rules not Applicable N=96 (19.6%)	
	Simple Rules assessed as **Benign**		Simple Rules assessed as **Malignant**		Simple Rules assessed as **Indeterminate**	
Simple Rules	**N = 274/489 (56.0%)**		**N = 119/489 (24.4%)**		**N = 96/489 (19.6%)**	
Benign Outcome	**N = 270/274 (98.5%)**		**N = 18/119 (15.1%)**		**N = 64/96 (66.7%)**	
	Diagnosis from:		**Diagnosis from:**		**Diagnosis from:**	
	Histology	142	Histology	17	Histology	46
	USS	128	USS (Resolved)	1	USS (Resolved x 5, Smaller x 3, Same x 10)	18
Malignant Outcome	**N = 4/274 (1.5%)**		**N = 101/119 (84.9%)**		**N = 32/96 (33.3%)**	
	Diagnosis from:		**Diagnosis from:**		**Diagnosis from:**	
	Histology:		Histology:		Histology:	
	- Borderline	1	- Borderline	10	- Borderline	9
	- Invasive Malignant	3	- Invasive Malignant	91	- Invasive Malignant	23
	USS	0	USS	0	USS	0

Table 8.4-2: Primary Analysis Both Reference Standards - Sensitivity, Specificity and Accuracy - IOTA Simple Rules Follow-Up Study. All patients N = 489, 28.0% malignancy. Simple Rules Applicable N = 393.

Test	Number	Sensitivity	Specificity	Accuracy
PR	489	94.2 (88.8 to 97.4)	96.3 (93.8 to 98.0)	95.7 (93.5 to 97.3)
SR	393	96.2 (90.5 to 99.0)	93.8 (90.3 to 96.3)	94.4 (91.6 to 96.5)
SR+PR †	489	93.4 (87.9 to 97.0)	94.3 (91.4 to 96.5)	94.1 (91.6 to 96.0)
SR+Mal §	489	97.1 (92.7 to 99.2)	76.7 (71.9 to 81.0)	82.4 (78.7 to 85.7)

† SR+PR – Simple rules and use of pattern recognition for all indeterminate tumours

§ SR+MA – Simple rules and malignancy *all indeterminate tumours*

Table 8.4-3: Primary Analysis Histology Only - Sensitivity, Specificity and Accuracy - IOTA Simple Rules Follow-Up Study. N = 342, 40.1% malignancy. Simple Rules Applicable N = 264.

Test	Number	Sensitivity	Specificity	Accuracy
PR	342	94.2 (88.8 to 97.4)	93.7 (89.4 to 96.6)	93.9 (90.8 to 96.2)
SR	264	96.2 (90.5 to 99.0)	89.3 (83.4 to 93.6)	92.0 (88.1 to 95.0)
SR+PR †	342	93.4 (87.9 to 97.0)	90.7 (85.9 to 94.3)	91.8 (88.4 to 94.5)
SR+Mal	342	97.1 (92.7 to 99.2)	69.3 (62.5 to 75.5)	80.4 (75.8 to 84.5)

† SR+PR – Simple rules and use of pattern recognition for all indeterminate tumours

§ SR+MA – Simple rules and malignancy *all indeterminate tumours*

Table 8.4-4: Primary Analysis - Comparison of Specificity between Histology and Follow-Up Ultrasound Patients - IOTA Simple Rules Follow-Up Study.

	Histology		**USS**		
Test	**Number**	**Specificity**	**Number**	**Specificity**	**P-value**
PR	195	93.8 (89.5 to 96.8)	147	100.0 (97.5 to 100.0)	0.006
SR	151	89.4 (83.4 to 93.8)	129	99.2 (95.8 to 100.0)	0.002
SR+PR †	195	90.8 (85.8 to 94.4)	147	99.3 (96.3 to 100.0)	0.003
SR+Mal §	195	69.2 (62.2 to 75.6)	147	87.1 (80.6 to 92.0)	0.001

† SR+PR – Simple rules and use of pattern recognition for all indeterminate tumours

§ SR+MA – Simple rules and malignancy *all indeterminate tumours*

Table 8.4-5: Secondary Bayesian Analysis Assuming 'High' performance Follow up USS (90% accuracy) – Sensitivity and Specificity - IOTA LR1 and LR2 Follow-Up Study. All patients N = 489, 28.0% malignancy.

Test	Number	Sensitivity	Specificity
PR	489	93.8 (88.9 to 97.0)	96.4 (94.1 to 98.0)
SR+PR †	489	93.2 (88.0 to 96.5)	94.4 (91.9 to 96.6)
SR+Mal §	489	96.7 (92.7 to 98.8)	76.8 (72.4 to 81.2)

Presented as medians and 95% credible intervals.

† SR+PR – Simple rules and use of pattern recognition for all indeterminate tumours

§ SR+MA – Simple rules and malignancy *all indeterminate tumours*

Table 8.4-6: Pattern Recognition versus Simple Rules - IOTA LR1 and LR2 Follow-Up Study - Rules Applicable (n = 393).

REF = Benign		**SR**	
		Benign	*Malign*
PR	*Benign*	269	8
	Malignant	1	10

P = 0.04 (Exact)

REF = Malign		**SR**	
		Benign	*Malign*
PR	*Benign*	2	1
	Malignant	2	100

P = 1.00 (Exact)

Table 8.4-7: LR1 versus Simple Rules - IOTA LR1 and LR2 Follow-Up Study - Rules Applicable (n = 393).

REF = Benign		**SR**	
		Benign	*Malignant*
LR1	*Benign*	251	1
	Malignant	19	17

P = 0.0001

REF = Malignant		**SR**	
		Benign	*Malignant*
LR1	*Benign*	2	0
	Malignant	2	101

P = 0.50 (Exact)

8.5 Study 5 - Use of the IOTA Simple Rules for the diagnosis of ovarian cancer: a Meta-Analysis

8.5.1 Introduction

The aim of this study was to perform a meta-analysis of the results of studies using the International Ovarian Tumour Analysis (IOTA) simple rules tool for the diagnosis of ovarian cancer.

8.5.2 Methods

8.5.2.1. Data sources

Articles were identified from a number of electronic databases. A search was performed on MEDLINE using the terms "simple rules and adnexal", "simple rules and ovarian" and "simple rules and ultrasound". A search was also performed on EMBASE using the terms "Simple Rules" after using terms "ovar tumour" and "Ultrasound". A Cochrane search was also performed for "ovary" and "tumour". Additionally, hand searches were performed on the bibliographies of included studies and any reviews found. Studies from 1st January 2008 until the 8th November 2013 were included for review. Our own data was included as one of the studies.

8.5.2.2. Study selection

Along with another reviewer, I selected articles from the search results. I read all abstracts. To be included in this meta-analysis, the study had to include data on use of the Simple Rules protocol when applied to adnexal masses for the diagnosis of ovarian cancer. All women must also have had surgery and a histological diagnosis for comparison as the gold standard or reference test. (Figure 8.5-1) Only prospective studies and preferably those with a clear indication that the sample selection was random or consecutive were included. This was to avoid selection bias and improve the applicability and generalisability of the results.

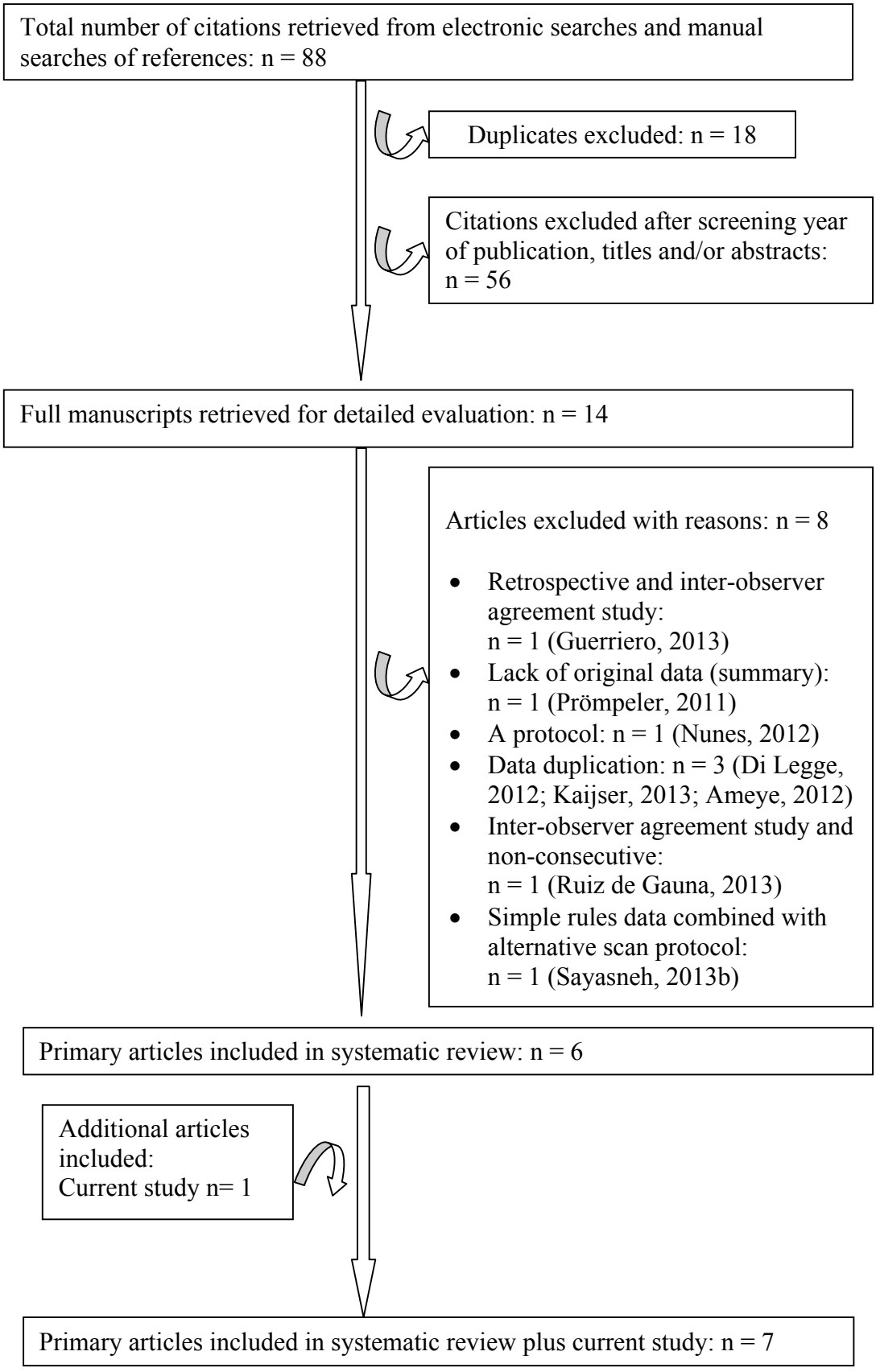
After choosing relevant articles, the full text articles were read and 2 by 2 contingency tables were created with the data extracted independently by myself and 1 other reviewer. QUADAS 2 assessments were performed on each included study by two independent reviewers (N.N. and X.F.) and discrepancies were resolved between them.

Studies that combined simple rules with another format of assessment were excluded because it was not possible to extract data on the simple rules alone as was the case with Sayasneh et al. (Sayasneh, 2013b) A further exclusion was Di Legge et al and as it incorporated data from IOTA Phase 1b (n=507) and Phase 2 (n=1938) both of which are already included in the data from Timmerman 2008 and Timmerman 2010. (Di Legge, 2012; Timmerman, 2008, Timmerman, 2010) Kaijser et al 10 was a summary article describing all the IOTA studies and giving a summary. (Kaijser, 2013) The authors included IOTA Phase 1, 1b and 2 (temporal and external validation studies) and this data is already included. Ameye et al included patients who were already included from other studies. (Ameye, 2012) These were IOTA Phase 1 data patients as well as Phase 2

data patients. The author also went on to exclude a significant proportion of the patients who could be classified according to simple descriptors. This study was therefore excluded.

Guerriero et al used five examiners to analyse stored data. This study was excluded as it was a retrospective study and as it had different results for each of the five examiners. (Guerriero, 2013) Ruiz de Gauna et al performed a similar study with two ultrasound operators analysing the tumours in real time and four doing so retrospectively in order to assess the inter-observer agreement and was therefore excluded as no single result was available. (Ruiz de Gauna, 2013) Prompeler was a summary and did not include any original data but instead presented the data from Timmerman 2008. (Prömpeler, 2011) Nunes et al is a published protocol for an ongoing study and has no data attached. (Nunes, 2012)

Figure 8.5-1: Flowchart of Study selection process for the Meta-analysis of use of Simple Rules for the identification of malignant adnexal tumours (Nunes, 2014)



8.5.2.3. Quality assessment

The quality of each study was assessed using the QUANDAS-2 tool looking at the 4 domains of patient selection, the index test used, the reference standard or gold standard and the flow and timing of the study. (Whiting, 2011) Quality characteristics were documented for the selected publications according to QUADAS 2 criteria. (Figure 8.5-2) (Table 8.5-1) Verification bias, which is also known as "workup" or "referral" bias, occurs if the decision to perform the reference test (surgery with resultant histology) is based on the result of the index test under assessment (simple rules). If the decision for surgery was based on the result of the simple rules model, verification bias was considered to have occurred. There are various types of verification bias. Partial verification bias occurs if only a selected subset of the population who underwent the index test undergo the reference test and the decision about which subset undergoes it, is dependent on the results of the test. Differential verification bias occurs when the index test result affects whether the participant receives the same reference standard test. Incorporation bias occurs if the index forms part of the reference standard.

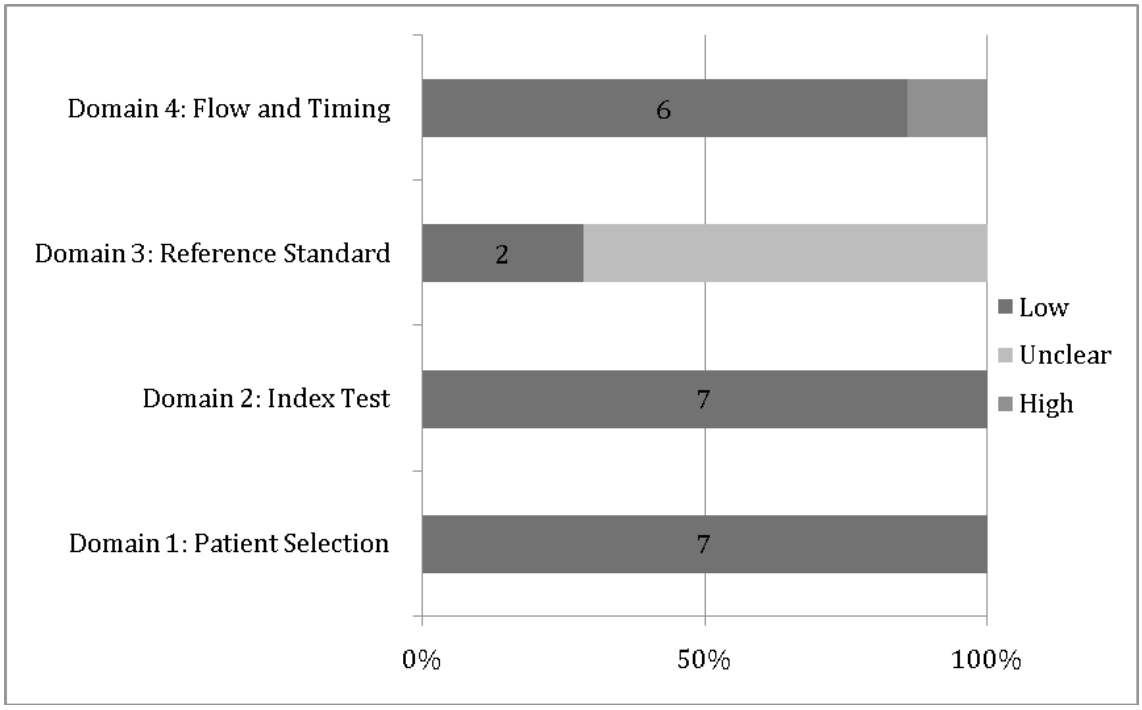
Only prospective studies with no significant evidence of the various types of verification, selection and incorporation bias were included. These patients must all have had transvaginal ultrasound scans and they should have had surgery within 120 days of performing the ultrasound scan.

Table 8.5-1: Simple Rules Study Quality QUADAS 2 Criteria (Nunes, 2014)

	Domain 1: Patient selection								Domain 2: Index test					Domain 3: Reference Standard							Domain 4: Flow and Timing							
Studies (First author, publication year)	Methods of patient selection described	Included patients described	Consecutive or random sample	Case control design avoided	Inappropriate exclusions avoided	Selection of patient introduced bias	Included patients do not match review question	**Domain 1: Patient selection**	Index test described	Index test results independent	Threshold pre-specified	Index test interpretation introduced bias	Index test does not match review question	**Domain 2: Index test**	Reference standard described	Reference standard correct	Reference standard results independent	Reference standard interpretation introduced bias	Target condition does not match review question	**Domain 3: Reference Standard**	Excluded patients described	described	Appropriate interval between tests	Reference standard for all patients	Same reference standard for all	All patients included in the analysis	Patient flow introduced bias	**Domain 4: Flow and Timing**
Timmerman2008	Y	Y	Y	Y	Y	L	L	**L**	Y	Y	Y	L	L	**L**	Y	Y	U	L	L	**U**	Y	Y	Y	Y	Y	Y	L	**L**
Timmerman2010	Y	Y	Y	Y	Y	L	L	**L**	Y	Y	Y	L	L	**L**	Y	Y	L	L	L	**L**	Y	Y	Y	Y	Y	Y	L	**L**
Fathallah 2011	Y	Y	Y	Y	Y	L	L	**L**	Y	Y	Y	L	L	**L**	Y	Y	U	L	L	**U**	Y	N	U	Y	Y	Y	U	**H**
Hartman 2012	Y	Y	Y	Y	Y	L	L	**L**	Y	Y	Y	L	L	**L**	Y	Y	U	L	L	**U**	Y	Y	Y	Y	Y	Y	L	**L**
Sayasneh 2013	Y	Y	Y	Y	Y	L	L	**L**	Y	Y	Y	L	L	**L**	Y	Y	U	L	L	**U**	Y	Y	Y	Y	Y	Y	L	**L**
Alcazar 2013	Y	Y	Y	Y	Y	L	L	**L**	Y	Y	Y	L	L	**L**	Y	Y	U	L	L	**U**	Y	Y	Y	Y	Y	Y	L	**L**
Nunes 2014	Y	Y	Y	Y	Y	L	L	**L**	Y	Y	Y	L	L	**L**	Y	Y	L	L	L	**L**	Y	Y	Y	Y	Y	Y	L	**L**

Legend: Y=Yes; U=Unclear; N=No; H=High risk bias; L=Low risk of bias

Figure 8.5-2: Simple Rules Study Quality QUADAS 2 Criteria (Nunes, 2014)



Low=low risk of bias; Unclear=risk of bias unclear; High=high risk of bias.

8.5.3 Statistical Analysis and Results

8.5.3.1. Statistical Techniques

Analysis was performed using the packages metan (Harris, 2007; Harris, 2008), metareg (Harbord, 2005; Harbord, 2008a) and metandi (Harbord, 2008b; Harbord, 2009) in Stata 13.1 ® (Stata Corp., College Station, TX, USA).

The sensitivity, specificity and accuracy of simple rules and pattern recognition were compared separately using McNemar's test for paired binary outcomes. These analyses are based on the 237 patients for whom the simple rules were applicable as well as the

full 303 population separately. The sensitivity, specificity and accuracy of simple rules were compared with the original and the validation IOTA studies separately using the Chi-squared ($\chi 2$) test for independent binary outcomes. (Timmerman, 2008; Timmerman, 2010)

Random effects meta-analysis was used to calculate univariate pooled estimates of sensitivity and specificity for the IOTA simple rules tool. (DerSimonian, 1986) A bivariate approach was also investigated to calculate these values (Harbord, 2007). Forest plots were constructed to summarise the results, and heterogeneity was quantified using the I^2 statistic. Meta-regression was used to investigate any heterogeneity present in the results and a funnel plot of the diagnostic odds ratio against study size was created to investigate the possibility of publication bias. (Thompson, 2002)

8.5.3.2. Results

There were 88 studies identified with the initial search. (Figure 8.5-1) After excluding duplicates (18) and studies that did not include the simple rules protocol (56), there were 14 studies left for detailed reading. These full text articles were then read fully. A further 8 studies were excluded because they were using simple rules combined with another ultrasound tool [1] (Sayasneh, 2013b) , a reanalysis of previously published data included here from another study and therefore of already included patients [3], (Di Legge, 2012; Kaijser, 2013; Ameye, 2012), a retrospective study looking at inter-observer variability [1] (Guerriero, 2013), a mixed prospective and retrospective assessment of inter observer variability 13 [1] (Ruiz de Gauna, 2013), a summary [1]

(Prömpeler, 2011) or a protocol [1] (Nunes, 2012). That left 6 studies added to our data. (Table 8.5-2)

All women in each study had the same index test, which was a simple rules assessment via a transvaginal ultrasound scan, and all had histology as the reference standard. Two of the seven studies did not state whether all pregnant patients were excluded though in one of those studies, two patients were excluded, because for each of them, the adnexal mass was an ectopic pregnancy. (Fathallah, 2011; Hartman, 2012) Exclusion criteria varied in the studies. Alcázar et al and Hartman et al appeared to have a high exclusion rates but they included patients who would have been excluded at the start of other studies such as patients who never attended for the first scan, patients who had no mass or those who chose expectant management or had surgery at a different institution. (Alcázar, 2013; Hartman, 2012) The exclusions were therefore appropriate and the lost to follow up rates were low. All studies therefore had a low risk of bias in relation to domain 1, ”Patient Selection”. In all studies, the index test was performed independently of the reference standard with a pre-determined threshold for diagnosis resulting in a low risk of bias in relation to domain 2 “Index Test”. One study had 5/282 (1.8%) participants without a histological diagnosis (2 cases of ovarian torsion untwisted at surgery, 3 cases of abscesses confirmed on microscopy and culture). (Sayasneh, 2013b) Only one study other than the author’s study declared blinding of the pathologist who performed the reference standard. (Fathallah, 2011) This caused domain 3 “Reference Standard” to have a large segment where the risk of bias was unclear (70%). Timing of surgery was unknown for only 1 study (Fathallah, 2011) whereas in all others, surgery occurred within 21 days (Alcázar, 2013), 112 days (Hartman, 2012) or 120 days (Timmerman, 2008; Timmerman, 2010; Sayasneh,

2013b). This therefore meant that domain 4 “Flow and Timing” had a low risk of bias in over 80% of the studies. (Figure 8.5-2) (Table 8.5-1)

Details of the prevalence of malignancy and statistical results for each study are shown. (Table 8.5-2) The pooled sensitivity value for all 7 studies was 0.93 (95%CI:0.90-0.96). (Figure 8.5-3) The I^2 value of 32% suggests moderate heterogeneity amongst the studies. The pooled specificity value was 0.95 (95%CI:0.93-0.97) and the I^2 value of 78% suggests considerable heterogeneity between the studies. (Figure 8.5-4) Very similar pooled values for sensitivity and specificity were obtained using the bivariate method (results not shown). A sub-analysis was done on only the externally validated studies and the pooled sensitivity decreased slightly to 0.92 (95%CI:0.88-0.96) whereas the pooled specificity increased slightly to 0.96 (95%CI:0.94-0.98). (Table 8.5-3) (Figure 8.5-5 and 8.5-6) Sub-analyses were done for pre and postmenopausal women. (Figures 8.5-7 to 8.5-10) Sensitivity was higher in postmenopausal women [0.94(95%CI:0.89-0.99)]when compared with premenopausal women [0.89(95%CI:0.82-0.95)] while specificity was similar but lower [0.94 (95%CI:0.88-0.99) and 0.97 (95%CI:0.94-0.99) respectively].

When sensitivity is plotted against specificity, there appears to be a decreasing relationship between sensitivity and specificity. (Figure 8.5-11) Meta-regression was subsequently performed for both sensitivity and specificity adjusting for prevalence. (Figures 8.5-12 & 8.5-13) The results suggest that the sensitivity and specificity values for the IOTA simple rules depend, at least in part, on the patient population to which it is being applied. In particular, the meta-regression results suggest that all the heterogeneity in sensitivity, and over 50% of the heterogeneity in specificity, is due to differences in prevalence across the studies.

A further meta-regression of sensitivity and specificity was performed on operator level but not all studies made it clear the level of their operators and for those studies in which it was know, there was not enough information to assess whether operator level was the cause of heterogeneity. Meta-regression was also used to investigate whether differences in the percentage of the population for whom the rules were applicable was related to the observed heterogeneity. However, this percentage only explained a small amount of the heterogeneity in the sensitivity results and none of the heterogeneity in the specificity results (results not included). We plotted odds ratio against study size but there was no evidence of publication bias. This assessment was also done for the external validation studies alone and this confirmed no evidence of publication bias.

Table 8.5-2: All Studies included in the Simple Rules Meta-analysis (Nunes, 2014)

	Study (First Author, Year of publication)	Popula-tion	Number of Patients where Rules are Applicable (RA) (% of Total Population)	Number of Malignant Tumours (% of Total Population)	Number of Benign Tumours (% of Total Population)	Prevalence of Malignancy in Population where Rules are Applicable (%)	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
1	Timmerman 2008	507	386 (76.1)			29.0	106	25	6	249	0.95 (0.89 to 0.98)	0.91 (0.87 to 0.94)
2	Timmerman 2010	1938	1501 (77.5)	542 (28.0)	1396 (72.0)	24.6	340	49	29	1083	0.92 (0.89 to 0.95)	0.96 (0.94 to 0.97)

3	Fathallah 2011	122	109 (89.3)	14 (11.5)	108 (88.5)	10.1	8	3	3	95	0.73 (0.39 to 0.94)	0.97 (0.91 to 0.99)
4	Hartman 2012	103	91 (88.3)	30 (29.1)	73 (70.9)	24.2	20	9	2	60	0.91 (0.71 to 0.99)	0.87 (0.77 to 0.94)
5	Sayasneh 2013	255	214 (83.9)	74 (29.0)	181 (71.0)	24.8	46	3	7	158	0.87 (0.75 to 0.95)	0.98 (0.95 to 1.00)
6	Alcázar 2013	340	270 (79.4)	55 (16.2)	285 (83.8)	12.2	29	6	4	231	0.88 (0.72 to 0.97)	0.97 (0.95 to 0.99)
7	Nunes 2014	303	237 (78.2)	135 (44.6)	168 (55.4)	44.3	101	15	4	117	0.96 (0.91 to 0.99)	0.89 (0.82 to 0.93)

Sensitivity and specificity given with 95% Confidence Intervals (CI). FN, false negative; FP, false positive; TN, true negative; TP, true positive

Figure 8.5-3: Forest Plot – Sensitivity (All Simple Rules Studies) (Nunes, 2014)

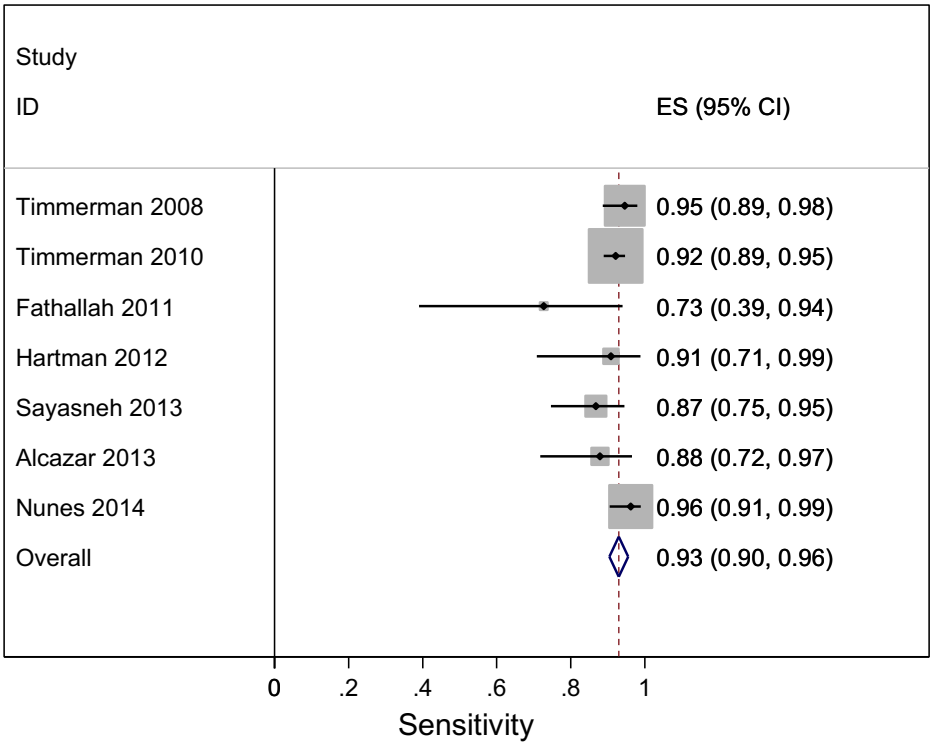


Figure 8.5-4: Forest Plot – Specificity (All Simple Rules Studies) (Nunes, 2014)

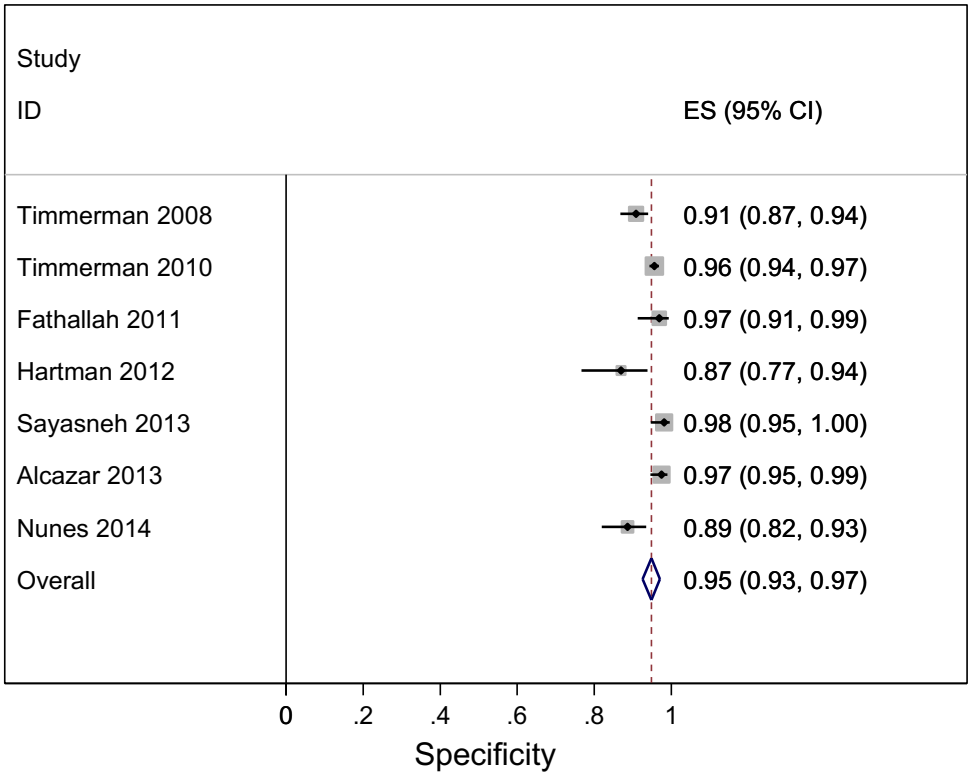


Table 8.5-3: External Validation Studies included in the Simple Rules Mata-analysis (Nunes, 2014)

All women				
Study	**Patient Number**	**Malignancy Prevalence**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
Timmerman 2010*	796	22.2%	0.91 (0.86 - 0.95)	0.96 (0.94 - 0.98)
Fathallah 2011 ¶	109	10.1%	0.73 (0.39 - 0.94)	0.97 (0.91 - 0.99)
Hartman 2012	91	24.2%	0.91 (0.71 - 0.99)	0.87 (0.77 - 0.94)
Sayasneh 2013	214	24.8%	0.87 (0.75 - 0.95)	0.98 (0.95 - 1.00)
Alcazar 2013	270	12.2%	0.88 (0.72 - 0.97)	0.97 (0.95 - 0.99)
Nunes 2014	237	44.3%	0.96 (0.91 - 0.99)	0.89 (0.82 - 0.93)
Premenopausal Women				
Timmerman 2010*	526	9.5%	0.90 (0.78 - 0.97)	0.97 (0.95 - 0.98)
Hartman 2012	39	23.1%	0.89 (0.52 - 1.00)	0.90 (0.73 - 0.98)
Sayasneh 2013	143	15.4%	0.82 (0.60 - 0.95)	1.00 (0.97 - 1.00)
Alcazar 2013	217	7.4%	0.88 (0.62 - 0.98)	0.97 (0.94 - 0.99)
Nunes 2014	130	22.3%	0.90 (0.73 - 0.98)	0.89 (0.81 - 0.94)
Postmenopausal Women				
Timmerman 2010*	270	47.0%	0.91 (0.85 - 0.96)	0.94 (0.88 - 0.97)
Hartman 2012	36	30.6%	0.91 (0.59 - 1.00)	0.80 (0.59 - 0.93)
Sayasneh 2013	71	43.7%	0.90 (0.74 - 0.98)	0.93 (0.80 - 0.98)
Alcazar 2013	53	32.1%	0.88 (0.64 - 0.99)	1.00 (0.90 - 1.00)
Nunes 2014	107	71.0%	0.99 (0.93 - 1.00)	0.87 (0.70 - 0.96)

* *External centre only.* ¶ *No menopausal status data from Fathallah et al.*

Figure 8.5-5: Forest Plot – Sensitivity (External Validation Simple Rules Studies) (Nunes, 2014)

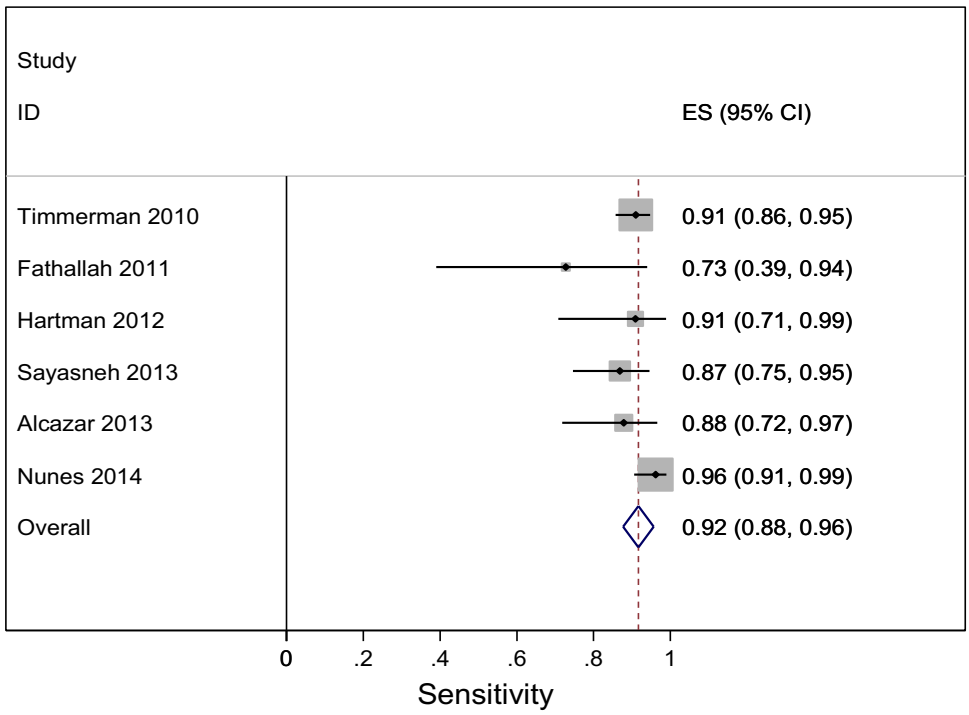


Figure 8.5-6: Forest Plot – Specificity (External Validation Simple Rules Studies) (Nunes, 2014)

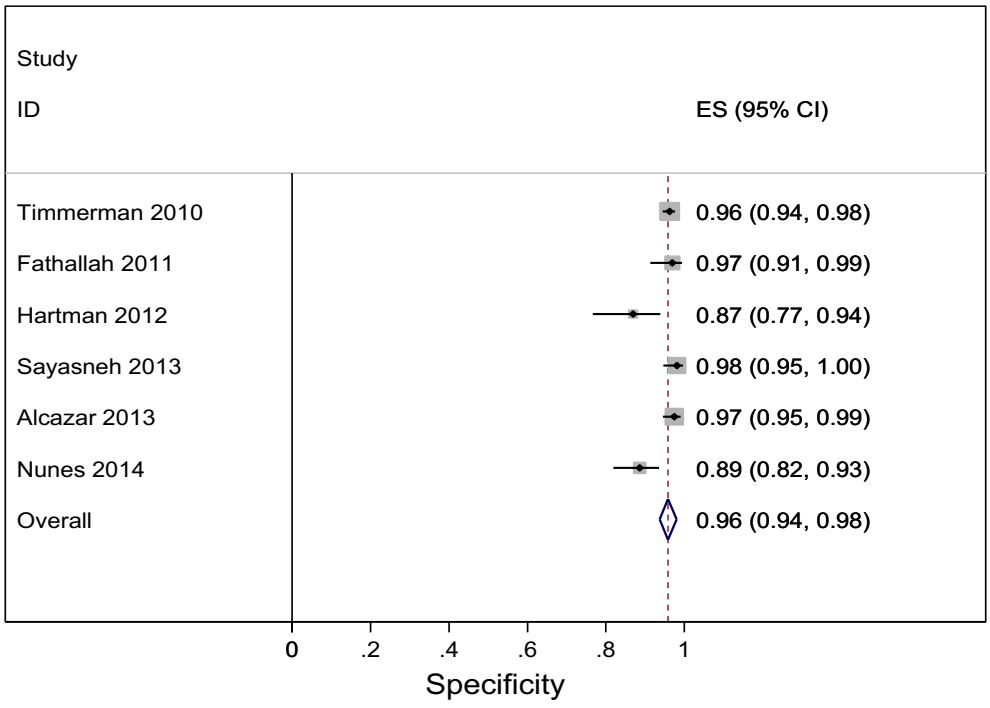


Figure 8.5-7: Forest Plot – Sensitivity (External Validation Simple Rules Studies Premenopausal Women) (Nunes, 2014)

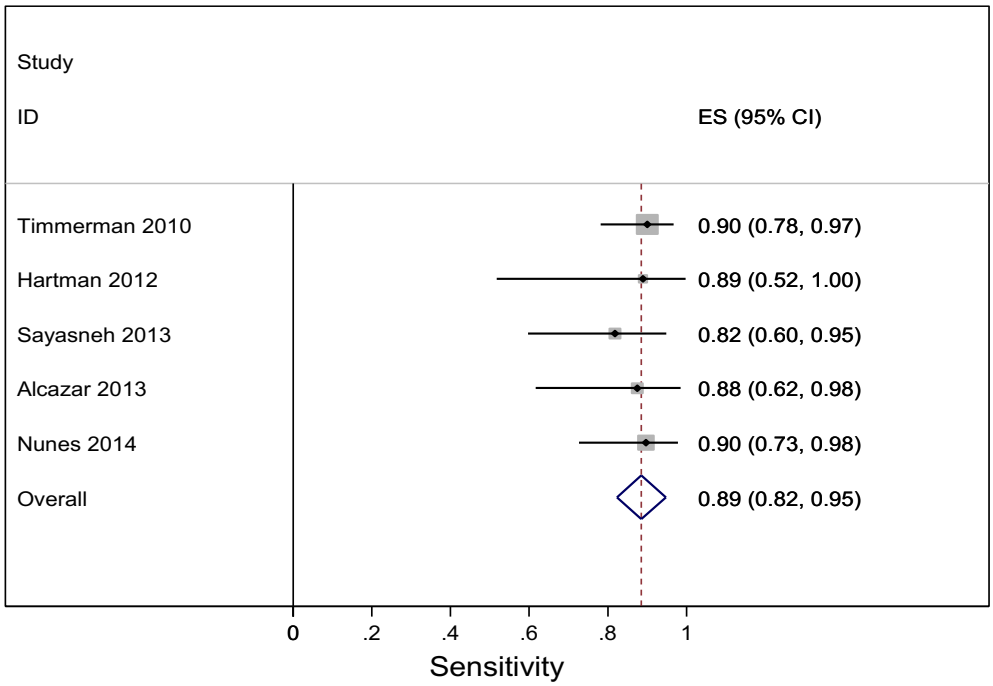


Figure 8.5-8: Forest Plot – Specificity (External Validation Simple Rules Studies Premenopausal Women) (Nunes, 2014)

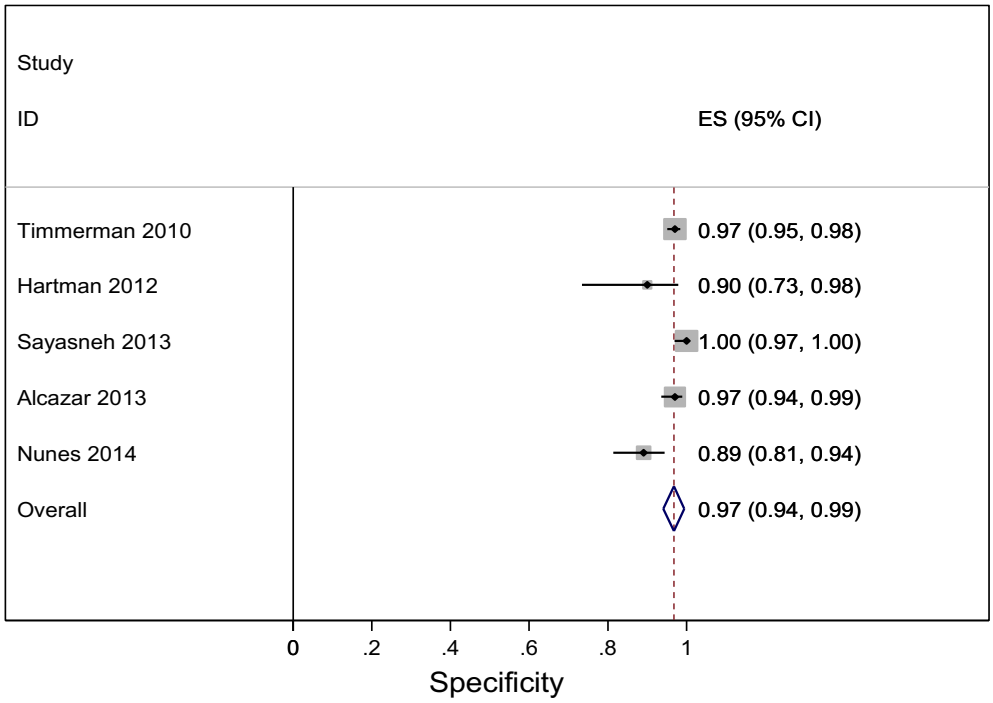


Figure 8.5-9: Forest Plot – Sensitivity (External Validation Simple Rules Studies Postmenopausal Women) (Nunes, 2014)

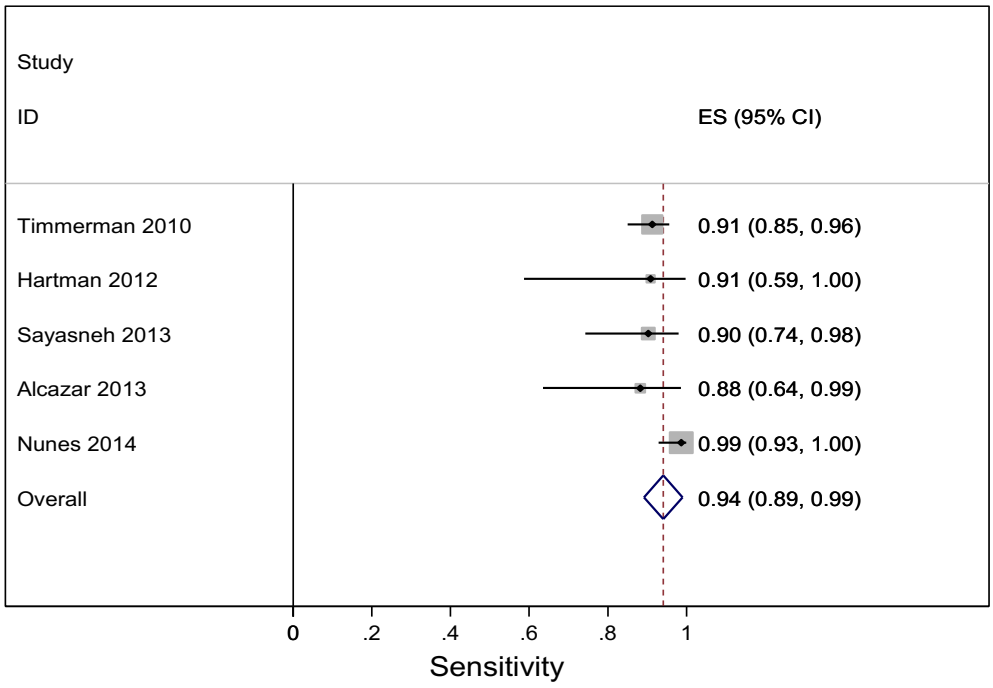


Figure 8.5-10: Forest Plot – Specificity (External Validation Simple Rules Studies Postmenopausal Women) (Nunes, 2014)

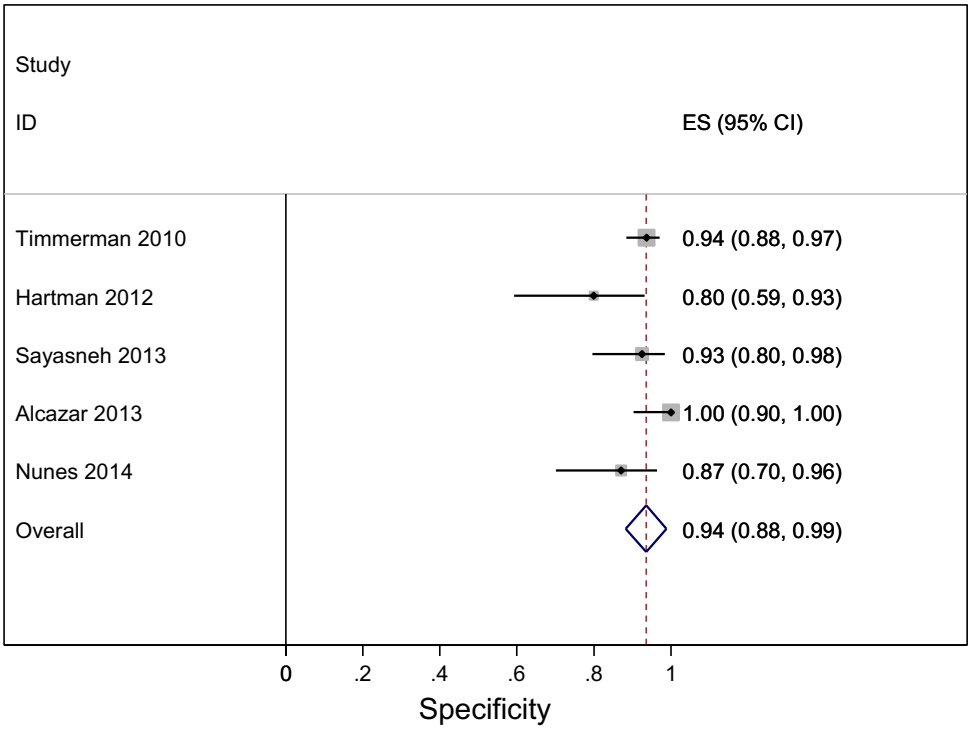


Figure 8.5-11: Sensitivity versus Specificity (All Simple Rules Studies) (Nunes, 2014)

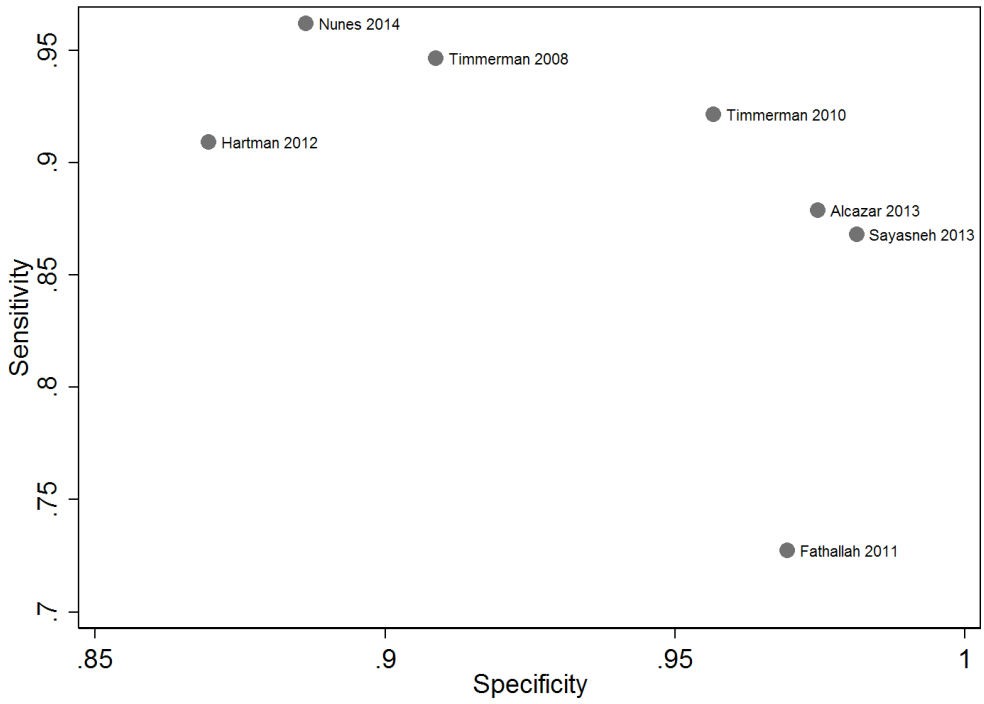


Figure 8.5-12: Meta-regression Sensitivity on Prevalence (All Simple Rules Studies) (Nunes, 2014)

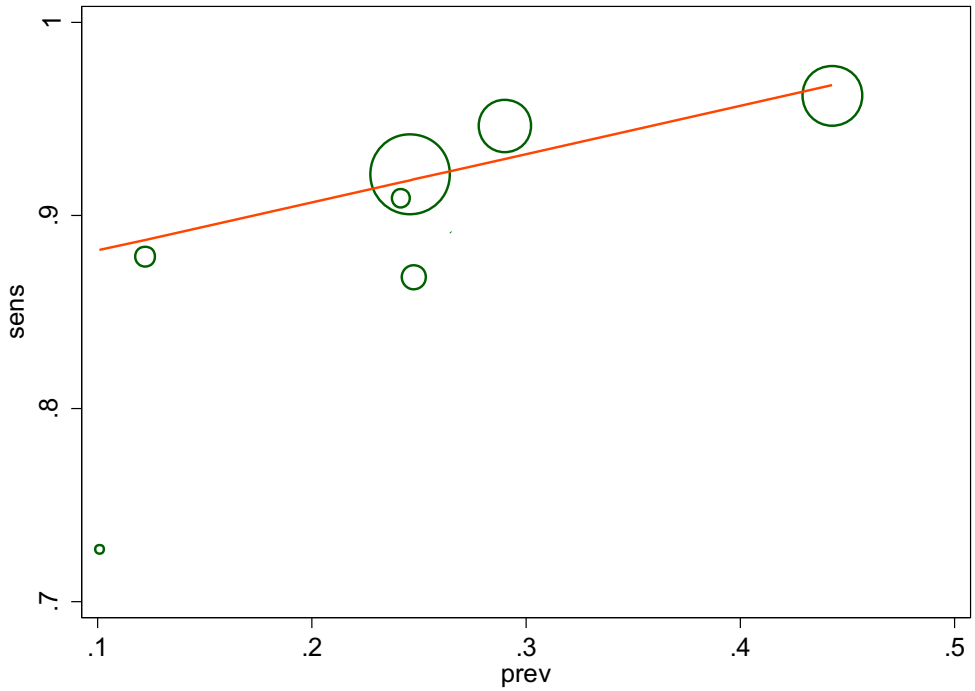
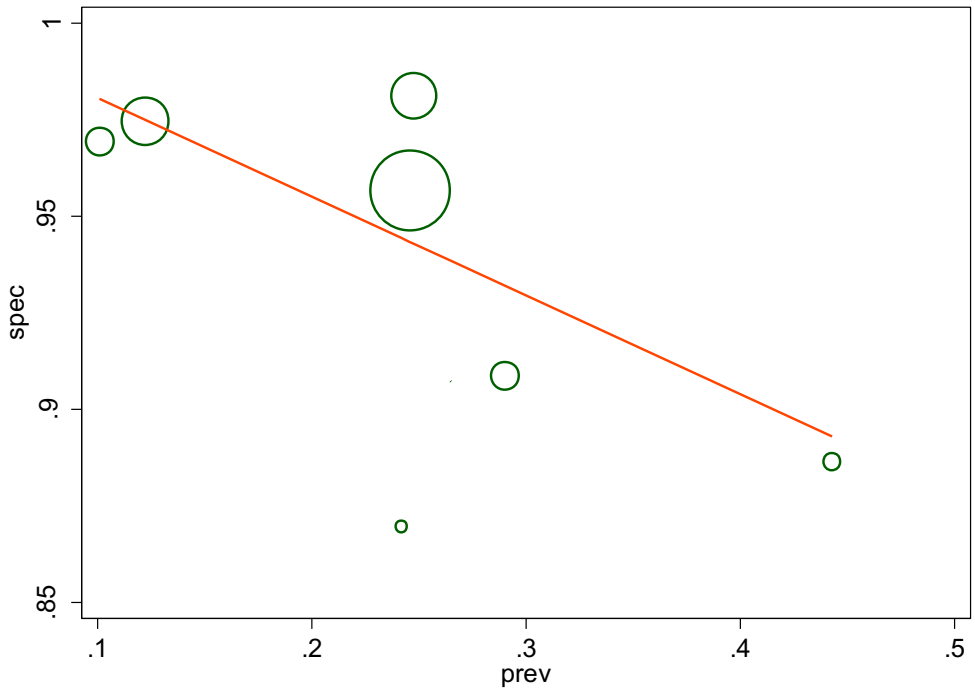


Figure 8.5-13: Meta-regression Specificity on Prevalence (All Simple Rules Studies) (Nunes, 2014)



8.6 Study 6 – A pilot study of comparison of the potential intervention rates when asymptomatic postmenopausal women with an adnexal tumour were assessed by two management protocols

8.6.1 Introduction

The incidence of ovarian cancer is known to increase with age. The incidence was found to be 0.4 -13% in premenopausal women and 45% in postmenopausal women who had a laparotomy for an ovarian tumour in a 10-year retrospective review. (Koonings, 1989) The risk of malignancy increased 12 fold for women in their sixties versus those in their twenties. The population in this study are likely to have been symptomatic as they are described as being almost exclusively self-referring women who had exploratory laparotomies.

Since 1989, there has been significant advancement of technology of imaging techniques. With the use of ultrasound and these various other imaging modalities such at CT and MRI to assess a variety of non-gynaecological conditions, there has been an increase in detection of adnexal cysts or tumours in asymptomatic postmenopausal women. These are cysts that would not have been detected 10 or 20 years ago. As postmenopausal women have a greater incidence of ovarian cancer as compared with premenopausal women, these discoveries have the potential to cause significant concern. Adnexal cysts have been found in up to 28% of asymptomatic women in previous studies. (Levine, 1992; Sladkevicius, 1995; Valentin, 2003)

A recent study has suggested that simple small unilocular cysts have a benign natural history. (Valentin, 2002) As cysts become larger or develop features of "complexity" such as septations, papillary projections, larger solid areas and increasing vascularity, it becomes more difficult to assess the risk of malignancy. Although the vast majority of incidentally detected cysts are benign, all women are routinely offered a blood test to measure serum CA 125 levels in order to facilitate the diagnosis of ovarian cancer. Women with raised CA 125 levels and those with a cyst of complex appearance are offered surgery for fear of missing an ovarian cancer. Many these interventions could be avoided if ultrasound assessment of an ovarian cyst could discriminate better between benign and malignant ovarian lesions. This would result in a significant decrease in morbidity and mortality associated with surgery, as postmenopausal women are more likely to suffer with co-morbidities that increase their operative and anaesthetic risks. This is also likely to result in substantial savings for the NHS due to fewer operations and fewer complications.

In the United Kingdom, the diagnostic evaluation of adnexal tumours in the Royal College of Obstetricians and Gynaecologists (RCOG) green top guidance is based on the calculation of RMI. The absolute value of the RMI score, the laboratory level of serum CA125 and the appearance of the tumours are used to determine the management plan. Women with an RMI of < 25 are managed by a general gynaecologist in any unit. Those with a simple unilateral cyst of <5cm and who have a CA 125 of < 30u/ml can be managed conservatively with 4 monthly ultrasound scans for 1 years followed by yearly scans for a total of 5 years. If the CA 125 ≥ 30u/ml or the cyst is larger (≥5cm) or has septations or solid areas then surgery is advised. For women who have a RMI of 25 – 250, it is advised they be managed in a cancer unit. Management in a tertiary cancer centre is the recommendation for those with a RMI >250, whatever the appearance of

the tumour. (RCOG, 2003) (Figure 8.6-1)

The simple rules as described above have 5 benign features and 5 malignant features and are purely dependant on the ultrasound scan without the use of any blood tests. Both the RMI within the RCOG guideline and the simple rules have a good detection rate for malignant tumours. This study's aim was to test the hypothesis that use of the simple rules can reduce operative rates for women with benign disease when compared with the current RCOG guidance without increasing the false negative rate.

There are several issues with using the RMI. It includes a blood test in addition to a scan, which can elicit additional distress to patients as well as additions to the cost. There is the additional time to await the result of the test and there is the need to ensure a system is in place to check the result, do the calculation and inform the patient. There is reduced tissue expression of CA 125 in 22% cases of ovarian cancers and normal serum CA 125 in 8-18% which is especially the case in borderline, Stage I and non-epithelial tumours. CA 125 is non-specific and can be raised in non-ovarian tumours such as endometrial, tubal, breast, lung and gastrointestinal. Serum levels can also be raised in many normal physiological states and benign conditions such ovulation, menstruation, pelvic infection, diarrhoea, endometriosis and adenomyosis. (Bast, 1983; Rosen, 2005) The advantages of the RMI are that it has extensive and wide use and it is well known and fairly easy to address the features on ultrasound scan It also allows clinicians to dissolve themselves of responsibility of the decision when a blood test is involved.

The aim of this study was to perform a pilot study to determine the potential intervention rates of the 2 management protocols when utilised to assess adnexal

tumours in asymptomatic post-menopausal women.

8.6.2 Methods

This prospective cross sectional observational study was conducted over a 1-year period between 1st December 2009 and 30th November 2010. Consecutive postmenopausal women who presented to the Gynaecology Diagnostic and Outpatient Treatment Unit at the University College Hospital, London, UK for assessment of an asymptomatic adnexal tumour or found to have an asymptomatic adnexal tumour, had their tumour assessed according to the 'Simple Rules' as well as according to the Royal College of Obstetricians and Gynaecologists (RCOG) Green top Guideline number 34 (October 2003) which uses the Risk of Malignancy Index (RMI). These postmenopausal women were judged to be asymptomatic if they did not present with pain localised to the area of the cyst or the lower abdomen.

The unit consultants then assessed the women fully and determined their plan of care. Therapeutic decisions were not based upon the results of either of the management protocols. Therefore, the study was not subject to institutional ethical review board approval. The potential intervention rates were then compared between the 2 management protocols.

8.6.3 Statistical Analysis and Results

At a preliminary analysis done at the 8-month mark, there were 67 asymptomatic women who were seen with an adnexal mass. (Study 7a) At that point, 41/67

(61.2%) were in the group where surgery was indicated according to the RCOG guideline while only 5/67 (7.5%) would be offered surgery according to the simple rules protocol. This data was used for the power calculation for the randomized controlled trial. (Study 8)

At the 1-year mark, there were 93 women analysed. (Study 7b) From the 153 women initially assessed, those who had simple unilateral cysts less than 2 cm were excluded (n=38), as according to the RCOG guideline these women do not need follow up. There were 8 women who never had their CA 125 blood tests, 1 who was a tertiary referral and 4 who died before surgery or before a follow up scan was performed. There were 9(5.8%) who did not attend or declined a follow-up scan.

Among the 93 women, the age range was 47 to 90 years with a mean age of 66 years. Parity ranged from 0 to 8 and time since the menopause ranged from 1 up to 40 years with the average being 14 years. Indications for the ultrasound scan were incidental detection of the adnexal mass during imaging done for a non-gynaecological cause (40/93 [43%]), post-menopausal bleeding (22/93 [24%]), “check-up” of a known cyst (10/93 [11%]), gastrointestinal complaints (6/93 [4%]), urogynaecological assessment (4/93 [3%]), or other causes (11/93 [12%]) such as vaginal discharge, cervical cysts seen at time of smear, experiencing a dragging sensation or having pain elsewhere in her abdomen. The RCOG guidance recommended surgery for 70/93 (75%), whereas simple rules recommended surgery for 11/93 (12%). (Table 8.6-1) There were 13/93 (14%) women with a raised CA125 (≥30). When looking purely at the RMI result 48/93 (52%) women had an RMI ≥ 25. (Table 8.6-2)

There were 16 patients who had surgery and 59 who had follow up ultrasound scans.

Amongst those who had expectant management, follow-up occurred between 6 and 21months with an average of 12 months and no missed malignancies were detected. Most tumours were unchanged (49/59 [83.1%]) and only 5/59 [8.5%] were larger (>20% increase in the largest diameter). (Table 8.6-3)

The women determined to have tumour of low risk by either protocol, all had benign tumours as qualified by histology or an ultrasound scan which continued to have benign features. (Table 8.6-4) There were 2 malignancies in the entire group. The first was assessed as malignant or of high risk by both protocols and the second was assessed as being of indeterminate or intermediate risk by both protocols. Appropriately, both women with a malignancy on histology were identified correctly for surgery by both protocols. The decision for surgery was made independently of either protocol.

Figure 8.6-1: Flow Diagram of Pilot Study. N = 93.

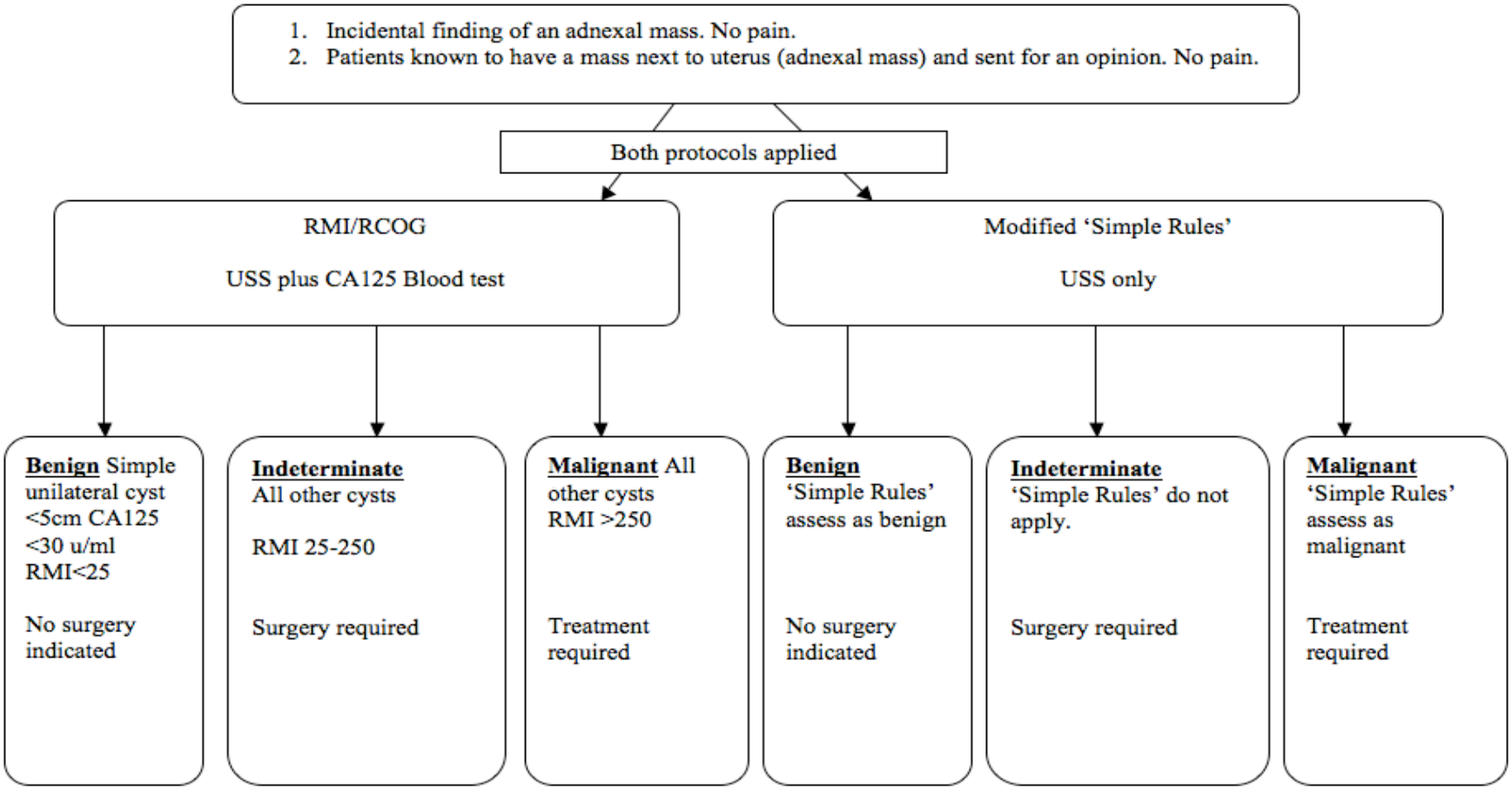


Table 8.6-1: Results of the Assessments for all Patients - RCOG/RMI and Simple Rules (Pilot Study). N = 93.

	RCOG/RMI	Simple Rules
Benign / Low Risk – No surgery	23 (24.7%)	82 (88.2%)
Indeterminate Risk - Surgery	64 (68.8%)	8 (8.6%)
Malignant / High Risk - Surgery	6 (6.5%)	3 (3.2%)
Surgery recommended	70 (75.3%)	11 (11.8%)

Table 8.6-2: Results of RMI for all Patients (Pilot Study). N = 93.

RMI	Number of patients (%)
<25	45 (48%)
≥ 25 - <250	42 (45%)
≥ 250	6 (7)

Table 8.6-3: Patient Outcomes (Pilot). N = 93.

Surgical Management	
	Number of Patients (%)
Benign	14 (87.5%)
Borderline	0
Malignant	2 (12.5%)
Total	N=16 (17.2% of total 93)

Expectant Management	
	Number of Patients (%)
Resolved	3 (3.9%)
Smaller	3 (3.9%)
Same	65 (84.4%)
Larger	6 (7.8%)
Total	N=77 (82.8% of total 93)

Table 8.6-4: Breakdown of Risk Categories as assessed by both Protocols (Pilot Study). N = 93.

	RCOG/RMI			**Simple rules**		
Low Risk	**Total**	**N = 18**		**Total**	**N = 82**	
	Surgery	0		Surgery	10	Benign (n=10) Malignant (n=0)
	Expectant	18	All Benign	Expectant	72	All Benign
Intermediate Risk	**Total**	**N = 64**		**Total**	**N = 8**	
	Surgery	13	Benign (n=12) Malignant (n=1)	Surgery	4	Benign (n=3) Malignant (n=1)
	Expectant	51	All Benign	Expectant	4	All Benign
High Risk	**Total**	**N = 6**		**Total**	**N = 3**	
	Surgery	3	Benign (n=2) Malignant (n=1)	Surgery	2	Benign (n=1) Malignant (n=1)
	Expectant	3	All Benign	Expectant	1	Benign

8.7 Study 7 – Comparison of the use of two ultrasound-based protocols for the management of asymptomatic post-menopausal women with adnexal tumours – A randomized controlled trial

8.7.1 Introduction

In recent years, there has been an increase in the number of women diagnosed with asymptomatic adnexal tumours. This is mainly due to liberal use of modern high-resolution abdominal and pelvic imaging such as ultrasound, MRI and CT. The diagnosis of an adnexal tumour often causes concerns to women and health professionals due to perceived difficulties in distinguishing between benign and malignant ovarian tumours.

The best survival for patients with ovarian cancer is achieved when treatment is organised and carried out by gynaecological oncologists who work in cancer centres. (Vernooij, 2007) However, patients with asymptomatic benign tumours can be managed expectantly or by minimally invasive surgery if symptomatic. General gynaecologists in their local hospitals can safely undertake these procedures. Achieving the right balance between desire to treat all ovarian cancers effectively and without delay against the need to avoid unnecessary interventions in women with benign and clinically insignificant lesions is particularly important in postmenopausal women. They are more likely to have ageing-related, pre-existing co-morbid conditions such as cardiac or pulmonary disease or diabetes mellitus that increase their risk of surgical and anaesthetic complications. They are also more likely to have had previous abdominal operations

increasing further their risk of surgical complications. (Ansaloni, 2010; Neufeld, 2013, Sztark, 2013; RCOG, 2003)

In 2003, the RCOG published a guideline on the management ovarian cysts in postmenopausal women (RCOG 2003). The guideline uses the RMI model-based management protocol to facilitate triaging of women with adnexal tumours for surgery and it provides advice whether the operation should be carried in a specialised cancer centre or by a general gynaecologist. In our previous study, we found that there may be significant differences in the number of postmenopausal women with adnexal cysts who are selected for surgery when the RMI based protocol is compared to 'Simple rules', which is a more recent method to diagnose ovarian cancer on ultrasound. This difference occurred despite previous meta-analyses showing that both tests have similar accuracy for the diagnosis of ovaran cancer. (Geomini – RMI, Nunes - Simple rules) (Geomini, 2009; Nunes, 2012)

The aim of this prospective randomised trial was to determine whether there are significant differences between the two management protocols in the proportion of postmenopausal women with incidentally detected adnexal tumours, who are triaged for and who receive surgical treatment.

8.7.2 Methods

This was a prospective non-blinded single-centre randomised controlled trial conducted in a general gynaecology clinic at University College Hospital, London, UK. Asymptomatic postmenopausal women who were diagnosed with adnexal tumours on

ultrasound scan or other imaging modalities were invited to join the study.

The North London Research Ethical Committee 2 ethics committee approved the randomised controlled trial (10/H0724/48) as well as the research and development committee at UCH.

8.7.2.1. Participant Identification

Eligible patients were identified when they attended the general gynaecology clinic and were found to have an adnexal tumour and no pain. An information leaflet about the study was given to all eligible women before assessment. A member of the research team, who took a detailed clinical history, did a physical examination and performed an ultrasound scan then reviewed them. If the diagnosis of an adnexal mass was confirmed women were invited to join the study. Written informed consent was obtained from all women who agreed to take part after which they would be randomized as described below and in section 7.6 and then proceed to have their assigned protocol assessment performed.

8.7.2.2. Inclusion and exclusion criteria

Women were eligible for inclusion if they were postmenopausal and found to have an adnexal tumour. They were considered asymptomatic if they did not present with pain localised to the area of the cyst or the lower pelvis. Postmenopausal women were defined as those who have had one year of spontaneous amenorrhoea at or above the age of 40 where no illness or medication may have caused the cessation of periods or

those at or above the age of 50 who have had a hysterectomy with ovarian conservation. Women aged 40 to 80 inclusive were eligible. We excluded all women who were unable or unwilling to give written consent and those with simple, unilateral, unilocular cysts of less than 2cm. (Table 8.7-1) Women referred from other hospitals as tertiary referrals to the Oncology clinic and those who were considered unfit for surgery were also excluded.

8.7.2.3. Randomization and Blinding

A blocked randomisation list with varying block sizes was generated by an independent statistician using a Stata 12.1 (Stata Corp., College Station, Texas, USA). The randomisation numbers were placed in consecutively numbered, sealed, opaque envelopes and kept in a box locked in a filing cabinet. This randomisation ensured allocation concealment. When a patient consented, a clinic nurse, who was not part of the research team, opened the next envelope and informed the recruiting doctor of woman's trial allocation.

Neither the patient nor the recruiting doctors were blinded to the protocol allocation but the operating surgeons and the pathologists were. The surgeons were informed whether the tumour was of high or intermediate risk of cancer, but not about the protocol which was used to determine that. The pathologist was informed of the operating findings only and their report was used to determine the nature of adnexal tumours.

8.7.2.4. Procedures: Trial Arms and Interventions

The women were randomised into either of two groups. The first group was assessed and managed in accordance with the current Royal College of Obstetricians and Gynaecologists (RMI/RCOG) protocol for the management of cysts in postmenopausal women. (RCOG, 2003) The second group of women were assessed using a structured approach to morphological analysis of ovarian tumours referred to as 'simple rules' (SR), which enables discrimination between benign and malignant lesions without the need to measure tumour markers. According to the literature both methods provide similar levels of accuracy for the diagnosis of ovarian cancer. (Jacobs, 1990; Geomini, 2009; Timmerman, 2008; Nunes, 2014)

In the RMI/RCOG arm all the women underwent an ultrasound examination and had a CA 125 blood test on the same day. The RMI was calculated as the product of the CA125 (U/ml) laboratory value, a score for menopausal status (1 if pre-menopausal and 3 if postmenopausal) and a greyscale ultrasound score of 0, 1 or 3 where 1 score is given each for bilaterality, ascites, multilocularity, solid areas and intra-abdominal metastases. (RMI = CA125 x M x U)

Women with simple unilateral unilocular cysts < 5cm, CA 125 <30 U/ml and RMI<25 were classified as low risk of cancer and were offered conservative management. Those with a RMI of 25 – 250 were considered to be at indeterminate risk and were offered surgical removal by gynaecologists. Those with a RMI<25 but with a cyst which is not simple, unilateral, unilocular and <5cm were also considered of indeterminate risk. Women with an RMI >250 were referred for surgery to the tertiary cancer centre as they were considered high risk. (Figure 8.7-1)

The second trial arm used a management protocol based on 'Simple Rules' (SR) diagnostic model. Women in this arm had an ultrasound scan assessment according to the simple rules protocol. The SR model uses 10 rules to assess adnexal masses. There are five rules that predict malignancy (M-rules): (1) irregular solid tumour; (2) ascites; (3) at least four papillary structures; (4) irregular multilocular–solid tumour with a largest diameter of at least 100 mm; and (5) very high colour content on colour Doppler examination (score 4). There are five rules to predict benignity (B-rules): (1) unilocular cyst; (2) presence of solid components where the largest solid component has a largest diameter of <7 mm; (3) acoustic shadows; (4) smooth multilocular tumour less than 100 mm in largest diameter; and (5) no detectable blood flow on Doppler examination (score 1). If one or more M-rules apply in the absence of a B-rule, the mass is classified as malignant. If one or more B-rules apply in the absence of an M-rule, the mass is classified as benign. If no rule applies or both M and B rules apply, the mass cannot be classified and is considered indeterminate. Ultrasound scans were performed in section 7.3 and 7.4.

Women with a cyst classified as benign using 'simple rules' were considered low risk and they were offered conservative management. Those classified as malignant were high risk and were managed in the tertiary oncology unit. Women with tumours of indeterminate nature were considered intermediate risk and were offered surgery by general gynaecologists.

Women selected for conservative management were offered 3-monthly ultrasound scans for 12 months. Those located to the RMI/RCOG arm had another CA125 test at their final follow up visit. At follow up visits women were asked about any change in their

symptoms. The tumours were assessed on ultrasound scan looking for any changes in their morphology and size.

The decision about the surgical approach; i.e. open versus minimally invasive, was made by the operating surgeon.

8.7.2.5. Outcome Measures

Primary Outcome

The primary outcome was the proportion of women who were selected for surgery using the two assessment protocols. We analysed both the assigned and the actual surgical intervention rates in both arms of the trial within 12 months from randomisation.

Secondary Outcomes

Secondary outcomes included the number of staging surgical procedures carried out by a gynaecological oncologist, number of women suffering surgical complications and the number of delayed diagnoses of ovarian cancer.

8.7.2.6. Statistical analysis

Pilot Study

A pilot study was conducted on a sample of 67 women. (Nunes, 2012) In this group of women 41/67 (61.2%; 95% CI: 48.5% to 72.9%) would be offered surgery according to the RCOG guideline while only 5/67 (7.5%: 95% CI: 2.5% to 16.6%) would be offered surgery according to the simple rules guidance.

Sample Size

We estimated that by using the 'Simple Rules' management protocol the intervention rate would be halved from 62% to 31%. To achieve a 5% significance level and 90% power, 59 patients were required in each group. Allowing for a 20% drop out rate, we planned to recruit 148 women.

Statistical tests

A Chi-square test was used to assess the significance of the difference between the two intervention rates. Fisher's exact test was used to assess the significance of the difference between the complication rates and gynaecological oncology referrals. The relative risk (RR) ratio with 95% confidence intervals (CI) was used to compare the numbers of women triaged for surgery and per protocol intervention rates.

An intention to treat analysis was done of all patients who fit the inclusion criteria. The analysis of surgery offered included all randomized patients and the analysis of those who had surgery included those who had surgery within 1 year or completed 1 year of follow-up ultrasound scans.

At the end of the study, an independent statistician (GA) analyzed the results. Statistical analysis was performed using Stata 14 software (Stata Corp, College Station, TX, USA).

Any adverse events were assessed for severity according to the University College Hospital Trial Coordinating Centre's criteria and recorded in the database.

8.7.2.7. Ethical Committee Approval

The North London Research Ethical Committee 2 ethics committee approved the randomized controlled trial (10/H0724/48) as well as the research and development committee at UCH and it was entered in the registry of randomized trials (ISRCTN89034131). http://www.controlled-trials.com/ISRCTN89034131/.

Figure 8.7-1: Flow Diagram of Patient selection and randomization for the RCT simple rules versus RCOG/RMI

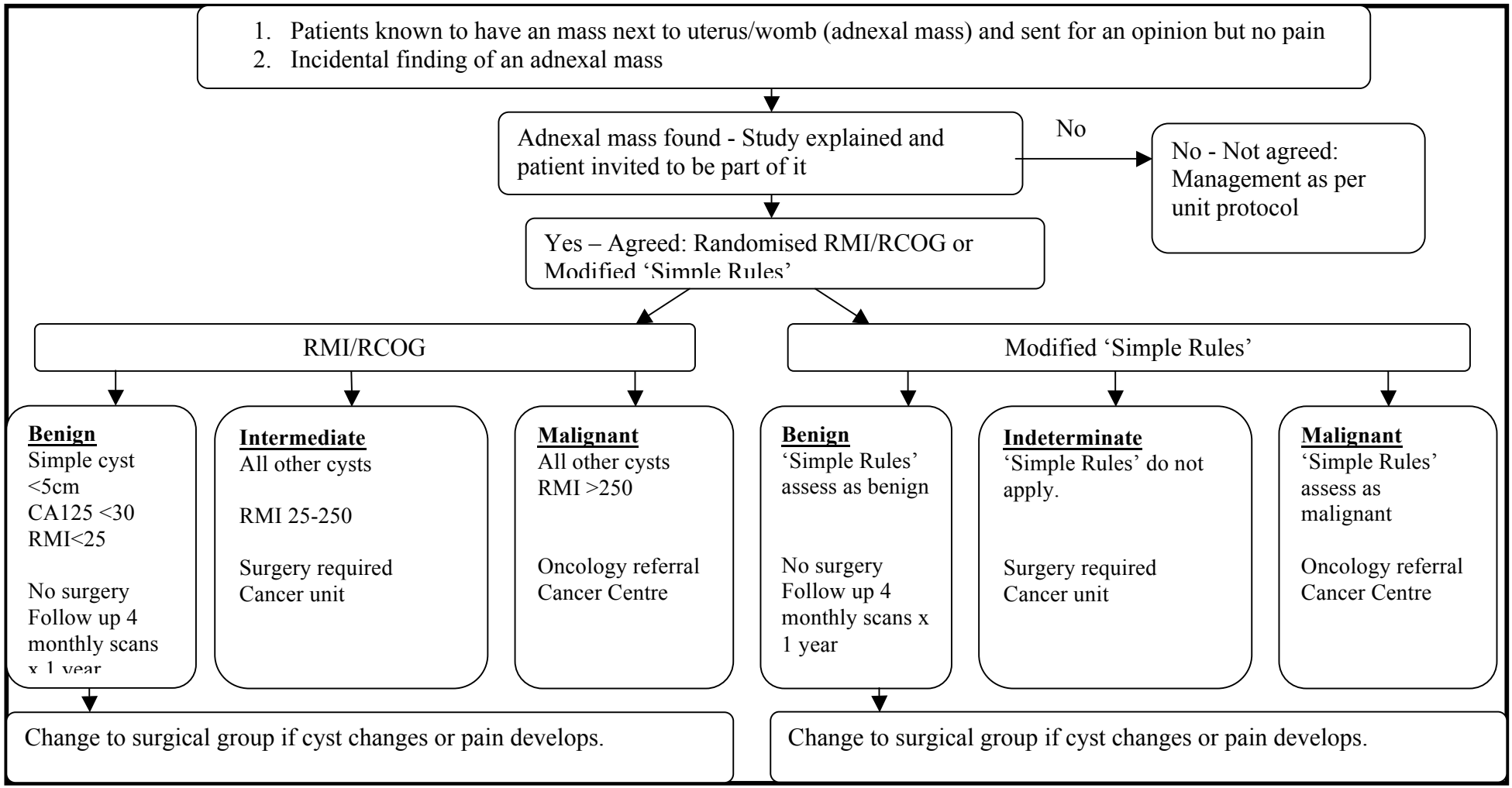
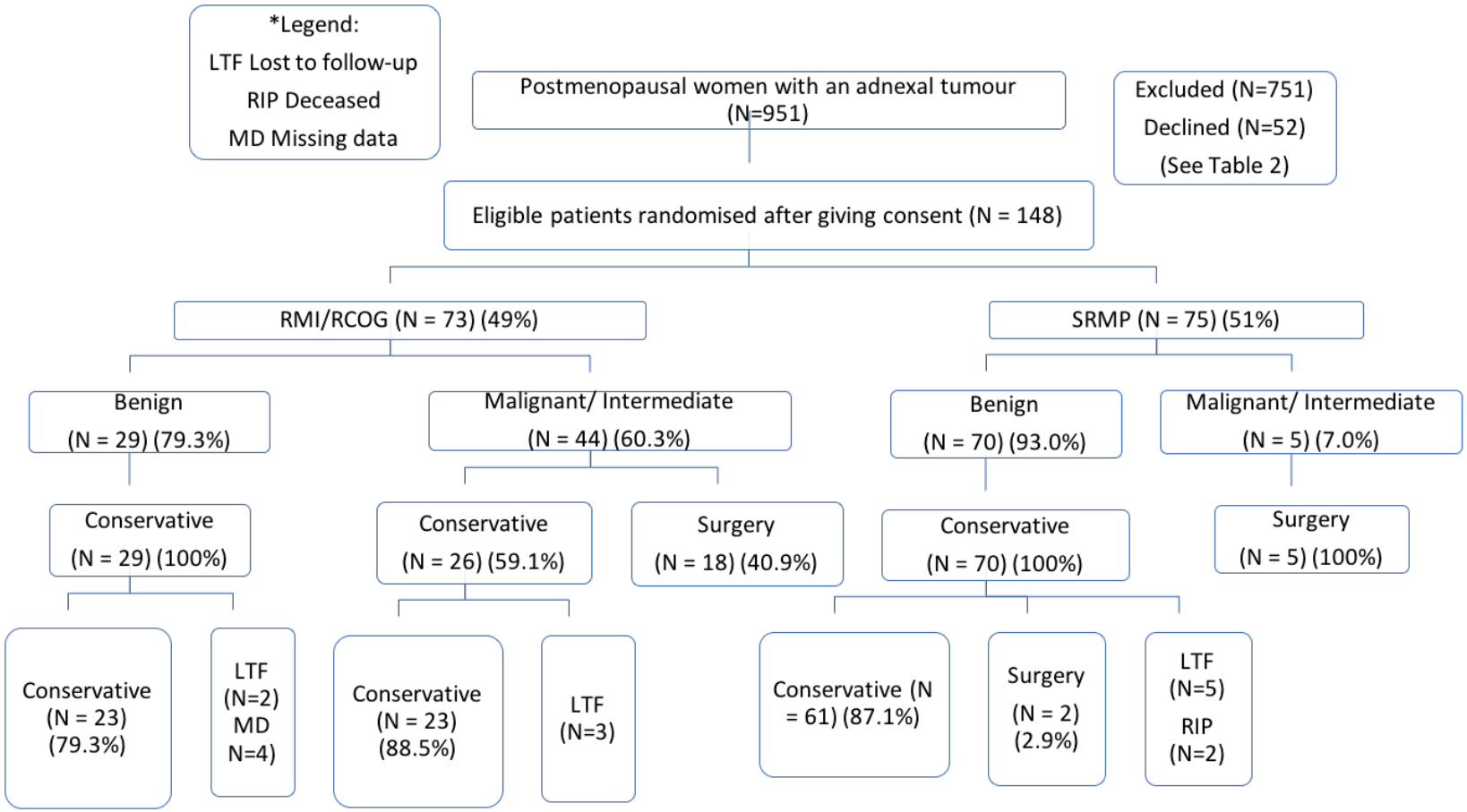


Figure 8.7-2: CONSORT Flow Diagram of Results – RCT (N = 148)



8.7.3 Statistical Analysis and Results

From April 2011 to June 2014, 951 postmenopausal women were diagnosed with adnexal tumours in our centre. 751 were excluded for various reasons. (Table 8.7-1) There were 200 women who fulfilled the inclusion criteria, 148 of whom agreed to take part in the study. 73 women were randomised to RMI/RCOG and 75 to SR. Their ages ranged from 49 to 80 with a mean age of 63. Both groups were similar in their demographic characteristics. (Table 8.7-2) Follow-up at 12 months was completed in June 2015.

8.7.3.1. Data Analysis

Primary Outcome

8.7.3.2. Intention to Treat Analysis

With an intention-to-treat analysis, a significantly higher proportion of women were assigned to surgical treatment using the RMI/RCOG protocol compared to SR [44/73 (60.3%: 95% CI: 48.1% to 71.5%) versus 5/75 (6.7%: 95% CI: 2.2% to 14.9%). ($P < 0.001$, $\chi 2$ test) (Table 8.7-3) The relative risk was 9.04 (95% CI: 3.80 to 21.5) so there was a 9 times higher chance of an assessment of Indeterminate or Malignant by RMI/RCOG compared to assessment via Simple Rules. The difference in these proportions is 53.6% (95% CI: 41.0% to 66.2%) so therefore, for every 100 patients we expect an additional 53 patients to be assessed as either Indeterminate or

Malignant by RMI/RCOG compared to assessment via SR.

Initial Management

Out of the 44 women triaged for surgery in the RMI/RCOG arm, 18 (40.9% 95% CI)) had surgery as planned, 7 (15.9% (95%) were not fit for surgery and 19 (43.2%95%) declined the operation for various reasons. Those judged as benign (29/73 [39.7%: 95% CI: 28.5 to 51.9]) all had conservative management. (Figure 8.7-2)

All five women assigned to surgery in the SR arm had operations. The remaining 70 women (93.3%: 95% CI: 85.1% to 97.8%) had their tumours classified as benign. 68/70 (97.1% 95%) of them were managed conservatively whilst the remaining 2/70(2.9%95%) had surgery: one woman had pelvic clearance for concomitant endometrial cancer and the other had a concurrent breast cancer and she was referred to the oncology unit by her breast physician for risk reducing surgery. In both of them the adnexal lesions were benign on histological examination.

A total of 136 (91.9%) women had surgery or had completed 12 months follow-up, whilst 12 (8.1%) women were lost to follow up. There were 18/68 (26.5%: 95% CI: 16.5% to 38.6%) women in the RMI/RCOG arm who had surgery versus 7/68 (10.3%: 95% CI: 4.2% to 20.1%) women in the SR arm. ($P < 0.015$, $\chi 2$ test). The relative risk was 2.57 (95% CI: 1.15 to 5.76) so there is a 2.6 times higher chance of surgery following assessment by RMI/RCOG protocol compared to assessment by SR. The difference in these proportions is 16.2% (95% CI: 3.4% to 28.9%) so therefore, for every 100 patients we expect an additional 16 patients to have surgery after assessment by RMI/RCOG compared to assessment via Simple Rules.

Secondary Outcomes

Two women, one in each arm were identified as having high-risk tumours and they both had staging procedures (one laparoscopy and one laparotomy) by gynaecological oncologists. In both cases the histological diagnosis was a borderline tumour. Additional six women with low or intermediate risk tumours were referred to the oncology centre for the following indications: concomitant breast cancer (1), concomitant endometrial cancer (1), metastatic bowel cancer (1), technically difficult surgery (1) and women's request (2). There were no significant differences between the proportions of women referred to the oncology units between the two groups of women (3/73 (4.1%) in RMI/RCOG and 5/75 (6.7%) in SR. The risk difference is 2.6% (-4.7% to 9.8%). The relative risk is 1.62 (0.40 to 6.54). (P=0.72)

Surgical complications occurred in 4/73 (5.5% (%% CI) women in the RMI/RCOG arm and 2/75 (2.7% 95%CI) in the SR arm. (P = 0.44) Most complications were minor. (Table 8.7-4) There were no delayed diagnoses of ovarian cancer detected up to 12 months of follow up.

Histological diagnoses are listed in Table 8.7-5. There were 25 women who had their surgery within the 1st year following the initial assessments. Six women with malignant tumours on histology (2 invasive tumours and 4 borderline tumours) were identified as high or intermediate risks by either diagnostic model. Five of them had surgery within two months of the initial visit whilst the remaining woman with a cardiac condition had her surgery delayed for eight months to optimise her general health.

23/26 women who were assessed as needing surgery but were either deemed surgically unfit (7) or declined operations (19) completed their 12 months' follow-up. Two of them went on to have surgery after one year. They were found to have benign lesions on histological examination. Of the remaining three women who did not attend for follow up visits, one had her 12-month scan elsewhere. The examination did not reveal any significant changes in the appearance of her adnexal lesion. We had no information about the remaining two women who were lost to follow up.

Table 8.7-1: Patients excluded pre-randomisation in RCT. N = 803

Reason for Exclusion	**Number**
Small - <2cm, unilocular, anechoic cyst	269
Oncology tertiary referral	179
Age <40 or >80	121
Pelvic pain	87
Declined	52
Severe pelvic adhesions	30
Surgery planned/ Urogynaecological management	24
Familial Cancer Clinic/UKCTOCS	21
Longstanding Patient – stable cyst	8
Cognitive disability	3
Transabdominal scan only	2
Ongoing VTE on treatment	1
Other	6
Total excluded	**803**

Table 8.7-2: Baseline patient characteristics by trial arm in the RMI/RCOG and the 'Simple Rules" arms of the RCT. N = 148.

		RMI / RCOG (n = 73)	**Simple Rules (n = 75)**
Age	(years)	62.7 (7.6) (range: 49 to 80)	63.8 (8.1) (range: 50 to 80)
Parity	0 1 2 ≥3	22 (30%) 11 (15%) 20 (27%) 20 (27%)	27 (36%) 8 (11%) 19 (25%) 21 (28%)
Time since menopause	(years)	12.5 (8.8) (range: 0 to 37)	13.7 (9.3) (range: 1 to 33)
	UK*	2	3
Current HRT use	yes	10 (14%)	11 (15%)
Past HRT use	yes no n/a UK*	12 (16%) 49 (67%) 10 (14%) 2 (3%)	19 (25%) 45 (60%) 11 (15%) 0

*Unknown

Table 8.7-3: Results: Patient Outcomes in RCT. N = 148.

Results / Outcome	**RMI/ RCOG**	**Simple rules**
Completed 4 months	68	73
Completed 6 months	64	65
Completed 8 months	63	62
Completed study	61	58
Protocol = Benign	29	70
Protocol = Indeterminate	43	4
Protocol = Malignant	1	1
Protocol = Surgery	**44**	**5**
Surgery ≤ 1 year	**18**	**7**
Declined surgery	26	0

Table 8.7-4: Complications experienced in RCT N = 25

Results / Outcome	**RMI/ RCOG**	**Simple rules**
Unable to remove other ovary due to adhesions	2	
Left ear bleeding post-operatively- ENT review	1	
Malignant cyst rupture, Wound infection	1	
Small bowel injury		1
Acute dyspnoea and fluid overload		1
Total surgical complications	**4 (5.5%)**	**2 (2.7%)**
Nil	13	5
Total	18	7

Table 8.7-5: Histological Diagnoses at 1 year in RCT N = 25

Histology		**RMI/ RCOG**	**Simple rules**
Benign	Cystadenoma/ Cystadenofibroma (2 with a Fibroma and 1 with a Brenner tumour)	8	3
	Endometriomas	3	
	Fibroma	2	
	Fibrothecoma	1	
	Benign stromal tumour and simple cyst	1	
	Inclusion cyst with hydrosalpinx	1	
	Total Benign	**16**	**3**
Borderline	Borderline serous	1	2
	Borderline mucinous (with a contralateral benign serous cystadenoma)		1
	Total Borderline	**1**	**3**
Invasive Malignancies	Papillary thyroid type carcinoma within a dermoid		1
	Metastatic colorectal cancer	1	
	Total Invasive Malignancies	**1**	**1**
Total Histology		**18**	7

Part IX – Discussions

9.1 Study 1 Discussion – A prospective validation of the IOTA Logistic Regression Models (LR1 and LR2) in comparison to Subjective Pattern Recognition for the diagnosis of ovarian cancer when utilised by an average ultrasound operator assessing adnexal lesions

Study 1A

This study has shown that the accuracy of the IOTA Logistic Regression Model (LR2) was similar in the hands of a non-expert operator compared to the findings in the original IOTA study. Although the AUCs were not significantly different between our study and the two other previous studies conducted by experts, there were some differences. The sensitivity was higher in our study than in the original study, but not in comparison to the prospective IOTA validation, which was also conducted by experts. (Timmerman, 2005; Timmerman, 2010) The specificity in our study was lower than that reported in other two previous studies, particularly in comparison to the validation study. These differences could be explained by the operator-effect on the performance of the model. It has been previously documented that the specificity of ultrasound diagnosis is higher when expert operators perform the examinations. (Yazbek, 2008; Timmerman, 2000) Ultrasound experts performed the assessment of ovarian tumours

during the previous IOTA studies. During data collection, the operators were also able to assess tumour characteristics using "pattern recognition". In many cases the operators could have determined the type of ovarian tumour before collecting the data using the IOTA protocol. This is a potential source of bias, which could have lead to an overestimate of the model specificity.

We considered that with time the experience of our Level II operator would increase, which could influence the performance of the IOTA model. Although both the sensitivity and specificity of the model were better in the second half of the study, the differences were not statistically significant. The Level II operator was under very clear instructions not to attempt subjective diagnosis of the nature of ovarian lesions, the aim of which was to minimise bias.

Our study had a larger component of malignant tumours; 53.2% (66/124) versus 24% (75/312) in the original IOTA study. (Timmerman, 2005) The validation studies also had a smaller population of malignant tumours; temporal 30.5% (287/941) and external 25.6% (255/997). (Timmerman, 2010) The higher prevalence in our study reflects the nature of work in a tertiary referral cancer centre with a high proportion of women attending with a strong suspicion of ovarian cancer. In addition, a large proportion of women with a presumed benign lesion are managed expectantly, which reduces the number of women with benign disease undergoing surgery. This higher prevalence of cancer may have had a positive effect on the positive predictive value, but it should not have affected the sensitivity or specificity of the test.

A very high sensitivity of the model, particularly in the non-expert hands, is reassuring and it indicates that the LR2 model could be used as a primary test in women with

adnexal tumours without fear of missing a significant number of malignant lesions. The LR-ve of the model is also high and therefore women with negative results could be managed conservatively or by using minimally invasive surgery.

The large number of false positive results in our study was mainly caused by the presence of solid areas within 15 of the 18 (83%) incorrectly classified benign lesions. In some cases, hyperechoic areas in benign cystic teratomas or precipitated debris in ovarian endometriomas were misclassified as solid components within the cyst. Ascites due to other non-malignant medical conditions also contributed to false positive results as the model assumes that the adnexal lesion is the cause of the ascites. The positive predictive value was 78.1%, which means that more than a fifth of presumed malignant lesions were in fact benign. Women with positive results suggestive of cancer would therefore require additional tests to check the accuracy of the diagnosis to avoid subjecting many women with benign lesions to unnecessary major gynaecological staging operations.

Study 1B

In the hands of a level II or non-expert operator, the AUC of both IOTA Logistic Regression Models (LR1 and LR2) performed similarly to the previous studies, which were conducted by experts. A closer scrutiny of the results showed that, although the AUCs did not differ, there were significant differences in both sensitivities and specificities between this and the previous studies. Typically, the logistic models tended to be more sensitive, but less specific when used by the non-expert in this study. I hypothesised that these differences were caused by the ability of the expert operators to

use pattern recognition while performing the IOTA logistic regression models. It has been demonstrated that the sensitivity of pattern recognition by the expert operator for diagnosis of cancer in this study was slightly lower compared with the models, but the specificity was much higher. This mirrors the comparisons in the performance of LR1 (and to some extent LR2) with the original studies. This suggests that the increased specificity of the models when tested by the experts could be due to it being performed by experts.

I considered whether the experience of the Level II operator would increase during the time of the study and therefore influence the performance of the IOTA model and hence compared the first half of the study with the second half. Although both the specificity and the AUC of the models were better in the second half of the study, the AUCs were not significantly different. As part of the study, I as the Level II operator did not attempt subjective diagnosis of the nature of the ovarian lesions. The goal was to help to avoid bias.

While the prevalence of cancer in our final surgical population was higher at 45.2% as compared with 24% (75/312) of the original IOTA study, the overall prevalence of cancer in our entire population of women was in fact 23.8%. Many women with benign lesions in our population opted for expectant management, which resulted in a relatively low proportion of women with benign disease undergoing surgery. Our higher proportion of malignant cases also reflects the nature of working in a tertiary referral centre. It is important to stress, however, that tertiary referrals accounted only for 31% of the total number of women examined by the non-expert operator.

We considered the possibility that prior knowledge that individual patients were high

risk could have contributed to the high false positive rate and lower specificity of the test. However, this had no effect on the performance of the pattern recognition, which is a test more open to bias, and therefore it is unlikely the population characteristics influenced the cancer risk assessment using logistic models.

Strength and Limitations

Strengths –

This study matched the original study exactly, which allowed us to make an accurate comparison of use of the model by a level II operator as compared with the original study.

Limitations –

This study did not compare malignant and borderline tumours separately. We had a single level II ultrasound operator and this would theoretically reduce the generalisability of this study.

9.2 Study 2 Discussion – A prospective validation of the IOTA Simple Rules for the diagnosis of ovarian cancer

When a level II operator performed the ultrasound scans, the simple rules protocol performed acceptably well. The rules were applicable in nearly 80% of women in our population, which is a similar percentage when compared with the original and validation IOTA studies, and when the rules were applicable the sensitivity was not significantly different. The specificity though was less than that in the 2010 validation study ($p< 0.001$) though similar to the original study ($p=0.48$).

Forty-four percent (44.6%) of our final population had malignant tumours, which may have affected our results. This reflects working in a tertiary referral unit but more so it reflects the fact that many of the women assessed to have benign tumours by the expert, chose not to have a surgical intervention (40% of total population). The overall malignancy rate in the entire population initially assessed was 24%, which is similar to the rates in the original (27%) and validation (25%) studies.

The 4 false negative tumours were either unilocular as in the case of the borderline tumour and / or had no demonstrable blood flow as was the case in all but the stage III tumour. The later tumour demonstrated acoustic shadowing, which caused it to be misclassified. These 4 tumours had none of the M features.

The 15 false positives had a range of features, which caused them to be misclassified by the simple rules. The actinomycosis, one of the endometriomas and the fibbrothecoma appeared to be irregular and solid. The other tumours were either >100mm and irregular

multilocular solid, had >4 papillary projections or had a high colour blood flow score. None of these patients had ascites present.

Pattern recognition performed by the expert operator was superior to the simple rules in cases where the rules were applicable mainly due to significant differences in specificity [88.6% vs. 93.2% (P=0.03)]. All ultrasound operators who use simple rules will need to determine what to do though for the tumours which cannot be classified by simple rules. Assuming malignancy for these will improve sensitivity but at the expense of a worsening specificity due to the increased false positive rate. The overall accuracy therefore deteriorates. Utilising an expert instead for this population, appears to be the best choice, as the overall accuracy is maintained. As experts are not widely available, simple rules provide a good alternative, which should be able to classify 78% of the tumours on average.

Strength and Limitations

Strengths –

The methodology matched the original study allowing an accurate assessment and comparison of the performance of the simple rules when performed by a level II versus a level III operator.

Limitations –

Similar to Study 1, we did not compare malignant and borderline tumours separately which should be considered for future studies. We had a single level II ultrasound operator and this would theoretically reduce the generalisability of this study.

9.3 Study 3 Discussion – A prospective evaluation of the IOTA Logistic Regression Models (LR1 and LR2) in comparison to Subjective Pattern Recognition for the diagnosis of ovarian cancer in the outpatient setting using two reference standards

Our study has shown that accuracy of LR1/LR2 for the diagnosis of ovarian cancer in the outpatient setting was similar to the previous studies, which were carried out on women who were scheduled for surgery. However, if LR1/LR2 had been used as the primary diagnostic test to guide the management decisions, rather than the pattern recognition, a significantly higher proportion of women would have been referred for treatment by gynaecological oncologists because of suspected ovarian cancer.

In previous original and subsequent validation studies the IOTA logistic regression models provided accurate diagnosis of ovarian cancer both in hands of expert and non-expert operators. (Timmerman, 2005; Timmerman, 2010; Nunes, 2012; Nunes, 2013; Kaijser, 2014) The models; however, had always been used in population of women who all had surgery. In view of that the results could be affected by selection bias and they cannot be extrapolated to low risk population majority of whom do not require surgical intervention. The results of previous studies can therefore only be used to help to select a surgeon (general or oncological) who should do the operation or the route of the surgery (laparoscopic or open). (Campbell, 2012) A more relevant question in clinical practice; however, is whether surgery is required at all.

A difficulty in conducting studies on populations of women with adnexal

tumours who are managed conservatively is the lack of agreement on the reference standard to define the nature of the lesion. In women with presumed benign lesions the only way to rule out an ovarian cancer in the absence of histological diagnosis is by arranging follow up visits for a certain length of time. The natural history of ovarian cancer is unknown and the decisions and about the length and frequency of follow visits are pragmatic and based on consensus of opinions, rather than science. Most ovarian cancer screening projects adopted the policy of six-monthly or annual visits in women with normal ultrasound findings, which is deemed to be sufficiently frequent to detect early disease before spreading beyond the ovaries. All women in our study had detectable lesions at the time of the initial scan. We therefore postulated that under these circumstance a 12 month follow up should be long enough to detect changes in the appearance and size of adnexal tumours which would be suggestive of their malignant nature. The absence of such changes was our second reference standard to discriminate between benign and malignant lesions.

During follow-up, it is possible that some woman could develop new abnormalities. This may erroneously be classified as a prior misdiagnosis rather than the new disease that it is. However, this limitation can be overcome.

There is no single cut-off point before which a cancer diagnosis is a misdiagnosis and after which it is a new malignancy. Nonetheless we used 12 months as this point in keeping with screening studies that suggest 1 year is sufficient time. (Tailor, 2003; Woodward, 2007) Several authors have found that in their studies, women with a prior benign-appearing adnexal tumour who subsequently develops an ovarian malignancy, is likely to have a borderline tumour or a slow-growing Type I tumour. (Horiuchi, 2003; Shuh, 2004; Sharma, 2012; Suh-Bergmann, 2014)

All the women in the study were assessed by one of the reference standards: (histology and a follow-up ultrasound scan in 12 months or more. This helped us to reduce the incomplete verification bias of including only surgical patients. Bias may though still arise if the results of the alternative reference standard are treated identically to the results from the preferred reference standard when they may not be identical. This is because the two reference standards may be of different quality and it is possible that follow up ultrasound scans would be less than100% accurate when compared to histology (preferred reference test). This information was considered when performing the statistical analysis, but in our study the results were nit different when the probability of ultrasound follow up (alternative test) was reduced to 90%. (de Groot, 2011)

Our results showed that in this population with a 28.0% malignancy rate the LR1 and LR2 models had a high sensitivity and a moderate specificity to diagnose ovarian cancer. The specificity of LR1 model was significant lower than pattern recognition ($P < 0.0001$) whilst the sensitivity was significantly different ($P=0.13$). This relatively high false positive rate would result on 62% more women being treated for potential ovarian cancer. However, four women with false negative diagnosis ovarian cancer (one invasive epithelial and three borderline) would receive earlier treatment.

Strength and Limitations

Strengths –

This study assessed and corrected for the potential bias of analysing only the women who had surgery. This is the first study on LR1/LR2 to do this.

Limitations –

The overall intervention rate in our study was relatively high which reflects the nature of work in a large clinical centre. Many women were referred following the diagnosis of an adnexal cyst in primary or community care. They often had larger lesions and were advised by their GPs that surgery was required even if the lesions were benign.

9.4 Study 4 Discussion – A prospective evaluation of the IOTA simple rules as a triage tool for the diagnosis of ovarian cancer in the outpatient setting using two reference standards

Also see discussion for study 3 (section 9.3) regarding use of follow-up ultrasound scans as a second and alternative reference test.

This study showed that simple rules performed better in the outpatient setting than IOTA models. In comparison to pattern recognition, there were no significant differences in the sensitivity and the specificity making this a useful and accurate tool. There are though, a large proportion of women (20%) for whom the rules do not apply and therefore who have indeterminate tumours. This poses significant limitations on the use of this model within a general gynaecological or ultrasound clinic as a secondary test should be readily available for the women for whom SR was unable to give a diagnosis. Pattern recognition thus far has proven itself to be the test that is able to do this.

There were fewer indeterminate tumours in the population who had expectant management (12.2% versus 22.8%) (P=0.0067) because most these tumours were benign and easier to classify as benign which makes the tool more useful in this population.

These analyses more accurately reflect the real clinical scenario faced in the general gynaecology clinic when a test needs to help the gynaecologist to determine whether surgery is needed at all and only after that, which type of surgeon should do it (general

or oncological) and the route (via laparoscopy or laparotomy). Simple rules was comparable to pattern recognition and was superior to LR1 and LR2.

Strength and Limitations

Strengths –

This is the first study on simple rules that includes women who had expectant management. This assesses and correct for a potential selection bias by only including women who had surgery.

Limitations –

There was single ultrasound operator, which may potentially affect the ability to apply the results in every unit.

9.5 Study 5 Discussion – Use of the IOTA Simple Rules for the diagnosis of ovarian cancer: a Meta-Analysis

This meta-analysis summarizes the currently available evidence concerning the accuracy of the simple rules ultrasound tool in the diagnosis of ovarian cancer. It has shown that the simple rules performed well overall for the diagnosis of ovarian cancer in hands of ultrasound operators of varying levels of expertise, with a pooled sensitivity of 0.93 (0.90 to 0.96) (I^2 of 32.1%) and a pooled specificity of 0.95 (0.93 to 0.97) (I^2 of 78.1%) when internal and external validation studies were included. The tool demonstrated a good discriminatory capacity in diagnosing ovarian cancer when the rules were applicable. The study with the lowest sensitivity and the study with the lowest specificity were those with the smallest sample sizes, which may have affected their results. When we only included the external validation studies the sensitivity decreased slightly and the specificity increased slightly. Amongst the premenopausal women the sensitivity and the overall accuracy was lower when compared with the postmenopausal population.

There appeared to be a strong and increasing relationship between the sensitivity and the proportion of women diagnosed with ovarian cancer in the study population. There was a weaker and decreasing relationship between specificity and the prevalence. Overall this suggests that when the rules are applicable, the sensitivity of simple rules increases and the specificity decreases with the increasing prevalence of malignancy in the study population. Daemen et al also saw this phenomenon and these authors reported that the specificity of the test was lower in centres with a higher prevalence, and suggested that these centres have more borderline and complex benign cases that

are harder to classify. (Daemen, 2011)

A recently published systematic review and meta-analysis included an analysis of simple rules. (Kaijser, 2013) They included all tumours unclassifiable by the simple rules as malignant and assessed data from 2 studies. (Timmerman, 2010; Sayasneh, 2013a) They found the same sensitivity as us but a lower specificity in both pre- and postmenopausal women. This is likely due to the inclusion of all indeterminately assessed tumours in the malignant group. When we did this with our current study, we found an increase in sensitivity and a significant fall in specificity.

Strength and Limitations

Strengths –

This study allows the limitations of the previous study to be overcome by including results from a range of units with a range of ultrasound scan abilities and experience. This makes the results now better generalisable.

Limitations –

While the process of literature review and meta-analysis is a practical and useful way of generating a more powerful estimate of diagnostic accuracy with less random error than the individual studies, it does present some challenges and have some limitations.

Primarily, the heterogeneity of the studies must be addressed because this may affect the justification for pooling the data into a single analysis. In this meta-analysis, diversity in study quality, nonconformity in study reporting, differences in the study population

characteristics and variation in the cancer prevalence in the study populations caused heterogeneity. Not all studies were clear on exclusion of pregnant women, which is a factor noted in the entire subset of IOTA studies. Possible reasons for exclusion of pregnant women are low risk of ovarian malignancy, increased vascularity of all tumours in pregnancy and delay in women undergoing surgery.

There were two studies that did not give the age range of their patients or the percentage of women who were postmenopausal making it impossible to determine congruity of patient characteristics. The prevalence of malignancy ranged from 10% to 44% across the studies. We overcame this limitation by assessing for the influence of this heterogeneity on the results. The study sizes also demonstrated diversity ranging from 103 to 1938 women.

Additionally, operator level was not clear in at least one study making assessment of influence of operator level not possible.

Summary

The simple rules protocol is easy to use in routine clinical practice and it is an accurate test to discriminate between benign and malignant ovarian lesions. The main limitation is the inability to apply the test to more than 20% of women who would need to undergo additional testing or an ultrasound examination by an expert who is able to use the pattern recognition method to determine the nature of the adnexal tumours.

9.6 Study 6 Discussion – A pilot study of comparison of the potential intervention rates when asymptomatic postmenopausal women with an adnexal tumour were assessed by two management protocols

In this pilot study, we found that none of the women with tumours classified as low risk with either protocol were malignant and all the malignancies were classified as being of high or intermediate/indeterminate risk. The number of women classified as having tumours of indeterminate or high risk was significantly higher using the RCOG/RMI protocol as compared to the simple rules protocol, which was predominantly due to tumours of intermediate risk. Intervention rates would therefore be over 6 times higher using the RCOG/RMI protocol as opposed to using the simple rules based protocol.

Further work was required to determine the risk of delayed diagnosis/missed cancers using simple rules, to determine the unnecessary operations done for benign tumours and to quantify the actual intervention rates when both protocols are in use in a general gynaecology clinic.

Strength and Limitations

Strengths –

This pilot study was very useful for the next study. It provided data to allow a power calculation to be done. The women had both tests conducted allowing each woman to be her own control and therefore allowing a direct comparison of how the each test would perform on the exact same tumour in each woman.

Limitations –

As the women had both tests performed we were unable to assess what final management the women would choose (surgical or expectant) when presented with one assessment or the other as neither result could be used. This would therefore need to be investigated in the next study, which was a randomised controlled trial.

9.7 Study 7 Discussion – Comparison of the use of two ultrasound-based protocols for the management of asymptomatic post-menopausal women with adnexal tumours – A randomized controlled trial

Our study has shown the proportion of postmenopausal women who are offered surgery is significantly higher when the RMI/RCOG protocol was used to triage women for surgery compared to the novel SR protocol. The number of women classified as having high-risk tumours was the same and the main difference between the protocols was the number of women who were classified as having intermediate risk lesions. It is reassuring that all malignant lesions (invasive and borderline) were correctly identified as requiring intervention and no woman with cancer had unnecessarily delayed treatment.

This significant difference in intervention rates is important and it indicates that in many postmenopausal women with adnexal tumours who are currently treated surgically in the UK, the operation may not be necessary and it could be avoided. Women classified as having tumours of intermediate risk are often anxious about the possibility of having an aggressive invasive tumour which may have adverse effect of their psychological health and social well-being. The cost of surgery is high and avoiding unnecessary operations may result is significant savings to health providers and to society as a whole.

The aim of our study was to assess differences in the intervention rates

between the two protocols based on different diagnostic models rather than to examine their accuracy for the diagnosis of ovarian cancer. This work has already been done and has shown that both methods are very sensitive in detecting malignant ovarian lesions. (Geomini 2009, Nunes 2014)

A relatively large number of women with intermediate risk lesions in the RMI/RCOG arm of the study declined surgery, which was a surprising finding. This could be partially explained by the counselling of women at the time of the initial scan which tends to place emphasis on reassurance and may downplay the risks of the observed lesion being malignant. For many of these women, the assessment of risk changed after the results of blood test to measure CA125 became available, typically a couple of days later. As all women were asymptomatic, it could have been that their natural inclination was to opt for conservative management relying more on the reassuring ultrasound findings rather taking into account the results of the blood test.. This trend was absent in the SR arm as the women assigned to this protocol were provided with instant information on the likely tumour nature based of the ultrasound findings alone.

A relatively large number of women were deemed unfit for surgery which confirms that the risks of operating on older women are increased and reiterates the need to develop effective strategies to reduce the number of avoidable procedures in this age group.
We decided to follow women up for 12 months as there is evidence from literature that this is a sufficient time span to detect changes in the morphology and size of the adnexal tumours which could indicate their malignant nature.

A study by Horiuchi in 2003 showed that most women who developed ovarian cancer in previously detected cysts had either borderline tumours or well-differentiated early

stage invasive lesions. These results support conservative approach to care of women with benign-looking adnexal lesions on initial ultrasound scans. In contrast women diagnosed with aggressive disseminated invasive ovarian cancers typically had no detectable precursor lesions on previous imaging. Shih and Kurman also reported similar findings in 2004. Similarly, Sharma et al, as part of the United Kingdom collaborative trial of ovarian cancer screening (UKCTOCS) trial, found the absolute risk of a diagnosis of an epithelial ovarian carcinoma (EOC) within 3 years in asymptomatic women with an adnexal tumour to be 1.08% (95% CI: 0.79% to 1.43%) and these were primarily borderline and Type I slow growing tumours whereas those who had normal ovaries and developed an EOC, were all found to have the more aggressive Type II tumours. (Sharma, 2012)

Strengths and Limitations

Strengths –

The main strength of our study is that this is the first randomised controlled trial looking at the use of the RMI/RCOG protocol which is widely used in routine clinical practice in the UK and compare it to a novel alternative management strategy. The novel diagnostic model had been extensively studied and all the results showed that it was suitable for the use in clinical practice. The study was carefully planned and power calculation was based on the results of a prospective audit, which had been carried out prior to starting the trial. Although it was not possible to blind women and all clinicians to the patient allocation, the pathologist and the surgeons were not aware which protocol was used in individual patients.

Limitations –

Our main weakness was that this was a single centre study, which can affect the applicability of the results in other units.

We did not blind the recruiting physician once the woman was assigned to an arm of the trial which could have influenced their approach to the conduct of ultrasound scan. We also did not blind the patients their trial allocation, which could have influenced their compliance with the protocol as they were less likely to follow a new rather than a traditional well-tested approach. Our data show; however, that the compliance was less with the standard compared to the novel management protocol. Simple rules require higher ultrasound skill that RMI which may affect the performance of the test in unit with lower ultrasound skills.

In conclusion, our study has shown that the introduction of simple rules in routine clinical practice could result in a significant reduction in the number of surgical procedures offered to postmenopausal women with incidental diagnoses of adnexal cysts on imaging. Further larger studies, including units with different levels of ultrasound expertise, are required to explore confirm these findings.

Implications for clinicians or policymakers

This study shows that relatively minor changes in the management protocols for women with adnexal tumours could result in significant differences in the number of surgical interventions. The diagnosis of suspected ovarian cancer often causes anxiety to women and their carers and they are typically urgently referred to gynaecological oncologists in

cancer centres. Algorithms based on models with low specificity would therefore cause significant pressure on sub-specialist services and further increase the cost of care. Further multicentre studies are required to confirm our findings and would facilitate development of more effective and rational care for postmenopausal women with adnexal tumours.

Part X – Conclusions and Further Research

This thesis has tested for the first time and shown that the IOTA logistic regression models LR1 and LR2 perform fairly well as a preoperative method, to determine if an adnexal tumour is benign or malignant, when utilised by an average ultrasound operator. The overall accuracy of the IOTA LR1 and LR2 models was maintained and they both demonstrated comparably good sensitivity though significantly poorer specificity when compared to when the experts used the models. Although the models' high sensitivity is reassuring the specificity is too low to allow the use of the LR1 and LR2 models as a sole test to diagnose ovarian cancer. In all women with diagnosis of ovarian cancer using LR1 or LR2 a secondary test should be used to reduce the rate of false positive findings. When available, expert operator should by used to re-examine all women with suspected cancer to reduce the rate of unnecessary interventions and referrals to regional oncology units.

This is the first work to assess use of the IOTA LR1 and LR2 models as a triaging test for use in the outpatient setting. We reduced incomplete verification bias by including all women who had histology and additionally, all women who have follow up ultrasound scans over 12 months (delayed cross sectional analysis). This has not been done before. It has demonstrated that the IOTA models maintained their high sensitivity when used in the outpatient setting. Specificity was relatively low which indicates that a significant proportion of women could be offered unnecessary surgery for suspected ovarian cancer. These results have therefore shown that LR1/LR2 can be used in the outpatient environment to triage women and determine who needs surgery. The

specificities of the models however, are too low to be used as a single tool for triaging women for surgery. It will have a lower risk of a delayed malignant diagnosis but a higher rate of a malignant misdiagnosis with the associated unnecessary anxiety and interventions for suspected cancer. Pattern recognition could be used as a second stage test to reduce the number of false positive findings and surgical interventions in this population of women. The overall intervention rate in our study was relatively high which reflects the nature of work in a large clinical centre. Further studies are required to assess the performance of IOTA models in hands of level 2 sonographers receiving direct referrals from primary care and reproductive health care physicians where the risk of ovarian cancer would be lower than in our population.

This thesis has confirmed that the preoperative diagnosis of an adnexal tumour made using the 'simple rules' also performed fairly well when used by an average operator and performed similarly to the original data. This suggests that it would be a useful screening test. It is fairly simple to use while focussing on the most useful tumour characteristics. There is though a large proportion of the population (25% on average) for whom the rules would not be applicable and who would need to consider surgery (assuming malignancy) or be offered a second test. Referral for examination by an expert using pattern recognition was the superior secondary test as opposed to assuming malignancy and offering surgery for all those unable to be defined by the simple rules as it demonstrated the greatest accuracy and higher specificity.

The meta-analysis of the simple rules included operators of varied experience and has confirmed that the simple rules can be used for the preoperative diagnosis of an adnexal mass. When the rules were applicable the model performed very well and this was particularly the case in postmenopausal women. Indeterminate tumours continued to

need a second line test.

Our dataset showed that when the simple rules model was used as a triage test in the outpatient setting with two reference tests (histology and follow up ultrasound scans over 12 months), it performed very well with a similar sensitivity and specificity as pattern recognition. It was also superior to LR1 and LR2. Use of pattern recognition for the indeterminate tumours maintained that good performance while assuming malignancy significantly reduced the specificity with an increased false positive rate. This improvement in performance was demonstrated because women, who do not have surgery, have been excluded from all previous studies, whereas we have included them. These women who have proven themselves to having benign tumours (delayed cross sectional analysis) are less likely to have indeterminate tumours as their tumours are more likely to be easily classifiable and described. This test has therefore proven to be a very useful and simple model that can be implemented in a general gynaecology clinic once it is determined which secondary test would be employed for women in whom the rules do not apply. Further work should explore the benefit of the use of other models or other tools in this population of women (20%).

The randomised controlled trial included in this thesis has confirmed the importance of test that can better discriminate between benign and malignant tumours. The RMI/RCOG protocol and the SR –based protocol both detected all the malignancies but the RMI/RCOG protocol had a significantly larger proportion of women being offered surgery for tumours of intermediate risk. These women ranged from those who should be offered surgery by the general gynaecologist to those who should be seen in a local cancer unit. Due to the greater sensitivity, specificity and accuracy seen in the postmenopausal population compared to the premenopausal population in the meta-

analysis (Nunes, 2014), changing the policy for postmenopausal women to the SR-based protocol is likely to benefit women, the NHS and society as a whole. We have seen that while women with benign adnexal tumours still have a risk of a future malignancy, this malignancy is likely to be borderline or a slow-growing Type I tumour which has a high chance of successful treatment. This compares to the women with normal ovaries, who if they develop a malignancy, this tumour is more likely to be an aggressive Type II EOC. Reducing surgery in asymptomatic women with benign tumours has many potential benefits. It would avoid the higher surgical and anaesthetic risks that these women face, it would potentially reduce the number of unnecessary referrals to regional gynaecological cancer centres, avoiding the congestion of these gynaecological oncology clinics and operating lists with benign disease and it could allow accelerated streamlined access and treatment for the women with an actual malignancy. Future work should concentrate on a financial evaluation of the benefit of simple rules in this population.

Finally, this thesis did show that in all the studies, pattern recognition was the superior test performed and increased training of expert ultrasound operators who can perform pattern recognition would increase the proportion of accurate diagnoses in the future.

References

- Aboulghar MA, Mansour RT, Serour GI. 1995. Ultrasonographically guided transvaginal aspiration of tuboovarian abscesses and pyosalpinges: an optional treatment for acute pelvic inflammatory disease. Am J Obstet Gynecol 172 (5), 1501-3.
- Alcázar JL, Jurado M. 1998. Using a logistic model to predict malignancy of adnexal masses based on menopausal status, ultrasound morphology, and color Doppler findings. Gynecol Oncol. 69 (2), 146-150.
- Alcázar JL, Pascual MÁ, Olartecoechea B, Graupera B, Aubá M, Ajossa S, Hereter L, Julve R, Gastón B, Peddes C, Sedda F, Piras A, Saba L, Guerriero S. 2013. IOTA simple rules for discriminating between benign and malignant adnexal masses: prospective external validation. Ultrasound Obstet Gynecol 42(4), 467-71.
- Ameye L, Timmerman D, Valentin L, Paladini D, Zhang J, Van Holsbeke C, Lissoni AA, Savelli L, Veldman J,Testa AC, Amant F, Van Huffel S, Bourne T. 2012. Clinically oriented three-step strategy for assessment ofadnexal pathology. Ultrasound Obstet Gynecol 40(5), 582-91.
- Ansaloni L, F. Catena, R. Chattat, D. Fortuna, C. Franceschi, P. Mascitti, R.M. Melotti. 2010. Risk factors and incidence of postoperative delirium in elderly patients after elective and emergency surgery. Br. J. Surg 97, 273–280.
- Antila R, Jalkanen J, Heikinheimo O. 2006. Comparison of secondary and primary ovarian malignancies reveals differences in their pre- and perioperative characteristics. Gynecol Oncol. 101 (1), 97-101.
- Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman

N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tulinius H, Thorlacius S, Eerola H, Nevanlinna H, Syrjäkoski K, Kallioniemi OP, Thompson D, Evans C, Peto J, Lalloo F, Evans DG, Easton DF. 2003. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 72 (5), 1117-30.

- Aslam N, Banerjee S, Carr JV, Savvas M, Hooper R, Jurkovic D. 2000a. Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer. Obstet Gynecol 96 (1), 75-80.
- Aslam N, Tailor A, Lawton F, Carr J, Savvas M, Jurkovic D. 2000b. Prospective evaluation of three different models for the preoperative diagnosis of ovarian cancer. Br J Obstet Gynaecol 107, 1347-53.
- Athey PA, Cooper NB. 1985. Sonographic features of paraovarian cysts. AJR Am J Roentgenol 144 (1), 83-6.
- Athey PA, Malone RS. 1987. Sonography of ovarian fibromas/thecomas. J Ultrasound Med 6 (8), 431–436.
- Axe SR, Klein VR, Woodruff JD. 1985. Choriocarcinoma of the ovary. Obstet Gynecol 66(1), 111–114.
- Ayhan A, Aksu T, Develioglu O, Tuncer ZS, Ayhan A. 1991. Complications and bilaterality of mature ovarian teratomas (clinicopathological evaluation of 286 cases). Aust N Z J Obstet Gynaecol 31 (1), 83–85.
- Ayhan A, Celik H, Taskiran C, Bozdag G, Aksu T. 2003. Oncologic and reproductive outcome after fertility-saving surgery in ovarian cancer. Eur J Gynaecol Oncol 24 (3-4), 223-32.

- Bailey CL, Ueland FR, Land GL, DePriest PD, Gallion HH, Kryscio RJ, van Nagell JR Jr. 1998. The malignant potential of small cystic ovarian tumors in women over 50 years of age. Gynecol Oncol 69 (1), 3-7.
- Bast RC Jr, Klug TL, St John E, Jenison E, Niloff JM, Lazarus H, Berkowitz RS, Leavitt T, Griffiths CT, Parker L, Zurawski VR Jr, Knapp RC. 1983. A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. N Engl J Med 309 (15), 883-7.
- Benedet JL, Bender H, Jones H 3rd, Ngan HY, Pecorelli S. 2000. FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. FIGO Committee on Gynecologic Oncology. Int J Gynaecol Obstet 70 (2), 209-62.
- Bossuyt PM, Reitsma JB, Bruns DE Gatsonis PP, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, De Vet HC; Standards for Reporting of Diagnostic Accuracy. 2003. Toward complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. BMJ 326 (7379), 41-4 (18 December 2013) http://www.stard-statement.org/
- Cancer Research UK. 2011. CancerStats. Ovarian Cancer – UK. http://publications.cancerresearchuk.org/downloads/product/CS_CS_OVARY.pdf
- Cancer Research UK. Cancer incidence and mortality in the UK for the 10 most common cancers. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer, Accessed 5th September 2015.
- Casey MJ, Synder C, Bewtra C, Narod SA, Watson P, Lynch HT. 2005. Intra-abdominal carcinomatosis after prophylactic oophorectomy in

women of hereditary breast ovarian cancer syndrome kindreds associated with BRCA1 and BRCA2 mutations. Gynecol Oncol 97 (2), 457-67.

- Caspi B, Appelman Z, Rabinerson D, Elchalal U, Zalel Y, Katz Z. 1996. Pathognomonic echo patterns of benign cystic teratomas of the ovary: classification, incidence and accuracy rate of sonographic diagnosis. Ultrasound Obstet Gynecol 7 (4), 275-9.
- Caspi B, Appelman Z, Rabinerson D, Zalel Y, Tulandi T, Shoham Z. 1997. The growth pattern of ovarian dermoid cysts: a prospective study in permenopausal and postmenopausal women. Fertil Steril 68 (3), 501–505.
- Cohen L, Sabbagha R. 1993. Echo patterns of benign cystic teratomas by transvaginal ultrasound. Ultrasound Obstet Gynecol 3 (2), 120-3.
- Collaborative Group on Epidemiological Studies of Ovarian Cancer, Beral V, Doll R, Hermon C, Peto R, Reeves G. 2008. Ovarian cancer and oral contraceptives: collaborative reanalysis of data from 45 epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls. Lancet 371 (9609), 303-14.
- Curtin JP, Malik R, Venkatraman ES, Barakat RR, Hoskins WJ. 1997. Stage IV ovarian cancer: impact of surgical debulking. Gynecol Oncol 64 (1), 9-12.
- Daemen A, Jurkovic D, Van Holsbeke C, Guerriero S, Testa AC, Czekierdowski A, Fruscio R, Paladini D, Neven P, Rossi A, Bourne T, De Moor B, Timmerman D. 2011. Effect of cancer prevalence on the use of risk-assessment cut-off levels and the performance of mathematical models to distinguish malignant from benign adnexal masses. Ultrasound Obstet Gynecol 37(2), 226-31.
- Dayhoff JE, DeLeo JM. 2001. Artificial neural networks: opening the black box. Cancer 91(8 Suppl), 1615-35.
- DeClerck BK, Post MD, Wisell JA. 2012. Cutaneous decidualized

endometriosis in a nonpregnant female: a potential pseudomalignancy. Am J Dermatopathol 34(5):541-3

- Dede M, Gungor S, Yenen MC, Alanbay I, Duru NK, Haşimi A. 2006. CA19-9 may have clinical significance in mature cystic teratomas of the ovary. Int J Gynecol Cancer 16 (1), 189-93.
- DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. Controlled Clinical Trials 7 (3), 177–188.
- de Groot JA, Dendukuri N, Janssen KJ, Reitsma JB, Bossuyt PM, Moons KG. 2011. Adjusting for differential-verification bias in diagnostic-accuracy studies: a Bayesian approach. Epidemiology 22 (2), 234-41.
- de Waal YR, Thomas CM, Oei AL, Sweep FC, Massuger LF. 2009. Secondary ovarian malignancies: frequency, origin, and characteristics. Int J Gynecol Cancer 19 (7), 1160-5.
- Di Legge A, Testa AC, Ameye L, Van Calster B, Lissoni AA, Leone FP, Savelli L, Franchi D,Czekierdowski A, Trio D, Van Holsbeke C, Ferrazzi E,Scambia G, Timmerman D, Valentin L. 2012. Lesion size affects diagnostic performance of IOTA logistic regression models, IOTA simple rules and risk of malignancy index in discriminating between benign and malignant adnexal masses. Ultrasound Obstet Gynecol 40(3), 345-54.
- Drapkin R, von Horsten HH, Lin Y, Mok SC, Crum CP, Welch WR, Hecht JL. 2005. Human epididymis protein 4 (HE4) is a secreted glycoprotein that is overexpressed by serous and endometrioid ovarian carcinomas. Cancer Res 65 (6), 2162-9.
- du Bois A, Ewald-Riegler N, de Gregorio N, Reuss A, Mahner S, Fotopoulou C, Kommoss F, Schmalfeldt B, Hilpert F, Fehm T, Burges A, Meier W, Hillemanns P, Hanker L, Hasenburg A, Strauss HG, Hellriegel M, Wimberger P,

Keyver-Paik MD, Baumann K, Canzler U, Wollschlaeger K, Forner D, Pfisterer J, Schröder W, Münstedt K, Richter B, Kommoss S, Hauptmann S; Arbeitsgmeinschaft Gynäkologische Onkologie (AGO) Study Group. 2013. Borderline tumours of the ovary: A cohort study of the Arbeitsgmeinschaft Gynäkologische Onkologie (AGO) Study Group. Eur J Cancer 49 (8), 1905-14.

- Eden J, Levit L, Berg A, Morton S. 2011. Finding what works in health care: Standards for systematic reviews. The National Academies Press, Washington DC.
- Edmondson RJ, Monaghan JM. 2001. The epidemiology of ovarian cancer. Int J Gynecol Cancer 11:423-9
- Edwards SJ, Lilford RJ, Braunholtz DA, Jackson JC, Hewison J, Thornton J. 1998. Ethical issues in the design and conduct of randomised controlled trials. Health Technol Assess 2 (15), i-vi, 1-132.
- EFSUMB. (Education and Practical Standards Committee, European Federation of Societies for Ultrasound in Medicine and Biology). 2006. Minimum training recommendations for the practice of medical ultrasound. Ultraschall Med 27 (1), 79-105.
- EFSUMB. (Education and Practical Standards Committee, European Federation of Societies for Ultrasound in Medicine and Biology). 2010. Minimum training requirements for the practice of Medical Ultrasound in Europe. Ultraschall in Med 31 (4), 426–427.
- Egger M, Smith GD, Phillips AN. 1997. Meta-analysis: principles and procedures. BMJ 315 (7121), 1533-7.
- Elattar A, Bryant A, Winter-Roach BA, Hatem M, Naik R. 2011. Optimal primary surgical treatment for advanced epithelial ovarian cancer. Cochrane Database Syst Rev (8):CD007565.

- Eltabbakh GH, Yadav PR, Morgan A. 1999. Clinical picture of women with early stage ovarian cancer. Gynecol Oncol 75 (3), 476-9.
- Fathallah K, Huchon C, Bats AS, Metzger U, Lefrère-Belda MA, Bensaid C, Lécuru F. 2011. External validation of simple ultrasound rules of Timmerman on 122 ovarian tumors. Gynecol Obstet Fertil 39(9), 477-81.
- Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 127 (12), 2893-917.
- Ferraro S, Braga F, Lanzoni M, Boracchi P, Biganzoli EM, Panteghini M. 2013. Serum Human Epididymis Protein 4 vs Carbohydrate Antigen 125 for Ovarian Cancer Diagnosis. A Systematic Review. J Clin Pathol 66 (4), 273-281.
- Fishman DA, Cohen LS. 2000. Is transvaginal ultrasound effective for screening asymptomatic women for the detection of early-stage epithelial ovarian carcinoma? Gynecol Oncol 77 (3), 347–349.
- Fried AM, Rhodes RA, Morehouse IR. 1993. Endometrioma: analysis and sonographic classification of 51 documented cases. South Med J 86 (3), 297-301.
- Fujii S, Konishi I, Suzuki A, Okamura H, Okazaki T, Mori T. 1985. Analysis of serum lactic dehydrogenase levels and its isoenzymes in ovarian dysgerminoma. Gynecol. Oncol 22 (1), 65–72.
- Galgano MT, Hampton GM, Frierson HF Jr. 2006. Comprehensive analysis of HE4 expression in normal and malignant human tissues. Mod Pathol 19 (6), 847–853.
- Genadry R, Parmley T, Woodruff JD. 1977. The origin and clinical behavior of the paraovarian tumor. Am J Obstet Gynecol 129 (8), 873–80.
- Geomini P, Kruitwagen R, Bremer GL, Cnossen J, Mol BW. 2009.

The accuracy of risk scores in predicting ovarian malignancy: a systematic review. Obstet Gynecol 113 (2 Pt 1), 384-94.

- Gotlieb WH, Chetrit A, Menczer J, Hirsh-Yechezkel G, Lubin F, Friedman E, Modan B, Ben-Baruch G; National Israel Ovarian Cancer Study Group. 2005. Demographic and genetic characteristics of patients with borderline ovarian tumors as compared to early stage invasive ovarian cancer. Gynecol Oncol 97 (3), 780-3.
- Gotlieb WH, Barchana M, Ben-Baruch G, Friedman E. 2006. Malignancies following bilateral salpingo-oophorectomy (BSO). Eur J Surg Oncol 32 (10), 1231-4.
- Granstrom C, Sundquist J, Hemminki K. 2008. Population attributable fractions for ovarian cancer in Swedish women by morphological type. Br J Cancer 98 (1), 199-205.
- Grant EG. 1992. Benign conditions of the ovary. In Nyberg DA, Hill LM, Böhm-Velez M, Mendelson EB (eds) Transvaginal Ultrasound. St Louis: Mosby Year Book, pp 187–208.
- Gramellini D, Fieni S, Sanapo L, Casilla G, Verrotti C, Nardelli GB. 2008. Diagnostic accuracy of IOTA ultrasound morphology in the hands of less experienced sonographers. Aust N Z J Obstet Gynaecol 48 (2), 195-201.
- Ghezzi F, Raio L, Cromi A, Duwe DG, Beretta P, Buttarelli M, Mueller MD. 2005. "Kissing ovaries": a sonographic sign of moderate to severe endometriosis. Fertil Steril 83 (1), 143-7.
- Granberg S, Wikland M, Jansson I. 1989. Macroscopic characterization of ovarian tumors and the relation to the histological diagnosis: criteria to be used for ultrasound evaluation. Gynecol Oncol 35 (2), 139-44.
- Guerriero S, Ajossa S, Garau N, Alcazar JL, Mais V, Melis GB.

2010. Diagnosis of pelvic adhesions in patients with endometrioma: the role of transvaginal ultrasonography. Fertil Steril 94 (2), 742-6.

- Guerriero S, Alcazar JL, Pascual MA, Ajossa S, Olartecoechea B, Hereter L. 2012. Preoperative diagnosis of metastatic ovarian cancer is related to origin of primary tumor. Ultrasound Obstet Gynecol 39 (5), 581-6.
- Guerriero S, Saba L, Ajossa S, Peddes C, Sedda F, PirasA, Olartecoechea B, Aubá M, Alcázar JL. 2013. Assessing the reproducibility of the IOTA simple ultrasound rules for classifying adnexal masses as benign or malignant using stored 3D volumes. Eur J Obstet Gynecol Reprod Biol 171(1), 157-60.
- Harbord RM, Steichen T. 2005. METAREG: Stata module to perform meta-analysis regression Boston College Department of Economics, Statistical Software Components series.
- Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. 2007. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 8 (2), 239-251.
- Harbord RM, Higgins JPT. 2008a. Meta-regression in Stata. The Stata Journal 8(4), 493-519.
- Harbord RM, Whiting P, Sterne J. 2008b. Metandi: Stata software for statistically rigorous meta-analysis of diagnostic accuracy studies. In First International Symposium on Methods for Evaluating Medical Tests, Birmingham, UK.
- Harbord RM, Whiting P. 2009. Metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. The Stata Journal 9(2), 211-229.
- Harrell FE. 2001. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis, Springer, New York.
- Harris RJ, Bradburn M, Deeks J, Harbord RM, Altman D, Steichen

T, Sterne J. 2007. METAN: Stata module for fixed and random effects meta-analysis. Boston College Department of Economics, Statistical Software Components series.

- Harris RJ, Bradburn M, Deeks J, Harbord RM, Altman D, Sterne JA. 2008. Metan: fixed- and random-effects meta-analysis. The Stata Journal 8(1), 3-28.
- Hartling L, Hamm M, Milne A, Vandermeer B, Santaguida PL, Ansari M, Tsertsvadze A, Hempel S, Shekelle P, Dryden DM. 2012. Validity and inter-rater reliability testing of quality assessment instruments. (Prepared by the University of Alberta Evidence-based Practice Center under Contract No. 290-2007-10021-I.) AHRQ Publication No. 12-EHC039-EF. Rockville, MD: Agency for Healthcare Research and Quality. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
- Hartman CA, Juliato CR, Sarian LO, Toledo MC, Jales RM, Morais SS, Pitta DD, Marussi EF, Derchain S. 2012. Ultrasound criteria and CA 125 as predictive variables of ovarian cancer in women with adnexal tumors. Ultrasound Obstet Gynecol 40(3), 360-6.
- Hata K, Hata T, Aoki S Takamiya O, Kitao M. 1989 Ultrasonographic evaluation of pelvic inflammatory disease. Nippon Sanka Fujinka Gakkai Zasshi 41 (7), 895–8.
- Haykin S 1994 Neural Networks: a comprehensive Foundation. New York: Macmillan Publishing Company.
- Heintz AP, Odicino F, Maisonneuve P, Quinn MA, Benedet JL, Creasman WT, Ngan HY, Pecorelli S, Beller U. 2006. Carcinoma of the ovary. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. Int J Gynaecol Obstet 95 (Suppl 1), 161-92.
- Hellström I, Raycraft J, Hayden-Ledbetter M, Ledbetter JA, Schummer M, McIntosh M, Drescher C, Urban N, Hellström KE. 2003. The HE4 (WFDC2)

protein is a biomarker for ovarian carcinoma. Cancer Res 63 (13), 3695–700.

- Hernán MA, Hernandez-Diaz S, Robins J. M. 2004. A structural approach to selection bias. Epidemiology 15 (5), 615-625.
- Hilger WS, Magrina JF, Magtibay PM. 2006. Laparoscopic management of the adnexal mass. Clin Obstet Gynecol 49 (3), 535-48.
- Hippisley-Cox J, Coupland C. 2011. Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. BMJ 344, d8009.
- Hirai M, Hirai Y, Tsuchida T, Takada T, Iwase H, Utsugi K, Sugiyama Y, Takeshima N, Furuta R, Takizawa K. 2011. Stage IA ovarian cancers: comparison of sonographic findings and histopathologic types between patients with normal and elevated serum cancer antigen 125 levels. J Ultrasound Med 30 (7), 943-52.
- Hoo WL, Yazbek J, Holland T, Mavrelos D, Tong EN, Jurkovic D. 2010. Expectant management of ultrasonically diagnosed ovarian dermoid cysts: is it possible to predict outcome? Ultrasound Obstet Gynecol 36 (2), 235-40.
- Horiuchi A, Itoh K, Shimizu M, Nakai I, Yamazaki T, Kimura K, Suzuki A, Shiozawa I, Ueda N, Konishi I. 2003. Toward understanding the natural history of ovarian carcinoma development: a clinicopathological approach. Gynecol Oncol 88, 309-17.
- Hutton L, Rankin R. 1979. The fat fluid level: another feature of dermoid tumors of the ovary. J Clin Ultrasound 7 (3), 215-6.
- IARC 2012 GLOBOCAN 2012. Cancer Incidence, Mortality and Prevalence Worldwide (2012 estimates)
- Im SS, Gordon AN, Buttin BM, Leath CA 3rd, Gostout BS, Shah C, Hatch KD, Wang J, Berman ML. 2005. Validation of referral guidelines for women with pelvic masses. Obstet Gynecol 105 (1), 35-41.

- International Federation of Gynecology and Obstetrics. 1971. Classification and staging of malignant tumors in the female pelvis. Acta Obstet Gynecol Scand 50 (1), 1–7.
- ISD (Information Services Division) Scotland Online NHS National Services Scotland. 2013. Cancer in Scotland 2011. Cancer Incidence, Mortality and Survival data
- Iyer VR, Lee SI. 2010. MRI, CT, and PET/CT for ovarian cancer detection and adnexal lesion characterization. AJR Am J Roentgenol 194 (2), 311-21.
- Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. 1990. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. Br J Obstet Gynecol 97 (10), 922-9.
- Jervis S, Song H, Lee A, Dicks E, Tyrer J, Harrington P, Easton DF, Jacobs IJ, Pharoah PP, Antoniou AC. 2014. Ovarian cancer familial relative risks by tumour subtypes and by known ovarian cancer genetic susceptibility variants. J Med Genet 51 (2),108-13.
- Junor EJ, Hole DJ, McNulty L, Mason M, Young J. 1999. Specialist gynaecologists and survival outcome in ovarian cancer: a Scottish national study of 1866 patients. Br J Obstet Gynaecol 160 (11), 1130-6.
- Jurkovic D, Valentin L, Vyas S. 2009. Gynaecological Ultrasound in Clinical Practice: Ultrasound imaging in the management of gynaecological conditions, RCOG Press, London.
- Kaijser J, Bourne T, Valentin L, Sayasneh A, VanHolsbeke C, Vergote I, Testa AC, Franchi D, Van CalsterB, Timmerman D. 2013a. Improving strategies for diagnosing ovarian cancer: a summary of the International Ovarian Tumor

Analysis (IOTA) studies. Ultrasound Obstet Gynecol 41(1), 9-20.

- Kaijser J, Sayasneh A, Van Hoorde K, Ghaem-MaghamiS, Bourne T, Timmerman D, Van Calster B. 2013b. Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: a systematic review and meta-analysis. Hum Reprod Update. Dec 9.
- Kawai M, Kano T, Kikkawa F, Morikawa Y, Oguchi H, Nakashima N, Ishizuka T, Kuzuya K, Ohta M, Arii Y, Tomoda Y. 1992. Seven tumor markers in benign and malignant germ cell tumors of the ovary. Gynecol Oncol 45 (3), 248-53.
- Kim JS, Woo SK, Suh SJ, Morettin LB. 1995. Sonographic diagnosis of paraovarian cysts: value of detecting a separate ipsilateral ovary. AJR Am J Roentgen 164 (6), 1441–4.
- Kinkel K, Lu Y, Mehdizade A, Pelte MF, Hricak H. 2005. Indeterminate ovarian mass at US: incremental value of second imaging test for characterization--meta-analysis and Bayesian analysis. Radiology 236 (1), 85-94.
- Kinkel K, Frei KA, Balleyguier C, Chapron C. 2006. Diagnosis of endometriosis with imaging: a review. Eur Radiol 16 (2), 285-98.
- Knottnerus JA, Dinant GJ. Medicine based evidence, a prerequisite for evidence based medicine. BMJ 1997; 315(7116) : 1109-10.
- Konishi I, Fujii S, Okamura H, Sakahara H, Endo K, Torizuka K, Suzuki A, Mori T. 1986. Analysis of serum CA125, CEA, AFP, LDH levels and LDH isoenzymes in patients with ovarian tumors—Correlation between tumor markers and histological types of ovarian tumors. Acta Obstet. Gynaecol. Jpn 38 (6), 827–836.
- Koonings PP, Campbell K, Mishell DR Jr, Grimes DA. 1989. Relative frequency of primary ovarian neoplasms: a 10-year review. Obstet Gynecol 74 (6), 921-6.

- Korbin CD, Brown DL, Welch WR. 1998. Paraovarian cystadenomas and cystadenofibromas: sonographic characteristics in 14 cases. Radiology 208 (2), 459–62.
- Kupfer MC, Schwimer SR, Lebovic J. 1992. Transvaginal sonographic appearance of endometriomata: spectrum of findings. J Ultrasound Med 11 (4), 129-33.
- Kurman RJ, Shih IM. 2010. The Origin and Pathogenesis of Epithelial Ovarian Cancer- a Proposed Unifying Theory. Am J Surg Pathol 34 (3), 433–443.
- Kurman RJ, Shih IeM. 2011. Molecular pathogenesis and extraovarian origin of epithelial ovarian cancer--shifting the paradigm. Hum Pathol 42 (7), 918-31.
- Kurman RJ. 2013. Origin and molecular pathogenesis of ovarian high-grade serous carcinoma. Ann Oncol 24 Suppl 10, 16-21.
- Lataifeh I, Marsden DE, Robertson G, Gebski V, Hacker NF. 2005. Presenting symptoms of epithelial ovarian cancer Aust N Z J Obstet Gynaecol 45:211-4
- Lee S, Nelson G, Duan Q, Magliocco AM, Duggan MA. 2013. Precursor lesions and prognostic factors in primary peritoneal serous carcinoma. Int J Gynecol Pathol 32 (6), 547-55.
- Levine D, Gosink BB, Wolf SI, Feldesman MR, Pretorius DH. 1992. Simple adnexal cyst: the natural history in postmenopausal women. Radiology 184 (3), 653–9
- Li J, Abushahin N, Pang S, Xiang L, Chambers SK, Fadare O, Kong B, Zheng W. 2011. Tubal origin of 'ovarian' low-grade serous carcinoma. Mod Pathol 24 (11), 1488-99.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that

evaluate health care interventions: explanation and elaboration. PLoS Med 6 (7), e1000100.

- Lv L, Yang K, Wu H, Lou, Peng Z. 2011. Pure choriocarcinoma of the ovary: a case report. J Gynecol Oncol 22(2), 135–139.
- Macedo AC, da Rosa MI, Lumertz S, Medeiros LR. 2014. Accuracy of serum human epididymis protein 4 in ovarian cancer diagnosis: a systematic review and meta-analysis. Int J Gynecol Cancer 24 (7), 1222-31.
- Machin D and Campbell MJ. 2005. Design of Studies for Medical Research, Wiley, West Sussex.
- Mais V, Ajossa S, Piras B, Marongiu D, Guerriero S, Melis GB. 1995. Treatment of nonendometriotic benign adnexal cysts: A randomized comparison of laparoscopy and laparotomy. Obstet Gynecol 86 (5), 770–4.
- Malander S, Ridderheim M, Måsbäck A, Loman N, Kristoffersson U, Olsson H, Nilbert M, Borg A. 2004. One in 10 ovarian cancer patients carry germ line BRCA1 or BRCA2 mutations: results of a prospective study in Southern Sweden. Eur J Cancer 40 (3), 422-8.
- Manchanda R, Abdelraheim A, Johnson M, Rosenthal AN, Benjamin E, Brunell C, Burnell M, Side L, Gessler S, Saridogan E, Oram D, Jacobs I, Menon U. 2011. Outcome of risk-reducing salpingo-oophorectomy in BRCA carriers and women of unknown mutation status. BJOG 118 (7), 814-24.
- Maneo A, Vignali M, Chiari S, Colombo A, Mangioni C, Landoni F. 2004. Are borderline tumors of the ovary safely treated by laparoscopy? Gynecol Oncol 94 (2), 387-92.
- Mecke H, Lehmann-Willenbrock E, Ibrahim M, Semm K.1992. Pelviscopic treatment of ovarian cysts in premenopausal women. Gynecol Obstet Invest 34 (1), 36–42.

- Medeiros LR, Rosa DD, da Rosa MI, Bozzetti MC. 2009. Accuracy of CA 125 in the diagnosis of ovarian tumors: a quantitative systematic review. Eur J Obstet Gynecol Reprod Biol 142 (2), 99-105.
- Meire HB, Farrant P, Guha T. 1978. Distinction of benign from malignant ovarian cysts by ultrasound. Br J Obstet Gynaecol 85 (12), 893-9.
- Miralles C, Orea M, España P, Provencio M, Sánchez A, Cantos B, Cubedo R, Carcereny E, Bonilla F, Gea T. 2003. Cancer antigen 125 associated with multiple benign and malignant pathologies. Ann Surg Oncol 10 (2), 150–4.
- Modesitt SC, Pavlik EJ, Ueland FR, DePriest PD, Kryscio RJ, van Nagell JR Jr. 2003. Risk of malignancy in unilocular ovarian cystic tumors less than 10 centimeters in diameter. Obstet Gynecol 102 (3), 594–599.
- Mogensen O, Mogensen B, Jakobsen A. 1989. CA 125 in the diagnosis of pelvic masses European Journal of Cancer and Clinical Oncology 25 (8), 1187-1190.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. 1999. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 354 (9193), 1896-900.
- Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6 (7), e1000097.
- Montagnana M, Danese E, Ruzzenente O, Bresciani V, Nuzzo T, Gelati M, Salvagno GL, Franchi M, Lippi G, Guidi GC. 2011. The ROMA (Risk of Ovarian Malignancy Algorithm) for estimating the risk of epithelial ovarian cancer in women presenting with pelvic mass: is it really useful? Clin Chem Lab Med 49 (3), 521-5.

- Moore RG, Brown AK, Miller MC, Skates S, Allard WJ, Verch T, Steinhoff M, Messerlian G, DiSilvestro P, Granai CO, Bast RC Jr. 2008. The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. Gynecol Oncol 108 (2), 402–408.
- Moore RG, McMeekin DS, Brown AK, DiSilvestro P, Miller MC, Allard WJ, Gajewski W, Kurman R, Bast RC Jr, Skates SJ. 2009. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. Gynecol Oncol 112 (1), 40-6.
- Moore RG, Miller MC, Disilvestro P, Landrum LM, Gajewski W, Ball JJ, Skates SJ. 2011. Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. Obstet Gynecol 118(2 Pt 1), 280-8.
- Nasser S, Arsenic R, Lohneis P, Kosian P, Sehouli J. 2014. A case of primary peritoneal carcinoma: evidence for a precursor in the fallopian tube. Anticancer Res 34 (1), 407-12.
- Neufeld KJ, Leoutsakos JM, Sieber FE, Wanamaker BL, Gibson Chambers JJ, Rao V, Schretlen DJ, Needham DM. 2013. Outcomes of early delirium diagnosis after general anesthesia in the elderly. Anesth Analg 117, 471–478.
- Nezhat F, Nezhat C, Welander CE, Benigno B. 1992. Four ovarian cancers diagnosed during laparoscopic management of 1011 women with adnexal masses. Am J Obstet Gynecol 167 (3), 790–796
- Northern Ireland Cancer Registry 2013 Cancer Registrations in Northern Ireland. Incidence and Survival 1993-2011 C56 http://www.qub.ac.uk/research-centres/nicr/CancerData/OnlineStatistics/Ovary/ (accessed 7th January 2013)
- Nunes N, Yazbek J, Ambler G, Hoo W, Naftalin J, Jurkovic D. 2012a. Prospective evaluation of the IOTA Logistic Regression Model

LR2 for the diagnosis of ovarian cancer. Ultrasound Obstet Gynecol 40 (3), 355-9.

- Nunes N, Foo X, Widschwendter M, Jurkovic D. 2012b. A randomised controlled trial comparing surgical intervention rates between two protocols for the management of asymptomatic adnexal tumours in postmenopausal women. BMJ Open 2 (6), pii: e002248.
- Nunes N, Ambler G, Hoo WL, Naftalin J, Foo X, Widschwendter M, Jurkovic D. 2013. A Prospective Validation of the IOTA Logistic Regression Models (LR1 and LR2) in Comparison to Subjective Pattern Recognition for the Diagnosis of Ovarian Cancer Int J Gynecol Cancer 23 (9), 1583-9.
- Nunes N, Ambler G, Foo X, Naftalin J, Widschwendter M, Jurkovic D. 2014. Use of the IOTA Simple Rules for the diagnosis of ovarian cancer: a Meta-Analysis. Ultrasound Obstet Gynecol 44 (5), 503-14.
- Office for National Statistics. 2012. Cancer Statistics registrations: Registrations of cancer diagnosed in 2010, England. Series MB1 no.41
- Papadias K, Kairi-Vassilatou E, Kontogiani-Katsaros K, Argeitis J, Kondis-Pafitis A, Greatsas G. 2005. Teratomas of the ovary: a clinico-pathological evaluation of 87 patients from one institution during a 10-year period. Eur J Gynaecol Oncol 26 (4), 446-8.
- Pavlik EJ, Ueland FR, Miller RW, Ubellacker JM, DeSimone CP, Elder J, Hoff J, Baldwin L, Kryscio RJ, van Nagell JR Jr. 2013. Frequency and disposition of ovarian abnormalities followed with serial transvaginal ultrasonography. Obstet Gynecol 122 (2), 210-7.
- Pereira A, Pérez-Medina T, Magrina JF, Magtibay PM, Rodríguez-Tapia A, Cuesta-Guardiola T, Peregrin I, Mendizabal E, Lizarraga S, Ortiz-Quintana L. 2016. The impact of debulking surgery in patients with node-

positive epithelial ovarian cancer: Analysis of prognostic factors related to overall survival and progression-free survival after an extended long-term follow-up period. Surg Oncol 25(1):49-59.

- Piver MS, Jishi MF, Tsukada Y, Nava G. 1993. Primary peritoneal carcinoma after prophylactic oophorectomy in women with a family history of ovarian cancer. A report of the Gilda Radner Familial Ovarian Cancer Registry. Cancer 71 (9), 2751-5.
- Pomel C, Jeyarajah A, Oram D, Shepherd J, Milliken D, Dauplat J, Reynolds K. 2007. Cytoreductive surgery in ovarian cancer. Cancer Imaging 7:210-5.
- Prömpeler HJ. 2011. Ultrasonographic clarification of adnexal findings. Radiologe 51(7), 568-80.
- Rasheed SM, Abdelmonem AM. 2011. Hydatid of Morgagni: a possible underestimated cause of unexplained infertility. Eur J Obstet Gynecol Reprod Biol 158 (1), 62–6.
- Robboy S, Mutter G, Prat J, Bentley R, Russell P, Anderson M. 2008. Robboy's Pathology of the Female Reproductive Tract (2nd edition) Churchill Livingstone, London.
- RCOG. 2003. Ovarian cysts in postmenopausal women (Green-top Guideline 34), London.
- RCOG. 2011. Management of Suspected Ovarian Masses in Premenopausal Women, (Green-top Guideline 62), London.
- Rosen DG, Wang L, Atkinson JN, Yu Y, Lu KH, Diamandis EP, Hellstrom I, Mok SC, Liu J, Bast RC Jr. 2005. Potential markers that complement expression of CA125 in epithelial ovarian cancer. Gynecol Oncol 99 (2), 267-77.
- Ruiz de Gauna B, Sanchez P, Pineda L, Utrilla-Layna J, Juez L, Alcázar JL. 2014. Inter-observer agreement with regard to describing adnexal

masses using the IOTA simple rules in a real-time setting and when using three-dimensional ultrasound volumes and digital clips. Ultrasound Obstet Gynecol 44 (1),95-9.

- Said MR, Bamigboye V. 2008. Twisted paraovarian cyst in a young girl. J Obstet Gynaecol 28 (5), 549-50.
- Salvesen KA, Lees C, Tutschek B. 2011. Basic European ultrasound training in obstetrics and gynecology: where are we and where do we go from here? Ultrasound Obstet Gynecol 36 (5), 525-9.
- Samaha M, Woodruff JD. 1985. Paratubal cysts: frequency, histogenesis, and associated clinical features. Obstet Gynecol 65 (5), 691-4.
- Sandri MT, Bottari F, Franchi D, Boveri S, Candiani M, Ronzoni S, Peiretti M, Radice D, Passerini R, Sideri M. 2013. Comparison of HE4, CA125 and ROMA algorithm in women with a pelvic mass: correlation with pathological outcome. Gynecol Oncol 128 (2), 233-8.
- Saba L, Guerriero S, Sulcis R, Virgilio B, Melis G, Mallarini G. 2009. Mature and immature ovarian teratomas: CT, US and MR imaging characteristics. Eur J Radiol 72 (3), 454-63.
- Saunders BA, Podzielinski I, Ware RA, Goodrich S, DeSimone CP, Ueland FR, Seamon L, Ubellacker J, Pavlik EJ, Kryscio RJ, van Nagell JR Jr. 2010. Risk of malignancy in sonographically confirmed septated cystic ovarian tumors. Gynecol Oncol 118 (3), 278-82.
- Savelli L, de Iaco P, Ghi T, Bovicelli L, Rosati F, Cacciatore B. 2004. Transvaginal sonographic appearance of peritoneal pseudocysts. Ultrasound Obstet Gynecol 23 (3), 284-8.
- Savelli L, Ghi T, De Iaco P, Ceccaroni M, Venturoli S, Cacciatore B. 2006. Paraovarian/paratubal cysts: comparison of transvaginal

sonographic and pathological findings to establish diagnostic criteria. Ultrasound Obstet Gynecol 28 (3), 330-4.

- Sayasneh A, Wynants L, Preisler J, Kaijser J, Johnson S, Stalder C, Husicka R, Abdallah Y, Raslan F, Drought A, Smith AA, Ghaem-Maghami S, Epstein E, Van Calster B, Timmerman D, Bourne T. 2013a. Multicentre external validation of IOTA prediction models and RMI by operators with varied training. Br J Cancer 108(12), 2448-54.
- Sayasneh A, Kaijser J, Preisler J, Johnson S, Stalder C,Husicka R, Guha S, Naji O, Abdallah Y, Raslan F, Drought A, Smith AA, Fotopoulou C, Ghaem-MaghamiS, Van Calster B, Timmerman D, Bourne T. 2013b. A multicenter prospective external validation of the diagnostic performance of IOTA simple descriptors and rules to characterize ovarian masses. Gynecol Oncol 130(1), 140-6.
- Schueler S, Schuetz GM, Dewey M. 2012. The revised QUADAS-2 tool. Ann Intern Med 156 (4), 323.
- Schulz KF, Altman DG, Moher D; CONSORT Group. 2010. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Obstet Gynecol 115 (5), 1063-70.
- Sedgwick P. 2012 Meta-analyses: tests of heterogeneity. BMJ 344, e3971
- Sedgwick P. 2012b. How to read a forest plot. BMJ 345, e8335.
- Seracchioli R, Venturoli S, Colombo FM, Govoni F, Missiroli S, Bagnoli A. 2001. Fertility and tumor recurrence rate after conservative laparoscopic management of young women with early-stage borderline ovarian tumors. Fertil Steril 76 (5), 999-1004.
- Serov SF, Scully RE, Sobin LH. 1973. Histological Typing of Ovarian Tumours. (WHO International Histological Classification of Tumours No. 9).

World Health Organization, Geneva, Switzerland.

- Shen-Gunther J, Mannel RS. 2002. Ascites as a predictor of ovarian malignancy. Gynecol Oncol 87 (1), 77-83.
- Shepherd JH. 1989. Revised FIGO staging for gynaecological cancer. Br J Obstet Gynaecol 96 (8), 889-92.
- Shepherd GM, Koch C. 1990. Introduction to synaptic circuits. In Shepherd GM, ed. The synaptic organization of the brain. New York: Oxford University Press 10, 5-8.
- Shih IeM, Kurman RJ. 2004. Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. Am J Pathol 164, 1511-8.
- Sibbald B, Roland M. 1998. Understanding controlled trials. Why are randomised controlled trials important? BMJ 316 (7126), 201.
- Silverberg SG, Bell DA, Kurman RJ, Seidman JD, Prat J, Ronnett BM, Copeland L, Silva E, Gorstein F, Young RH. 2004. Borderline ovarian tumors: key points and workshop summary. Hum Pathol 35 (8), 910-7.
- Sladkevicius P, Valentin L, Marsal K. 1995. Transvaginal gray-scale and Doppler ultrasound examinations of the uterus and ovaries in healthy postmenopausal women. Ultrasound Obstet Gynecol 6 (2), 81–90
- Sohaib SA, Mills TD, Sahdev A, Webb JA, Vantrappen PO, Jacobs IJ, Reznek RH. 2005. The role of magnetic resonance imaging and ultrasound in patients with adnexal masses. Clin Radiol 60, 340–348
- Sokalska A, Timmerman D, Testa AC, Van Holsbeke C, Lissoni AA, Leone FP, Jurkovic D, Valentin L. 2009. Diagnostic accuracy of transvaginal ultrasound examination for assigning a specific diagnosis to adnexal masses. Ultrasound Obstet Gynecol 34 (4), 462-70.
- Suh-Burgmann E, Hung Y-Y, Kinney W. 2014. Outcomes from

ultrasound follow-up of small complex adnexal masses in women over 50. American Journal of Obstetrics and Gynecology 211 (6), 623.e1623.e7.

- Stein AL, Koonings PP, Schlaerth JB, Grimes DA, d'Ablaing G 3rd. 1990. Relative frequency of malignant paraovarian tumors: should paraovarian tumors be aspirated? Obstet Gynecol 75 (6), 1029–31.
- Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, Carpenter J, Rücker G, Harbord RM, Schmid CH, Tetzlaff J, Deeks JJ, Peters J, Macaskill P, Schwarzer G, Duval S, Altman DG, Moher D, Higgins JP. 2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 343, d4002.
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 283 (15), 2008-12.
- Suzuki S, Furukawa H, Kyozuka H, Watanabe T, Takahashi H, Fujimori K. 2013. Two cases of paraovarian tumor of borderline malignancy. J Obstet Gynaecol 39, 437–41.
- Sztark F, Le Goff M, André D, Ritchie K, Dartigues JF, Helmer C. 2013. Exposure to general anaesthesia could increase the risk of dementia in elderly: 18AP1- 4. Eur J of Anaesth. Perioperative Care of the Elderly 30, 245-245.
- Tailor A, Jurkovic D, Bourne TH, Collins WP, Campbell S. 1997. Sonographic prediction of malignancy in adnexal masses using multivariate logistic regression analysis. Ultrasound Obstet Gynecol 10 (1), 41-7.
- Tailor A, Jurkovic D, Bourne TH, Collins WP, Campbell S. 1999. Sonographic prediction of malignancy in adnexal masses using an artificial

neural network. Br J Obstet Gynaecol 106 (1), 21-30.

- Tailor A, Bourne TH, Campbell S, Okokon E, Dew T, Collins WP. 2003. Results from an ultrasound-based familial ovarian cancer screening clinic: a 10-year observational study. Ultrasound Obstet Gynecol 21 (4), 378-85.
- Tazegül A, Seçilmiş Kerimoğlu O, Incesu FN, Doğan NU, Yılmaz SA, Celik C. 2013. A Case Presentation: Decidualized Endometrioma Mimicking Ovarian Cancer during Pregnancy. Case Rep Obstet Gynecol Article ID 728291.
- Testa AC, Ferrandina G, Timmerman D, Savelli L, Ludovisi M, Van Holsbeke C, Malaggese M, Scambia G, Valentin L. 2007. Imaging in gynecological disease (1): ultrasound features of metastases in the ovaries differ depending on the origin of the primary tumor. Ultrasound Obstet Gynecol 29 (5), 505-11.
- Thompson SG, Higgins JPT. 2002a. Quantifying heterogeneity in a meta-analysis. Statistics in Medicine 21(11), 1539–1558.
- Thompson SG, Higgins JPT. 2002b. How should meta-regression analyses be undertaken and interpreted? Statistics in Medicine 21(11), 1559–1573.
- Timor-Tritsch LE, Lerner JP, Monteagudo A, Santos R. 1993. Transvaginal ultrasonographic characterization of ovarian masses by means of color flow-directed Doppler measurements and a morphologic scoring system. Am J Obstet Gynecol 168 (3 Pt 1), 909-13.
- Tinelli R, Malzoni M, Cosentino F, Perone C, Tinelli A, Malvasi A, Cicinelli E. 2009. Feasibility, safety, and efficacy of conservative laparoscopic treatment of borderline ovarian tumors. Fertil Steril 92 (2), 736-41.
- Timmerman D, Verrelst H, Bourne TH, De Moor B, Collins WP, Vergote I, Vandewalle J. 1999a. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound Obstet Gynecol 13, 17-25.

- Timmerman D, Schwarzler P, Collins WP, Claerbout F, Coenen M, Amant F, Vergote I, Bourne TH. 1999b. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. Ultrasound Obstet Gynecol 13, 11-16.
- Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I; International Ovarian Tumor Analysis (IOTA) Group. 2000. 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 16 (5), 500-5.
- Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, Van Calster B, Collins WP, Vergote I, Van Huffel S, Valentin L; International Ovarian Tumor Analysis Group. 2005. 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 23 (34), 8794-801.
- Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, Paladini D, Van Calster B, Vergote I, Van Huffel S, Valentin L. 2008. Simple ultrasound-based rules for the diagnosis of ovarian cancer. Ultrasound. Obstet Gynecol 31 (6), 681–90.
- Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D, Lissoni AA, Van Holsbeke C, Fruscio R, Czekierdowski A, Jurkovic D, Savelli L, Vergote I, Bourne T, Van Huffel S, Valentin L.2010a. 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 36 (2), 226-34.
- Timmerman D, Ameye L, Fischerova D, Epstein E, Benedetto Melis G,

Guerriero S, Van Holsbeke C, Savelli L, Fruscio R, Lissoni AA, Testa AC, Veldman J, Vergote I, Van Huffel S, Bourne T, Valentin L. 2010b. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA Group. Br Med J 341, c6839.

- Topalak O, Saygili U, Soyturk M, Karaca N, Batur Y, Uslu T, Erten O. 2002. Serum, pleural effusion, and ascites CA-125 levels in ovarian cancer and nonovarian benign and malignant diseases: a comparative study. Gynecol Oncol 85 (1), 108-13.
- Valentin L. 1999a. Prospective cross-validation of Doppler ultrasound examination and gray-scale ultrasound imaging for discrimination of benign and malignant pelvic masses. Ultrasound Obstet Gynecol 14, 273–83.
- Valentin L. 199b. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. Ultrasound Obstet Gynecol 14, 338-47.
- Valentin L, Hagen B, Tingulstad S, Eik-Nes S. 2001. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. Ultrasound Obstet Gynecol 18 (4), 357-65.
- Valentin L, Akrawi D. 2002. The natural history of adnexal cysts incidentally detected at transvaginal ultrasound examination in postmenopausal women. Ultrasound Obstet Gynecol 20 (2), 174-80.
- Valentin L, Skoog L, Epstein E. 2003. Frequency and type of adnexal lesions in autopsy material from postmenopausal women: ultrasound study with histological correlation. Ultrasound Obstet Gynecol 22 (3), 284-9.
- Valentin L. 2004. Use of morphology to characterize and manage common

adnexal masses. Best Pract Res Clin Obstet Gynaecol 18 (1), 71-89.

- Valentin L, Ameye L, Jurkovic D, Metzger U, Lécuru F, Van Huffel S, Timmerman D. 2006. Which extrauterine pelvic masses are difficult to correctly classify as benign or malignant on the basis of ultrasound findings and is there a way of making a correct diagnosis? Ultrasound Obstet Gynecol 27 (4), 438-44.
- Valentin L, Ameye L, Testa A, Lécuru F, Bernard JP, Paladini D, Van Huffel S, Timmerman D. 2006b. Ultrasound characteristics of different types of adnexal malignancies. Gynecol Oncol 102 (1), 41-8.
- Valentin L, Ameye L, Franchi D, Guerriero S, Jurkovic D, Savelli L, Fischerova D, Lissoni A, Van Holsbeke C, Fruscio R, Van Huffel S, Testa A, Timmerman D. 2013. Risk of malignancy in unilocular cysts: a study of 1148 adnexal masses classified as unilocular cysts at transvaginal ultrasound and review of the literature. Ultrasound Obstet Gynecol 41 (1), 80-9.
- Van Calster B, Timmerman D, Bourne T, Testa AC, Van Holsbeke C, Domali E, Jurkovic D, Neven P, Van Huffel S, Valentin L. 2007. Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125. J Natl Cancer Inst 99 (22), 1706-14.
- Van Holsbeke C, Zhang J, Van Belle V, Paladini D, Guerriero S, Czekierdowski A, Muggah H, Ombelet W, Jurkovic D, Testa AC, Valentin L, Van Huffel S, Bourne T, Timmerman D. 2010a. Acoustic streaming cannot discriminate reliably between endometriomas and other types of adnexal lesion: a multicenter study of 633 adnexal masses. Ultrasound Obstet Gynecol 35 (3), 349-53.
- Van Holsbeke C, Van Calster B, Guerriero S, Savelli L, Paladini D, Lissoni AA, Czekierdowski A, Fischerova D, Zhang J, Mestdagh G, Testa AC, Bourne T, Valentin L, Timmerman D. 2010b. Endometriomas: their ultrasound characteristics. Ultrasound Obstet Gynecol 35 (6), 730-40.

- Van Holsbeke C, Daemen A, Yazbek J, Holland TK, Bourne T, Mesens T, Lannoo L, Boes AS, Joos A, Van De Vijver A, Roggen N, de Moor B, de Jonge E, Testa AC, Valentin L, Jurkovic D, Timmerman D. 2010c. Ultrasound experience substantially impacts on diagnostic performance and confidence when adnexal masses are classified using pattern recognition. Gynecol Obstet Invest 69 (3), 160-8.
- Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M; STROBE Initiative. 2007. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 4 (10), e297.
- Varras M, Polyzos D, Perouli E, Noti P, Pantazis I, Akrivis Ch. 2003. Tubo-ovarian abscesses: spectrum of sonographic findings with surgical and pathological correlations. Clin Exp Obstet Gynecol 30 (2-3), 117-21.
- Vernooij F, Heintz P, Witteveen E, van der Graaf Y. 2007. The outcomes of ovarian cancer treatment are better when provided by gynecologic oncologists and in specialized hospitals: a systematic review. Gynecol Oncol 105(3), 801-12.
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. 2007b. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 4 (10), e296.
- Webb PM, Purdie DM, Grover S, Jordan S, Dick ML, Green AC. 2004. Symptoms and diagnosis of borderline, early and advanced epithelial ovarian cancer. Gynecol Oncol 92 (1), 232-9.
- Wells G, Shea B, O'Connell J, Robertson J, Peterson D, Welch V, Losos M, Tugwell P. 2000. The Newcastle- Ottawa Scale (NOS) for assessing

the quality of nonrandomised studies in meta-analysis. 3rd Symposium on Systematic Reviews: Beyond the Basics; July 3–5; Oxford.

- Wells G, Brodsky L, O'Connell D, Robertson J. 2003. Evaluation of the Newcastle-Ottawa Scale (NOS): an assessment tool for evaluating the quality of non-randomized studies. XI Cochrane Colloquium: Evidence, Health Care and Culture; Oct 26–31; Barcelona, Spain.
- Wells G, Shea B, O'connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. (Accessed 29 April 2014) URL:http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
- Welsh Cancer Intelligence and Surveillance Unit. 2013. Cancer Incidence in Wales, 2007-2011. No. SA13/01.
- Westhoff C, Pike M, Vessey M. 1988. Benign ovarian teratomas: a population-based case-control study. Br J Cancer, 58 (1), 93–98
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155 (8), 529-36.
- Woodward ER, Sleightholme HV, Considine AM, Williamson S, McHugo JM, Cruger DG. 2007. Annual surveillance by CA125 and transvaginal ultrasound for ovarian cancer in both high-risk and population risk women is ineffective. BJOG 114 (12), 1500-9.
- World Medical Association. 1948. Declaration of Geneva. Adopted by the General Assembly of World Medical Association at Geneva Switzerland, September 1948.
- World Medical Association. 1949. International code of medical

ethics. World Medical Association Bulletin 1 (3), 109, 111.

- World Medical Association. 1995. World Medical Association Declaration on the Rights of the Patient, as amended by the 47th WMA General Assembly, Bali, Indonesia, September 1995.
- World Medical Association. 2013. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. JAMA 310 (20), 2191-2194.
- Yazbek J, Raju KS, Ben-Nagi J, Holland T, Hillaby K, Jurkovic D. 2007a. Accuracy of ultrasound subjective 'pattern recognition' for the diagnosis of borderline ovarian tumors. Ultrasound Obstet Gynecol 29 (5), 489-95.
- Yazbek J, Helmy S, Ben-Nagi J, Holland T, Sawyer E, Jurkovic D. 2007b. Value of preoperative ultrasound examination in the selection of women with adnexal masses for laparoscopic surgery. Ultrasound Obstet Gynecol 30 (6), 883–888.
- Yazbek J, Raju SK, Ben-Nagi J, Holland TK, Hillaby K, Jurkovic D. 2008. Effect of quality of gynaecological ultrasonography on management of patients with suspected ovarian cancer: a randomised controlled trial. Lancet Oncol 9 (2), 124-31.
- Yazbek J, Ameye L, Testa AC, Valentin L, Timmerman D, Holland TK, Van Holsbeke C, Jurkovic D. 2010. Confidence of expert ultrasound operators in making a diagnosis of adnexal tumor: effect on diagnostic accuracy and interobserver agreement. Ultrasound Obstet Gynecol 35 (1), 89-93.
- Young RH, Scully RE. 1991. Metastatic tumors in the ovary: a problem-oriented approach and review of the recent literature. Semin Diagn Pathol 8 (4), 250-76.
- Young RH. 2006. From krukenberg to today: the ever present problems posed by metastatic tumors in the ovary: Part I. Historical perspective,

general principles, mucinous tumors including the krukenberg tumor. Adv Anat Pathol 13 (5), 205-27.

- Young RH. 2007. From Krukenberg to today: the ever present problems posed by metastatic tumors in the ovary. Part II. Adv Anat Pathol 14 (3), 149-77.

Appendix 1: Publications

2017 Nunes N, Ambler G, Foo X, Naftalin J, Derdelis G, Widschwendter M, Jurkovic D. Comparison of two protocols for the management of asymptomatic postmenopausal women with adnexal tumours - a randomised controlled trial of RMI/RCOG vs Simple Rules. Br J Cancer. 2017 Feb 28;116(5):584-591.

2014 Nunes N, Ambler G, Foo X, Naftalin J, Widschwendter M, Jurkovic D. 2014. Use of the IOTA Simple Rules for the diagnosis of ovarian cancer: a Meta-Analysis. Ultrasound Obstet Gynecol 44 (5), 503-14.

2013 Nunes N, Ambler G, Hoo WL, Naftalin J, Foo X, Widschwendter M, Jurkovic D. 2013. A Prospective Validation of the IOTA Logistic Regression Models (LR1 and LR2) in Comparison to Subjective Pattern Recognition for the Diagnosis of Ovarian Cancer. Int J Gynecol Cancer 23 (9), 1583-9.

Permission to reproduce this data has been granted by Wolters Kluwer Health, Lippincott Williams & Wilkins.

2012 Nunes N, Foo X, Widschwendter M, Jurkovic D. 2012. A randomised controlled trial comparing surgical intervention rates between two protocols for the management of asymptomatic adnexal tumours in postmenopausal women. BMJ Open 2 (6), pii: e002248.

Permission to reproduce this data has been granted by BMJ Publishing Group Ltd.

2012 Nunes N, Yazbek J, Ambler G, Hoo W, Naftalin J, Jurkovic D. 2012. Prospective evaluation of the IOTA Logistic Regression Model LR2 for the diagnosis of ovarian cancer. Ultrasound Obstet Gynecol 40 (3), 355-9.

Appendix 2: RCT Proforma Simple Rules

RMI RCOG versus 'Simple Rules'

'Simple Rules'

Study number _____ Scan date _____

Symptoms:

Pain Pressure symptoms

Age: Time since the menopause: Parity:

'Simple Rules'

	M Rules			**B Rules**	
M1	Irregular solid tumour		**B1**	Unilocular	
M2	Ascites present		**B2**	Solid component largest <7mm	
M3	>/ 4 papilary projections		**B3**	Acoustic shadows	
M4	Irreg multiloc - solid tumour >/ 100mm		**B4**	Smooth multilocular tumour <100mm	
M5	Blood flow score 4		**B5**	No blood flow - score 1	

If one or more M-rules apply in the absence of a B-rule, the mass is classified as malignant.

If one or more B-rules apply in the absence of an M-rule, the mass is classified as benign.

If there are both or neither M and B rules that apply, the mass is unclassified.

Classified: **Benign** **Indeterminate** **Malignant**

Follow up visit 1: Unchanged Changed How _____

Follow up visit 2: Unchanged Changed How _____

Follow up visit 3: Unchanged Changed How _____

Proforma Simple Rules Version 1.1 07.04.2011

Appendix 3: RCT Proforma RMI/RCOG

RMI RCOG versus 'Simple Rules'

RMI

Study number _____ Scan date _____

CA125	
Bilateral	
Ascites	
Multolocular	
Solid	
Metastases	

RMI: _____

Cyst: Simple Unilateral Unilocular >2 - <5cm, CA125<30, RMI<25:

☐ Yes ☐ No

Classified: **Benign** **Indeterminate** **Malignant**

Follow up visit 1: Unchanged Changed How _____

Follow up visit 2: Unchanged Changed How _____

Follow up visit 3: Unchanged Changed How _____

Proforma RMI Version 1.1 07.04.2011