Pre- and peri-operative assessment of the adnexal mass

Peggy Geomini
© P.M.A.J. Geomini, 2008

NUR no: 883

No part of this thesis may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission on the copyright owner.

Cover: Dierk Hendriks
Lay-out: Henk en Hannie Dinnissen.
Printed by Verhagen Grafische Media, Veldhoven, The Netherlands.
Pre- and peri-operative assessment of the adnexal mass

PROEFSCHRIFT

ter verkrijging van de graad van doctor,
an de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. Mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 19 december 2008 om 10.00 uur
door

Peggy Marie Antoinette Joseph Geomini
Geboren 27 juni 1973 te Venlo
Promotores: Prof. dr. J.L.H. Evers,
Academisch Ziekenhuis Maastricht
Prof. dr. B.W.J. Mol,
Academisch Medisch Centrum Amsterdam

Copromotores: Dr. G.L. Bremer,
Maaslandziekenhuis Sittard
Prof. dr. R.F.P.M. Kruitwagen,
Academisch Ziekenhuis Maastricht

Beoordelingscommissie: Prof. dr. G.G.M. Essed,
Academisch Ziekenhuis Maastricht
Prof. dr. R.G.H. Beets-Tan,
Academisch Ziekenhuis Maastricht
Prof. dr. M.P.M. Burger,
Academisch Medisch Centrum Amsterdam
Dr. H.J. Hoogland,
Universiteit Maastricht
Prof. dr. D. Timmerman,
Universiteits Ziekenhuis Leuven
Chapter 1  General introduction  7

Chapter 2  The accuracy of risk scores in predicting ovarian malignancy; a systematic review. Submitted  15

Chapter 3  The pre-operative assessment of the adnexal mass: the accuracy of clinical estimates versus clinical prediction rules. BJOG 2003; 110:519-23.  39

Chapter 4  Evaluation of adnexal masses with threedimensional ultrasonography: a prospective study. Obstet Gynecol 2006; 108:1167-75.  51

Chapter 5  Is threedimensional ultrasonography of additional value in the assessment of adnexal masses? Gynecol Oncol 2007; 106:153-9.  67


Chapter 7  The impact of size of the adnexal mass on the accuracy of frozen section diagnosis. Gynecol Oncol 2005; 99:362-6.  97
| Chapter 8 | Patients' preferences in mode of surgery of an adnexal mass. J Psychosom Obstet and Gynaecol, accepted | 109 |
| Chapter 9 | Should we centralize care for the patient suspected of having ovarian malignancy? Submitted | 121 |
| Chapter 10 | Considerations, summary, clinical implications and future research. Beschouwing, samenvatting, klinische implicaties en toekomstig onderzoek. | 135 |
| Chapter 11 | Dankwoord en CV. | 171 |
Chapter 1

General introduction
Chapter 1

Introduction

Adnexal masses are a common reason for visiting a gynaecologist. Sometimes adnexal masses cause non-specific complaints like gastro-intestinal disturbances, increased circumference of the belly or problems at micturition or defecation. These symptoms occur only late in the course of disease. Mostly however, adnexal masses do not cause any signs or symptoms to the patient. They may be a chance finding at pelvic ultrasound. The main problem in the clinical management of the adnexal mass is its risk of malignancy. Because of this risk of malignancy, adnexal masses have to be assessed carefully prior to surgery. Treatment of a malignant adnexal mass is completely different from that of a benign adnexal mass.

According to data of the Integral Cancer Centre in The Netherlands, in 2003, ovarian malignancy was diagnosed in 1,096 women, which corresponds with an incidence of 13.4 in 100,000 women. In 2003, 928 women died because of an ovarian malignancy. The main reason for this high mortality rate is the fact that the majority of women with ovarian cancer have advanced-stage disease at the time of diagnosis. With respect to detailed information on survival, in The Netherlands, data from three regional cancer centres are available (Comprehensive Cancer Centre North, South and Amsterdam). Figure 1 shows the relative survival curve by stage between 1990 and 2002 in the Comprehensive Cancer Centre South.\(^1\) Five-year survival in stage I disease is about 80%, whereas five-year survival in stage II disease is 50%, which declines to 20% or less in stage III-IV disease. Similar data can be obtained from the FIGO Annual Reports. The 26\(^{th}\) FIGO annual report presented data from 7,314 cases (obvious malignant neoplasia and low potential malignancy) treated between 1999 and 2001 in centres all over the world. This report showed an overall five-year survival rate of ovarian malignancy of 50%, varying from 90% in stage I A to 19% in stage IV.\(^2\)

In case of early-stage ovarian malignancy (stage I A-II A) optimal staging is of utmost importance. Optimal staging includes, besides two sided salpingo-oophorectomy, hysterectomy and omentectomy, also peritoneal biopsies and lymph node sampling. The importance of optimal staging is demonstrated by the Adjuvant Chemo Therapy in Ovarian Neoplasm Trial (ACTION trial).\(^3\) In this study, adjuvant chemotherapy was associated with statistically significantly improved recurrence-free survival in patients with early-stage ovarian cancer. However, the benefit of adjuvant chemotherapy appeared to be limited to patients with non-optimal staging, i.e. patients with a higher risk of unappreciated residual disease. The degree of differentiation and completeness of staging are the most important prognostic factors in low stage ovarian malignancy.\(^4\)

In case of advanced-stage ovarian malignancy (stage IIB-IV), the standard of care is upfront surgery (primary cytoreduction) followed by intravenous administration of
Figure 1. *Ovarian malignancy* (relative survival curve 1990 - 2002 by stage)

Survival curve 1990 - 2002 by stage (data from Comprehensive Cancer Centre South).

taxane- and platinum-based chemotherapy. Patients with residual tumour one cm or less in diameter after primary surgery have higher survival rates than those with more extensive residual disease.\textsuperscript{5} In case the initial debulking is inadequate, a secondary surgical cytoreduction after three courses of chemotherapy needs to be considered, although literature is not unanimous regarding its value.\textsuperscript{6,7} For patients with coexisting disease that precludes the performance of a safe cytoreductive procedure, it is reasonable to establish the diagnosis of presumed ovarian cancer by biopsy or cytology, followed by the administration of platinum-based chemotherapy (so called neoadjuvant chemotherapy). If the patient has a response to treatment and becomes a more appropriate surgical candidate, it is rational to attempt tumour debulking after three cycles of chemotherapy.\textsuperscript{6,8-10} The degree of differentiation, Karnofsky-index, diameter of rest tumour after debulking and chemotherapy are important factors for the prognosis of advanced ovarian malignancy.\textsuperscript{11}

In conclusion, in ovarian malignancy, it is important to determine the correct disease stage and to remove as much tumour as possible in patients. Several studies showed that surgery by a gynaecological oncologist compared to surgery by a general gynaecologist more often occurred according to surgical guidelines and tumour was removed more often completely as a result of which survival of ovarian malignancy improved.\textsuperscript{12-14} In
The Netherlands, there is at present debate about the necessity for centralization of care for patients with ovarian cancer.

Treatment of benign adnexal masses is often also surgical, if only to establish or confirm the diagnosis. In case of benign adnexal masses, a laparoscopic approach is preferable to a laparotomic approach because of less postoperative complications like fever or bladder retention, less postoperative pain and shorter admission time. Moreover, in benign disease, ovarian tissue can be spared by performing a cystectomy instead of an adnectomy.

The optimal surgical route in case of an adnexal mass is determined by the risk of malignancy. It is therefore important to assess the risk of ovarian malignancy both pre-operatively and during surgery. For the differentiation between benign and malignant ovarian masses, many diagnostic parameters are available. In the pre-operative setting, the available parameters are the patients' history, physical and gynaecological examination, serum tumour markers, transabdominal or transvaginal ultrasonography including colour- and power Doppler examination, computed tomography and magnetic resonance imaging. In order to improve the pre-operative assessment of adnexal masses, most of these parameters have been combined with patient characteristics in diagnostic models. Although initial publications reported an almost perfect performance of these models, external validation showed their diagnostic performance to be less good.16-17

A potentially new diagnostic tool in the assessment of an adnexal mass is three dimensional (3D) ultrasonography. This technique visualizes the adnexal mass in all three planes (longitudinal, transverse and coronal), thus facilitating spatial evaluation as opposed to 2D ultrasonography, in which only two planes can be visualized. With 3D ultrasonography a volume of ultrasonographic data is acquired and stored, while with 2D ultrasonography only slices can be assessed. The integration of the acquired information enables computer assisted reconstruction of the organ being scanned. The addition of 3D power Doppler imaging allows visualization and quantitative judgement of vascularity in all three planes. Early publications on 3D ultrasonic examination of adnexal masses have suggested that the specificity of ultrasonographic imaging might improve with the addition of 3D power Doppler.18,19 These studies reported that masses with central vascular flow, vascular flow within excrescences or flow within septations are more likely to be malignant, as well as masses including vessels with a chaotic architecture or complex branching pattern.

During surgery, an additional diagnostic procedure for the assessment of the adnexal mass is frozen section diagnosis. In patients in whom the presence of malignancy prior to surgery is uncertain, frozen section diagnosis might be of use. Depending on the result of the frozen section diagnosis, the surgeon will decide to finish the surgical
procedure after removal of the adnexal mass, if surgical pathology confirms the mass to be benign, or to perform further surgery i.e. adequate staging or optimal debulking in case of malignancy.

This thesis deals with the diagnosis of malignancy in adnexal masses, both pre-operative as well as during surgery. The aim of the thesis is to answer the following questions:

1. How accurate are existing prediction models for the pre-operative assessment of adnexal masses, and how do these models compare to clinical estimates?
2. Is threedimensional ultrasonography of additional value in the assessment of adnexal masses?
3. What is the accuracy of frozen section diagnosis and which factors have impact on the accuracy of frozen section diagnosis?
4. What are patients’ preferences regarding the use of frozen section diagnosis and regarding the mode of surgery?
5. What are the implications of centralized and regular care for women with an ovarian malignancy in The Netherlands?
Outline of the thesis

To address the research questions we conducted the following studies:

Chapter 2 presents a meta-analysis of existing prediction models for the assessment of adnexal masses, in which the accuracy of different diagnostic models for the presence of malignancy is diagnosed.

Chapter 3 focuses on a study of the accuracy of existing diagnostic models for the assessment of adnexal masses in comparison with clinical estimates by oncologists, gynaecologists and residents.

Chapter 4 reports the results of a study in which the value of three-dimensional ultrasonography is examined. Women scheduled for surgical treatment of an adnexal mass were included in a multicenter prospective study. All women underwent two-dimensional and three-dimensional ultrasonographic examination in the week prior to surgery. All parameters were compared in women with benign tumours, borderline tumours, and malignant tumours using receiver operating characteristic analysis and likelihood ratios.

Chapter 5 compares a model based on patient characteristics and two-dimensional ultrasonographic parameters to a model based on patient characteristics, two-dimensional and three-dimensional ultrasonographic parameters.

Chapter 6 presents a meta-analysis of studies reporting on the accuracy of frozen section diagnosis. Because both the false positive and false negative frozen section diagnoses will generate some regret under patients, we additionally assessed expected regret reduction with frozen section diagnosis.

Chapter 7 reports the results of a retrospective study of the impact of tumour size on the accuracy of frozen section diagnosis. In this study, all women who had undergone frozen section diagnosis for an adnexal mass were included.

Chapter 8 studies patients’ preferences for frozen section diagnosis as an alternative to primary staging in case of an adnexal mass. Structured interviews were used to assess women’s preferences. In addition, these women were asked about their preference for laparoscopic or laparotomic approach of an adnexal mass, in view of possible negative side effects of laparoscopy in case of unexpected ovarian malignancy.
In chapter 9, we assessed the implications of centralized and regular care for women with an ovarian malignancy in The Netherlands, and compared it with a strategy in which a risk assessment was performed prior to surgery.

Chapter 10 contains considerations, a summary and implementations for future research.

References

1. www.ikcnet.nl
Chapter 1


Chapter 2

The accuracy of risk scores in predicting ovarian malignancy; a systematic review

Peggy M.A.J. Geomini
Roy F.P.M. Kruitwagen
Gérard L. Bremer
Ben Willem J. Mol
Abstract

Objective: Pre-operative assessment of the potential malignancy of an ovarian mass is important for optimal surgical treatment. The risk of malignancy of ovarian masses can be assessed pre-operatively with prediction models, in which multiple factors are included. The purpose of the present study was to perform a systematic review of the literature on the accuracy of prediction models in the pre-operative assessment of the adnexal mass.

Methods: Studies were identified through the Medline databases and EMBASE from inception to March 2008. Language restrictions were not applied. Two independent reviewers selected studies and extracted study characteristics, study quality and test accuracy. Accuracy data were used to form two by two contingency tables of the results of the risk score versus definitive histology. We used bivariate meta-analysis to estimate pooled sensitivities and specificities with 95% confidence intervals (CI) as well as summary Receiver Operating Characteristic (sROC) curves.

Results: There were 109 accuracy studies reporting on 21,750 adnexal mass (5,826 malignancies), that met the selection criteria. These studies reported on 83 different prediction models, one of which the model developed by Sassone was the most evaluated. All described prediction models have acceptable sensitivity and specificity. However, the Risk of Malignancy Index (RMI) I and II appeared to be the best predictors.

Conclusion: Based on our review, the RMI should be the test of choice in the pre-operative assessment of the adnexal mass.
Introduction

In the management of a patient with an adnexal mass, it is important to differentiate between malignant and benign pathology prior to surgery. Accurate assessment is of interest for patient counselling and pre-operative preparations, as well as for the choice for the optimal surgical route during the operation. In case of malignancy, optimal surgical staging and cytoreductive surgery are of utmost importance for survival.\textsuperscript{1,2} In case of benign pathology, a laparoscopically performed cystectomy is usually sufficient.\textsuperscript{3}

At present, several parameters are available to distinguish benign and malignant masses. Gray-scale ultrasonographic parameters are frequently used to assess the risk of malignancy. Other factors that are used to assess the risk of malignancy are colour Doppler ultrasound, as well as gynecologic examination, patient characteristics like menopausal state or age, tumor markers like CA125 and many others.\textsuperscript{4-6} In order to improve the pre-operative assessment of adnexal masses, most of these parameters have been combined in diagnostic models, which have been evaluated in a number of primary studies.\textsuperscript{7-13}

Existing reviews compare a selection of the described prediction models.\textsuperscript{9,14-16} Mol et al. performed an external validation of 21 existing models and concluded that their prognostic performance was not as good as that reported in original publications.\textsuperscript{16} Ferrazzi et al. compared the accuracy of five different scoring systems, four existing systems and one newly developed and came to a similar conclusion.\textsuperscript{9} Manjunath et al. compared three risk of malignancy indices, defined as the product of the ultrasound score, the menopausal score and the absolute value of serum CA125 levels.\textsuperscript{15} The Risk of Malignancy Index (RMI I) was originally described by Jacobs et al. in 1990.\textsuperscript{12} Tinguistad modified the RMI and developed a RMI II and RMI III.\textsuperscript{17,18} Manjunath did not find statistically differences in the performance of these three RMIs. The International Ovarian Tumor Analysis (IOTA-) Group collected in a prospective multicenter study, ultrasonographic and demographic data of > 1,000 patients with an adnexal mass.\textsuperscript{19} Based on this dataset, Van Holsbeke et al. performed an external validation of 19 mathematical models.\textsuperscript{20} The conclusion of this validation study was that the best result was obtained with a relevance vector machine with radial basis function kernel.

At present, there are no systematically conducted reviews of test accuracy of prediction models for differentiation between benign and malignant adnexal masses. The purpose of the present study was to perform a systematic review of the literature on the subject.
Methods

Search strategy and data extraction

We performed a computerized search to identify all registered articles evaluating diagnostic models comparing the risk of malignancy of ovarian pathology with the final histopathological diagnosis. We searched Medline and EMBASE from inception to March 2008.

The Medline search was performed using the following keywords: ["ovarian neoplasms"[MeSH] NOT "therapeutics"[MeSH] AND "model"] and ["ovarian neoplasms"[MeSH] NOT "therapeutics"[MeSH] AND "prediction"]. The EMBASE search was performed using the following keywords: [ovary tumor AND prediction], [ovary tumor AND Mathematical model] and [ovary tumor AND statistical model]. One of the investigators (PG) read all abstracts of the articles that were identified by the computer search. To be included for further reading, studies should contain data on the accuracy of models predicting the risk of malignancy in ovarian masses. A model should at least combine two parameters. In addition, cross-references of all selected articles were checked for other articles meeting the inclusion criteria. If they were applicable, these studies were added to the analysis.

After reading the selected full text publications, two by two tables comparing results of the prediction model and the definitive histological diagnosis were constructed independently by two out of three of the authors (GB, RK, PG). Calculations were limited to studies in which existing models were validated. Studies in which models were developed were not included. In the event of disagreement, the judgement of another author (BM) was decisive. The studies in which it was not possible to construct two by two tables were excluded from further processing.

From the selected publications the following quality characteristics were registered: (1) study design (cohort study versus case-control study), (2) type of data collection (prospective versus retrospective), (3) sampling method (consecutive versus other), (4) blinding of pathologist (present or absent), and (5) verification bias. Verification bias occurs if the decision to perform the reference test is based on the result of the test under study. Verification bias was default if the decision to obtain a histological diagnosis was based on the result of the prediction model.

Data analysis

For each study, we constructed a two by two table cross-classifying the risk score results and the presence of malignancy in an adnexal mass. To visualize data, we plotted for each model combinations of sensitivity and specificity in Receiver Operating
Characteristic (ROC) plots. A bivariate meta-regression model was used to calculate pooled estimates of sensitivity and specificity for risk score cut-off values and to fit a summary ROC (sROC) curve. This method has been extensively described elsewhere. Briefly, rather than using a single outcome measure per study, like the diagnostic odds ratio, the bivariate model preserves the twodimensional nature of diagnostic data in a single model. This model incorporates the correlation that may exist between sensitivity and specificity within studies due to possible differences in threshold between studies. The bivariate model uses a random effects approach for both sensitivity and specificity, allowing for heterogeneity beyond chance due to clinical or methodological differences between studies. In addition, the model acknowledges the difference in precision by which sensitivity and specificity have been measured in each study. This means that studies with a larger number of patients with a malignancy receive more weight in the calculation of the pooled estimate of sensitivity, while studies with more patients with a benign adnexal mass are more influential in the pooling of specificity.

In case the sensitivity-specificity points of multiple studies reporting on one model were grouped around an imaginable summary point-estimate, the pooled sensitivity and specificity were estimated with the bivariate model. In case the sensitivity-specificity points were grouped alongside an imaginable underlying ROC-curve (i.e. studies with high sensitivity had relatively low specificity and vice versa), a summary ROC-curve was estimated, again using the bivariate model. The bivariate models were fitted using the Proc NLMixed (SAS 9.1 for Windows (SAS Institute Inc, Cary, NC, USA).

Results

The results of the search are shown in figure 1. The computerized Medline en EMBASE search detected 1,161 publications. From the cross-references, 116 other studies were identified and selected for further reading. After reading abstracts 1,057 were excluded because they did not report accuracy of models which predict risk of malignancy in ovarian masses. As such, 220 studies were selected of which 114 had to be excluded for different reasons. Another three studies were found in the reference lists. Therefore, 109 studies were available for final analysis. The included studies reported on 21,750 adnexal masses, of which 15,490 were benign, 5,826 were malignant and 434 were reported as borderline malignancy. However, the number of borderline malignancies was not always described separately. The 109 included studies reported on 83 different prediction models. Figure 2 shows quality characteristics from the selected publications. We classified 95% of the studies as cohort studies. In 56% of the studies, data were collected prospectively, in 17% retrospectively, whereas in the
Figure 1. Results electronic search

Total citations identified from electronic searches to capture primary articles on accuracy of models which predict risk of malignancy in ovarian masses;
- first hits  n = 1161
- related articles n = 116

Citations excluded after screening titles and / or abstracts: n = 1057

Primary articles on risk scoring retrieved for detailed evaluation: n = 220

Articles excluded with reasons
- no test accuracy study n = 69
- lack of data to construct 2x2 table n = 39
- lack of original data i.e. letters n = 2
- not traceable n = 4
- total excluded n = 114

Primary articles included in systematic review
- from electronic search: n = 106
- from reference lists: n = 3

Figure 2. Quality characteristics of selected publications
other 26% of the studies the type of data collection was not described. In 45% of the studies adnexal masses were included consecutively. Blinding of the pathologist was not described in 77% of the studies. In 14% of the studies, verification bias could not be excluded.

Eighteen studies reported on the model originally described by Sassone.\textsuperscript{9,13,16,26-41} This model is based on a weighted sum of the following gray scale ultrasonographic variables: inner wall structure, wall thickness, septa and echogenicity. In two studies, the model described by Sassone was tested in combination with Doppler indices and CA125.\textsuperscript{34,36} In the 18 studies, 2,670 masses were included (1,726 benign, 889 malignant and 55 borderline tumours). Figure 3A shows the receiver operating characteristic sheet with results of 18 studies that reported on the prediction model of Sassone et al. The scoring range varied between 4 and 15, of which the cut-off point 9 was used most often. When the analysis was limited to estimates of sensitivity and specificity at a cut-off point of 9, the pooled estimate for sensitivity was 84% (95% CI 76 - 93) for a specificity of 80% (95% CI 73 - 88).

Alcazar used the result of the morphologic score of Sassone as a variable in a model based on logistic regression.\textsuperscript{7} His newly created model also contained colour Doppler findings. Alcazar’s model was evaluated in four publications which together report on 393 masses.\textsuperscript{7,42-44} Sensitivity and specificity ranged rather widely, thus limiting the possibility of providing pooled estimates (figure 3B).

Lerner described another modification of the model of Sassone and replaced wall thickness by shadowing.\textsuperscript{45} The model of Lerner was evaluated in eight studies, all of which reported on a cut-off point of 3.\textsuperscript{9,16,33,37,46-49} When the analysis was limited to estimates of sensitivity and specificity at this cut-off point, the pooled estimate for sensitivity was 90% (95% CI 87 - 98) for a specificity of 63% (95% CI 40 - 81). A summary ROC-curve was not estimated as there was no clear shift of increasing sensitivity and decreasing specificity or vice versa (figure 3C).

Two other morphologic scoring systems are the model described by DePriest et al. which is evaluated in ten studies\textsuperscript{8,9,16,28,29,31,33,37,50,51} and the model of Ferrazzi et al. which is evaluated in eight studies.\textsuperscript{9,16,26,28,33,52-55} The model described by DePriest et al. uses the weighted sum of tumour volume, wall structure and septal structure.\textsuperscript{8} The weighted sum of these three variables varied between 0 and 12, with the cut-off point most often set at 5.\textsuperscript{9,16,28,29,31,33,37,50} At this cut-off point, the pooled estimate for sensitivity was 91% (95% CI 85 - 97) for a specificity of 69% (95% CI 60 - 78) Again, a summary ROC-curve was not estimated (figure 3D).

The model described by Ferrazzi et al. is based on a weighted sum of wall structure, septa, vegetations and echogenicity.\textsuperscript{9} The scoring range varied between 4 and 18. At a cutoff point of 9, the model of Ferrazzi had a pooled estimate for sensitivity of 88%
Figure 2. Summary ROC curves of validated prediction models
(95% CI 71 - 96) and specificity of 74% (95% CI 59 - 89).\textsuperscript{16,26,28,33,53-55} (figure 3E)

There was one study in which malignancy was suspected in case the model of Ferrazzi scored ≥ 9 in combination with a CA125 level ≥ 35 kU/l (sens 87%, spec 95%).\textsuperscript{55}

Another study defined malignancy as a score ≥ 9 and presence of an additional factor like presence of ascites, mean diameter ≥ 10 cm, immobile or bilateral masses, RI ≤ 0.6, and serum CA125 ≥ 35 kU/l (sens 90%, spec 97%).\textsuperscript{52}

Granberg and Finkler published descriptive morphologic scoring systems.\textsuperscript{10,11} The scoring system of Granberg is claimed to be validated in three studies, however in two out of three studies the used cutoff value was unclear.\textsuperscript{9,16,33} The sensitivity-specificity points for the model of Finkler, validated in three studies, were strongly scattered in the ROC-sheet.\textsuperscript{10,16,56,57} The summary ROC-curve is shown in figure 3F.

The Risk of Malignancy Index was first described by Jacobs in 1990, and has since than been evaluated in 15 studies.\textsuperscript{12,14-17,20,53,58-66} This risk of malignancy index is also referred to as the RMI I, and uses the product of the serum CA125 level (U/ml), the ultrasound scan result (expressed as a score 0, 1 or 3) and the menopausal state (1 if premenopausal and 3 if postmenopausal).\textsuperscript{12} The test results vary between 0 and infinity. Jacobs described a cut off level of 200 with a sensitivity of 85% and a specificity of 97%.\textsuperscript{12} However most studies evaluate a range of cut off levels varying between 25 and 250. When 200 was used as cut-off level, the pooled estimate for sensitivity was 78% (95% CI 71 - 85) for a specificity of 87% (95% CI 83 - 91).\textsuperscript{14-17,20,58-65} At a cut-off level of 50, the pooled estimate for sensitivity was 91% (95% CI 85 - 97) for a specificity of 74% (95% CI 69 - 80). Figure 3G shows the results of all the individual studies, as well as the estimated summary ROC-curve.

In 1996, Tingulstad described an adjustment of the RMI, named the RMI II.\textsuperscript{17} The RMI II is based on the same product as the RMI I, except that the score for menopause is 1 in case of a premenopausal status and 4 in case of a postmenopausal status and the ultrasound score is expressed as a score 1 or 4. The score of the RMI II varies between 1 and ∞. The RMI II is evaluated in seven studies.\textsuperscript{14-16,20,60,67,68} When 200 was used as cut-off level, the pooled estimate for sensitivity was 79% (95% CI 71 - 87) for a specificity of 81% (95% CI 72 - 90).\textsuperscript{14-16,20,60,67,68} Figure 3H shows the results of all the individual studies, as well as the estimated summary ROC-curve.

Finally, also a RMI III and RMI IV have been developed.\textsuperscript{18,69} The RMI III an RMI IV both apply different ultrasound scores compared to RMI I and RMI II. The RMI III is evaluated in one study and showed at validation a sensitivity and specificity of 74% and 91%, respectively.\textsuperscript{15} The RMI IV has not been validated in other studies.

The model of Tailor is based on logistic regression analysis and integrates age, Doppler, i.e. time-averaged maximum mean velocity (TAMXV) and a papillary projections score as variables.\textsuperscript{70} This model has been evaluated in six
publications. The score varies between 0 and 100%, and the cut-off level is set on 50%. When 50 was used as cut-off level, the pooled estimate for sensitivity was 60% (95% CI 20 - 100) for a specificity of 93% (95% CI 82 - 100). (figure 31) When 25 was used as cutoff level, the pooled estimate for sensitivity was 78% (95% CI 33 - 100) for a specificity of 77% (95% CI 35 - 100).

Figure 4 shows all estimated summary ROC-curves of the validated prediction models. The figure indicates that the Risk of Malignancy Index I has the highest accuracy.

Other models have been validated only once or twice in another study. The results of these validation studies are shown in table 1. The logistic regression model II of Timmerman et al. has been validated in three studies, however at different cut-off levels. All other predictions models were only described in original publications without validation. Table 2 shows the results of these non validated prediction models.

Discussion

We performed a systematic review of the literature on the accuracy of prediction models in the pre-operative assessment of the adnexal mass. There were 109 studies

Figure 4. Estimated summary ROC-curves of all validated prediction models
reporting on 21,750 adnexal masses (5,826 malignancies) that met the selection criteria. Quality of the included studies was moderate. In 45% of the studies adnexal masses were included consecutive, however in 41% of the studies type of inclusion was not described clearly. In 14% of the studies verification bias could not be excluded, since not every mass was assessed histologically. Whether the pathologist was blinded for the result of the prediction model was almost never described.

The included studies reported on 83 different prediction models. All described prediction models had acceptable sensitivity en specificity. The Risk of Malignancy Index (RMI) I and II appeared to be the best predictors.

We differentiated four types of prediction models. First, descriptive numerical prediction models, which are based on description of morphologic characteristics of an adnexal mass assessed by ultrasonography. In these models, each characteristic is scored. These scores are added, with the total score expressing the risk of malignancy. The model of Sassone is the most frequent validated example of such a scoring system.\(^{13}\) At a cut-off point of nine, pooled estimation for sensitivity was 84% for a specificity of 80%. Lerner described a modification of Sassone’s prediction model.\(^{45}\) We found Lerner’s model to be less specific compared to that of Sassone, as were other validated descriptive numerical prediction models (Depriest \(^8\), Ferrazi \(^9\), Granberg \(^11\) and Finkler \(^10\)). As the criteria of Sassone’s prediction model are explicit and rigidly defined, this scoring system is easy to apply in clinical practice.

A second group of prediction models is based on logistic regression analysis. Logistic regression analysis is a mathematical procedure for systematically developing an algorithm to assess the probability that a particular event has occurred. To assess the probability of malignancy by logistic regression a complex formula has to be filled in. Several prediction models based on logistic regression analysis have been described in literature,\(^7,19,26,28,44,46,61,64,70,72,73,74,77,78,79,88\) but only a few models have been validated externally.\(^7,19,28,44,61,70,75,77,80,81,83,85,87,88\) Tailor’s model and Alcazar’s model have both been validated in four studies.\(^7,70\) For Tailor’s model, the pooled estimate for sensitivity was 60% for a specificity of 93% at a cut-off value of 50.\(^14,16,43,71\) For Alcazar’s model, sensitivity and specificity ranged rather widely, thus limiting the possibility of providing pooled estimates.\(^7,42,44\)

The formula of the risk of malignancy index (RMI), originally described by Jacobs (RMI I) and modified by Tingulstat (RMI II and III), is a simplification of a formula provided by logistic regression analysis.\(^{12,17,18}\) The RMI uses the product of three variables: ultrasound score, menopausal state and absolute value of serum CA125. The score of the RMI varies between zero and infinity for RMI I and between one and infinity for RMI II and III. The cut-off level most frequently applied for all three RMI’s is 200. When 200 was used as cut-off level, the pooled estimate for sensitivity was 78%
<table>
<thead>
<tr>
<th>Model (reference of original author)</th>
<th>Type of model</th>
<th>Validated by</th>
<th>Cut-off level</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression I of Timmerman(^{80})</td>
<td>Logistic regression</td>
<td>Timmerman(^{80})*(^{**})</td>
<td>25%</td>
<td>15</td>
<td>9</td>
<td>1</td>
<td>32</td>
<td>94</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Van Holsbeke(^{20})</td>
<td>25%</td>
<td>189</td>
<td>16</td>
<td>53</td>
<td>551</td>
<td>78</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Valentin(^{71})</td>
<td>25%</td>
<td>21</td>
<td>17</td>
<td>8</td>
<td>36</td>
<td>72</td>
<td>68</td>
</tr>
<tr>
<td>Logistic regression II of Timmerman (^{81})*</td>
<td>Logistic regression</td>
<td>Timmerman(^{61})*(^{**})</td>
<td>60%</td>
<td>15</td>
<td>7</td>
<td>1</td>
<td>34</td>
<td>94</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Van Holsbeke(^{20})</td>
<td>60%</td>
<td>203</td>
<td>140</td>
<td>39</td>
<td>427</td>
<td>84</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aslam(^{43})</td>
<td>50%</td>
<td>24</td>
<td>6</td>
<td>9</td>
<td>61</td>
<td>73</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mol(^{16})</td>
<td>45%</td>
<td>27</td>
<td>62</td>
<td>3</td>
<td>78</td>
<td>90</td>
<td>56</td>
</tr>
<tr>
<td>Artificial Neural Network I</td>
<td>Artificial neural network</td>
<td>Timmerman(^{81})*(^{**})</td>
<td>45%</td>
<td>14</td>
<td>3</td>
<td>2</td>
<td>38</td>
<td>88</td>
<td>93</td>
</tr>
<tr>
<td>Timmerman(^{61})</td>
<td></td>
<td>Van Holsbeke(^{20})</td>
<td>45%</td>
<td>197</td>
<td>126</td>
<td>45</td>
<td>441</td>
<td>81</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mol(^{16})</td>
<td>60%</td>
<td>27</td>
<td>56</td>
<td>3</td>
<td>84</td>
<td>90</td>
<td>60</td>
</tr>
<tr>
<td>Artificial Neural Network II</td>
<td>Artificial neural network</td>
<td>Timmerman(^{61})*(^{**})</td>
<td>60%</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>39</td>
<td>94</td>
<td>95</td>
</tr>
<tr>
<td>Timmerman(^{61})</td>
<td></td>
<td>Van Holsbeke(^{20})</td>
<td>60%</td>
<td>238</td>
<td>374</td>
<td>4</td>
<td>193</td>
<td>98</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mol(^{16})</td>
<td>60%</td>
<td>27</td>
<td>76</td>
<td>3</td>
<td>64</td>
<td>90</td>
<td>46</td>
</tr>
<tr>
<td>Alcazar(^{28})</td>
<td>Logistic regression</td>
<td>Alcazar(^{28})*(^{**})</td>
<td>6%</td>
<td>31</td>
<td>3</td>
<td>0</td>
<td>56</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Guerrieri(^{99})</td>
<td>6%</td>
<td>35</td>
<td>8</td>
<td>2</td>
<td>58</td>
<td>95</td>
<td>88</td>
</tr>
<tr>
<td>Prompeleger(^{76})</td>
<td>Logistic regression</td>
<td>Mol(^{16})</td>
<td>10%</td>
<td>27</td>
<td>91</td>
<td>3</td>
<td>49</td>
<td>90</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Van Holsbeke(^{20})</td>
<td>10%</td>
<td>153</td>
<td>112</td>
<td>113</td>
<td>688</td>
<td>56</td>
<td>86</td>
</tr>
<tr>
<td>Kurjak(^{100})</td>
<td>Combination of parameters</td>
<td>Mol(^{16})</td>
<td>4%</td>
<td>26</td>
<td>83</td>
<td>4</td>
<td>57</td>
<td>87</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vos(^{101})</td>
<td>4%</td>
<td>13</td>
<td>8</td>
<td>5</td>
<td>66</td>
<td>72</td>
<td>89</td>
</tr>
<tr>
<td>Minaretzi(^{75})</td>
<td>Logistic regression</td>
<td>Mol(^{16})</td>
<td>27%</td>
<td>81</td>
<td>3</td>
<td>59</td>
<td>90</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Van Holsbeke(^{20})</td>
<td>159 (80)</td>
<td>107</td>
<td>720</td>
<td>60</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gadduc(^{102})</td>
<td>Morphologic scoring system</td>
<td>Chiu(^{103})</td>
<td>10%</td>
<td>25</td>
<td>4</td>
<td>19</td>
<td>190</td>
<td>57</td>
<td>98</td>
</tr>
<tr>
<td>Mainz score(^{104})</td>
<td>Morphologic scoring system</td>
<td>Merz(^{105})</td>
<td>10%</td>
<td>25</td>
<td>30</td>
<td>3</td>
<td>256</td>
<td>89</td>
<td>90</td>
</tr>
<tr>
<td>Kurjak(^{106})</td>
<td>Morphologic scoring system</td>
<td>Kupesic(^{107})</td>
<td>5%</td>
<td>17</td>
<td>4</td>
<td>1</td>
<td>54</td>
<td>94</td>
<td>93</td>
</tr>
<tr>
<td>Valent(^{108})</td>
<td>Morphologic scoring system</td>
<td>Canuso(^{11})</td>
<td>3%</td>
<td>21</td>
<td>39</td>
<td>0</td>
<td>62</td>
<td>100</td>
<td>61</td>
</tr>
<tr>
<td>Yamashita(^{81})</td>
<td>Logistic regression</td>
<td>Mol(^{16})</td>
<td>27%</td>
<td>71</td>
<td>3</td>
<td>69</td>
<td>90</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Maggino(^{109})</td>
<td>Combination of parameters</td>
<td>Mol(^{16})</td>
<td>18%</td>
<td>27</td>
<td>12</td>
<td>113</td>
<td>60</td>
<td>81</td>
<td></td>
</tr>
</tbody>
</table>
Table 1.  (Continued)

<table>
<thead>
<tr>
<th>Model (reference of original author)</th>
<th>Type of model</th>
<th>Validated by</th>
<th>Cut-off level</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Szpurek 90</td>
<td>Artificial neural network</td>
<td>Szpurek 90**</td>
<td>75%</td>
<td>60</td>
<td>4</td>
<td>11</td>
<td>111</td>
<td>85</td>
<td>97</td>
</tr>
<tr>
<td>Schelling 77</td>
<td>Logistic regression</td>
<td>Schelling 77**</td>
<td>50%</td>
<td>36</td>
<td>13</td>
<td>3</td>
<td>205</td>
<td>92</td>
<td>94</td>
</tr>
<tr>
<td>Jokubkiene 43</td>
<td>Logistic regression</td>
<td>Van Holsbeke 20</td>
<td>12%</td>
<td>210</td>
<td>123</td>
<td>56</td>
<td>677</td>
<td>79</td>
<td>85</td>
</tr>
<tr>
<td>Lu 84</td>
<td>LS-SVM linear</td>
<td>Van Holsbeke 20</td>
<td>30%</td>
<td>218</td>
<td>141</td>
<td>24</td>
<td>426</td>
<td>90</td>
<td>75</td>
</tr>
<tr>
<td>Lu 84</td>
<td>LS-SVM non-linear</td>
<td>Van Holsbeke 20</td>
<td>30%</td>
<td>216</td>
<td>135</td>
<td>26</td>
<td>432</td>
<td>89</td>
<td>76</td>
</tr>
<tr>
<td>Lu 84</td>
<td>RVM, linear</td>
<td>Van Holsbeke 20</td>
<td>30%</td>
<td>219</td>
<td>144</td>
<td>23</td>
<td>432</td>
<td>90</td>
<td>75</td>
</tr>
<tr>
<td>Lu 84</td>
<td>RVM, non-linear</td>
<td>Van Holsbeke 20</td>
<td>30%</td>
<td>218</td>
<td>146</td>
<td>24</td>
<td>421</td>
<td>90</td>
<td>74</td>
</tr>
<tr>
<td>Lu 89</td>
<td>Artificial neural network</td>
<td>Van Holsbeke 20</td>
<td>30%</td>
<td>164</td>
<td>52</td>
<td>78</td>
<td>515</td>
<td>68</td>
<td>91</td>
</tr>
<tr>
<td>Van Calster 66</td>
<td>LS-SVM, add RBF</td>
<td>Van Calster 66*</td>
<td>12%</td>
<td>68</td>
<td>43</td>
<td>7</td>
<td>194</td>
<td>91</td>
<td>82</td>
</tr>
<tr>
<td>Van Calster 66</td>
<td>RVM, add RBF</td>
<td>Van Calster 66*</td>
<td>15%</td>
<td>70</td>
<td>45</td>
<td>5</td>
<td>192</td>
<td>93</td>
<td>81</td>
</tr>
<tr>
<td>Tailor 91</td>
<td>Artificial neural network</td>
<td>Tailor 91**</td>
<td>50%</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Moszynski 85</td>
<td>Logistic regression</td>
<td>Moszynski 85**</td>
<td>24%</td>
<td>12</td>
<td>7</td>
<td>5</td>
<td>57</td>
<td>77</td>
<td>83</td>
</tr>
<tr>
<td>Moszynski 85</td>
<td>Artificial neural network</td>
<td>Moszynski 85**</td>
<td>27%</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>63</td>
<td>87</td>
<td>91</td>
</tr>
<tr>
<td>Valentin 88</td>
<td>Logistic regression</td>
<td>Valentin 88**</td>
<td>50%</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>75</td>
<td>55</td>
</tr>
<tr>
<td>Zhang 92</td>
<td>Artificial neural network</td>
<td>Zhang 92**</td>
<td>50%</td>
<td>72</td>
<td>7</td>
<td>9</td>
<td>79</td>
<td>89</td>
<td>92</td>
</tr>
<tr>
<td>Zhang 93</td>
<td>Artificial neural network</td>
<td>Zhang 93**</td>
<td>50%</td>
<td>37</td>
<td>2</td>
<td>5</td>
<td>96</td>
<td>88</td>
<td>98</td>
</tr>
<tr>
<td>Itakura 51</td>
<td>Combination of parameters</td>
<td>Itakura 51**</td>
<td>26%</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>34</td>
<td>84</td>
<td>98</td>
</tr>
<tr>
<td>Timmerman 19</td>
<td>Logistic regression</td>
<td>Timmerman 19**</td>
<td>75%</td>
<td>41</td>
<td>1</td>
<td>34</td>
<td>236</td>
<td>55</td>
<td>99</td>
</tr>
<tr>
<td>Smolen 87</td>
<td>Logistic regression</td>
<td>Smolen 87**</td>
<td>50%</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>66</td>
<td>79</td>
</tr>
<tr>
<td>Marret 44</td>
<td>Logistic regression</td>
<td>Marret 44**</td>
<td>10%</td>
<td>12</td>
<td>6</td>
<td>3</td>
<td>47</td>
<td>80</td>
<td>89</td>
</tr>
<tr>
<td>Tay 110</td>
<td>Morphologic scoring system</td>
<td>Tay 111</td>
<td>7%</td>
<td>20</td>
<td>29</td>
<td>2</td>
<td>52</td>
<td>91</td>
<td>64</td>
</tr>
</tbody>
</table>

*Logistic regression II of Timmerman 61 has been validated by three authors at different cut-off levels, **Development and validation by the same author.
<table>
<thead>
<tr>
<th>Author</th>
<th>Type of model</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity %</th>
<th>Specificity %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>Artificial neural network</td>
<td>74</td>
<td>25</td>
<td>7</td>
<td>110</td>
<td>91</td>
<td>81</td>
</tr>
<tr>
<td>Seboui</td>
<td>Combination of parameters</td>
<td>11</td>
<td>5</td>
<td>18</td>
<td>119</td>
<td>38</td>
<td>96</td>
</tr>
<tr>
<td>Manceusso</td>
<td>Combination of parameters</td>
<td>14</td>
<td>4</td>
<td>0</td>
<td>107</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>Guerriero</td>
<td>Combination of parameters</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>73</td>
<td>100</td>
<td>97</td>
</tr>
<tr>
<td>Roman</td>
<td>Combination of parameters</td>
<td>22</td>
<td>4</td>
<td>5</td>
<td>150</td>
<td>81</td>
<td>97</td>
</tr>
<tr>
<td>Roman</td>
<td>Combination of parameters</td>
<td>7</td>
<td>2</td>
<td>9</td>
<td>27</td>
<td>44</td>
<td>93</td>
</tr>
<tr>
<td>Chou</td>
<td>Combination of parameters</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>83</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Botsas</td>
<td>Combination of parameters</td>
<td>18</td>
<td>8</td>
<td>4</td>
<td>32</td>
<td>82</td>
<td>80</td>
</tr>
<tr>
<td>Bay</td>
<td>Combination of parameters</td>
<td>30</td>
<td>3</td>
<td>4</td>
<td>95</td>
<td>90</td>
<td>97</td>
</tr>
<tr>
<td>Antonic</td>
<td>Combination of parameters</td>
<td>18</td>
<td>28</td>
<td>0</td>
<td>25</td>
<td>100</td>
<td>47</td>
</tr>
<tr>
<td>Franchi</td>
<td>Combination of parameters</td>
<td>17</td>
<td>54</td>
<td>20</td>
<td>38</td>
<td>46</td>
<td>41</td>
</tr>
<tr>
<td>Jacobs</td>
<td>Combination of parameters</td>
<td>28</td>
<td>7</td>
<td>14</td>
<td>94</td>
<td>67*</td>
<td>93*</td>
</tr>
<tr>
<td>Szpurek</td>
<td>Logistic regression</td>
<td>224</td>
<td>27</td>
<td>31</td>
<td>404</td>
<td>88</td>
<td>94</td>
</tr>
<tr>
<td>Smolen</td>
<td>Logistic regression</td>
<td>43</td>
<td>16</td>
<td>6</td>
<td>143</td>
<td>88*</td>
<td>90*</td>
</tr>
<tr>
<td>Alcazar</td>
<td>Logistic regression</td>
<td>31</td>
<td>6</td>
<td>1</td>
<td>97</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>Kebat</td>
<td>Logistic regression</td>
<td>21</td>
<td>3</td>
<td>3</td>
<td>107</td>
<td>88</td>
<td>97</td>
</tr>
<tr>
<td>Brown</td>
<td>Logistic regression</td>
<td>26</td>
<td>13</td>
<td>2</td>
<td>170</td>
<td>93*</td>
<td>93*</td>
</tr>
<tr>
<td>Kishi</td>
<td>Logistic regression</td>
<td>98</td>
<td>15</td>
<td>2</td>
<td>120</td>
<td>98</td>
<td>89</td>
</tr>
<tr>
<td>Jakubkienie</td>
<td>Logistic regression</td>
<td>27</td>
<td>6</td>
<td>0</td>
<td>73</td>
<td>100</td>
<td>92</td>
</tr>
<tr>
<td>Moore</td>
<td>Logistic regression</td>
<td>51</td>
<td>8</td>
<td>16</td>
<td>158</td>
<td>76</td>
<td>95</td>
</tr>
<tr>
<td>Ulusoy</td>
<td>Logistic regression</td>
<td>67</td>
<td>22</td>
<td>39</td>
<td>168</td>
<td>63</td>
<td>88</td>
</tr>
<tr>
<td>Balbi</td>
<td>Logistic regression</td>
<td>17</td>
<td>5</td>
<td>5</td>
<td>45</td>
<td>77</td>
<td>90</td>
</tr>
<tr>
<td>Mousavi</td>
<td>Logistic regression</td>
<td>52</td>
<td>7</td>
<td>1</td>
<td>41</td>
<td>98</td>
<td>85</td>
</tr>
<tr>
<td>Rampone</td>
<td>Morphologic scoring system</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>48</td>
<td>100</td>
<td>94</td>
</tr>
<tr>
<td>Szpurek</td>
<td>Morphologic scoring system</td>
<td>221</td>
<td>99</td>
<td>34</td>
<td>332</td>
<td>87</td>
<td>77</td>
</tr>
<tr>
<td>Ueland</td>
<td>Morphologic scoring system</td>
<td>52</td>
<td>75</td>
<td>1</td>
<td>314</td>
<td>98*</td>
<td>81*</td>
</tr>
<tr>
<td>Benjipibal</td>
<td>Morphologic scoring system</td>
<td>34</td>
<td>15</td>
<td>1</td>
<td>70</td>
<td>97*</td>
<td>82*</td>
</tr>
<tr>
<td>Twickler</td>
<td>Morphologic scoring system</td>
<td>26</td>
<td>36</td>
<td>5</td>
<td>178</td>
<td>84*</td>
<td>83*</td>
</tr>
<tr>
<td>Merce</td>
<td>Morphologic scoring system</td>
<td>21</td>
<td>19</td>
<td>1</td>
<td>88</td>
<td>95</td>
<td>82</td>
</tr>
<tr>
<td>Kawal</td>
<td>Morphologic scoring system</td>
<td>8</td>
<td>6</td>
<td>1</td>
<td>9</td>
<td>89</td>
<td>60</td>
</tr>
<tr>
<td>Szpurek</td>
<td>Morphologic scoring system</td>
<td>143</td>
<td>20</td>
<td>22</td>
<td>279</td>
<td>87</td>
<td>93</td>
</tr>
<tr>
<td>Brazari</td>
<td>Morphologic scoring system</td>
<td>24</td>
<td>4</td>
<td>2</td>
<td>29</td>
<td>92</td>
<td>88</td>
</tr>
<tr>
<td>Szpurek</td>
<td>Morphologic scoring system</td>
<td>143</td>
<td>20</td>
<td>22</td>
<td>279</td>
<td>87</td>
<td>93</td>
</tr>
<tr>
<td>Canuso</td>
<td>Morphologic scoring system</td>
<td>21</td>
<td>8</td>
<td>0</td>
<td>93</td>
<td>100</td>
<td>92</td>
</tr>
<tr>
<td>Daskalakis</td>
<td>Morphologic scoring system</td>
<td>21</td>
<td>9</td>
<td>3</td>
<td>94</td>
<td>88</td>
<td>91</td>
</tr>
<tr>
<td>Kapesic</td>
<td>Multiplication of parameters</td>
<td>17</td>
<td>0</td>
<td>1</td>
<td>58</td>
<td>94</td>
<td>100</td>
</tr>
</tbody>
</table>

* study provides sensitivity and specificity at different cut-off values; best cut-off level is shown
(RMI I) and 79% (RMI II) for a specificity of 87% (RMI I) and 81% (RMI II) respectively. Because of its simplicity combined with good test characteristics, the RMI is useful in clinical practice.

A third type of prediction models is the artificial neural network, which distinguishes itself from logistic regression analysis by its possibility to investigate the possible existence of non-linear interaction between variables. Neural networks provide very complex formulas. Timmerman developed two artificial neural networks (ANN I and ANN II) to predict malignancy in an adnexal mass.\textsuperscript{61} At external validation, ANN II performed better compared to ANN I. Sensitivity and specificity of ANN I were 81% and 78% respectively 90% and 60% in two validation studies.\textsuperscript{16,20} Meanwhile some more artificial neural networks have been developed.\textsuperscript{46,85,89-93} However, external validation is limited. Because formula’s provided by artificial neural networks are very complex to fill in, these are less useful in clinical practice.

A fourth type of mathematical models are least squares support vector machines (LS-SVM) and relevance vector machines (RVM), which are very complex models. Experience with these models is very limited.\textsuperscript{66,94} However, Van Holsbeke found a RVM with radial basis function kernel to perform the best out of 19 externally validated mathematical models.\textsuperscript{20}

The outcome of ovarian cancer is better when a gynaecological oncologist provides treatment than a general gynaecologist.\textsuperscript{95} Consequently, patients suspected for having ovarian cancer should be operated by a gynecologic oncologist. The American College of Obstetricians and Gynecologists (ACOG) formulated in conjunction with the Society of Gynecologic Oncologists (SGO) criteria that should be used to consider referral to or consultation with a gynaecological oncologist.\textsuperscript{96} These referral guidelines were evaluated by Dearking et al.\textsuperscript{97} The guidelines, which are not based on mathematical models, had a positive predictive value of 39.6% for premenopausal women and a positive predictive value of 64.6% for postmenopausal women. Bailey et al. examined whether RMI can be used for adequate referral of ovarian cancer to a tertiary center.\textsuperscript{98} In this retrospective study, patients referred to a tertiary centre were included. CA125 measurements were performed in different laboratories and ultrasonographic examinations were performed by different ultrasonographers. Bailey found a positive predictive value of 86.8%.

In conclusion, there are several validated prediction models available for the pre-operative assessment of the adnexal mass. All validated prediction models have an acceptable sensitivity and specificity. The RMI I as described by Jacobs is a simple scoring system, appears to be very accurate and is useful in clinical practice, and should therefore be the test of choice in the pre-operative assessment of the adnexal mass.
References


Risk scores in predicting ovarian malignancy; a systematic review


42. Aleazar JL, Jurado M. Prospective evaluation of a logistic model based on sonographic morphologic and color Doppler findings developed to predict adnexal malignancy. J Ultrasound Med 1999; 18:837-42.


88. Valentin L, Ameve L, Jurkovic D, Metzger U, Lecuru F, Van HS et al. Which extraterine 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 2006; 27:438-44.


95. Vermeij F, Heintz P, Witteveen E, van der GY. The outcomes of ovarian cancer treatment are better when provided by gynecologic oncologists and in specialized hospitals: a systematic review. Gynecol Oncol 2007; 105:901-12.


Chapter 3

The pre-operative assessment of the adnexal mass: the accuracy of clinical estimates versus clinical prediction rules

Dorry Boll
Peggy M.A.J. Geomini
Hans A.M. Brölfmann
Edith A. Sijmons
Peter M. Heintz
Ben Willem J. Mol

Chapter 3

Abstract

Objective: The aim of this study was to evaluate the reproducibility of the clinical judgment of gynaecologists, gynaecologists in training and gynaecological oncologists and to compare the predictive performance of the off hand assessment with the predictive performance of existing mathematical models.

Methods: We offered 45 gynaecologists five different sets of 34 cases, with data on female age and menopausal status, a written description of the ultrasonography, Doppler flow measurement and serum CA125 measurement. Nine observers for every set were asked to express their perception on the chance that the adnexal mass was malignant.

Results: Statistically significant differences could be found between the accuracy of the risk assessments made by the clinicians and the accuracy of the risk assessments made by prediction models.

Conclusion: This study demonstrates that at this moment there is no need to introduce complicated predictive scoring systems like neural networks or logistic regression models for the pre-operative assessment of the adnexal masses.
Introduction

The adnexal mass is still a diagnostic dilemma in daily gynaecological practice. In case a surgical procedure is planned, it is important to distinguish a benign from a malignant mass prior to surgery. In practice, the diagnosis of ovarian cancer is made during surgery and often leads to inadequate surgical exploration by junior or inexperienced surgeons. This appears to be an important factor for the prognosis of the patient with ovarian cancer.1-3

In order to increase the accuracy of the pre-operative assessment of the adnexal mass, many authors have developed scoring systems. At the presentation of these models, sensitivity and specificity were claimed to be as high as 95% for both.4,5 However, external validation of these scoring systems showed that diagnostic accuracy was not as good as it appeared to be at initial presentation.4 At external validation, sensitivity and specificity dropped to 90% and 50% for logistic regression models, whereas neural networks showed a sensitivity of 90% for a specificity of 60% at external validation.

In view of these data, it is questionable whether the predictive performance of these mathematical models is superior to the off hand assessment that is made by clinicians when assessing a patient with an adnexal mass in absence of these formal models. The comparison between off hand risk assessment by the clinician and the risk assessment as made by clinical prediction models is important, since clinical prediction rules require more complicated calculations or the use of a computer, and many clinicians are not familiar with such procedures. Thus, in order to justify introduction of these clinical prediction rules, they must have a better accuracy as compared to the off hand risk assessment that is made by a gynaecologists.

We studied the clinical judgment of gynaecologists in women undergoing surgery for an adnexal mass. The reproducibility of the assessment made by the gynaecologists was evaluated, and the predictive performance of the off hand assessment was compared to the predictive performance of existing models.

Methods

We used a dataset of 170 women who underwent surgery for an adnexal mass, this dataset has been reported on previously.4 The women were prospectively included between January 1991 and December 1998 in Máxima Medical Centre in Veldhoven, The Netherlands. Máxima Medical Centre is a general teaching hospital with 500 beds, and has a general gynaecology department without a specialist referral function for gynaecologic oncology. Each case consisted data on female age and menopausal status, a written description of the transvaginal ultrasonography, transvaginal Doppler flow
measurement and serum CA125 measurement. Clinical examination was not included in any of the prediction models that were evaluated. Consequently, we did not add the results of clinical examination to the questionnaire. Ultrasonography and Doppler flow measurement had been performed by two gynaecologists with a vaginal 5.0 MHz convex array probe (Aloka SSD-1400 and ATL-4 Ultramark, Almere, The Netherlands). In case of large adnexal masses the examination was extended by abdominal ultrasound examination. CA125 measurements had been performed using a chemoluminescent immunometric assay (Immulite 2000 OM-MA, DPC, Diagnostic Product Corporation, Los Angeles, USA) and were expressed in IU/L. Malignancy was defined according to criteria of histological typing of ovarian tumours, according to the WHO 1973 and staged according to criteria recommended by the International Federation of Gynaecology and Obstetrics (FIGO 1988).6

Between 1991 and 1998, 170 patients were operated for an adnexal mass. The median age of these patients was 46 years, with a range from 20 to 89 years. Sixty-one (36%) of the 170 patients were postmenopausal. On ultrasound, 120 (71%) of the masses were translucent. Papillary formations were seen in 50 (29%), of the adnexal masses, 51 (30%) had septa smaller than 3 mm, and 25 (15%) had septa of more than 3 mm. The median size of the adnexal masses was 7 cm (range 3 cm to 28 cm). Ascites was present in 87 (51%) of the patients, and 37 (22%) of the patients had an abnormal Doppler flow. The median level of CA125 was 11 (range 1 to 1635). Forty-two patients (25%) had a CA125 above 30 IU/L.

Overall, thirty patients (18%) turned out to have a malignancy, whereas the remaining patients had benign ovarian lesions. Among the malignant diagnoses, there were five serous carcinomas, 14 mucinous carcinomas, five endometrioid carcinomas, five clearcell carcinomas and one metastatic tumour from the breast. Seventeen (57%) of the patients had FIGO stage I disease, five (17%) had FIGO stage II disease and eight (27%) had FIGO stage III disease.

From this dataset of 170 cases, we randomly selected 34 cases to form five different sets. We offered these sets to 45 gynaecologists, 15 gynaecological oncologists, 15 general gynaecologists and 15 senior residents. A senior resident is a medical doctor who is doing a six year training program to become a gynaecologist. We have sent the questionnaires to senior residents who had at least two years of training behind. Every set was sent to three gynaecological oncologists, three general gynaecologists and three senior residents. So there were nine observers for every set. All the participating gynaecologists were unaware of the histological status of the patient at the moment they were completing the forms. After presentation of each case summary, the gynaecologist was asked to express his or her perception on the chance that the adnexal mass was malignant on a scale from 0% to 100%. Each participant was also asked to judge
whether the subsequent surgical approach would be laparoscopy, laparotomy by Pfannenstiel’s incision or laparotomy by midline incision.

**Statistical analysis**

We assessed the reproducibility of the risk estimates that were made by the participants, using Intraclass Correlation Coefficients (ICC). The ICC expresses the degree to which the total variance can be attributed to the true variance: true differences between subjects. It not only assesses the strength of correlation between two measurements but also detects systemic errors. We calculated an ICC for all participants, and compared ICC’s between the three subgroups of participants, i.e. gynaecological oncologists, general gynaecologists and residents. We also assessed the reproducibility of the surgical approach that was chosen. This was done by calculation of a weighted k-coefficient. A k-statistic can vary between 0 and 1: a k-statistic of 0 indicates no agreement beyond chance, whereas a k-statistic of 1 indicates perfect agreement between observers. The reproducibility in case of k-statistic between 0 and 0.2 is regarded as ‘slight’, between 0.2 and 0.4 as ‘fair’, between 0.4 and 0.6 as ‘moderate’, between 0.6 and 0.8 as ‘substantial’, and between 0.8 and 1.0 as ‘almost perfect’.

In a next step, we assessed the accuracy of the judgment of the clinicians in the prediction of malignancy. To do so, we performed Receiver Operating Characteristic (ROC) analyses, in which the test result, i.e. the risk of malignancy as estimated by the gynaecologists, was compared to the presence or absence of malignancy, as established at histological examination by a pathologist. We constructed a ROC-curve, and calculated an area under the ROC-curve. The area under the ROC-curve expresses the performance of a diagnostic test taking values in the range between 0.5 and 1. An area under the ROC-curve of 0.5 implies that the diagnostic test under study has a discriminative capacity that does not exceed chance, whereas an area under the ROC-curve of 1 implies that the discriminative capacity of the test under study is perfect. ROC-analysis was performed for all participants together, as well as for gynaecological oncologists, general gynaecologists and residents separately.

Subsequently, we compared these ROC-curves to curves that were constructed for the performance of prediction models, i.e. logistic regression models or neural networks.

To assess the relationship between diagnostic accuracy and reproducibility, we stratified the results of the interpretation of the adnexal mass for the amount of agreement. To do so, the results of the three categories of observers (gynaecological oncologists, general gynaecologists and residents) were added, so there were nine observations on each patient with an adnexal mass.
Finally, we compared the risk assessment made by the gynaecologists to the choice of surgical approach. The distribution of the risk of malignancy was plotted against the choice of surgical approach, i.e. laparoscopically, laparotomy par Pfannenstiel or par midline incision.

Results

The response was 42 sets out of 45 (93%). Two senior residents and one general gynaecologist did not respond. The ICC for all participants was 0.82 (lower 95% CI 0.77). For the gynaecological oncologists the ICC was 0.65 (lower 95% CI 0.57), whereas for general gynaecologists and residents the ICC’s were 0.59 (lower 95% CI 0.52) and 0.61 (lower 95% CI 0.54), respectively. The overall weighted k-value for the choice of the surgical approach was 0.52 (95% CI 0.46 - 0.58). For the subgroups of oncologists, general gynaecologists and residents these weighted k-values were 0.50 (95% CI 0.39 - 0.60), 0.50 (95% CI 0.40 - 0.61) and 0.57 (95% CI 0.47 - 0.66), respectively.

Figure 1 shows the ROC-curves for the three categories of participants. The areas under the ROC-curve were 0.82 (se 0.02), 0.84 (se 0.02) and 0.82 (se 0.02) for oncologists, general gynaecologists and residents, respectively. These curves did not differ statistically significant (p-value 0.2). Figure 1 also shows the areas under the ROC-curve for the logistic model and neural network. Again, no statistically significant differences could be found between the accuracy of the risk assessments made by the clinicians and the accuracy of the risk assessments made by the prediction models (p-value 0.2).

Figure 2 shows a histogram for the judgment, in which the diagnosis on each adnexal mass as made by the observers is stratified for the amount of agreement between observers. The white parts of the bars represent the presence of a benign adnexal mass, whereas the black parts represent malignancy. Figure 2A shows the ideal situation: if, according to the reference standard, disease is present, all observers consider disease to be present, and vice. The nine observers strongly agreed in their interpretation in a majority of the benign masses. There was one mass which turned out to be malignant, that was diagnosed by none of the nine observers; there was one mass that turned out to be malignant, that was diagnosed by one of the nine observers; and there was one mass that turned out to be malignant that was diagnosed by two of the nine observers. Six adnexal masses that were malignant were qualified as such by all nine observers. In the remaining 21 malignant masses, there was no agreement, although in each of these cases a large majority of the nine observers qualified the mass as being malignant.

Figure 3 shows the distribution of the estimated probabilities of malignancy as
The accuracy of clinical estimates versus clinical prediction rules

Figure 1.

Receiver Operating Characteristic curves demonstrating the capacity of residents, general gynaecologists and gynaecological oncologists to predict malignancy of the adnexal mass. As comparison, the ROC-curves that represent the performance of prediction models based on logistic regression models and neural networks are shown.

Figure 2.

Relation between diagnostic accuracy and reproducibility. Each bar represents the number of observers that judged an adnexal mass as malignant or benign. The black bars represent malignancy, whereas the white bars represent benign adnexal masses.

A. Test with perfect performance;
B. Situation for adnexal mass.
Chapter 3

Figure 3.

Distribution of the estimated probabilities of malignancy as estimated by residents, general gynaecologists and gynaecological oncologists.
estimated by residents, general gynaecologists and gynaecological oncologists. The residents opted for laparoscopy in 49 cases, for Pfannenstiel incision in 69 cases, and for midline incision in 52 cases. The median estimated probabilities of malignancy in these three categories were 1%, 6% and 50%, respectively. The general gynaecologists opted for laparoscopy in 65 cases, for Pfannenstiel incision in 55 cases, and for midline incision in 50 cases. The median estimated probabilities of malignancy in these three categories were 2%, 5% and 30%, respectively. The gynaecological oncologists opted for laparoscopy in 85 cases, for Pfannenstiel incision in 30 cases, and for midline incision in 54 cases. The median estimated probabilities of malignancy in these three categories were 1.5%, 4% and 20%, respectively.

Table 1 shows the choice for laparoscopy, Pfannenstiel or midline incision in relation to the presence or absence of malignancy for the three groups of participants. If one would consider the choice for midline incision as a ‘positive’ test result and the choice for a Pfannenstiel incision or laparoscopy as a ‘negative’ test result, than the performance of the three groups would be comparable, with sensitivities varying between 70% and 77%, and specificities varying between 77% and 79%.

<table>
<thead>
<tr>
<th>Table 1.</th>
<th>The choice for midline incision in case of malignancy as positive test result and the choice for laparoscopy/Pfannenstiel as negative test result, gives the sensitivity and specificity for the residents, general gynaecologists and the gynaecological oncologists.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Residents</td>
</tr>
<tr>
<td></td>
<td>Malignancy</td>
</tr>
<tr>
<td>Midline incision</td>
<td>23</td>
</tr>
<tr>
<td>Pfannenstiel</td>
<td>7</td>
</tr>
<tr>
<td>Laparoscopy</td>
<td>0</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>23/30 = 77%</td>
</tr>
<tr>
<td>Specificity</td>
<td>111/140 = 79%</td>
</tr>
</tbody>
</table>

Discussion

In the above study, we compared the clinical judgment of gynaecologists to the predictive performance of existing prediction models (logistic regression models and neural networks) in the pre-operative assessment of the adnexal mass. The predictive performance of the clinical judgment of the participating gynaecologists was comparable to the performance of existing prediction models.
Chapter 3

There are several explanations for the disappointing performance of existing logistic regression models and neural networks. First, external validation, i.e. testing of the model in another population as in which the model has been developed, usually demonstrates that the performance of the models is decreased as compared to the performance that is reported initially. A second explanation for the decrease in performance as compared to the performance that was initially reported might be that the performance of ultrasonography at model development has been better than the performance of the ultrasonography at validation. However, at Máxima Medical Centre two specialized gynaecologists performed ultrasonography. Thus, although their performance might be not as perfect as the performance of the ultrasonography in the study of Timmerman et al., we feel that their performance is very representative for the performance of ultrasonography in a general gynaecologic clinic.

We found that intuitive estimation of probabilities by experienced clinicians was as good as the prediction by mathematical models. This is not entirely surprising, after all this is what expertise is about. However, somewhat disturbingly, clinical residents were as good as experienced clinicians. One might hypothesize that the cases were so straightforward that no additional benefit could be accrued from either expertise or modelling. However, the case series were derived from a random selection of 170 consecutive patients in our hospital, and all cases were used 3 times by each group of observers. Moreover, the information that was provided to the clinicians was limited to that information that was designed to be important to prediction models. Thus, additional data, for example results of the physical examination, could not improve the performance of the mathematical models, since this information has not been included in those models.

We found no difference between the accuracy of the offhand estimates made by clinicians on one hand, and the accuracy of the prediction models on the other hand. As a consequence, we think that the decision for referral of a patient with an adnexal mass can be made by gynaecologists in a general hospital, without consultation of a gynaecological oncologist. When the decision to refer patients to a gynaecological oncologist is made in case surgery is planned via midline incision, 70 to 75% of the patients with a malignancy would be referred correctly, whereas almost 20% of the patients without a malignancy would be referred incorrectly. In view of the prevalence of malignancy in women with an adnexal mass of 18%, this would implicate that about 45% of the patients referred to a specialized centre would be referred correctly, whereas 55% of these patients would be referred and operated on by midline incision, despite the absence of malignancy.

In case the clinician decided for midline incision the median estimated probabilities of malignancy were 50%, 30% and 20% for residents, general gynaecologists and
gynaecological oncologists, respectively. Apparently, gynaecological oncologist decided at a lower probability of malignancy for a midline incision as compared to general gynaecologists and residents, although the difference with the gynaecologists is only a small one.

This study demonstrates that at this moment there is no need to introduce complicated predictive scoring systems like neural networks or logistic regression models for the pre-operative assessment of the adnexal masses. Using data from patients’ history, ultrasonography and CA125 measurement the clinician is capable to make a decision with a similar accuracy as compared to the accuracy obtained from predictive models.

References

Chapter 4

Evaluation of adnexal masses with threedimensional ultrasonography: a prospective study

Peggy M.A.J. Geomini
Kirsten B. Kluivers
Evelien Moret
Gérard L. Bremer
Roy F.P.M. Kruitwagen
Ben Willem J. Mol

Abstract

Objective: The objective of this study was to estimate whether 3D ultrasonography and 3D power Doppler investigation can potentially contribute to the differentiation between benign and malignant ovarian masses.

Methods: Women scheduled for surgical treatment of an adnexal mass were included in a multicenter prospective study. All women underwent 2D and 3D ultrasonographic examination in the week prior to surgery. All parameters were compared in women with benign tumours, borderline tumours, and malignant tumours using receiver operating characteristic analysis and likelihood ratios.

Results: We included 181 women, of whom 144 had a benign mass, 26 had a malignancy and 11 had a borderline tumour. At 3D ultrasonography the most striking difference was found in the presence of central vessels in an adnexal mass. Central vessels assessed by 3D ultrasonography were present in 15% (21 / 144) of the benign masses, 69% (18 / 26) of the malignant masses and 27% (3 / 11) of the masses with borderline malignancy. The likelihood ratios for presence of central vessels for a mass being malignant and/or borderline was 4.9 (95% CI 2.1 - 12). Mean gray index and flow index were also significantly different between the groups, but other features were not.

Conclusion: The central localization of vessels in an adnexal mass, as observed by 3D ultrasonography, the mean gray index and the flow index are potentially important parameters to distinguish benign from malignant adnexal masses.
Introduction

The pre-operative assessment of the potential malignancy of an adnexal mass is still a diagnostic challenge. Accurate pre-operative assessment facilitates an optimal choice of surgery. Optimal surgical staging and cytoreductive surgery is of utmost importance in case of malignancy. In case of benign pathology, a laparoscopically performed cystectomy is usually sufficient.

At present, several parameters are available to distinguish benign and malignant masses. The gray-scale twodimensional (2D) ultrasonographic parameters that are used mostly are tumour diameter or volume, septation, presence of papillary projections, echogenicity and the presence of free fluid. Ovarian tumour blood flow can be evaluated by B-mode colour Doppler ultrasound and waveform analysis.\textsuperscript{1} The resistance to flow is lower in malignant tumours in comparison with benign tumours. Frequently used Doppler parameters are resistance index (RI), pulsatility index (PI) and Peak Systolic Velocity (PSV). In order to improve the pre-operative assessment of adnexal masses, most of these parameters have been combined with patient characteristics in diagnostic models. Although initial publication reported an almost perfect performance of these models, external validation showed their diagnostic performance to be less good.\textsuperscript{2,3}

A potentially new diagnostic tool in the assessment of the adnexal mass is threedimensional (3D) ultrasonography. This technique visualizes the adnexal mass in all three planes (coronal, sagittal and frontal), thus facilitating spatial evaluation in stead of 2D ultrasonography, in which only two planes can be visualized. With 3D ultrasonography a volume of ultrasonographic data is acquired and stored, while with 2D ultrasonography only slices can be assessed. The integration of the acquired information enables reconstruction of the organ being scanned. The addition of 3D power Doppler imaging allows visualization and quantitative judgement of vascularity in all three planes.

Early publications on 3D ultrasound examination of adnexal masses have suggested that the specificity of ultrasonographic imaging might improve with the addition of 3D power Doppler.\textsuperscript{4,5} These studies reported that masses with central vascular flow, vascular flow within excrescences or flow within septation are more likely to be malignant, as well as masses including vessels with a chaotic architecture or complex branching pattern.

The objective of this prospective study was to estimate whether data acquired by 3D ultrasonography and 3D power Doppler investigation could potentially contribute to the differentiation between benign and malignant ovarian masses.
Methods

The study was performed in five gynaecologic units in hospitals in the south of The Netherlands. In Máxima Medical Centre (MMC), consecutive patients were included. To enlarge the study population also patients from four surrounding hospitals were included. Since the 3D ultrasonography was performed at MMC, women from surrounding hospitals were only included in case they were prepared to travel to MMC.

In this prospective study all women who were scheduled for surgery because of an adnexal mass between January 2003 and March 2005 were eligible for the study. The study was approved by the institutional review board of MMC. All women received verbal and written information about the study. After informed consent had been obtained, women were scheduled for 2D and 3D gray scale ultrasonography and for 2D and 3D power Doppler ultrasonography in the week prior to surgery. The 2D and 3D ultrasound were performed using a 7.5-MHz transvaginal transducer (Voluson 730, Kretztechnik, Austria). All ultrasound examinations were performed transvaginal. Power Doppler settings were kept constant for every patient: frequency 3 - 9 MHz, pulse repetition frequency 0.6 kHz, gain -4.0 dB and wall motion filter "low1".

All examinations were performed by three of the authors (PG, KK, EM). The acquired 3D ultrasonographic data were stored on a hard disk to enable full evaluation at a later point. Subsequently, all recordings were reviewed systematically by one of the authors (PG), without knowledge of the final histopathologic diagnosis.

To perform volume measurements the virtual organ computer-aided analysis (VOCAL™) was used. Volume measurements were performed by manual delineation of the outside of the adnexal mass using the nine degrees rotation step (see figure 1). In practice the dataset was rotated clockwise through 180 degrees and the outside of the mass was delineated every nine degrees until a calculated volume was generated. The volume was stored and analysed using specially designed software (3D View function). Analogous to 2D pixels, the stored volume ultrasound information obtained using 3D ultrasonography is defined by voxels (smallest measurable volume). Gray scale voxels contain all 3D gray scale information grades from white to black. Voxels with a power Doppler signal are defined as colour voxels. The range of colour voxels can vary between 0 and 100. As described by Parleitner et al., the “histogram” facility of 3D View function calculates three vascularization indices: vascularization index (VI), flow index (FI) and vascularization flow index (VFI). The vascularization index is the ratio of the number of colour voxels to the number of all voxels in the defined contour and is expressed as a percentage value. The flow index (scale 0 - 100) is the mean value of the colour voxels and is thought to express the average intensity of flow in the vessels. The vascularization flow index measures the ratio of the mean value of
Figure 1.

Virtual organ computer-aided analysis (VOCAL™). This figure shows a multiplanar display of an adnexal mass. Longitudinal and transverse views are shown in the upper left and right image, the coronal plane is shown in the lower left image. The resultant 3D model can be seen in the lower right image.

Figure 2.

3D power Doppler images. (a) peripheral vessels, (b) central vessels, (c) central and peripheral vessels combined in one tumour.
the colour voxels and all the voxels in the defined contour (VI*FI/100) and is a feature of both vascularization and flow. It identifies the extremes between low vascularization and low blood flow on one side, and high vascularization and high blood flow, on the other. With the "histogram" facility of 3D View function also the mean value of the gray scale voxels can be calculated, and is expressed as the mean gray index (scale 0 - 100). The mean gray index is a quantification of the echogenicity of an adnexal mass. When the mean gray index is close to zero the adnexal mass is mainly sonoluent, whereas a mean gray index close to 100 refers to a high echogenic mass.

For each patient, the following 3D gray scale ultrasonographic and 3D power Doppler data were recorded: tumour volume, mean gray index, localization of vessels, vascularization index, flow index and vascularization flow index. The localization of vessels was ordered into four categories (see figure 2). The first category ("no vessels") contained masses without any sign of vascularization, neither peripheral (in the wall of the mass) nor central (within the mass). The second category ("peripheral vessels") contained cystic masses with vascularization of the wall of the mass and no vascularization within the mass. The third group ("central vessels") contained masses with central vascularization, but no vascularization of the wall of the masses. The central vascularization could be within septa, solid areas or papillary formations. The fourth group ("peripheral and central vessels") contained masses which had both central vascularization and peripheral vascularization.

In case it was impossible to include the adnexal mass into one volume box because of its size, the volume of the tumour could not be measured. A representative part of the adnexal mass was assessed to define the localization of the vessels and to calculate the vascularization indices.

For each patient, the following 2D gray scale ultrasonographic data were recorded: presence and thickness of septations, presence of papillary formations and free fluid, echogenicity and the mean diameter of the mass, and the following power Doppler parameters: resistance index (RI), pulsatility index (PI) and peak systolic velocity (PSV).

Beside the ultrasonographic parameters we recorded serum CA125 level, menopausal status and age. The CA125 was measured with chemoluminescent immunometric assay (Immulse 2000 OM-MA, DPC, Diagnostic Product Corporation, Los Angeles, USA).

In case of suspicion on malignancy, surgery was performed by a median laparotomy and combined with frozen section diagnosis. In case frozen section diagnosis showed malignancy a complete debulking was pursued. The risk for malignancy was assessed by means of the results of a clinical examination, 2D gray scale ultrasonographic information (septation, presence of papillary formations, echogenicity and presence of
free fluid), 2D power Doppler parameters (resistance index), age of the patient and level of serum CA 125. In case the mentioned parameters did not lead to any suspicion on malignancy a cystectomy was performed in case the woman was younger than 45 years and an ovariectomy was performed in case the patient had an age higher than 45 years. All adnexal masses were sent to the laboratory of the Department of Pathology and examined by one of the experienced pathologists. A definite paraffin diagnosis was obtained.

The distribution of each of the 2D and 3D parameters were compared between the three categories. Distribution of continuous variables was calculated by constructing normal plots. Depending on normality, we calculated means or medians. In case of a normal distribution, statistical comparison was made with analysis of variance. In case of non-normal distributions, we used the Kruskal-Wallis test.

Subsequently, for each of the continuous variables we performed a Receiver Operating Characteristic (ROC) analysis, in which malignant and borderline adnexal masses were considered to be the disease, and benign masses to be non-disease. We calculated an area under the ROC-curve (AUC). The AUC can have in general a value between 0.5 and 1; an AUC of 0.5 expresses no discriminative capacity at all, whereas an AUC of 1 expresses perfect discriminative capacity. Categorical variables were compared using chi-square analysis, and further assessed by calculating likelihood ratios (LR). A likelihood ratio of a test result was the ratio of the probability of the test result in women with malignant disease (borderline included) and the probability of the test result in women with benign disease. A likelihood ratio above 1 increases the probability of malignancy whereas a LR below 1 decreases the probability of disease. In general, a LR between 1 and 2 is considered as not useful, a LR between 2 and 5 as moderately useful, whereas the test becomes useful at higher likelihood ratios. These valubales however, should be considered together with the prevalence of disease and subsequent therapeutic consequences.

Results

Between January 2003 and March 2005, 181 women were included, 164 in Máxima Medical Centre (Veldhoven / Eindhoven), 12 in Tweesteden Hospital (Tilburg), two in Elkerliek Hospital (Helmond) and three in Elisabeth Hospital (Tilburg). At surgery, there appeared to be 144 (80%) benign adnexal masses, 26 (14%) malignant adnexal masses and 11 (6%) borderline malignant masses. Among the 26 malignant masses there were 16 serous adenocarcinomas, two mucinous adenocarcinomas, two endometroid carcinomas, three clearcell carcinomas, two undifferentiated adenocarcinomas of the ovary and one coeloma carcinoma.
Because of initial suspicion on malignancy, surgery was performed by a median laparotomy and combined with frozen section diagnosis on 29 of the 144 women with benign pathology. Forty-five percent of the women with borderline tumours (5 / 11) were operated laparoscopically and all but one of the women with malignant tumours had a laparotomy.

Baseline characteristics are shown in table 1. The mean age was 46 years (SD 16.0, range 15 - 85 years) for women with benign masses, 61 years (SD 12.8, range 42 - 87 years) for women with malignant masses, and 51 years (SD 15.0, range 38 - 83 years) for women with a borderline malignancy (p ≤ 0.001). Among the 144 women with a benign mass, 54 were postmenopausal (38%), versus five out of 11 (45%) among women with borderline malignancy and 18 of 26 (69%) among women with malignancy (p = 0.005).

The median level of serum CA125 in women with malignant masses was 155 kU/l (range 2.1 - 13000), whereas the median level of serum CA125 was respectively 11 kU/l (range 1.4 - 450) and 29 kU/l (range 5.1 - 146) for benign masses and borderline masses, respectively (p < 0.001). Twenty-five out of 140 benign masses (18%) showed a CA125 ≥ 35 kU/l and 10/140 showed a CA125 ≥ 100 kU/l. Among these ten benign masses with a serum CA125 ≥ 100 kU/l were seven endometriomas, two fibromas and one mucinous cystadenoma. Twelve out of 37 malignant and borderline tumours (32%) showed a serum CA125 < 35 kU/l.

Table 2 and 3 show the distribution of the 2D and 3D parameters with their 95% confidence intervals (CI). Ten adnexal masses (six malignant tumours, one borderline

<table>
<thead>
<tr>
<th>Table 1. Baseline characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>mean age (years) (SD ‡)</td>
</tr>
<tr>
<td>median serum CA125 (kU/l) (range)</td>
</tr>
<tr>
<td>menopausal state</td>
</tr>
<tr>
<td>premenopausal (n)</td>
</tr>
<tr>
<td>postmenopausal (n)</td>
</tr>
</tbody>
</table>

§ standard deviation; † absolute number; ‡ confidence interval; † area under Receiver Operator Characteristics curve; § analysis of variance; * Kruskal-Wallis test; * chi-square analysis
tumour and three benign tumours) could not be included in one volume box because of their sizes. Consequently their volumes could not be calculated. In eight out of these ten cases it was possible to assess a representative part of the tumour to define the localization of the vessels and to calculate the vascularization indices.

The volume of the mass did not differ between benign, borderline and malignant masses (p-value 0.17, area under the ROC-curve 0.53). The malignant masses had a higher mean gray index as compared to the benign and borderline masses (p-value 0.006, area under the ROC-curve 0.69). The likelihood ratio increases with an increasing mean gray index, thus indicating that a more dense mass is more at risk for malignancy (table 3).

The malignant masses also had a higher flow index compared to the benign and borderline masses (p-value 0.04, area under the ROC-curve 0.65). The likelihood ratio for malignancy increased with an increasing flow index, implicating that a mass with a high flow index (> 30) is more at risk for malignancy (table 3). The mean vascularization index and vascularization flow index did not differ significantly between the three categories, indicating that these tests do not have value.

Table 3 also shows the localization of vessels. Central vessels were present in 69% (18 / 26) of the malignant masses, 27% (3 / 11) of the masses of borderline malignancy and 15% (21 / 144) of the benign masses. The likelihood ratios for absence of vessels or for presence of peripheral vessels only were 0.42 (95% CI 0.13 - 1.3) and 0.54 (95% CI 0.35 - 0.86) respectively. When central vessels were present the risk of a malignancy increased, as indicated by likelihood ratios of 3.3 (95% CI 1.3 - 6.8) and 4.9 (95% CI 2.1 - 12) for central vessels only and central and peripheral vessels in the same mass. Table 4 shows which benign masses had central vascularization.

Among the 181 included tumours 75 masses were sonolucent without septation or sonolucent with septa smaller than 3 mm. If we exclude these so called simple cysts, the presence of central vessels and a flow index > 30 showed likelihood ratios of 4.4 (95% CI 1.8 - 9.9) and 1.8 (95% CI 1.2 - 2.7), respectively.

The presence of random echogenicity (2D) and central vessels (3D) were the most sensitive predictors, since they were present in respectively 22 and 21 of the 37 women with borderline malignancy or invasive malignancy. Sensitivity and specificity of random echogenicity were 59% (95% CI 42 - 75) and 80% (95% CI 73 - 86) and sensitivity and specificity of presence of central vessels were 57% (95% CI 40 - 73) and 85% (95% CI 80 - 91). Presence of septation and their thickness, presence of papillary formations, the mean diameter and the volume of the mass, PI, PSV, VI and VFI were not predictive. Premenopausal status, absence of free fluid, an anechoic performance of the mass and the presence of only peripheral (and not central) vessels reduced the risk for malignancy.
<table>
<thead>
<tr>
<th></th>
<th>Benign n = 144</th>
<th>Borderline n = 11</th>
<th>Malignant n = 26</th>
<th>Likelihood ratio</th>
<th>95% CI</th>
<th>AUC $^5$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean diameter mass (cm) (SD) $^6$</td>
<td>6.9 (3.3)</td>
<td>9.1 (4.1)</td>
<td>6.6 (3.8)</td>
<td>-</td>
<td>-</td>
<td>0.51</td>
<td>0.52$^6$</td>
</tr>
<tr>
<td>septation</td>
<td>6</td>
<td>1</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no septation (n$^1$)</td>
<td>88 (62%)</td>
<td>4 (36%)</td>
<td>14 (54%)</td>
<td>0.79</td>
<td>0.55 - 1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>presence of septa (n)</td>
<td>55 (38%)</td>
<td>7 (64%)</td>
<td>12 (46%)</td>
<td>1.3</td>
<td>0.92 - 1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>median thickness septa (mm) (range)</td>
<td>0 (0-9)</td>
<td>2.3 (0-4.5)</td>
<td>0 (0-3.7)</td>
<td>-</td>
<td>-</td>
<td>0.57</td>
<td>0.15$^9$</td>
</tr>
<tr>
<td>papillary formations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>0.89</td>
<td>0.75 - 1.04</td>
</tr>
<tr>
<td>no papillary formations (n)</td>
<td>132 (93%)</td>
<td>7 (78%)</td>
<td>21 (84%)</td>
<td>0.89</td>
<td>0.75 - 1.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>papillary formations (n)</td>
<td>10 (7%)</td>
<td>2 (22%)</td>
<td>4 (16%)</td>
<td>2.5</td>
<td>0.98 - 6.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>free fluid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.74</td>
<td>0.60 - 0.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>absence of free fluid (n)</td>
<td>135 (94%)</td>
<td>10 (91%)</td>
<td>16 (62%)</td>
<td></td>
<td></td>
<td>2.3 - 12</td>
<td></td>
</tr>
<tr>
<td>presence of free fluid (n)</td>
<td>8 (6%)</td>
<td>1 (9%)</td>
<td>10 (38%)</td>
<td>5.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>morphologic characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.001*</td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anechoic (n)</td>
<td>79 (55%)</td>
<td>7 (64%)</td>
<td>4 (15%)</td>
<td>0.54</td>
<td>0.32 - 0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uniform echogenic (n)</td>
<td>36 (25%)</td>
<td>0</td>
<td>4 (15%)</td>
<td>0.43</td>
<td>0.16 - 1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>random echogenic (n)</td>
<td>29 (20%)</td>
<td>4 (36%)</td>
<td>18 (70%)</td>
<td>3.0</td>
<td>1.9 - 4.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vessels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>14</td>
<td>2</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>no vessels or not suspect (n) $^††$</td>
<td>124 (95%)</td>
<td>8 (89%)</td>
<td>15 (71%)</td>
<td>0.80</td>
<td>0.66 - 0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>suspect (= RI &lt; 0.5) (n) $^††$</td>
<td>6 (53%)</td>
<td>1 (11%)</td>
<td>6 (29%)</td>
<td>5.1</td>
<td>1.8 - 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean resistance index (SD)</td>
<td>0.67 (0.18)</td>
<td>0.63 (0.13)</td>
<td>0.56 (0.17)</td>
<td>-</td>
<td>-</td>
<td>0.65</td>
<td>0.04$^4$</td>
</tr>
<tr>
<td>missing values**</td>
<td>106</td>
<td>5</td>
<td>8</td>
<td>-</td>
<td>-</td>
<td>0.63</td>
<td>0.13$^5$</td>
</tr>
<tr>
<td>pulsatility index (SD)</td>
<td>1.5 (0.75)</td>
<td>1.3 (0.26)</td>
<td>1.1 (0.64)</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values**</td>
<td>119</td>
<td>5</td>
<td>11</td>
<td>-</td>
<td>-</td>
<td>0.51</td>
<td>0.93$^3$</td>
</tr>
<tr>
<td>peak systolic velocity (SD)</td>
<td>17.0 (10.5)</td>
<td>16.1 (8.4)</td>
<td>17.1 (10.0)</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values**</td>
<td>113</td>
<td>5</td>
<td>7</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^6$ standard deviation, $^1$ absolute number, $^2$ confidence interval, $^5$ area under Receiver Operator Characteristics curve, $^††$ not suspect vessels had resistance index 0.5, suspect vessels had resistance index < 0.5, $^*analysis of variance, ^*Kruskal-Wallis test, ^*chi-square analysis, **mostly due to absence of vessels.
<table>
<thead>
<tr>
<th>Table 3. 3D parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>median volume mass (cc)</td>
</tr>
<tr>
<td>(range)</td>
</tr>
<tr>
<td>missing values</td>
</tr>
<tr>
<td>mean mean gray index¹ (SD⁶)</td>
</tr>
<tr>
<td>missing values</td>
</tr>
<tr>
<td>mean gray index &lt; 15 (n²)</td>
</tr>
<tr>
<td>mean gray index 16-20 (n)</td>
</tr>
<tr>
<td>mean gray index 21-25 (n)</td>
</tr>
<tr>
<td>mean gray index ≥ 26 (n)</td>
</tr>
<tr>
<td>vascularization indices (SD)</td>
</tr>
<tr>
<td>missing values</td>
</tr>
<tr>
<td>median vascularization index² (%)</td>
</tr>
<tr>
<td>(range)</td>
</tr>
<tr>
<td>mean flow index³ (SD)</td>
</tr>
<tr>
<td>flow index &lt; 20 (n)</td>
</tr>
<tr>
<td>flow index 21-25 (n)</td>
</tr>
<tr>
<td>flow index 26-30 (n)</td>
</tr>
<tr>
<td>flow index ≥ 31 (n)</td>
</tr>
<tr>
<td>median vascularization flow index⁴ (%)</td>
</tr>
<tr>
<td>(range)</td>
</tr>
<tr>
<td>localization vessels</td>
</tr>
<tr>
<td>missing values</td>
</tr>
<tr>
<td>no vessels (n)</td>
</tr>
<tr>
<td>peripheral vessels (n)</td>
</tr>
<tr>
<td>central vessels (n)</td>
</tr>
<tr>
<td>peripheral and central vessels (n)</td>
</tr>
</tbody>
</table>

¹ standard deviation, ² absolute number, ³ confidence interval, ⁴ area under Receiver Operator Characteristics curve, ⁵ analysis of variance
⁶ Kruskal-Wallis test, ⁷ chi-square analysis, ⁸ mean gray index is a quantification of the echogenicity, ⁹ vascularization index expresses the percentage of present vessels in a mass, ¹⁰ flow index express the average intensity of flow in the vessels, ¹¹ vascularization flow index is a feature of both vascularization and flow
Table 4. Histologic diagnosis of benign masses (n = 144)

<table>
<thead>
<tr>
<th>histology</th>
<th>n</th>
<th>Number of tumours with central vessel localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>serous cystadenoma</td>
<td>33</td>
<td>4</td>
</tr>
<tr>
<td>dermoid cyst</td>
<td>32</td>
<td>3</td>
</tr>
<tr>
<td>endometrioma</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>mucinous cystadenoma</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>fibroma</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>luteal cyst</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>hydrosalpinx</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>twisted adnex</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>abscess</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>uterine horn</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>myoma</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>other non ovarian tissue</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>not specified</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

Discussion

A crucial step for the continuous growth of tumours is the induction of vasculature. Any increase in tumour mass beyond 2 - 3 mm³ requires the proliferation and morphogenesis of vascular endothelial cells. Malignant neoplasms are highly vascular. Vascular flow within an adnexal mass is often described in terms like resistance index (RI), pulsatility index (PI) and Peak Systolic Velocity (PSV). Another possibility to assess potential malignancy is to describe the localization of the vessels or the branching pattern. In malignant masses central vascular flow with a chaotic branching pattern is expected. Guerriero et al. assessed the use of 2D gray scale imaging combined with a 2D power Doppler evaluation for the presence or absence of vascularity within solid portions or excretions. They concluded that the combination of gray scale imaging with a descriptive power Doppler imaging of the localization of the vessels performs better than the resistance index.

By means of 3D power Doppler ultrasonography it is possible to visualize vessels in all three planes (coronal, sagittal and frontal), thus facilitating spatial evaluation. Assessing the localization of vessels by 3D power Doppler ultrasonography is a simple procedure.

Since the introduction of 3D transvaginal ultrasonography, several studies have been published assessing the potential contribution of 3D ultrasonography in the differentiation between benign and malignant ovarian masses (table 5). The authors use morphologic criteria or the presence of central or chaotic vessels to define malignancy. Most of the studies show improvement of the specificity of ultrasonographic imaging by addition of 3D ultrasonography.
In this prospective study, 181 adnexal masses were evaluated and we assessed the localization of vessels by means of 3D power Doppler ultrasonography. Central vessels were seen more often in malignancies (69%) compared to benign masses (15%) and masses with borderline malignancy (27%). The presence of central and peripheral vessels in an adnexal mass had a likelihood ratio of 4.9 (95% CI 2.1 - 11). Our results are in accordance with the results of Alcazar.12 The criteria for suspicion on malignancy used in the study of Alcazar et al. are comparable with our criteria. We suspected malignancy in case blood flow was central anywhere in an adnexal mass, also in case blood flow was detected in septations. However Alcazar assessed localization of blood flow by 2D and 3D power Doppler examination and 3D power-Doppler imaging did not have a better performance than 2D power Doppler imaging. We did not assess blood flow localization by 2D power Doppler examination, so we cannot confirm these results.

Malignant neoplasms are highly vascularized. The neovascularization is of poor quality with low compliance and low resistance. Because of the low resistance of the vasculature the flow in malignant masses will be increased. In this study, we found an increased flow index in the malignant masses (32.6%) as compared to the borderline masses (30.8%) and benign masses (28.5%). The flow index represents the intensity of

### Table 5. Previous studies on the accuracy of 2D and 3D ultrasound in the assessment of adnexal masses

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Malignant</th>
<th>Definition of malignancy</th>
<th>2D sensitivity</th>
<th>2D specificity</th>
<th>3D sensitivity</th>
<th>3D specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonilla12</td>
<td>1995</td>
<td>57</td>
<td>morphologic criteria</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Hata13</td>
<td>1999</td>
<td>20</td>
<td>morphologic criteria</td>
<td>100</td>
<td>38</td>
<td>100</td>
<td>92</td>
</tr>
<tr>
<td>Kurjak5</td>
<td>2000</td>
<td>90</td>
<td>morphologic criteria, vessel architecture, branching pattern</td>
<td>67</td>
<td>94</td>
<td>78</td>
<td>98</td>
</tr>
<tr>
<td>Cohen4</td>
<td>2001</td>
<td>71</td>
<td>central vascular grow</td>
<td>100</td>
<td>54</td>
<td>100</td>
<td>75</td>
</tr>
<tr>
<td>Alcazar14</td>
<td>2003</td>
<td>44</td>
<td>morphologic criteria</td>
<td>90</td>
<td>61</td>
<td>100</td>
<td>78</td>
</tr>
<tr>
<td>Alcazar10</td>
<td>2005</td>
<td>96</td>
<td>central vascular grow</td>
<td>98</td>
<td>88</td>
<td>98</td>
<td>79</td>
</tr>
<tr>
<td>Geomini</td>
<td>2006</td>
<td>181</td>
<td>morphologic criteria, power Doppler, CA125 (2D), central vascular grow (3D)</td>
<td>91</td>
<td>63</td>
<td>57</td>
<td>85</td>
</tr>
</tbody>
</table>
flow in an adnexal mass. However, the LR of a flow index of more than 30% is close to one, indicating that it has no value in the differentiation between benign and malignant masses.

The mean gray index was significantly higher in the malignant adnexal masses compared to benign and borderline masses. Adnexal masses which contain papillary formations and masses which are mixed or high echogenic are suspicious for malignancy. These features result in a high mean gray index. However also benign teratomas or hemorrhagic masses often show mixed echogenicity. This probably explains the low likelihood ratio of the mean gray index.

In this study, which to our knowledge is the largest study on this subject so far, we estimated whether data acquired by 3D ultrasonography and 3D power Doppler investigation could potentially contribute to the differentiation between benign and malignant ovarian masses. The central localization of vessels in an adnexal mass, the mean gray index and the flow index as observed by 3D ultrasonography, are potentially important parameters to distinguish benign from malignant adnexal masses. As stated before assessing localization of vessels by 3D imaging is a simple procedure. Another advantage of 3D ultrasonography compared to 2D ultrasonography is the objective aspect of the mean gray index and flow index. The mean gray index and flow index are objective quantitative parameters based on the number of gray scale and colour voxels whereas in 2D ultrasonography the echogenicity of an adnexal mass and the amount of flow are always subjective descriptions of the examiner.

Further research steps are analyses in which the additional value of 3D parameters is estimated. A larger study population with a larger part of malignant masses would be helpful.
References


Chapter 5

Is threedimensional ultrasonography of additional value in the assessment of adnexal masses?

Peggy M.A.J. Geomini
Sjors F.P.J. Coppus
Kirsten B. Kluivers
Gérard L. Bremer
Roy F.P.M. Kruitwagen
Ben Willem J. Mol

Chapter 5

Abstract

Objective: Patient characteristics, CA125 level and twodimensional (2D) ultrasonography can be used to predict the probability of malignancy of an ovarian mass. Threedimensional (3D) ultrasonography might also contribute to the prediction of malignancy. We evaluated whether addition of 3D features to a diagnostic model could improve the discriminative capacity of the model.

Methods: This multicenter prospective study was approved by the institutional review board. Women with an adnexal mass scheduled for surgery underwent 2D and 3D ultrasonographic examination in the week prior to surgery. Stepwise logistic regression was used to construct two models for the prediction of malignancy: a model based on patient characteristics, level of CA125 and 2D ultrasonography and a second model based on patient characteristics, level of CA125, 2D and 3D ultrasonography. Receiver operating characteristic (ROC) analysis was used to compare the capacity of the two models to discriminate between benign and malignant adnexal masses.

Results: We included 181 women with an adnexal mass, of which 144 were benign and 37 showed malignancy on histopathology. The 3D model discriminated better between benign and malignant adnexal masses than the 2D model (areas under the ROC-curve of 0.92 and 0.82, respectively, p = 0.02). The calibration of both models was good.

Conclusion: In the assessment of the ovarian mass, the use of 3D ultrasonography significantly improves the prediction of malignancy as compared to patient characteristics and 2D ultrasonography.
Value of threedimensional ultrasonography in the assessment of adnexal masses

Introduction

The pre-operative assessment of an adnexal mass is still a diagnostic challenge. Accurate pre-operative assessment for presence of malignancy facilitates an optimal choice of surgery. Optimal surgical staging and cytoreductive surgery is of utmost importance in case of malignancy. The decision to start adjuvant chemotherapy or not depends on the stage of disease.\(^1\) Optimal cytoreductive surgery has a major impact on survival.\(^2\) In case of benign pathology, a laparoscopically performed cystectomy is usually sufficient.

At present, several parameters are available to distinguish benign and malignant masses. Age, menopausal state, CA125 level, twodimensional (2D) gray scale and power Doppler parameters are all known to contribute to the pre-operative diagnosis. These parameters have been combined in diagnostic models. Although initial publication reported an almost perfect performance of these models, external validation showed their diagnostic performance to be less optimal. In case sensitivity was maximalized until almost 100\%, the highest specificities varied between 0.45 and 0.60.\(^5\)\(^6\)

A potentially new diagnostic tool in the assessment of the adnexal mass is threedimensional (3D) ultrasonography, which offers the possibility of volume measurements and quantification of echogenicity of the adnexal mass. Threedimensional power Doppler examination offers the possibility to visualize and quantify blood flow in adnexal masses.\(^7\)

A recent study has shown that the central localization of vessels in an adnexal mass, the mean gray index and the flow index as observed by 3D ultrasonography are potentially important parameters to distinguish benign from malignant adnexal masses.\(^8\) In a diagnostic cohort study, these features were associated with the probability of malignancy. However, 3D ultrasonography is not routinely available in daily clinical practice and before introducing this technique in daily practice it is important to know whether the use of 3D ultrasonography has added value on the result of medical history, CA125 level and 2D ultrasonography.

The aim of the present study was therefore to evaluate the additional value of 3D ultrasonography in the prediction of malignancy in ovarian masses. To do so we estimated probabilities of malignancy with multivariable logistic regression analysis using patient characteristics, level of CA125 and 2D ultrasonography and evaluated whether these predictions improved after adding results from 3D ultrasonography.

69
Methods

Patients

The study was performed in five gynaecological units in hospitals in the south of The Netherlands between 2003 and 2005 and approved by the institutional review board. The study design has been described previously. In short, after informed consent had been obtained, all women who were scheduled for surgery because of an adnexal mass were examined by 2D and 3D gray scale and power Doppler ultrasonography in the week prior to surgery. Surgery was performed by median laparotomy, laparoscopic ovariectomy and laparoscopic cystectomy, depending on the choice of the gynaecologist and the preference of the patient prior to the procedure. All adnexal masses were sent to the pathologist for a definite paraffin histological diagnosis.

Ultrasonography

The 2D and 3D ultrasonography was performed transvaginally using a 7.5-MHz transvaginal transducer (Voluson 730, Kretztechnik, Austria). Power Doppler settings were kept constant for every patient: frequency 3 - 9 MHz, pulse repetition frequency 0.6 kHz, gain -4.0 dB and wall motion filter “low1”. The acquired 3D ultrasonographic data were stored on a hard disk to enable full evaluation later. Subsequently, all recordings were reviewed systematically by one of the authors, without knowledge of the final histopathologic diagnosis.

Statistics

We used stepwise logistic regression analysis to construct a model for the prediction of malignancy, using patient characteristics, CA125 level, 2D and 3D ultrasonography data. Patient characteristics considered were age and menopausal state. Twodimensional ultrasonography parameters considered were the presence of septations, presence of papillary formations, presence of free fluid, mean diameter of the cyst, suspected vessels as defined by a resistance index (RI) < 0.5, and the echogenicity of the tumour. Three-dimensional ultrasonography parameters evaluated were cyst volume, mean gray index, vascularization index, flow index, vascularization flow index and localization of vessels.

First, we checked the assumption of linearity between the continuous variables age, CA125, tumour volume on 2D and 3D ultrasonography, vascularization index, flow index and vascularization flow index on the one hand and malignancy on the other hand visualized with smoothed piecewise polynomials (splines). Based on these spline functions, non-linear associations of these continuous variables were redefined to better approach linearity. Subclasses within categorical variables were dichotomized. For
dichotomous and continuous variables, univariate beta coefficients ($\beta$), odds ratios (OR's) and 95% confidence intervals (95% CI) as well as $p$-values were calculated. In the univariate analysis, the $p$-value for statistical significance was set at $\geq 0.05$.

Subsequently, multivariable logistic regression analysis with a stepwise backwards selection procedure was used to construct the prediction model. Since the erroneous exclusion of a variable is more deleterious for a model than including too many factors, we used a significance level of 50% for entry in the multivariate model, and a significance level of 20% to stay in the model.\textsuperscript{10,11}

To adjust for overfitting, we performed internal validation with bootstrapping. Bootstrapping is a technique to evaluate the robustness of the model by simulating sampling variation in repeated samples from the study population. Cases are selected randomly from the original sample to create a new sample of the same size. As cases are drawn with replacement, some cases may be randomly selected multiple times and others not at all.\textsuperscript{12} We generated 500 bootstrap data sets in which the same multivariate logistic regression model was estimated. By analyzing the difference between the model based on the original dataset and the bootstrap estimates a shrinkage factor was calculated, and used to correct the original model coefficients and associated probabilities for overfit.

We developed two models. The first model was based on age and CA125 level as well as results from the 2D ultrasonography. In the second model, the results of the 3D ultrasonography were added. This second model was developed to see whether addition of 3D parameters improved the diagnostic performance of the model. To do so, we plotted the probability of malignancy as predicted by the model without the 3D parameters ("2D model") against the probability of malignancy predicted by the model with the 3D parameters added to the patient characteristics, level of CA125 and 2D ultrasonography ("3D model"). The capacity of both models to distinguish between benign and malignant adnexal masses was also assessed and compared by calculating areas under the ROC curve (AUC).\textsuperscript{13} Calibration of the model, which is the agreement between predicted and observed probabilities of malignancy, was evaluated with the goodness-of-fit Hosmer and Lemeshow test statistic, and by visual inspection of the calibration plots. Data were analyzed using SAS 9.1 (SAS Institute Inc., Cary, NC, USA), SPSS 12.0.1 (SPSS Inc., Chicago, IL, USA) and S-PLUS 2000 (MathSoft Inc.).
Chapter 5

Results

Between January 2003 and March 2005, 181 women were included of which 144 (80%) appeared to have a benign adnexal mass and 37 (20%) a malignant adnexal of which 11 were classified as borderline malignancy. Among the 26 invasive malignant masses there were 16 serous adenocarcinomas, two mucinous adenocarcinomas, two endometroid carcinomas, three clearcell carcinomas, two undifferentiated adenocarcinomas of the ovary and one coeloma carcinoma.

Spline analysis showed that the risk of malignancy increased linear with age and CA125 level. The 3D flow index value (FI) (scale 0 - 100) was associated with a higher probability of malignancy until a level of 30 (figure 1a). The 3D vascularization flow index (VFI) (scale 0 - 100%) showed a steep increase in probability of disease until a level of 2%, to generate no additional risk when the VFI was above this cut-off value (figure 1b). Based on these spline-analyses, we redefined flow index (linear until 30) and vascularization flow index (linear until 2%) for further analysis.

Table 1 shows the results of the uni- and multivariable analysis. In the univariable analysis age, serum CA125, menopausal status, ascites, random echogenicity, presence

Figure 1a.

![Spline function, visualizing the association between the continuous variable 3D flow index (FI3D) and the probability of malignancy (log scale).]

Figure 1b.

![Spline function, visualizing the association between the continuous variable 3D vascularization flow index (VFI3D) and the probability of malignancy (log scale).]
Figure 2. *Receiver Operating Characteristic curves of the 2D and the 3D model*

The AUC of the models was 0.82 and 0.92 for the 2D and 3D model, respectively (p = 0.02).

Figure 3. *Predicted probabilities of 2D and 3D model*

Predicted probability on malignancy by our 2D model compared to predicted probability on malignancy by our 3D model. ● malignant adnexal masses; ○ benign adnexal masses.
### Table 1. Univariable and multivariable analyses of baseline characteristics, 2D and 3D parameters

<table>
<thead>
<tr>
<th></th>
<th>Benign n = 144</th>
<th>Malignant n = 37</th>
<th>Univariable analysis</th>
<th>Multivariable analysis (2D model)</th>
<th>Multivariable analysis (3D model)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean age (years) (SD(^2))</td>
<td>46 (16)</td>
<td>58 (14)</td>
<td>1.05 1.03-1.08 &lt;0.001</td>
<td>1.05 1.00-1.07 0.07</td>
<td>1.05 1.01-1.10 0.02</td>
</tr>
<tr>
<td>median serum CA125 (kU/L) (range)</td>
<td>11 (1.4-450)</td>
<td>74 (2.1-13000)</td>
<td>1.01 1.00-1.01 0.001</td>
<td>1.01 1.00-1.01 0.15</td>
<td>1.01 1.00-1.01 0.04</td>
</tr>
<tr>
<td>menopausal state</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>premenopausal</td>
<td>90 (63%)</td>
<td>14 (38%)</td>
<td>2.74 1.30-5.77 0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>postmenopausal</td>
<td>54 (37%)</td>
<td>23 (63%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2D parameters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean diameter cyst (cm) (SD(^2))</td>
<td>6.9 (3.3)</td>
<td>7.3 (4.2)</td>
<td>1.04 0.93-1.15 0.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>septation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no septation (n=)</td>
<td>88 (62%)</td>
<td>18 (49%)</td>
<td>1.69 0.83-3.30 0.16</td>
<td>2.03 0.69-6.0 0.19</td>
<td></td>
</tr>
<tr>
<td>presence of septa (n)</td>
<td>55 (38%)</td>
<td>19 (51%)</td>
<td>1.08 0.88-1.34 0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>median thickness septa (mm) (range)</td>
<td>0 (0-9)</td>
<td>0.8 (0-4.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>papillary formations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no papillary formations (n)</td>
<td>132 (93%)</td>
<td>28 (82%)</td>
<td>2.55 0.87-7.48 0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>papillary formations (n)</td>
<td>10 (7%)</td>
<td>6 (18%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>free fluid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>absence of free fluid (n)</td>
<td>135 (94%)</td>
<td>26 (70%)</td>
<td>7.14 2.62-19.5 &lt;0.001</td>
<td>3.80 0.86-16.9 0.08</td>
<td>10.3 1.48-72.1 0.02</td>
</tr>
<tr>
<td>presence of free fluid (n)</td>
<td>8 (6%)</td>
<td>11 (30%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>morphologic characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anechoic (n)</td>
<td>79 (55%)</td>
<td>11 (30%)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uniform echogenic (n)</td>
<td>36 (25%)</td>
<td>4 (11%)</td>
<td>0.80 0.24-2.68 0.72</td>
<td>0.14 0.02-1.03 0.05</td>
<td></td>
</tr>
<tr>
<td>random echogenic (n)</td>
<td>29 (20%)</td>
<td>22 (59%)</td>
<td>5.45 2.35-12.6 &lt;0.001</td>
<td>3.35 1.15-9.80 0.03</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. (continuation)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Benign</th>
<th>Malignant</th>
<th>Univariable Analysis</th>
<th>Multivariable Analysis (2D model)</th>
<th>Multivariable analysis (3D model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>n</em> = 144</td>
<td><em>n</em> = 37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>vessels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no vessels or not suspect (a) ††</td>
<td>132</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>suspect (&gt;RI&lt;0.5) (a) ††</td>
<td>9</td>
<td>12</td>
<td></td>
<td>7.65</td>
<td>2.90-20.2</td>
</tr>
<tr>
<td>mean resistance index (SD)</td>
<td>0.67</td>
<td>0.57</td>
<td>0.036</td>
<td>0.001-0.92</td>
<td>0.04</td>
</tr>
<tr>
<td>missing values**</td>
<td>106</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pulsatility index (SD)</td>
<td>1.5</td>
<td>1.1</td>
<td>0.46</td>
<td>0.17-1.3</td>
<td>0.13</td>
</tr>
<tr>
<td>missing values**</td>
<td>119</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>peak systolic velocity (SD)</td>
<td>17.0</td>
<td>16.8</td>
<td>1.00</td>
<td>0.95-1.05</td>
<td>0.93</td>
</tr>
<tr>
<td>missing values**</td>
<td>113</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**3D parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Benign</th>
<th>Malignant</th>
<th>Univariable Analysis</th>
<th>Multivariable Analysis (2D model)</th>
<th>Multivariable analysis (3D model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>median volume cyst (cc) (range)</td>
<td>124</td>
<td>144</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean mean gray index (SD)</td>
<td>18</td>
<td>22</td>
<td>1.05</td>
<td>1.01-1.09</td>
<td>0.02</td>
</tr>
<tr>
<td>missing values</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vascualization indices (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing values</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>median vascualization index (%) (range)</td>
<td>0.94</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• vascularization index per % &lt; 6</td>
<td>0.92</td>
<td>0.78-1.09</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• vascularization index per % &gt; 6</td>
<td>0.96</td>
<td>0.87-1.06</td>
<td>0.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean flow index (SD)</td>
<td>28.5</td>
<td>32.1</td>
<td>0.83</td>
<td>0.74-0.94</td>
<td>0.003</td>
</tr>
<tr>
<td>• flow index per unit &lt; 30</td>
<td>1.07</td>
<td>0.99-1.15</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• flow index per unit &gt;30</td>
<td>0.29</td>
<td>0.45</td>
<td></td>
<td>0.70</td>
<td>0.56-0.88</td>
</tr>
<tr>
<td>median vascualization flow index (%) (range)</td>
<td>0.29</td>
<td>0.45</td>
<td></td>
<td>0.70</td>
<td>0.56-0.88</td>
</tr>
<tr>
<td>• vfi per % &lt; 2</td>
<td>0.74</td>
<td>0.45-1.23</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• vfi per % &gt; 2</td>
<td>0.96</td>
<td>0.74-1.25</td>
<td>0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Localization</td>
<td>Benign n=144</td>
<td>Malignant n=37</td>
<td>Univariable analysis</td>
<td>Multivariable analysis (2D model)</td>
<td>Multivariable analysis (3D model)</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>----------------</td>
<td>----------------------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>odds ratio 95% CI</td>
<td>p-value</td>
<td>odds ratio 95% CI</td>
</tr>
<tr>
<td>missing values</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>0.62</td>
<td>6.53</td>
</tr>
<tr>
<td>no vessels (n)</td>
<td>30 (21%)</td>
<td>3 (8%)</td>
<td>1.40 0.37-5.24</td>
<td>0.003</td>
<td>1.42-30.1</td>
</tr>
<tr>
<td>peripheral vessels (n)</td>
<td>93 (65%)</td>
<td>13 (35%)</td>
<td>13 (35%)</td>
<td>12.5 2.77-56.4</td>
<td>0.001</td>
</tr>
<tr>
<td>central vessels (n)</td>
<td>13 (9%)</td>
<td>11 (30%)</td>
<td>11 (30%)</td>
<td>11 (30%)</td>
<td>11 (30%)</td>
</tr>
<tr>
<td>peripheral and central vessels (n)</td>
<td>8 (5%)</td>
<td>10 (27%)</td>
<td>10 (27%)</td>
<td>10 (27%)</td>
<td>10 (27%)</td>
</tr>
</tbody>
</table>

± standard deviation †absolute number ‡confidence interval
††not suspect vessels had resistance index ≥ 0.5; suspect vessels had resistance index < 0.5, **mostly due to absence of vessels
of vessels with RI < 0.5, resistance index, mean gray index, flow index until 30, presence of central vessels as well as the combination of peripheral and central vessels on 3D ultrasonography were all associated with the presence of malignancy.

In the multivariable analysis using baseline characteristics and 2D ultrasonography, age, CA125 level, presence of vessels suspected for malignancy, random echogenicity, presence of septations and ascites were selected as independent predictors of malignancy at a p-level of < 0.20. For the 3D model, age, CA125 level, ascites, uniform echogenicity, flow index (until a level of 30), vascularization flow index (until a level of 2%), volume of the mass on 3D ultrasonography, mean gray index and the visualization of central vessels showed an association with the probability of malignancy.

Since internal validation with bootstrapping showed a 19% overfit for the 2D model, and a 36% overfit for the 3D model, we corrected the regression coefficients of the model variables with these shrinkage factors. This reduces extremeness of predictions and causes the results to be more generalisable. The AUC was 0.82 (95% CI 0.73 - 0.92) and 0.92 (95% CI 0.87 - 0.98) for the 2D and 3D model respectively (p = 0.02) (figure 2).

Figure 3 shows the classification of patients according to their predicted probabilities as calculated from the both models. Calibration of the models was good, as was confirmed with the Hosmer and Lemeshow goodness-of-fit test statistic. This test statistic had a p-value of 0.59 for the 2D model and a p-value of 0.92 for the 3D model.

Discussion

In this study, we assessed whether 3D ultrasonography had additional value in the assessment of an adnexal mass as compared to 2D ultrasonography. We found that a model that incorporated data of 3D ultrasonography predicted significantly better than a model that included data of 2D ultrasonography and patient characteristics only.

We want to stress that the findings must be considered in the context of a pilot study in light of our relatively small sample size. We included nine parameters in the 3D model. As a general rule, about ten cases of malignancy are needed for every factor that is included in the model, which implicates that about 90 malignant masses would be needed. Moreover, external validation, an essential and final step in the development of prediction models, is needed before these models can be used in clinical practice, because prediction models tend to be overoptimistic when applied in other populations than the one in which they were developed.

Several studies on predicting malignancy in a patient with an adnexal mass by means of 3D ultrasonography have been published. Most of these studies conclude that 3D
ultrasonography has a higher specificity in predicting malignancy of adnexal masses compared to 2D ultrasonography. However, such isolated test research does not reflect daily clinical practice in which the findings of ultrasonography are always interpreted in the context of patient characteristics and laboratory results, and the additional value of 3D ultrasonography next to patients characteristics and 2D ultrasonography can only be assessed with multivariable analyses.24

Many authors have published prediction models based on 2D ultrasonography, patient characteristics, and tumourmarkers. By multivariable analysis using baseline characteristics, CA125 and ultrasonography, we developed and compared a 2D model and a 3D model. Before prediction models can be applied in clinical practice, they need internal and external validation. Internal validation was performed in the present study by bootstrapping to correct the models for overoptimism. Several authors have published models for the interpretation of data of 2D ultrasonography in clinical practice. A comparison of our newly developed 3D model with existing 2D models would be erroneous, since the 3D model could only be validated on the data-set in which it was developed. Therefore, we limited our comparison to the 2D model and the 3D model that we developed in the present study.

As the addition of 3D ultrasonography improves the accuracy of the pre-operative diagnosis of malignancy of the adnexal mass, the next question is whether this improvement is clinically relevant. As can be derived from the ROC-curves in figure 3, both models combine a specificity of 100% to a sensitivity of almost 40%, whereas at a 100% sensitivity a model with 3D ultrasonography improves the specificity from 30% to 50% as compared to a model with 2D ultrasonography. In view of the prevalence of malignancy of 20% in our population, this means that in a fictive population of 100 women with an adnexal mass, a model with 2D ultrasonography can detect eight malignant masses without any false positive diagnoses, whereas the diagnosis malignancy can be rejected in 24 women without any false negative diagnosis. In the remaining 68 women, health might be gained by the performance of 3D ultrasonography. Although our study population (n = 181) is large compared to previous studies reporting on 3D ultrasonography, the number of malignancies was relatively small (n = 37). To examine the value of 3D ultrasonography in this subgroup, a large prospective multicenter study would be necessary with a higher percentage of malignancies. Such data are needed to establish whether the improvement in diagnostic accuracy can be translated in more laparoscopies, less frozen section diagnostic procedures on laparotomy and less relaparotomies after an unclear first diagnosis.

Many models have been developed in the past to increase the accuracy of the pre-operative assessment of the adnexal mass. However the clinical judgment of gynaecologists appears to be as good as the capacity of complicated predictive scoring
systems like neural networks or logistic regression models for the pre-operative assessment of the adnexal masses. In a previous study we constructed ROC-curves for clinical offhand assessment of adnexal masses and calculated AUC’s of 0.82 - 0.84. Clinical offhand assessment seems to perform as well as multivariate models, including our 2D model, however, the newly constructed model incorporating 3D ultrasonography in addition to patient characteristics and 2D results showed an increase of the AUC to 0.92.

In conclusion, the aim of the present study was to evaluate the contribution of 3D ultrasonography in the prediction of malignancy in ovarian masses. We plotted the probability of malignancy as predicted by the model without the 3D parameters against the probability of malignancy predicted by the model with the 3D parameters added to the patient characteristics, level of CA125 and 2D ultrasonography. Both models had an acceptable capacity to distinguish between benign and malignant adnexal masses, but the prediction with the 3D model was significantly better than prediction with the 2D model. Based on the present dataset 3D ultrasonography seems to be a very promising technique to estimate the risk of malignancy in the pre-operative diagnostic workup of adnexal masses.

References

Chapter 6

Diagnostic accuracy of frozen section diagnosis of the adnexal mass: a meta-analysis

Peggy M.A.J. Geomini
Gérard L. Bremer
Roy F.P.M. Kruitwagen
Ben Willem J. Mol

Chapter 6

Abstract

Objective: Frozen section diagnosis is a diagnostic procedure for the assessment of the adnexal mass during surgery. The purpose of the present study was to perform a systematic review of the literature on the accuracy of frozen section diagnosis in the assessment of the adnexal mass.

Methods: We performed a computerized Medline and EMBASE search to identify all registered articles published between January 1966 and June 2003 comparing frozen section diagnosis of ovarian pathology to the final histopathological diagnosis. For each study we calculated the prevalence of malignant and borderline tumours and the sensitivity and specificity of the frozen section diagnosis using the final histopathological diagnosis as reference. We performed the calculations in two ways. In the first analysis, tumours that were found to be borderline were considered as malignant in the two by two table. In the second analysis, tumours that were found to be borderline were considered as benign in the two by two table.

Results: Eighteen studies were included for analysis. When the diagnosis borderline was classified as malignant, the sensitivity of frozen section diagnosis varied between 65% and 97%, and the specificity between 97% and 100%. When the diagnosis borderline was considered to be benign, the sensitivity varied between 71% and 100%, for a specificity varying between 98% and 100%.

Conclusion: The accuracy of frozen section diagnosis for the assessment of the ovarian mass is good, with acceptable sensitivities for almost perfect specificities. Future studies on patients preferences for the different outcomes as well as economic analysis are needed for definite position of this diagnostic technique.
Introduction

To provide optimal management of a patient with an adnexal mass it is important to differentiate between malignant and benign pathology prior to and during surgery. Accurate assessment is of interest for patient counselling and pre-operative preparations, as well as for the choice for the optimal surgical route during the operation. In case of doubt on the state of malignancy prior to surgery, frozen section diagnosis might give additional information during surgery. In case an adnexal mass is malignant, optimal debulking and surgical staging are of the utmost importance. In case an adnexal mass is of benign origin, one-sided salpingo-oophorectomy or cystectomy will suffice.

Pre-operative discrimination between the malignant and the benign adnexal mass can be made with patient characteristics such as age, menopausal status, ultrasonography, and the CA125 level in the peripheral blood. Some authors have developed prediction models that combine patient characteristics, ultrasonographic results and serum markers. Although such models are thought to improve pre-operative assessment of the adnexal mass, we have recently demonstrated that clinical judgement of the adnexal mass is as accurate as the use of formal prediction rules. However, both formal prognostic models as well as clinical judgement have demonstrated to lack perfect diagnostic accuracy. The areas under the ROC-curve for the clinical judgement were approximately 0.82. No statistically significant differences could be found between the accuracy of the risk assessments made by clinicians and the accuracy of the risk assessments made by mathematical models as the logistic model and neural network.

An additional diagnostic procedure for the assessment of the adnexal mass during surgery is frozen section diagnosis. In patients in whom the presence of malignancy prior to surgery is uncertain, frozen section diagnosis might be of use. This technique has been known since 1891, but studies on its diagnostic accuracy have only been performed since two decades ago. We are not aware of any study that assesses the literature on the subject in a systematic way. The purpose of the present study was therefore to perform a systematic review of the literature on the accuracy of frozen section diagnosis in the assessment of the adnexal mass.
Methods

Search strategy and data extraction

We performed a computerized Medline and EMBASE search to identify all registered articles published between January 1966 and June 2003 comparing the frozen section diagnosis of ovarian pathology with the final histopathological diagnosis. Key words were all words with the phrase ("frozen section") AND ("ovar*") AND ("accuracy"). Moreover, for MEDLINE, we used the phrases ("frozen section") AND ("ovar*") AND ("accuracy") in combination with the diagnostic filter of MEDLINE. Final definition of the MEDLINE search was as follows: (accuracy AND ovar* AND frozen section) AND (sensitivity and specificity [MESH] OR (predictive [WORD] AND value*[WORD])). In addition, cross-references of all selected articles were checked.

Articles were excluded whenever data were insufficient to construct a two by two table of the frozen section diagnosis and the final histopathological diagnosis. Two by two tables were constructed independently by two of the authors (PG and GB). In event of disagreement, the judgement of a third author (BM) was decisive.

Each of the included studies was scored by two of the authors (PG and GB) on the following characteristics of study design: 1. sampling (consecutive patients versus other); 2. data collection (prospective versus other); 3. study design (cohort study versus case-control study); 4. masking for the frozen section diagnosis while a definite diagnosis was made; 5. verification bias; 6. selection bias. Verification bias occurs if the decision to perform the reference test is based on the result of the test under study. Verification bias was default if the decision to perform a paraffin diagnosis was based on the result of the frozen section diagnosis. Selection bias can be present if not all cases or a random subset presenting with the relevant condition during a specified period are included. Selection bias was default if not all performed frozen section diagnosis in a specified period or a random subset were included.\textsuperscript{12}

Assessment of diagnostic performance

The analysis was conducted according to a methodology that has been described in detail elsewhere.\textsuperscript{13} In brief, for each study we calculated the prevalence of malignant and borderline tumours and the sensitivity, specificity and likelihood ratios of the frozen section diagnosis using the final histopathological diagnosis as reference. In adnexeal masses in which a diagnosis at frozen section could not be made, the result of frozen section diagnosis was considered to be benign, and cases were considered as such in the further analysis.

We performed the calculations in two ways. In the first analysis, tumours that were
found to be borderline were considered as malignant in the two by two table. In the second analysis, tumours that were found to be borderline at the final assessment were considered as benign in the two by two table. All sensitivity-specificity points were plotted in a receiver operating characteristic (ROC) curve. We used logistic regression in order to evaluate if characteristics of each individual study, as mentioned above (i.e. consecutive or non-consecutive sampling, prospective or retrospective data collection, cohort or case-control study, selection bias and verification bias), were associated with the discriminative capacity of frozen section diagnosis as reported in each study. Thereafter, tests for homogeneity by means of the chi-square-test statistic were performed for sensitivity and specificity separately. If homogeneity could not be rejected for both sensitivity and specificity, a more precise summary point estimate of sensitivity and specificity, with its 95% confidence intervals (CI), was calculated. Sample sizes were used as the weight of each study.14-15 To explore if the differences in sensitivity-specificity combinations were due to the use of different cut-off levels of the test under study, a Spearman correlation coefficient was calculated to assess the association between sensitivity and specificity.

Assessment of expected regret reduction with frozen section diagnosis

To conclude whether test characteristics justify the use for frozen section diagnosis you have to consider the consequences of misclassification due to false positive or false negative results of the frozen section diagnosis, as well as the prevalence of malignant disease.16 If we consider true positive (malignancy at frozen section diagnosis and final diagnosis) and true negative (benign disease at frozen section diagnosis and final diagnosis) as the optimal test results that do not generate regret, both the false positive and false negative diagnoses will generate some regret. The ratio between these two types of regret can be quantified by answering the question how much worse a patient will experience a false-positive frozen section diagnosis (radical surgery for benign disease) as compared to a false-negative frozen section diagnosis (inadequate surgery and a second procedure in case of malignant disease). A second issue that affects the expected value of frozen section diagnosis is the prevalence of malignancy. In case the probability of malignancy is virtually zero prior to surgery, it is unlikely that frozen section diagnosis will be of value. On the other hand, a patient in whom malignancy is virtually certain prior to surgery, will also not benefit from frozen section diagnosis. The risk and consequences of spill during surgery are left out of consideration.

The regret expected from three strategies can be compared in a decision frame work advocated by Pauker and Kassirer.16 The three strategies to consider are (I) radical surgery for all patients, (II) ovariectomy for all patients and (III) frozen section diagnosis, and, depending on the frozen section diagnosis, either radical surgery or one-
sided ovariectomy. If the prevalence of malignant disease is represented as \( P \), the expected regret from one-sided ovariectomy in case of malignancy as \( L \), and the expected regret from radical surgery in case of benign disease as the regret ratio \( RR \), than the expected regret can be expressed for each of the three strategies.

## Results

The computerized search detected 28 studies, of which eight studies were excluded after reading the abstract because they did not report on the association between frozen section diagnosis and the final histopathological diagnosis. From the cross-references of the 20 selected articles, seven additional studies were identified for further reading. Nine articles were inappropriate for analysis for the following reasons. In two studies, gynecologic frozen section diagnosis was compared with the final diagnosis, but the results of adnexal masses could not be analyzed separately from other gynecological tumours.\(^{17,18}\) From three studies two by two tables could not be derived,\(^{19-21}\) In three other studies only borderline malignancy was considered.\(^{22-24}\) One study was excluded because it reported only on cases with benign disease.\(^{25}\) There was no discordance between the two authors who judged the pre-selected articles.

All 18 included studies were retrospective cohort studies (table 1).\(^ {11,26-41}\) All studies but two included consecutive patients undergoing frozen section diagnosis.\(^ {11,39}\) All studies were performed during a certain period with the exception of one study.\(^ {39}\) The study of Slavutin was limited to frozen section diagnosis of serous epithelial tumours of the ovary.\(^ {11}\) In this study, the original frozen section was reviewed without knowledge of the previously given frozen section or final diagnosis. In all other studies presence of masking was not described separately.

Obiakor reported on a retrospective study in which 311 patients undergoing frozen section diagnosis were included.\(^ {26}\) The author excluded eight patients from the study, because at frozen section analysis a diagnosis could not be made. Thus, since data on eight patients were missing, presence of selection bias cannot be excluded in this non-consecutive study.

Six of the 16 consecutive studies reported that in some patients the diagnosis was deferred at frozen section diagnosis (table 2).\(^ {27,29,33,37-39}\) The percentage of patients in which the diagnosis was deferred varied between 1% and 6%, whereas in one study the diagnosis was even deferred in 20% of the patients.\(^ {39}\) The authors mention technical difficulties like sampling errors because lack of time, decreased pathological processing compared with paraffin diagnosis, inflammatory conditions, haemorrhage or necrosis and non-representative tissue. Another reason for deferring frozen section diagnosis was lack of communication between pathologist and surgeon.
Table 1. Included studies and reported test characteristics

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>No of patients</th>
<th>Malignant masses</th>
<th>Borderline masses</th>
<th>Frozen section benign</th>
<th>Frozen section borderline</th>
<th>Frozen section malignant</th>
<th>Borderline considered as malignant</th>
<th>Borderline considered as benign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slavutin</td>
<td>1979</td>
<td>55</td>
<td>19 (35%)</td>
<td>2 (4%)</td>
<td>33 (1)</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>5 (5)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Bastos</td>
<td>1983</td>
<td>120</td>
<td>35 (29%)</td>
<td>0</td>
<td>85 (5)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>30 (30)</td>
</tr>
<tr>
<td>Twaalthoven</td>
<td>1991</td>
<td>176</td>
<td>61 (35%)</td>
<td>18 (10%)</td>
<td>96 (10)</td>
<td>3 (1)</td>
<td>8 (4)</td>
<td>4 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Obiakor</td>
<td>1991</td>
<td>303</td>
<td>73 (24%)</td>
<td>6 (2%)</td>
<td>224 (6)</td>
<td>4 (0)</td>
<td>0 (0)</td>
<td>1 (1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Cheung</td>
<td>1992</td>
<td>159</td>
<td>38 (24%)</td>
<td>27 (17%)</td>
<td>94 (9)</td>
<td>0 (0)</td>
<td>16 (1)</td>
<td>1 (0)</td>
<td>2 (37)</td>
</tr>
<tr>
<td>Zhang</td>
<td>1993</td>
<td>136</td>
<td>62 (46%)</td>
<td>22 (16%)</td>
<td>52 (5)</td>
<td>1 (0)</td>
<td>17 (3)</td>
<td>3 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Hamed</td>
<td>1993</td>
<td>324</td>
<td>55 (17%)</td>
<td>10 (3%)</td>
<td>258 (2)</td>
<td>0 (0)</td>
<td>17 (0)</td>
<td>0 (0)</td>
<td>1 (55)</td>
</tr>
<tr>
<td>Spann</td>
<td>1994</td>
<td>307</td>
<td>44 (14%)</td>
<td>0</td>
<td>262 (0)</td>
<td>12 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Rose</td>
<td>1994</td>
<td>383</td>
<td>120 (31%)</td>
<td>29 (8%)</td>
<td>231 (15)</td>
<td>5 (3)</td>
<td>13 (4)</td>
<td>0 (1)</td>
<td>1 (111)</td>
</tr>
<tr>
<td>Michael</td>
<td>1996</td>
<td>63</td>
<td>26 (41%)</td>
<td>8 (13%)</td>
<td>29 (5)</td>
<td>7 (0)</td>
<td>3 (0)</td>
<td>0 (0)</td>
<td>19 (19)</td>
</tr>
<tr>
<td>Pulos</td>
<td>1997</td>
<td>294</td>
<td>38 (13%)</td>
<td>40 (14%)</td>
<td>213 (8)</td>
<td>1 (3)</td>
<td>31 (10)</td>
<td>0 (1)</td>
<td>27 (27)</td>
</tr>
<tr>
<td>Lim</td>
<td>1997</td>
<td>173</td>
<td>37 (21%)</td>
<td>8 (5%)</td>
<td>127 (1)</td>
<td>3 (1)</td>
<td>7 (0)</td>
<td>0 (0)</td>
<td>34 (34)</td>
</tr>
<tr>
<td>Wang</td>
<td>1998</td>
<td>299</td>
<td>75 (25%)</td>
<td>23 (8%)</td>
<td>201 (5)</td>
<td>3 (0)</td>
<td>18 (3)</td>
<td>0 (0)</td>
<td>69 (69)</td>
</tr>
<tr>
<td>Ushutsum</td>
<td>1998</td>
<td>360</td>
<td>90 (25%)</td>
<td>4 (1%)</td>
<td>258 (3)</td>
<td>6 (0)</td>
<td>1 (0)</td>
<td>0 (0)</td>
<td>84 (84)</td>
</tr>
<tr>
<td>Yee</td>
<td>1998</td>
<td>316</td>
<td>46 (15%)</td>
<td>20 (6%)</td>
<td>249 (8)</td>
<td>3 (1)</td>
<td>12 (3)</td>
<td>0 (0)</td>
<td>40 (40)</td>
</tr>
<tr>
<td>Cetin</td>
<td>1999</td>
<td>489</td>
<td>71 (15%)</td>
<td>23 (5%)</td>
<td>392 (6)</td>
<td>1 (3)</td>
<td>14 (3)</td>
<td>3 (0)</td>
<td>67 (67)</td>
</tr>
<tr>
<td>Wakahara</td>
<td>2001</td>
<td>187</td>
<td>54 (29%)</td>
<td>15 (8%)</td>
<td>117 (5)</td>
<td>0 (1)</td>
<td>10 (0)</td>
<td>0 (0)</td>
<td>54 (54)</td>
</tr>
<tr>
<td>Pinto</td>
<td>2001</td>
<td>243</td>
<td>69 (28%)</td>
<td>18 (7%)</td>
<td>156 (6)</td>
<td>2 (0)</td>
<td>11 (3)</td>
<td>0 (1)</td>
<td>64 (64)</td>
</tr>
</tbody>
</table>

No = number; par. = paraffin; ben. = benign; bor. = borderline; mal. = malignant; sens. = sensitivity; spec. = specificity
Chapter 6

Table 2. Deferred cases at frozen section diagnosis

<table>
<thead>
<tr>
<th>No patients</th>
<th>Deferred diagnosis (%)</th>
<th>Paraffin diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>benzine</td>
<td>borderline</td>
</tr>
<tr>
<td>Wang²⁹</td>
<td>299</td>
<td>3</td>
</tr>
<tr>
<td>Usubutun³³</td>
<td>360</td>
<td>12</td>
</tr>
<tr>
<td>Twaalthoven³⁷</td>
<td>176</td>
<td>11</td>
</tr>
<tr>
<td>Lim³⁸</td>
<td>173</td>
<td>2</td>
</tr>
<tr>
<td>Michael³⁹</td>
<td>63</td>
<td>16</td>
</tr>
<tr>
<td>Bastos³⁷</td>
<td>120</td>
<td>2</td>
</tr>
</tbody>
</table>

No = number

Figure 1. Included studies ordered according to study size

Borderline tumour considered as malignant

---

Borderline tumour considered as benign

---

Author | No. of patients |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Slavutin¹¹</td>
<td>55</td>
</tr>
<tr>
<td>Michael³⁹</td>
<td>63</td>
</tr>
<tr>
<td>Bastos³⁷</td>
<td>120</td>
</tr>
<tr>
<td>Zhang³⁰</td>
<td>136</td>
</tr>
<tr>
<td>Cheung²⁸</td>
<td>159</td>
</tr>
<tr>
<td>Lim³⁸</td>
<td>173</td>
</tr>
<tr>
<td>Twaalthoven³⁷</td>
<td>176</td>
</tr>
<tr>
<td>Wakahara²⁴</td>
<td>187</td>
</tr>
<tr>
<td>Pinto³⁴</td>
<td>243</td>
</tr>
<tr>
<td>Pule³¹</td>
<td>294</td>
</tr>
<tr>
<td>Wang³⁹</td>
<td>299</td>
</tr>
<tr>
<td>Obiakor²⁶</td>
<td>303</td>
</tr>
<tr>
<td>Spann³⁰</td>
<td>307</td>
</tr>
<tr>
<td>Yeo³⁶</td>
<td>316</td>
</tr>
<tr>
<td>Hamed⁴²</td>
<td>324</td>
</tr>
<tr>
<td>Usubutun³³</td>
<td>360</td>
</tr>
<tr>
<td>Rose³²</td>
<td>383</td>
</tr>
<tr>
<td>Cuello³¹</td>
<td>489</td>
</tr>
</tbody>
</table>

---

Sensitivity / specificity

---

Sensitivity / specificity

---

88
Almost no study described criteria to decide for frozen section diagnosis. In the study of Bastos frozen section diagnose was performed in all patients with the diagnosis of adnexal mass, including masses that were evident malignant. A judgement on the presence of selection bias was not possible in the remaining included studies. Verification bias was not detected in the included studies.

Two studies did not include borderline tumours as a separate entity in their analyses. Both studies classified tumours of low malignant potential (borderline) as malignant without argumentation.

**Assessment of diagnostic performance**

We did not find a significant association between study characteristics and the diagnostic accuracy of frozen section diagnosis. When a diagnosis borderline malignancy was considered to be malignant, the sensitivity of frozen section diagnosis varied between 65% and 97% (table 1). The specificity varied between 97% and 100%. For the sensitivity, homogeneity was rejected (p-value < 0.01), whereas for specificity it could not be rejected (p-value 0.51). The pooled estimate for the specificity was 99.4% (95% CI 99.0 - 99.6). The Spearman correlation coefficient between the sensitivity and specificity was 0.03, indicating no correlation. Thus, we could not estimate a summary ROC-curve.

When a diagnosis borderline malignancy was considered as benign, the sensitivity varied between 71% and 100%. The specificity varied between 99.3% and 100%, except for one study that reported a specificity of 98.3%. Again, for the sensitivity homogeneity was rejected (p-value < 0.001), whereas for specificity homogeneity could not be rejected (p-value 0.85). The pooled estimate for the specificity was 99.6% (95% CI 99.3 - 99.8). The Spearman correlation coefficient between the sensitivity and specificity was -0.13, indicating a poor correlation, which did not allow estimation of a summary ROC-curve.

Figure 1 shows the sensitivity and specificity reported in each study with their 95% confidence intervals (CI), ordered according to study size. From the two studies that reported a sensitivity for borderline and true malignancy below 75%, one was among the smallest studies. From the four studies that reported a sensitivity for malignancy only below 75%, two were among the smallest studies. However, exclusion of studies with less than 100 patients from the calculations for homogeneity did not alter the results.

**Assessment of expected regret reduction with frozen section diagnosis**

Table 3 demonstrates expected regret reduction from frozen section diagnosis as compared to radical surgery or ovariectomy in a hypothetical cohort of 1,000 patients.
Table 3. Expected regret reduction obtained from a strategy with frozen section diagnosis as compared to either radical surgery in all patients or ovariectomy in a hypothetical cohort of 1000 patients undergoing surgery for an adnexal mass. For each of the three strategies, the expected regret was calculated from the prevalence of malignancy and valuation of a false positive and a false negative diagnosis of frozen section diagnosis. The regret of a false negative diagnosis was set as 1, and the regret of a false positive diagnosis was linked with a regret ratio to the value of a false negative diagnosis. For example, a regret ratio of 10 implicates that a false positive diagnosis is expected to generate 10 times more regret than a false negative diagnosis. In the table, the prevalence of disease is varied between 5% and 90%, whereas the regret ratio is varied between 0.5 and 100. From the tables, it becomes apparent that frozen section diagnosis is expected to reduce regret in a majority of patients. Ovariectomy is superior in case the prevalence is less than 40% and the regret ratio exceeds 100.

<table>
<thead>
<tr>
<th></th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>-41</td>
<td>-85</td>
<td>-174</td>
<td>-262</td>
<td>-350</td>
<td>-439</td>
<td>-527</td>
<td>-615</td>
<td>-703</td>
<td>-792</td>
</tr>
<tr>
<td>10 ovarietomy</td>
<td>-34</td>
<td>-128</td>
<td>-222</td>
<td>-316</td>
<td>-410</td>
<td>-504</td>
<td>-598</td>
<td>-692</td>
<td>-786</td>
<td></td>
</tr>
<tr>
<td>50 ovarietomy ovarietomy ovarietomy ovarietomy ovarietomy ovarietomy ovarietomy</td>
<td>-54</td>
<td>-172</td>
<td>-290</td>
<td>-408</td>
<td>-526</td>
<td>-644</td>
<td>-762</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 ovarietomy ovarietomy ovarietomy ovarietomy ovarietomy ovarietomy ovarietomy</td>
<td>-140</td>
<td>-288</td>
<td>-436</td>
<td>-584</td>
<td>-732</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Borderline considered as malignant; sensitivity 88%, specificity 99.4%

undergoing surgery for an adnexal mass. For each of the three strategies (strategy I: radical surgery for all patients; strategy II: ovariectomy for all patients; strategy III: frozen section diagnosis, and, depending on the frozen section diagnosis, either radical surgery or one-sided ovariectomy), the expected regret was calculated from the prevalence of malignancy and valuation of a false positive and a false negative diagnosis of frozen section diagnosis (see addendum). The regret of a false negative diagnosis was set as 1, and the regret of a false positive diagnosis was linked with a regret ratio to the value of a false negative diagnosis. For example, a regret ratio of 10 implicates that a false positive diagnosis is expected to generate 10 times more regret than a false negative diagnosis. In the table, the prevalence of disease is varied between 5% and 90%, whereas the regret ratio is varied between 0.5 and 100.
Discussion

We performed a systematic review of the literature on the accuracy of frozen section diagnosis in the assessment of the adnexal mass. The specificity was found to be higher than 99%, whereas the sensitivity varied between 65% and 100%. We could no find an explanation for the heterogeneity among reported sensitivities. We cannot judge the influence of publication bias.

Both frozen section diagnosis and paraffin diagnosis of the adnexal mass are classified as benign, borderline malignant, or malignant. Since the meta-analytic method that we used required classification of patients in a two by two table, we had to classify borderline tumours as benign or malignant. A three by three table would be an alternative way to present our results. However, current meta-analytic methods do not facilitate interpretation of more than two disease categories in one analysis. One should realize that the clinical decision during surgery has two dimensions: either you consider the disease as malignant and you stage, or you consider the disease as benign you do not stage. Thus, a two by two table is useful if we consider the clinical situation.

Another question is how the clinician deals with borderline disease, and this has in our opinion two dimensions. In absence on clear evidence on this subject, there are wide variances in local practice. Our study offers the clinician the possibility to deal with this issue: in the first analysis by considering borderline tumours as malignant and in the second analysis by considering borderline tumours as benign. The clinician can choose which of these situations is appropriate for a specific patient. In case a borderline malignancy is treated with a bilateral salpingo-oophorectomy and hysterectomy with subsequent surgical staging, a borderline malignancy at frozen section diagnosis has to be considered as malignant disease. However, if borderline disease does not require such extensive surgery, it might be better to classify the disease as benign.

In this context, it is important to realize that there is no evidence on the optimal treatment of the borderline adnexal mass. Although many guidelines recommend radical surgery, one could argue whether such mutilative treatment is justified, specially in the situation where the woman is young and still has a wish to preserve her fertility. Although this discussion is important for further management of the borderline adnexal mass, it is of less importance for the assessment of the accuracy of frozen section diagnosis, since we found only subtle differences in sensitivity of the frozen section diagnosis when borderline disease was classified as either benign or malignant.

There were six studies that reported explicitly that a diagnosis was deferred at frozen section diagnosis. Since it would generate bias if we would exclude these cases from our study, we included them, and considered the result as frozen section diagnosis as benign. Such a choice is justified, if one realizes that in case malignancy is not found
at frozen section diagnosis, subsequent radical surgical treatment is not justified. Classification of a deferred diagnosis at the frozen section procedure is also justified if one takes a closer look at the numbers found in the present meta-analysis. Of the 46 deferred frozen section diagnoses, final diagnosis showed malignancy in 17 tumours (37%), borderline malignancy in 10 tumours (22%), and benign disease in 19 tumours (41%) (table 2). These numbers implicate that, if borderline disease is classified as malignant, the likelihood of a deferred diagnosis in case of malignancy is 7% (27 deferred cases among 385 cases of malignancy and borderline disease in the six studies26,28,32,36-38) for a likelihood of 3.7% of a deferred diagnosis in case of benign disease (19 deferred cases among 507 cases of benign disease in the six studies27, 29, 33, 37-39). Thus, the likelihood ratio for malignancy in case of deferred diagnosis is 7% / 3.7% = 1.9, a value much too small to justify radical surgery in case of a deferred diagnosis.

Leaving the deferred frozen section diagnosis out of the two by two table would cause bias and inflate test accuracy, since the test would only decide in case there is certainty about the disease status. However, exclusion of studies that reported deferred frozen section diagnosis the calculations for homogeneity did not alter the results.

We found homogeneity among the reported specificity of frozen section diagnosis, whereas estimates of the sensitivity of the test varied between 65% and 100%. This heterogeneity is unlikely due to differences in study design, since most studies were retrospective cohort studies reporting on a consecutive series of patients. Moreover, we could not find a clear association between prevalence of disease, year of publication or study size on one hand and sensitivity on the other hand. One explanation for the imperfect sensitivity of frozen section diagnosis might be the size of the tumour. It is not unlikely that malignant disease is not detected in case the adnexal mass is large. However, individual patient data could not be retrieved from the included papers. This issue should be addressed in future studies.

In general histological characteristics of borderline tumours are stratification of the epithelial lining of the papillae in two or three layers, formation of microscopic papillary projections arising from the epithelial lining of the papillae, epithelial pleomorphism, atipicality, mitotic activity and the absence of stromal invasion. When a tumour shows even minimal invasive growth it is considered to be malignant, and should be treated accordingly. However in the included studies the authors do not describe the used definition of borderline tumours.

Histological variation may occur more commonly in mucinous than in serous tumours, making the accuracy of frozen section diagnosis in case of mucinous tumours less predictable. For a clear diagnosis and to rule out invasiveness it is necessary to section the neoplasm extensively. In case of borderline malignancy, mucinous tumours
and large cysts it is difficult for frozen section examination to meet the sampling requirements and frozen section diagnosis will be less sensitive. Some of the included studies implicate this sort of explanation, however the data were insufficient to use in this meta-analysis.

Despite the limitations mentioned above, we found excellent test characteristics for frozen section diagnosis. The question arises whether these test characteristics justify the use of frozen section diagnosis in clinical practice. To answer this question we calculated expected regret reduction based on the consequences of misclassification due to false positive or false negative results of the frozen section diagnosis and on the prevalence of malignant disease. Table 3 demonstrates expected regret reduction from frozen section diagnosis as compared to radical surgery of ovariectomy in a hypothetical cohort of 1000 patients undergoing surgery for an adnexal mass. For example, a regret ratio of 10 implicates that a false positive diagnosis is expected to generate 10 times more regret than a false negative diagnosis. In the table, the prevalence of disease is varied between 5% and 90%, whereas the regret ratio is varied between 0.5 and 100. From the tables, it becomes apparent that frozen section diagnosis is expected to reduce regret in a majority of patients. Ovariectomy is superior in case the prevalence is less than 40% and the regret ratio exceeds 100. Sensitivity analysis showed that even if the sensitivity dropped to 30%, frozen section diagnosis was beneficial in case the prevalence of malignancy exceeded 15%.

We are not aware of any studies that assess the valuation of radical surgery in absence of malignancy as compared to the valuation of the need for a second look laparotomy after inadequate surgery for malignancy at the initial procedure as is found in other fields.43 However, one could speculate that if the initial surgery is performed by midline incision, the ratio between these valuations will be around one. As can be seen in table 3, frozen section diagnosis is superior over the other two strategies for virtually each prevalence of disease. However, if the initial surgery is performed laparoscopically, a false positive frozen section diagnosis (radical surgery where laparoscopy would be sufficient) will generate more regret than a false negative frozen section diagnosis (a second look laparotomy after an initial laparoscopy). A similar change in regret ratio can be expected in case the woman is premenopausal, and even has the wish to preserve her reproductive capacity. In that situation, unnecessary radical surgery will generate much more regret as compared to a similar procedure in postmenopausal women. As can be seen from table III, an increase of the regret ratio to 50 reduces the expected benefit from frozen section diagnosis. Thus, studies in which patient preferences for the different situations are assessed, are warranted.

In conclusion, the present meta-analysis has demonstrated that the accuracy of frozen section diagnosis for the assessment of the ovarian mass is good, with acceptable
sensitivities for almost perfect specificities. Future studies on patients preferences for the different outcomes as well as economic analysis are needed for definite position of this diagnostic technique.

Addendum

The regret expected from three strategies can be compared in a decision frame work advocated by Pauker and Kassirer. The three strategies to consider are (I) radical surgery for all patients, (II) ovariectomy for all patients and (III) frozen section diagnosis, and, depending on the frozen section diagnosis, either radical surgery or one-sided ovariectomy. If the prevalence of malignant disease is represented as P, the expected regret from one-sided ovariectomy in case of malignancy as 1, and the expected regret from radical surgery in case of benign disease as the regret ratio RR, than the expected regret can be expressed for each of the three strategies.

Strategy I:  \( \text{Radical surgery (expected regret)} \ (P, RR) = P \times 0 + (1 - P) \times RR \)

Strategy II:  \( \text{One-sided ovariectomy (expected regret)} \ (P) = P \times 1 \)

Strategy III:  \( \text{Frozen section diagnosis (expected regret)} \ (P, RR, sens, spec) = \\
P \times sens \times 0 + P \times (1 - sens) \times 1 + (1 - P) \times spec \times 0 + (1 - P) \times (1 - spec) \times RR \)

For example, if the prevalence of disease is 30% and if the regret ratio is 10, than the expected regret from strategy I (radical surgery) equals \((1 - 30\%) \times 10 = 0.7\). For strategy II (one-sided ovariectomy), the expected regret equals 0.3. If frozen section diagnosis is applied with a sensitivity of 88% and a specificity of 99.4%, the expected regret equals \(0.3 \times 0.12 + (1 - 0.3) \times 0.006 \times 10 = 0.078\). Thus, use of frozen section diagnosis reduces the expected regret in one patient with 0.222, which equals the regret reduction that is expected in a fictive cohort of 1000 patients with a prevalence of 30% and a regret ratio of 10 in table 3.

References


95


Chapter 7

The impact of size of the adnexal mass on the accuracy of frozen section diagnosis

Peggy M.A.J. Geomini
Laura D. Zuurendonk
Gérard L. Bremer
Jan de Graaff
Roy F.P.M. Kruitwagen
Ben Willem J. Mol

Chapter 7

Abstract

Objective: We recently showed that frozen section diagnosis has an almost perfect specificity for the diagnosis of malignancy in an adnexal mass, whereas the sensitivity was, though good, not perfect. The aim of the present study was to examine whether the accuracy of frozen section diagnosis is affected by the size of the adnexal masses.

Methods: We included women that underwent frozen section diagnosis for adnexal mass surgery. For each patient we recorded serum CA125 level, menopausal status, tumour size, and histologic classification both at frozen section diagnosis and at definitive histological examination. We calculated sensitivity and specificity of frozen section diagnoses, both in tumours below and above 10 cm diameter.

Results: We included 257 patients, of whom 142 had a benign tumour, 28 had a borderline tumour and 87 had a malignant tumour at definitive histological assessment. In case frozen section diagnosis showed malignancy, this was always confirmed at final histological assessment. In women with a tumour < 10 cm, there was only one false negative diagnosis in 50 women with a benign frozen section diagnosis, whereas there were 11 false negative diagnoses in 97 women in women with a tumour ≥ 10 cm. The corresponding likelihood ratios of a benign diagnosis for presence of malignancy were 0.15 for tumours ≥ 10 cm and 0.03 for tumours < 10 cm, respectively.

Conclusion: The accuracy of frozen section diagnosis is dependent on tumour size. In adnexal masses ≥ 10 cm, a benign result of the frozen section diagnosis is less reliable than in women with a tumour < 10 cm.
Introduction

In the management of a patient with an adnexal mass, it is important to differentiate between malignant and benign pathology both prior to as well as during surgery. Accurate assessment prior to surgery is of interest for patient counselling and pre-operative preparations, as well as for the choice for the optimal surgical route during the operation. Accurate assessment during surgery is of interest for the subsequent surgical treatment, and frozen section diagnosis is a useful tool for the latter problem.

Recently, we performed a systematic review on the accuracy of frozen section diagnosis in the assessment of the adnexal mass. After analysis of 18 studies, the specificity was found to be higher than 99%, whereas the sensitivity varied between 65% and 100%. We could not find an explanation for the heterogeneity among reported sensitivities.

The size of the adnexal mass might be of influence on the accuracy of the frozen section diagnosis. Whereas malignancy is always detected in case the size of the adnexal mass is relatively small, frozen section diagnosis might miss malignancy in case the adnexal mass has a larger volume. Since this hypothesis could not be evaluated in our previous meta-analysis, we examined in this retrospective study whether size of the adnexal mass had impact on the accuracy of frozen section diagnosis.

Materials and methods

The study was performed in Máxima Medical Centre in Veldhoven, Tweesteden Hospital in Tilburg and Saint Elisabeth Hospital in Tilburg. All three hospitals are large teaching hospitals in the south of The Netherlands. We included all women who had undergone frozen section diagnosis for an adnexal mass between January 2000 and December 2003 retrospectively.

The indication for frozen section diagnosis was the suspicion on malignancy in pre-operative examination by ultrasound, level of CA125 and/or gynaecologic examination. In case frozen section diagnosis was performed, this was registered in the database that holds all surgical interventions performed in the three hospitals. We selected all patients who were marked as having had frozen section diagnosis for ovarian surgery.

During surgery, the complete adnexal mass was sent fresh to the laboratory of the Department of Pathology and examined by one of the experienced pathologists. The most suspected parts of the adnexal mass were selected for frozen section diagnosis. The pathologist measured the tumour size in centimetres and the largest diameter was used for analysis.

The definitive paraffin diagnosis and the frozen section diagnosis were made either
Chapter 7

by the same pathologist or by two separate pathologists. At the moment of examining
the paraffin slices the pathologist knew the result of the frozen section diagnosis. The
malignancy status of frozen section diagnosis and permanent diagnosis was divided into
benign, borderline malignancy or malignancy. For each case the following data were
recorded: pre-operative level of serum CA125, menopausal status, tumour size at
surgery, and histological classification both at frozen section diagnosis and at final
histology. The level of CA125 was measured with chemoluminouscent immunometric
assay (Immulite 2000 OM-MA, DPC, Diagnostic Product Corporation, Los Angeles,
USA).

Analysis

We tabulated the results of the frozen section diagnosis against the results of the final
paraffin diagnosis. We calculated sensitivity and specificity in two ways, first by
considering borderline tumours as benign, and second by classifying them as malignant.
Subsequently, we calculated the mean tumour size and CA125 level for each of the
combinations of frozen section diagnosis and final diagnosis.

Results

From January 2000 to December 2003, 257 consecutive frozen section diagnoses of
adnexal masses were performed. All cases were included in this retrospective study. The
permanent paraffin diagnosis showed benign pathology in 142 cases (55%), malignancy
in 87 cases (34%) and borderline malignancy in 28 tumours (11%). Two deferred cases
appeared to be benign at final assessment.

The median age of the included patients was 56 years (range 21 - 90 years). Among
166 postmenopausal patients, there were 63 with a malignancy (38%) and 12 with a
borderline tumour (8%). Of the 90 premenopausal women, 23 (26%) had a malignant
paraffin diagnosis and 16 (18%) had a tumour of borderline malignancy. In one case the
menopausal status was unknown.

The results of the frozen section diagnoses and the paraffin diagnoses are compared
in table 1. All of the 69 frozen section diagnoses that showed malignancy were
confirmed to be malignant at the permanent paraffin diagnosis. When in the paraffin
diagnosis borderline malignancy was considered as malignant, the likelihood ratios of
benign, borderline and malignant result of the frozen section diagnosis were 0.11, 5.9
and infinite, respectively. When a diagnosis borderline malignancy was considered as
benign, these values were 0.10, 0.70 and infinite, respectively.

Table 2 shows the mean tumour size per diagnostic category. The mean size of the
adnexal mass in patients with a malignant frozen section diagnosis was 12.7 cm (SD 6.8
cm, range 2.0 to 40 cm). In patients with a borderline frozen section diagnosis adnexal masses were somewhat smaller, but there were no significant differences between those patients who had benign, malignant or borderline histology at final assessment. However, for benign frozen section diagnosis, there was a remarkably difference between the tumour size of those patients that had benign histology, borderline histology and malignant histology, i.e. 13.1, 17.2 and 19.6 cm, respectively (p-value 0.02). The smallest tumour size of a patient with a false-negative frozen section diagnosis (i.e. benign frozen section and malignant at final assessment) was eight cm. There were 81 women with a tumour < 10 cm, of which 25 (31%) had a malignancy

### Table 1. Correlation of status of malignancy between frozen section diagnosis and permanent section diagnosis

<table>
<thead>
<tr>
<th>Frozen section diagnosis</th>
<th>Permanent section diagnosis</th>
<th>likelihood ratio* (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>benign</td>
<td>benign</td>
<td>0.11 (0.1 - 0.2)</td>
</tr>
<tr>
<td></td>
<td>borderline</td>
<td>5.9 (2.8 - 13.0)</td>
</tr>
<tr>
<td></td>
<td>malignant</td>
<td>∞ (16.0 - ∞)</td>
</tr>
<tr>
<td>deferred</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>totals</td>
<td>142</td>
<td>257</td>
</tr>
</tbody>
</table>

*borderline malignancy considered as malignant
#CI means confidence interval

### Table 2. Mean tumour size in cm (SD) and range per diagnostic category

<table>
<thead>
<tr>
<th>Frozen section diagnosis</th>
<th>Permanent section diagnosis</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>benign</td>
<td>benign</td>
<td>19.6 (8.7)</td>
</tr>
<tr>
<td></td>
<td>n = 133</td>
<td>n = 7</td>
</tr>
<tr>
<td></td>
<td>2.0 - 39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>borderline</td>
<td>14.9 (8.2)</td>
</tr>
<tr>
<td></td>
<td>n = 6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.0 - 33.0</td>
<td></td>
</tr>
<tr>
<td>borderline</td>
<td>malignant</td>
<td>14.3 (7.6)</td>
</tr>
<tr>
<td></td>
<td>n = 23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0 - 35.0</td>
<td></td>
</tr>
<tr>
<td>malignant</td>
<td>2.0 - 40.0</td>
<td></td>
</tr>
<tr>
<td>deferred</td>
<td>17.5 (10.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.0 - 28.0</td>
<td></td>
</tr>
</tbody>
</table>

*3 missing values
#ns means not significant
and six (7%) had a borderline tumour. In this group, 50 had a benign frozen section diagnosis, of whom one turned out to be malignant. Nine women had a borderline frozen section diagnosis, in which six were indeed borderline at paraffin diagnosis, two were malignant and one was benign. All 22 women with malignancy at frozen section diagnosis had a malignancy at the final diagnosis. In women with a tumour < 10 cm, the likelihood ratios for benign frozen section diagnosis, borderline frozen section diagnosis, and malignant frozen section diagnosis were 0.03 (95% CI 0.01 - 0.23), 12.9 (95% CI 1.7 - 98.2) and infinite (95% CI 7.1 - ∞), respectively.

There were 173 women with a tumour ≥ 10 cm, of which 59 (34%) had a malignancy and 22 (13%) had a borderline tumour. In this group, 97 had a benign frozen section

<table>
<thead>
<tr>
<th>Permanent section diagnosis</th>
<th>Number</th>
<th>Frozen section diagnosis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>benign</td>
<td>borderline</td>
</tr>
<tr>
<td>mucinous cystadenoma</td>
<td>39</td>
<td>37</td>
<td>2</td>
</tr>
<tr>
<td>serous cystadenoma</td>
<td>39</td>
<td></td>
<td>*36</td>
</tr>
<tr>
<td>endometrioid tumour</td>
<td>14</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>dermoid cyst</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>fibroma</td>
<td>23</td>
<td>*21</td>
<td></td>
</tr>
<tr>
<td>luteal cyst</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>benign Brenner tumour</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>thecoma</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>inflammation</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>benign, not specified</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

* inclusive 1 deferred frozen section diagnosis

<table>
<thead>
<tr>
<th>Frozen section diagnosis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>benign</td>
<td>borderline</td>
</tr>
<tr>
<td>benign</td>
<td>4</td>
</tr>
<tr>
<td>borderline, not specified</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frozen section diagnosis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>benign</td>
<td>borderline</td>
</tr>
<tr>
<td>benign</td>
<td>5</td>
</tr>
<tr>
<td>borderline</td>
<td>1</td>
</tr>
<tr>
<td>malignant</td>
<td>2</td>
</tr>
<tr>
<td>malignant</td>
<td>11</td>
</tr>
<tr>
<td>undifferentiated carcinoma</td>
<td>6</td>
</tr>
<tr>
<td>malignant Brenner tumour</td>
<td>2</td>
</tr>
<tr>
<td>carcinosarcoma11</td>
<td></td>
</tr>
<tr>
<td>Sertoli Leydig cell tumour</td>
<td>1</td>
</tr>
<tr>
<td>Krukenberg tumour</td>
<td>1</td>
</tr>
<tr>
<td>melanoma</td>
<td>1</td>
</tr>
</tbody>
</table>

102
Table 4. Mean CA125 in kU/l (SD) and range per diagnostic category

<table>
<thead>
<tr>
<th>Frozen section diagnosis</th>
<th>Permanent section diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>benign</td>
</tr>
<tr>
<td>benign</td>
<td>58 (156)*</td>
</tr>
<tr>
<td>n = 133</td>
<td>n = 5</td>
</tr>
<tr>
<td>4 - 1137</td>
<td>4 - 410</td>
</tr>
<tr>
<td>borderline</td>
<td>55 (0)*</td>
</tr>
<tr>
<td>n = 7</td>
<td>n = 23</td>
</tr>
<tr>
<td>55 - 55</td>
<td>6 - 169</td>
</tr>
<tr>
<td>malignant</td>
<td>n = 0</td>
</tr>
<tr>
<td></td>
<td>n = 69</td>
</tr>
<tr>
<td>deferred</td>
<td>141 (49)</td>
</tr>
<tr>
<td></td>
<td>92 - 190</td>
</tr>
</tbody>
</table>

# 4 missing values; * 1 missing value

diagnosis, of whom five turned out to be borderline and six turned out to be malignant. Thirty-two women had a borderline frozen section diagnosis, in which 17 were indeed borderline at paraffin diagnosis, nine were malignant and six were benign. All 44 women with malignancy at frozen section diagnosis had a malignancy at the final diagnosis. In women with a tumour ≥ 10 cm, the likelihood ratios for benign frozen section diagnosis, borderline frozen section diagnosis, and malignant frozen section diagnosis were 0.15 (95% CI 0.08 - 0.25), 4.9 (95% CI 2.1 - 11.4) and infinite (95% CI 9.7 - ∞), respectively.

In table 3 the histologic classification of the different diagnostic categories is shown. Nine out of 12 false negative diagnoses concerned a mucinous cystadenoma. One case concerned a Krukenberg tumour. In this case frozen section diagnosis showed a benign fibrothecoma, permanent paraffin diagnosis showed signet-ring cells which are referred to a Krukenberg tumour. Another case concerned a metastasis of a malignant melanoma. Frozen section diagnosis showed malignancy not specified.

We found 72 mucinous tumours and 185 non-mucinous tumours. The sensitivity and specificity for mucinous tumours (n = 72) were 73% (95% CI 55 - 87) and 95% (95% CI 83 - 100), whereas the sensitivity and specificity for non-mucinous tumours (n = 185) were 96% (95% CI 87 - 100) and 95% (95% CI 89 - 98). After stratifying for tumour size, we calculated a sensitivity of 69% (95% CI 49 - 85) in mucinous tumours ≥ 10 cm (n = 63) with a specificity of 94% (95% CI 80 - 99).
In table 4 the mean CA125 per diagnostic category is shown. The mean CA125 in the patients with a malignant paraffin diagnosis was 778 kU/l (SD 2086 kU/l, range 5 - 13000), whereas for women with benign and borderline paraffin diagnosis these levels were 59 and 57 kU/l, respectively. In women with a benign frozen section diagnosis, there were no differences between level of CA125 of those patients that had benign histology, borderline histology and malignant histology, i.e. 58, 84 and 22 kU/l, respectively (p-value 0.70). In women with a borderline diagnosis, there were no statistically significant differences between benign, borderline and malignant histology (p-value 0.13).

Discussion

Accurate differentiation between malignant and benign adnexal masses during surgery is of interest for subsequent surgical treatment. Frozen section diagnosis is a useful tool for differentiation between benign and malignant pathology during surgery. In a recently performed systematic review on the accuracy of frozen section diagnosis in the assessment of the adnexal mass the sensitivity of frozen section diagnosis was, though good, not perfect. In the aim of the present study was to examine whether the accuracy of frozen section diagnosis is affected by the size of the adnexal masses. The frozen section diagnoses were not reviewed. We wanted to examine the influence of tumour size on accuracy of frozen section diagnosis as it is performed in daily practice.

In this retrospective study, malignancy at frozen section diagnosis was always confirmed at the final histological assessment. In women with a tumour < 10 cm, there was only one false negative diagnosis in 50 women with a benign frozen section diagnosis, whereas there were 11 false negative diagnoses in 97 women with a tumour ≥ 10 cm. The corresponding likelihood ratios of a benign diagnosis for presence of malignancy were 0.15 for tumours ≥ 10 cm and 0.03 for tumours < 10 cm, respectively.

These findings have consequences for clinical management. In case a tumour has a diameter of more than 10 cm and the frozen section diagnosis is benign, still 11% (11 / 97) of the women turn out to have a malignant or borderline tumour. This means one out of nine patients needs further surgery despite of a negative frozen section diagnosis in case a tumour has a diameter of more than 10 cm.

Other studies confirm the influence of the size of adnexal masses on the accuracy of frozen section diagnosis. In a recent study of Tangjitgamol et al, in which 212 frozen section diagnoses were reviewed (35.9% malignant tumours), the authors found 18 cases of discordant diagnoses. No incorrect frozen section diagnoses were found in women with a mass less than 10 cm, while 47.6% of incorrect diagnoses occurred when masses were 20 cm or larger. Similarly Puls et al. reported in 294 frozen section
diagnoses a decrease in the sensitivity with an increase in the size of ovarian tumours. In tumours larger than 1360 gram, the sensitivity dropped to 69%.\textsuperscript{3}

The influence of the size of the adnexal mass on the accuracy of frozen section diagnosis can be explained by the nature of the procedure of frozen section diagnosis. One of the major problems of frozen section diagnosis is the difficulty of sampling. Because frozen section diagnosis should be a rapid procedure during surgery and the diagnosis is expected as soon as possible there is inadequate time to make as much slices as is performed in a paraffin procedure. Wang et al. reported on 792 consecutive gynaecological frozen section diagnoses, and found a correlation between frozen section numbers performed per case and the compatibility of the diagnoses in mucinous ovarian tumours.\textsuperscript{4} The authors suggest that at least one frozen section for every 10 cm in diameter of mucinous ovarian tumour should be recommended. Unfortunately, our data do not provide the number of slices that have been made during frozen section diagnosis, so we cannot show a correlation between number of slices that have been made, the size of the tumour and the accuracy of frozen section diagnosis.

It is plausible to presume that the level of experience of the pathologist who is performing the frozen section diagnosis will influence the accuracy of frozen section diagnosis. Also the quality of the frozen section slides will influence the accuracy of frozen section diagnosis. However our data do not provide enough information to examine these parameters.

Besides the size of the tumour, the histologic type can be a problem for frozen section procedures. In this study we found 12 false negative frozen section diagnoses. Five benign frozen section diagnoses appeared to be of borderline malignancy and seven benign frozen section diagnoses appeared to be malignant at final diagnosis. Nine out of these 12 false negative diagnoses concerned a mucinous cystadenoma. We found an overall sensitivity for mucinous tumours (n = 72) of 73%, which dropped to 69% in mucinous tumours ≥ 10 cm (n = 63). Forty-one mucinous tumours ≥ 10 cm showed benign frozen section diagnosis, nine of them were false negative. This means the risk of a false negative result and consequently a second operation is 23% in case of a mucinous tumours ≥ 10 cm.

Mucinous tumours are known as the largest of all ovarian tumours. In this study 93% (63 / 68) of the mucinous tumours were >10 cm. Mucinous tumours are multiloculated, cystic masses that contain thick mucinous material. Microscopically, the cysts are lined by a single layer of columnar cells with mucin-containing cytoplasm and small, basally located nuclei. Borderline mucinous tumours are characterized by papillary projections with cell stratification and mild to moderate atypia; destructive stromal invasion is absent. Extensive histologic sampling is required to exclude invasive carcinoma because these tumours may be very heterogeneous. Mucinous carcinoma may present as
a focal solid area in an otherwise cystic benign or borderline mucinous tumour. The histological characteristics of mucinous tumours make them less suitable for frozen section diagnosis.

The problem of false negative frozen section diagnosis in case of mucinous tumours is described in literature at several occasions. Wang et al. examined 23 borderline ovarian tumours, and found frozen section diagnosis in mucinous borderline tumours had a lower sensitivity than in non-mucinous borderline tumours. Pulls et al. describe the poorest accuracy and sensitivity in patients with largest mucinous tumours (n = 34). Rose et al. describes a sensitivity for borderline serous tumours of 64.3%, in contrast to a sensitivity for borderline mucinous tumours of 30.8%. Twaalhofen et al. compared 176 frozen section diagnosis to final paraffin diagnosis and found false negative results in 12 cases. In nine out of these 12 cases a tumour of mucinous histologic cell type was present and ten of these twelve tumours had a diameter of more than 10 cm. Our results are similar as those reported in literature.

In this study we also examined whether the level of CA125 affects the accuracy of frozen section diagnosis. CA125 is an antigen characteristic for fetal coelom. Serum CA125 is frequently used in diagnostics of epithelial ovarian cancer, although the sensitivity of the test is low. The application of serum CA125 seems to be more trustful during follow-up of (epithelial) ovarian cancer after surgery and/or chemotherapy. As expected we found a high level of serum CA125 in true positive frozen section diagnoses (mean 7778 kU/l). In the eight false negative frozen section diagnoses, which appeared to be malignant at final assessment the serum CA125 was much lower (mean 22 kU/l). The eight false negative frozen section diagnoses concerned five mucinous adenocarcinomas, two ovarian carcinomas not specified and one Krukenberg tumour. It is known that mucinous types of ovarian malignancy express less CA125 antigen, which explains the low mean level of CA125 in this group. In literature the CA125 level is not mentioned as influencing factor for accuracy of frozen section diagnosis.

A major issue encountered by gynaecological pathologists is accuracy of frozen section diagnosis of primary versus metastatic tumours. One of the false negative diagnoses concerned a Krukenberg tumour. In this case frozen section diagnosis showed a benign fibrothecoma, permanent paraffin diagnosis showed signet-ring cells which refer to a Krukenberg tumour.

Frozen section diagnosis appears to be an accurate procedure to differentiate between malignant and benign ovarian masses. Tumour size influences the accuracy of frozen section diagnosis. In adnexal masses ≥ 10 cm, a benign result of the frozen section diagnosis is less reliable than in women with a tumour < 10 cm. Mucinous tumours are troublesome in frozen section diagnosis because of its relative large size and its histological characteristics. The level of CA125 does not contribute to the
differentiation between benign and malignant mucinous tumours, because of its low expression in mucinous tumours. It would be useful if the pathologist is able to differentiate mucinous tumours from others tumours during frozen section diagnosis, as the non-malignant findings in mucinous tumours are less reliable.

References

Chapter 8

Patients' preferences in mode of surgery of an adnexal mass

Peggy M.A.J. Geomini
Gérard L. Bremer
Roy F.P.M. Kruitwagen
Brent C. Opmeer
Ben Willem J. Mol

J of Psychosom Obstet and Gynaecol, accepted.
Abstract

Objective: We assessed the preferences of women with an adnexal mass for the mode of surgery.

Methods: A structured interview was designed, in which women scheduled for surgery for an adnexal mass were confronted with fictive scenarios of the different approaches. Women were asked at what probability of a false negative test result of frozen section diagnosis they would prefer frozen section diagnosis over primary radical surgery. Furthermore, the women were asked at what probability of ovarian malignancy they would prefer laparoscopy over laparotomy.

Results: We interviewed 43 women. When the probability of frozen section diagnosis being false negative was set at 90%, 97% of the women preferred primary radical surgery. The mean threshold at which women switched their preference from primary radical surgery to frozen section diagnosis was at a risk of 49% on a false negative test result of frozen section diagnosis. In the choice between laparoscopy over laparotomy, the mean threshold at which the women switched their preference from laparoscopy to laparotomy was at a risk of 55% on ovarian malignancy.

Conclusion: Women scheduled for surgery of an adnexal mass at low risk of ovarian malignancy, prefer frozen section diagnosis over primary radical surgery and prefer laparoscopy over laparotomy.
Introduction

Accurate assessment of an adnexal mass prior to surgery facilitates an optimal choice for the type of surgery. Optimal surgical staging and cytoreductive surgery is of utmost importance in case of malignancy. In case of benign pathology, a laparoscopically performed cystectomy is usually sufficient. In patients in whom the presence of malignancy prior to surgery is uncertain, frozen section diagnosis might be of use. The technique of frozen section diagnosis has been known since 1891, but studies on its diagnostic accuracy have only been performed in the last two decades.\(^1\)

Recently, we performed a systematic review on the accuracy of frozen section diagnosis in the assessment of the adnexal mass.\(^2\) Meta-analysis of 18 studies showed that the specificity varied between 97% and 100%, whereas the sensitivity varied between 65% and 97%. The heterogeneity among reported sensitivities might be explained by tumour size, as false negative diagnoses especially occurred in women with an adnexal mass \(\geq 10\) cm.\(^3\) A false negative test result of frozen section diagnosis, i.e. the situation in which frozen section diagnosis is benign and paraffin diagnosis is malignant, implicates that the patient has to be operated a second time to perform optimal surgery. Apart from the risks of a second surgical procedure, a false negative frozen section diagnosis can also cause disappointment for the patient who was reassured after the first time. An alternative strategy could be to perform primary radical surgery in every patient operated for an adnexal mass without using frozen section diagnosis. This implicates that all patients with adnexal masses would undergo a median laparotomy with both sided salpingo-oophorectomy, hysterectomy and omentectomy, peritoneal biopsies and lymph node sampling.

Another point of discussion is whether to perform laparoscopy or laparotomy in case of an adnexal mass. Although laparoscopy provides advantages such as short hospital admission, small scars and quick resumption of daily activities, laparotomy is the preferred treatment in case of malignancy as the laparoscopic approach might be associated with cell seeding with possible negative consequences for the prognosis.

When a choice has to be made between two medical strategies, it is generally acknowledged that patients' preferences should be incorporated into medical decision making.\(^4\) Several studies have shown that patients' perspectives on the burden and benefits of therapy can differ from those of health professionals.\(^5\) Since we are not aware of studies that address patients' views on this subject, we performed a study to document patients' preferences for the mode of surgery of an adnexal mass.
Patients and methods

Women with an adnexal mass were invited to participate in the study after they were scheduled for surgery. This study was performed in Máxima Medical Centre (MMC), which is a large teaching hospital in the south of The Netherlands. As interviews were performed prior to surgery, the type of pathology (benign or malignant) was unknown at the time of the interview. Inclusion of patients depended upon willingness to participate in the study. A structured interview was designed to assess patients' preferences for primary radical surgery relative to applying the technique of frozen section diagnosis conditional on the risk of malignancy and the risk of a false negative frozen section diagnosis. Because of the extremely high specificity of frozen section diagnosis, we did not focus on the risk of a false positive diagnosis. Further, the interview focussed on patients' preference for laparoscopy relative to laparotomy in case of an adnexal mass. All interviews were performed by the same research nurse in an outpatient setting. Interviews were repeated six weeks after surgery. The study was approved by the institutional review board of MMC. All women received verbal and written information about the study. Women were interviewed after informed consent had been obtained.

First, women were asked whether they preferred to be informed extensively about diagnostic and therapeutic options to make their own decisions, or whether they preferred not to be informed and let the doctor decide what to do.

Second, women were informed about two possible alternatives. For primary radical surgery, we explained that both ovaries would be removed by median laparotomy even as the uterus and omentum. Optimal surgery also implied biopsies all over the peritoneal cavity and lymph node sampling. We pointed at the risks and the relatively long hospital stay associated with this extensive type of surgery. We also explained that this surgical strategy implies an iatrogenic climacterium praecox in case of premenopausal women.

For frozen section diagnosis, we explained that a pathologist judges during surgery whether an adnexal mass is benign or malignant. In case the pathologist judges the mass to be benign, surgery could be limited to removal of the adnexal mass. In case the mass is judged malignant, optimal surgery would be performed as described above. We informed the women that initial frozen section diagnosis is confirmed or invalidated by paraffin diagnosis, which result is available approximately one week after surgery. We pointed at the risk of a false negative frozen section diagnosis, i.e. the situation where frozen section diagnosis suggests a benign lesion, but subsequent paraffin diagnosis appears to be malignant. For this scenario, we emphasized the disappointment that a false negative frozen section diagnosis would generate for the patient one week after
initial surgery, as well as the consequence of a second surgical procedure. Women were asked whether they would prefer frozen section diagnosis over primary radical surgery in a scenario where there was a risk of 90% on a false negative test result of frozen section diagnosis. If the women preferred primary radical surgery, the probability of a false negative frozen section diagnosis was decreased in 10% steps until the women switched in their preference to frozen section diagnosis (threshold). We also asked the women to assess several advantages and disadvantages of primary radical surgery and frozen section diagnosis.

Third, we explained the women that adnexal masses can be operated by laparoscopy or by laparotomy. Although laparoscopy has advantages (short hospital admission, small scarves and quick resumption of daily activities), women might have a preference for laparotomy in case of malignancy, because of uncertain risks of a laparoscopic approach to malignancy with possible negative consequences regarding the prognosis. To determine the threshold, the women were asked whether they would prefer laparoscopy over laparotomy in a scenario with one percent risk on malignancy. If the women preferred laparoscopy, the probability of malignancy was increased in 10% steps until the women switched in their preference to laparotomy. Finally, the women were asked whether they would feel inconvenient about a laparotomic approach if a laparoscopic approach had been equally adequate and safe.

In the analysis, the cumulative percentage of women preferring frozen section diagnosis to primary radical surgery was plotted against the negative predictive value of frozen section diagnosis. Similarly, we plotted the cumulative percentage of women preferring laparotomy to laparoscopy against the probability of presence of a malignancy.

Results

Between April 2006 and May 2007, we interviewed 43 women. The mean age was 43 years (range 20 - 71 years). One woman did not have surgery because her adnexal mass disappeared spontaneously. From the 42 women who had surgery, 38 women appeared to have benign adnexal masses (11 teratomas, seven endometriomas, nine serous cystadenomas, six mucinous cystadenomas, six not further specified cysts, one hydatide cyst and one abscess). One woman had a myoma in stead of an adnexal mass and three women had a borderline malignancy.

Frozen section diagnosis was performed in three cases, and showed two borderline malignancies and one benign serous cystadenoma. Paraffin diagnoses confirmed the frozen section diagnoses in all three cases. All three patients had a median laparotomy. There were no false negative frozen section diagnoses in our study population. Among
the 42 women, 36 were operated by laparoscopy (20 cystectomies including one borderline malignancy, 15 ovarietctomies of which four both sided and one myomectomy) and six by laparotomy (all both sided ovarietctomies, including two borderline malignancies).

**Figure 1.**

- Amount of information women prefer to receive about advantages and disadvantages of diagnostic and therapeutic options.
- Amount of influence women prefer to have on diagnostic and therapeutic decisions of the doctor.

The amount of information about advantages and disadvantages of the operation and the amount of influence women want to have on diagnostic and therapeutic decisions of the doctor.

**Figure 2.**

Women's thresholds for frozen section diagnosis at variable probabilities of false negative frozen section diagnosis.
Patients' preferences in mode of surgery of an adnexal mass

Figure 1 illustrates the amount of information and amount of influence patients preferred to have on diagnostic and therapeutic procedures. More than half of the interviewed women wanted to be informed on advantages and disadvantages of diagnostic and therapeutic options as much as possible. Most women wanted the doctor taking decisions on diagnostic and therapeutic procedures taking into account the preferences of the patients as much as possible.

In a scenario where the risk of a false negative test result of frozen section diagnosis was 90%, i.e. the situation where nine out of ten women with a negative frozen section diagnosis would hear that malignancy was present, 97% of the women preferred primary radical surgery over frozen section diagnosis. The probability of a false negative frozen section diagnosis was decreased in 10% steps until these women switched their preference to frozen section diagnosis. At the moment the probability of a false negative test result was decreased to 50%, 80% of the women had switched their preference to frozen section diagnosis. Nine percent of the women still preferred primary radical surgery over frozen section diagnosis in a scenario with a risk of 10% on a false negative frozen section diagnosis. The mean threshold at which the women switched their preference from primary radical surgery to frozen section diagnosis was a risk of 49% on a false negative test result of frozen section diagnosis (figure 2).

Figure 3 illustrates women's assessment of (dis)advantages of primary radical surgery and frozen section diagnosis. Avoiding a climacterium praecox and certainty about the definitive test result appeared to be very important for most of the interviewed women, while a second operation was not a problem for most women. The length of hospital admission or the size of a scar were factors of less importance to women.

Finally, we examined whether women preferred a laparoscopic or a laparotomic approach in case of an adnexal mass. In a scenario with one percent risk on malignancy of an adnexal mass, 98% of the women preferred laparoscopy to laparotomy. The probability of malignancy was increased in 10% steps until these women switched their preference to laparotomy. At a probability of malignancy of 50%, 62% of the women had switched their preference to laparotomy. Eleven percent of the women still preferred laparoscopy to laparotomy in a scenario with a risk of 90% on malignancy. Figure 4 shows the thresholds at which the women switched their preference from laparoscopy to laparotomy (mean threshold was at a risk of 55% on ovarian malignancy). Eighty-four percent (36 / 43) of the interviewed women revealed they would feel inconvenient about a laparotomic approach if a laparoscopic approach had been even adequate and safe.

Six weeks after surgery the women were interviewed a second time (n = 33). As the answers did not differ from the first interview, we do not present them here.
Figure 3.

Women's valuation on (dis)advantages of primary surgery and frozen section diagnosis.

Figure 4.

Women's thresholds for laparotomy at variable probabilities of ovarian malignancy.
Discussion

In this study, we assessed preferences of patients with an adnexal mass for the mode of surgery. We interviewed 43 patients with an adnexal mass and informed these women about the procedure of frozen section diagnosis and about the possibility of a laparoscopic approach. In realistic scenarios, women preferred laparoscopy over laparotomy and frozen section diagnosis over primary radical surgery. On average women accepted a risk of almost 50% on a false negative test result of frozen section diagnosis. Most women preferred a laparoscopic approach to a laparotomic approach if possible, one in five even in scenarios where the probability of malignancy exceeds 75%. A possible bias in this study is the relative low mean age of 43 years of the study population (43 years) with only 11 women being 50 years or older.

In a meta-analysis reviewing studies on accuracy of frozen section diagnosis, we found the sensitivity of frozen section diagnosis to vary between 65% and 97%, and the specificity varying between 97% and 100%.

In case the prevalence of malignancy is 20%, such sensitivity and specificity results in a negative predictive value of frozen section diagnosis of 92% (assuming a sensitivity of 65% and a specificity of 97%). From these data, we conclude that women prefer frozen section diagnosis over primary radical surgery as long as there is no 100% certainty of malignancy prior to surgery. Even in case of a prevalence of malignancy of 70%, the negative predictive value of frozen section diagnosis will be more than 50% (i.e. 55%). Women scheduled for surgery for an adnexal mass should be informed about the possibility of frozen section diagnosis if there is any (pre-operative) suspicion on malignancy, rather than to perform primary radical surgery without any evidence of malignancy.

We do not think frozen section diagnosis needs to be performed in all women with adnexal masses. In the current study, we performed frozen section diagnosis in 7% of 42. To discriminate between malignant and benign masses pre-operatively prediction models can be applied. The Risk of Malignancy Index (RMI) as originally described by Jacobs et al. is an extensionally external validated prediction model with a good sensitivity and specificity. A strength of the model is its simplicity. The model uses the product of the serum CA125 level (U/ml), the ultrasound scan result (expressed as a score 0, 1 or 3) and the menopausal state (1 if premenopausal and 3 if postmenopausal). In the original publication of Jacobs et al. the likelihood ratio for malignancy of patients predicted to have benign and malignant disease by an RMI of 200 was 0.2 and 42, respectively. In case of a RMI above 200, frozen section diagnosis should be considered during surgery, especially in case of premenopausal women.

Adnexal masses can be removed laparoscopically as well as by laparotomy. In case of benign adnexal masses, a laparoscopic approach is clinically preferable to a
laparotomic approach because of less postoperative complications like fever or bladder retention, less postoperative pain and shorter admission time. In case of malignant masses, evidence is not unambiguous whether a laparoscopic or a laparotomic approach is preferable. Abdominal wall metastases at trocar sites following laparoscopy have been described, although their impact on survival is not clear. The negative effects of carbon dioxide on the peritoneum with risk on tumour dissemination has been described in animal studies, but until now there is no clinical evidence for this problem in case of ovarian malignancy. The Dutch guidelines advocate laparotomy in case of high risk on ovarian malignancy as long as possible negative effects of laparoscopy in case of ovarian malignancy cannot be ruled out. In contrast to the preference of the interviewed women to have a laparoscopy in case of a risk of 55% on malignancy, in practice, a gynaecologist will probably prefer to perform a laparotomy in case of high risk on malignancy. However, we are not aware of any studies considering the preference of a gynaecologist.

We propound the dilemma of laparoscopy versus laparotomy to patients scheduled for surgery because of an adnexal mass. These patients preferred laparoscopy to laparotomy, in some cases even up to 90% risk of malignancy. We are aware of the fact we interviewed a low risk population on ovarian malignancy, in which less than 10% of the women suffered a borderline malignancy of the ovary and none of the interviewed women had an invasive ovarian malignancy. It is possible that a high-risk population will have different preferences. In a retrospective study on vulva carcinoma, women with proven malignancy did not want to take any risk and preferred laparotomy to laparoscopy.

Furthermore, the interviews took place after scheduling for surgery, after the patient had been counselled by the surgeon. As a consequence, the preference of the interviewed women could be influenced by the meaning of the surgeon. The fact that answers before surgery and six weeks after surgery did not differ might be explained by the fact that no invasive ovarian malignancies were diagnosed in our study sample.

In conclusion, women with an adnexal mass prefer frozen section diagnosis over primary radical surgery and accept a certain risk of a false negative result of frozen section diagnosis. Women with low risk on malignancy prefer laparoscopy over laparotomy.
Patients’ preferences in mode of surgery of an adnexal mass

References

11. Kruitwagen RF, Swinkels BM, Keyser KG, Doesburg WH, Schijff CP. Incidence and effect on survival of abdominal wall metastases at trocar or puncture sites following laparoscopy or paracentesis in women with ovarian cancer. Gynecol Oncol. 1996; 60:233-7.
Chapter 9

Should we centralize care for the patient suspected of having ovarian malignancy?

Peggy M.A.J. Geomini
Roy F.P.M. Kruitwagen
Gérard L. Bremer
Leon Massuger
Ben Willem J. Mol

submitted
Chapter 9

Abstract

Objective: Outcome of ovarian cancer is better when surgery is provided by a gynaecological oncologist than by a general gynaecologist. However, when all patients with an adnexal mass have to be operated by gynaecological oncologists, this requires a change in the organization of care, which generates additional costs. In this study, we assess the implications of centralized and regular care for women with an ovarian malignancy in The Netherlands.

Methods: We performed a cost-effectiveness analysis in which we integrated epidemiologic data, data on the survival of ovarian cancer patients in relation to surgical modality and estimates of diagnostic test accuracy to assess malignancy of the adnexal mass prior to surgery. We considered three strategies. In the first strategy, patients were operated by a general gynaecologist (general care strategy). In the second strategy, patients were operated by a gynaecological oncologist (specialized care strategy). In the third strategy, evaluation of the adnexal mass took place prior to surgery by means of the Risk of Malignancy Index (diagnostic strategy). Patients at high risk for malignancy were supposed to be operated in a specialized care setting, whereas low risk patients were supposed to be operated in a general care setting. All patients were supposed to receive adjuvant chemotherapy, except patients with stage I disease who had been operated by a gynaecological oncologist. For each strategy we calculated life expectancy and incremental costs per life year gained (LYG).

Results: Life expectancy of a patient with an ovarian malignancy in the general strategy was 2.7 years, in the diagnostic strategy 3.0 years and in the specialized strategy 3.1 years. The diagnostic strategy appeared to be dominant over the general strategy. The incremental costs to gain one additional life year with specialized surgery as compared to the diagnostic strategy were € 63,477 per LYG.

Discussion: In women with an adnexal mass, a diagnostic strategy prior to the decision for surgery by a general gynaecologist or a gynaecological oncologist provides the best balance between costs and effects.
Centralization of care for patient suspected of ovarian malignancy

Introduction

In The Netherlands, the treatment of ovarian cancer is not centralized. Most patients with ovarian cancer are treated by general gynaecologists with a special interest in gynaecologic oncology or gynaecological oncologists. About 16% of the patients are treated in academic hospitals. The remaining patients are operated in a general (teaching) hospital and depended on regional agreements, a gynaecological oncologist can be invited to operate the patient on site. There is at present debate about the necessity for centralization of care for patients with ovarian cancer. In a recent review on the relationship between care setting and care outcomes in the treatment of ovarian cancer, the outcome of ovarian cancer was estimated better when treatment was provided by a gynaecological oncologist or in a specialized hospital.\textsuperscript{1} Long-term survival was better after treatment in a specialized hospital by a gynaecological oncologist. Such an approach led to a five to eight month median survival benefit for patients with advanced stage disease. In early stage disease, gynaecological oncologists staged cancer more often adequate than general gynaecologists did.

The observed survival for FIGO I patients that are adequately staged is better than the observed survival in inadequately staged patients, but this observation is probably based on stage migration from stage I to higher stages after optimal staging. So far, no evidence is found that adequate staging surgery itself leads to a survival benefit for patients with early stage disease, as long as inadequately staged patients classified as stage I obtain adjuvant chemotherapy.\textsuperscript{2,3}

According to data of the Integral Cancer Centre in The Netherlands, in 2003 epithelial ovarian cancer was diagnosed in almost 1,100 women. About 60% of the cases of ovarian cancer were diagnosed in an advanced stage of disease, indicating that in 2003 approximately 700 women with an ovarian malignancy stage III or IV needed debulking. The remaining 40% with stage I and stage II needed only salpingo-ovariectomy (stage I) respectively salpingo-ovaricectomy and hysterectomy (stage II), under the condition that they had been staged optimally (n = 417 in 2003).

The number of women that had surgery because of a benign adnexal mass was much higher. Assuming an overall prevalence of 10 - 15% of malignancy in adnexal masses, in The Netherlands about 6,500 to 9,000 women annually have surgery because of a benign adnexal mass. Since the pre-operative assessment of an adnexal mass is difficult, it is inevitable that when one aims that gynaecological oncologists should operate all women with an ovarian malignancy, those gynaecological oncologists will also operate a part of the women with benign adnexal masses. However, when all women with an adnexal mass would be operated by a gynaecological oncologist, this would probably lead to high costs. Below, we assess the implications of centralized and regular care for
women with an ovarian malignancy in The Netherlands, and compare it with a strategy in which risk assessment with the Risk of Malignancy Index is performed prior to surgery.

Methods

We compared the costs and effects of three strategies. In the first strategy, patients were operated by a general gynaecologist in a general hospital (general care setting). In the second strategy, patients were operated by a gynaecological oncologist whether or not in a specialized centre (specialized care setting). In the third strategy, systematic diagnostic evaluation of the adnexal mass took place prior to surgery. Patients with a high suspicion of malignancy were supposed to be operated in a specialized setting, whereas patients with a low suspicion of malignancy were supposed to be operated in a general setting (diagnostic strategy).

We combined epidemiologic data from Dutch national registries with estimates of diagnostic test accuracy of tests that are used to distinguish women with a malignant adnexal mass from women with a benign adnexal mass and data on the survival of ovarian cancer patients in relation to the way they were operated. In The Netherlands, since 1971 a nationwide network of histo- and cytopathology (PALGA) has been active, in which reports generated by all 70 pathology departments are registered. We requested from PALGA the number of pathology reports on benign, borderline and malignant adnexal masses in 2003. Reports on borderline malignancy and non-epithelial ovarian malignancy were excluded. For the count of reports on benign pathology, we only included reports in case the report described an enlarged adnex. Reports on normal ovary in case of prophylactic ovariectomy because of BRCA gene mutations or mama carcinoma were excluded. In addition, reports on ovaries which were removed because of an endometrial carcinoma, were excluded. Finally, reports on premenopausal women with simple cysts (unilocular transone cysts without papillary formations and absence of ascites) were excluded, because a gynaecological oncologist will never operate them due to a negligible risk on malignancy. Because these data were not reducible from the PALGA dataset, we had to estimate the part of reports on premenopausal women with simple cysts. Based on a validity study for the use of threedimensional ultrasound with consecutive series of women with an adnexal mass, we estimated the part of reports on premenopausal women with simple cysts on 5%.

Data on survival were obtained from studies that reported on the impact of the type of surgeon, i.e. a gynaecological oncologist or a general gynaecologist, on survival of ovarian cancer patients. We supposed median survival in early stage disease to be equal in the general care and specialized care setting on condition that all patients with stage
I disease operated in a general care setting received adjuvant chemotherapy. Patients with stage I disease operated in the specialized care setting did not receive adjuvant chemotherapy. Mean survival for stage I/II was set on 60 months for both settings.

Based on studies reporting on the impact of the type of surgeon, i.e. a gynaecological oncologist or a general gynaecologist we also estimated survival in advanced stage disease. Mean survival for stage III/IV was estimated 15 months in the general care setting and 22 months in the specialized care setting. Table 1 shows the estimated survival time in relation to stage of disease. Since we anticipated no difference in survival between women with benign disease being operated by a general gynaecologist or a gynaecological oncologist, the survival of women with benign disease was not taken into account.

The analysis was performed for all women that had surgery for an adnexal mass in The Netherlands. We first calculated the life expectancy in the first two strategies, i.e. the general care strategy and the specialized care strategy. Life expectancy was calculated per stage by multiplying the absolute numbers of patients with the mean expected survival expressed in years.

For the third strategy, we considered the use of a diagnostic test in which patients scheduled for surgery for an adnexal mass were classified as being at low risk or at high risk for a malignancy (diagnostic strategy). Testing was performed with the risk of malignancy index (RMI). The RMI was first described by Jacobs et al. and has been validated in several studies. In the baseline analysis we considered the sensitivity of the RMI to be 78% for a specificity of 87% (chapter 2, the accuracy of risk scores in predicting ovarian malignancy; a systematic review). Based on these test characteristics, we calculated for the Dutch population of patients operated for an adnexal mass in 2003 the number of patients that were correctly classified as having a

Table 1. Expected mean survival in months

<table>
<thead>
<tr>
<th>Stage</th>
<th>Specialized care setting</th>
<th>General care setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>60 *</td>
<td>60 **</td>
</tr>
<tr>
<td>II***</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>III***</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>IV***</td>
<td>22</td>
<td>15</td>
</tr>
</tbody>
</table>

All patients receive adjuvant chemotherapy, except patients with stage I disease operated by a gynaecological oncologist.
* no adjuvant chemotherapy
** all patients received adjuvant chemotherapy
*** all patients with stage II, III and IV disease received adjuvant chemotherapy
malignancy (true positives), the number of patients that were incorrectly classified as having a malignancy (false positives), the number of patients correctly classified as having benign disease (true negatives), and the number of patients incorrectly classified as having benign disease (false negatives). Patients with a positive diagnosis, i.e. the true positives and the false positives were supposed to be operated in a specialized care setting by a gynaecological oncologist, whereas patients with a negative diagnosis, i.e. the true negatives and the false negatives were supposed to be operated in a general care setting by a general gynaecologist. We calculated the life expectancy for the third strategy, i.e. a diagnostic evaluation of the adnexal mass, followed by specialized surgery in case of a high suspicion of malignancy, and otherwise surgery in a general care setting. As life expectancy in women with a benign mass was considered not to be affected by the type of surgery, we did not consider the life expectancy of women with benign disease in the analysis.

The economic analysis was performed from a societal perspective. The time horizon was limited to 5 years. Discounting for life years and costs in the future was not applied. Table 2 shows the costs that we used. Since 2005, wages of doctors in The Netherlands are based on diagnostic-treatment combinations, which represent a fixed total price (doctor's costs) for treatment, examination, admission and surgery for every separate diagnosis. Doctors' costs of treatment of benign adnexal masses are less than costs of treatment of malignant adnexal masses (€ 554 versus € 837). In case surgery was

<table>
<thead>
<tr>
<th>Item</th>
<th>Amount per patient</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-operative assessment of an adnexal mass</td>
<td>€ 110</td>
<td>doctor's costs of the diagnostic-treatment combination in case a patient with an adnexal mass is treated conservative</td>
</tr>
<tr>
<td>treatment of benign adnexal mass</td>
<td>€ 554</td>
<td>doctors' costs for diagnosis and surgical treatment</td>
</tr>
<tr>
<td>treatment of malignant adnexal mass</td>
<td>€ 837</td>
<td>doctors' costs for diagnosis and surgical treatment</td>
</tr>
<tr>
<td>additional costs in case treatment by a gynaecological oncologist</td>
<td>€ 1,275</td>
<td>based on the basic salary of a gynaecological oncologist in relation to the number of patients he treats annually</td>
</tr>
<tr>
<td>adjuvant chemotherapy (six cycles)</td>
<td>€ 14,177</td>
<td>• chemothepaetics (cisplatin 175 mg/m²/cycle and paclitaxel 75 mg/m²/cycle): € 1917 per cycle* &lt;br&gt;• admission one day care, one per cycle: € 375 per cycle &lt;br&gt;• doctors' costs: € 425 per diagnosis</td>
</tr>
</tbody>
</table>
performed in presence of a gynaecological oncologist, additional costs were calculated. These costs were based on the basic salary of a gynaecological oncologist in relation to the number of patients he treats annually (€ 1,275 per procedure). In the diagnostic procedure, we calculated costs for pre-operative assessment of the adnexal mass. We calculated € 110 per adnexal mass, which represents the doctors' costs of the diagnostic-treatment combination in case a patient with an adnexal mass is treated conservative and not scheduled for surgery. Hospital costs and non-medical costs were considered to be similar in both groups. Costs of six cycle's adjuvant chemotherapy were considered to be € 14,177 per patient, which included chemotherapeutics (based on doses using weight and height of average Dutch women with a weight of 78 kg and a height 173 cm), hospital admission and doctors' costs. A correction for discounting was not applied.

Surgery in a general care setting was considered to be the reference strategy. We calculated the incremental costs to gain one additional life year (LYG) for specialized surgery and for a diagnostic strategy with risk assessment prior to surgery. Sensitivity analysis was performed for the most important variables, i.e. diagnostic accuracy, costs of diagnostic tests and treatments, prevalence of malignancy and life expectancy. Moreover, we analyzed a variant in which none of the patients with stage I disease received adjuvant chemotherapy. In this scenario, survival in early stage disease was supposed to be less in the general setting compared to the specialized setting. This difference was due to the fact that about 15% of the patients that are operated in a general setting, i.e. without lymphadenectomy, were classified incorrectly as stage I disease. Due to undetected positive lymph nodes, the mean survival in these patients would be conform patients in stage III. Consequently, for 15% of patients with stage I disease operated in the general care setting, mean survival was supposed to be 15 months, whereas for the other 85% of the patients with stage I disease operated in the general care setting, mean survival was supposed to be 60 months. In contrast, in all patients with stage I disease operated in the specialized care setting, mean survival was set on 60 months. Mean survival for patients with stage II disease was set on 60 months.

Results

In the PALGA database, we found all 1,096 reports of the epithelial ovarian malignancies diagnosed in 2003 in The Netherlands, of which 179 were diagnosed in academic centres and the other 917 in non-academic centres. We found 6,844 reports on enlarged ovaries with benign pathology. We included 6,502 reports on benign pathology, of which the reports on simple cysts in premenopausal women were excluded (estimated on 5%). Consequently, we calculated a prevalence of epithelial
<table>
<thead>
<tr>
<th></th>
<th>Specialized strategy</th>
<th>General strategy</th>
<th>Diagnostic strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TP*</td>
<td>FN*</td>
<td>FP*</td>
</tr>
<tr>
<td>Stage I</td>
<td>239</td>
<td>68</td>
<td>164</td>
</tr>
<tr>
<td>Stage II</td>
<td>86</td>
<td>24</td>
<td>59</td>
</tr>
<tr>
<td>Stage III</td>
<td>385</td>
<td>108</td>
<td>263</td>
</tr>
<tr>
<td>Stage IV</td>
<td>145</td>
<td>41</td>
<td>99</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Survival (years) and costs in 1000 euros per patient:

- Specialized strategy: 3.1 years, 3342 euros
- General strategy: 2.7 years, 2639 euros
- Diagnostic strategy: 3.0 years, 2544 euros

All patients receive adjuvant chemotherapy, except patients with stage I disease operated by a gynaecological oncologist.

*Diagnostic strategy using the Risk of Malignancy Index with sensitivity 78% and specificity 87%.

*TP = true positives; FN = true negatives; FP = false positives; TN = true negatives.
ovarian malignancy of 14% (1096 / 7598). According to data of the Dutch Integral Cancer Centers 28% of the ovarian malignancies was diagnosed in stage I, 10% in stage II, 45% in stage III and 17% in stage IV. The calculations were performed for this cohort of 7,598 women.

Table 3 shows life expectancy and costs for the three strategies. The mean life expectancy of patients with an ovarian malignancy would be 2.7 years per patient for a cost of € 2,639 per patient (doctors' costs and costs of adjuvant chemotherapy) in the general strategy. Life expectancy was 3.0 years per patient for a cost of € 2,544 per patient in the diagnostic strategy, making this strategy dominant over the general strategy. In the specialized strategy, life expectancy increased to 3.1 years per patient for a cost of € 3,342 per patient. The incremental costs to gain one additional life year (LYG) for specialized care strategy as compared to the general strategy were € 12,298. Figure 1 shows the costs in relation to the life years gained for each of the three strategies.

As the costs in the diagnostic strategy were less than the costs in the general setting, whereas life expectancy was better, we could not calculate the incremental costs to gain one additional life year for the diagnostic strategy with risk assessment prior to surgery. The incremental costs per LYG for the specialized care strategy as compared to the diagnostic strategy were € 63,477.

**Figure 1.** Plot showing the expected survival and expected costs in a cohort of 7,598 women operated for an adnexal mass.
Figure 2A. *Plot showing the cost-effectiveness of the diagnostic strategy as compared to general care for different combinations of sensitivity and specificity. The black dot represents the sensitivity - specificity combination of the RMI I.*

In the baseline analysis, the RMI had a sensitivity of 78% for a specificity of 87%. Figure 2 shows the impact of variation of the test accuracy on the cost-effectiveness of the diagnostic strategy. Figure 2A shows that the diagnostic strategy is dominant over general care at realistic combinations of sensitivity and specificity. Moreover, at virtually every combination of sensitivity and specificity, the costs of the diagnostic strategy to gain one additional life year are below €10,000. Figure 2B shows that the specialized care strategy costs much more than €50,000 to save one additional life year as compared to the diagnostic strategy, as long as the sensitivity of the diagnostic test that is used is high enough. However, when the sensitivity drops below 60%, the cost-effectiveness of the specialized strategy drops below the €50,000 to gain one additional life year.
Figure 2B. Plot showing the cost-effectiveness of the specialized care as compared to the diagnostic strategy for different combinations of sensitivity and specificity. The black dot represents the sensitivity - specificity combination of the RMI I.

Cost-effectiveness of specialized strategy compared to diagnostic strategy
for different combinations of sensitivity and specificity
(cost in € 1000 to gain one additional life year)

Further sensitivity analysis showed that the prevalence of malignancy had to increase to 58%, in order to make the specialized strategy dominant over the diagnostic strategy. At a prevalence of 17.3%, the cost of the specialized strategy as compared to the diagnostic strategy were € 50,000 to gain one additional life year. The diagnostic strategy was dominant over general care as long as the prevalence of malignancy was above 10%.

At the baseline prevalence of 14%, the diagnostic strategy stayed dominant over the general strategy as long as the costs of diagnostics were less than € 205.

In a scenario in which patients with stage I disease did not receive adjuvant chemotherapy and consequently survival was lower for patients with stage I disease operated by a general gynaecologist, the incremental costs for specialized strategy and for the diagnostic strategy were € 15,939 per LYG and € 408 per LYG, respectively.

In comparison to the diagnostic strategy, the incremental costs to gain one additional life year in the specialized strategy were € 52,579 per LYG.
Discussion

We assessed the implications of specialized and regular care for women with an ovarian malignancy in The Netherlands, and compared it with a strategy in which risk assessment was performed using the Risk of Malignancy Index before surgery. Life expectancy of patients with an ovarian malignancy in the general strategy was 2.7 years per patient, in diagnostic strategy 3.0 years per patient and in the specialized strategy 3.1 years per patient. The diagnostic strategy appeared to be dominant to the general strategy. Compared to the general strategy, the incremental costs to gain one additional life year (LYG) for specialized surgery were €12,298. Compared to the diagnostic strategy, the incremental costs to gain one additional life year for specialized surgery were €63,477.

In this analysis, we performed an economic analysis based on a cohort of women who were primarily diagnosed for ovarian malignancy in 2003. Because we did not know the real costs and survival data of this cohort, we had to make some assumptions. To estimate survival time of the 1096 women with ovarian malignancy we used data from recent literature.\textsuperscript{1,5-7} We assumed survival of early-stage patients (stage I and II) respectively advanced-stage patients (stage III and IV) to be the same. We did not differentiate between substages (Ia, Ib, Ic, IIa, etc). Furthermore, we supposed that surgery in case of benign adnexal masses had no influence on survival. As small differences beyond the five-year survival would have strong implications of the outcome of our analysis, we limited it to five years. We used mean survival in stead of median survival, as the use of the mean facilitated the calculations.

Since 2005, tariffs of hospital care in The Netherlands are based on diagnostic-treatment combinations which represent a fixed total price for treatment, examination, admission and surgery for every separate diagnosis. In this economic analysis, doctors’ costs of surgery of ovarian cancer were based on the price of the diagnostic-treatment combination "ovarian cancer treated by extensive surgery", which can be declared by every gynaecologist. Doctors’ costs of surgery of a benign adnexal mass were based on the price of the diagnostic-treatment combination "benign adnexal mass removed surgically". To express the extra costs in case patients were operated in presence of a gynaecological oncologist, we calculated an the additional cost of a gynaecological oncologist as €1,275 per procedure. The costs of chemotherapy were based on the fixed price of the diagnostic-treatment combination "adjuvant chemotherapy in case of ovarian malignancy" plus the real costs of the chemotherapeutic medication and the costs of six day-time admissions. Contingent prices for side effects or complications were not calculated.

However, despite of the uncertainties of some assumptions included in this economic
Centralization of care for patient suspected of ovarian malignancy

analysis, we illustrated clearly that costs in a general strategy (all patients with an adnexal mass operated by a general gynaecologist) were low compared to costs in a specialized strategy (all patients with an adnexal mass operated by an gynaecological oncologist) detrimental to survival. On the other side, treatment of all adnexal masses in specialized hospitals by gynaecological oncologists would be very expensive.

Bristow et all performed a cost-effectiveness analysis for centralization of care for patients with advanced stage ovarian cancer and concluded that centralized referral of patients with ovarian cancer was a cost effective healthcare strategy ($3,800 per QALY)\(^\text{10}\). In this study the difficulty of assessment of an adnexal mass pre-operatively, whether it is benign or malignant, is passed even as the difficulty of pre-operative assessment of the stage of disease. In contrast to the study of Bristow, which was based on a hypothetical model, Grevig et al performed a cost-effectiveness analysis for centralization of ovarian cancer based on actual patient data.\(^\text{11}\) Grevig found treatment of ovarian cancer patients (early stage and advanced stage disease) in semi specialized hospital settings was the most cost-effective strategy. Treatment of ovarian cancer in semi-specialized hospitals yielded 0.15 QALY compared to treatment in general hospitals, resulting in an incremental cost-effectiveness ratio of €5,700. Treatment of patients in tertiary referral hospitals resulted in an incremental cost-effectiveness ratio of €93,700. If the optimal debulking rate in specialized hospitals would increase from 46% to 70%, ovarian care in these hospitals also would become a cost-effective strategy (incremental cost-effectiveness ratio of €25,000). However, in this study, costs of unjustified referrals of patients with a benign mass to specialized hospitals (false positive pre-operative assessment of an adnexal mass), and loss of survival years because of unjustified not referred patients with malignant adnexal masses (false negative pre-operative assessment of an adnexal mass), were not included.

In our analysis, a diagnostic approach as triage between treatment by a general gynaecologist or a specialized gynaecological oncologist was very cost-effective. Several diagnostic models for assessment of an adnexal mass have been described in literature, some of these have been validated extensively.\(^\text{8,12,13}\) One the most extensive validated models is the Risk of Malignancy Index originally described by Jacobs and modified by Tinguistad.\(^\text{8,13,14}\) In a systematic review, we assessed accuracy of both models and we calculated a pooled estimate for sensitivity of 78% for a specificity of 87%. We calculated that the diagnostic strategy is cost-effective in case the applied test, for a specificity of 87%, has a sensitivity of > 64%

We assessed the implications of centralized and regular care for women with an ovarian malignancy in The Netherlands and showed centralized care is cost-effective on condition that pre-operative assessment of an adnexal mass has been performed
References


Considerations, summary, clinical implications and future research

Beschouwingen, samenvatting, klinische implicaties en toekomstig onderzoek
Considerations and summary

Adnexal masses are a common reason for visiting a gynaecologist. Mostly, adnexal masses do not cause serious disturbances to the patient. The main problem in the clinical management of the adnexal mass is its risk of malignancy. Because of this risk of malignancy, adnexal masses have to be assessed carefully. Treatment of a malignant adnexal mass is completely different from that of a benign adnexal mass.

In case of early-stage ovarian malignancy, optimal staging is of utmost importance, because optimal staging serves to identify those patients who can be followed without adjuvant chemotherapy. In case of advanced-stage ovarian malignancy, the standard of care is primary surgical cytoreduction aiming at complete, or at least optimal debulking (i.e., residual tumor one centimetre or less in diameter), followed by intravenous administration of taxane- and platinum-based chemotherapy. Several studies have suggested that surgery performed by a gynaecological oncologist improved survival of both early-stage and advanced-stage ovarian malignancy as compared to surgery performed by a general gynaecologist.

In case of a benign adnexal mass ovarian tissue can be spared by performing a cystectomy instead of an adnectomy. Moreover, a laparoscopic approach is often preferable to a laparotomic approach.

Because of therapeutic consequences, it is important to assess the risk of ovarian malignancy both pre-operatively and during surgery. For the differentiation between benign and malignant ovarian masses, many diagnostic parameters are available. In the pre-operative setting among these are patients' characteristics, serum tumour markers, ultrasonography, computed tomography and magnetic resonance imaging. In an attempt to improve the accuracy of the pre-operative assessment of adnexal masses, most of these parameters have been combined in diagnostic models. A potentially new diagnostic tool in the assessment of the adnexal mass is threedimensional (3D) ultrasonography. This technique visualizes the adnexal mass in all three planes (longitudinal, transverse and coronal), with computer assisted 3D reconstruction allowing for spatial evaluation as opposed to 2D ultrasonography, in which only two planes can be visualized.

During surgery, an additional diagnostic procedure for the assessment of the adnexal mass is frozen section surgical pathology.

This thesis deals with the diagnosis of malignancy in adnexal masses, both pre-operatively as well as during surgery.
How accurate are existing prediction models for the pre-operative assessment of adnexal masses, and are these models more accurate compared to clinical estimates?

Chapter 2 presents a review of existing prediction models. The risk of malignancy of ovarian masses can be assessed pre-operatively with prediction models, in which multiple factors are included. We performed a systematic review of the literature on the accuracy of prediction models in the pre-operative assessment of the adnexal mass. Studies were identified through the Medline databases and EMBASE from inception to March 2008. There were 109 accuracy studies reporting on 21,750 adnexal masses (5,826 malignancies), that met the selection criteria. These studies reported on 83 different prediction models. We differentiated four types of prediction models. First, descriptive numerical prediction models, which are based on a weighted sum of morphologic characteristics of an adnexal mass assessed by ultrasonography. The model of Sassone is the most frequent validated example of such a scoring system. At a cut-off point of nine, pooled estimation for sensitivity was 84% for a specificity of 80%. A second group of prediction models is based on logistic regression analysis. Logistic regression analysis is a mathematical procedure for systematically developing an algorithm to assess the probability that a particular event has occurred. The formula of the Risk of Malignancy Index (RMI), originally described by Jacobs (RMI I) and modified by Tingulstad (RMI II and III), is a simplification of a formula provided by logistic regression analysis. The RMI uses the product of three variables: ultrasound score, menopausal state and absolute value of serum CA125. The score of the RMI varies between zero and infinity for RMI I and between one and infinity for RMI II and III. The cut-off level most frequently applied for all three RMI's is 200. When 200 was used as cut-off level, the pooled estimate for sensitivity was 78% (RMI I) and 79% (RMI II) for a specificity of 87% (RMI I) and 81% (RMI II) respectively. A third type of prediction models is the artificial neural network, which distinguishes itself from logistic regression analysis by its possibility to investigate the possible existence of non-linear interaction between variables. A fourth type of mathematical models are least squares support vector machines (LS-SVM) and relevance vector machines (RVM), which are very complex models. Experience with these models is very limited.

All described prediction models had acceptable sensitivity and specificity. However, the Risk of Malignancy Index (RMI I and II) appeared to be the best predictors.

Since the outcome of ovarian cancer is better when a gynaecological oncologist provides treatment than a general gynaecologist, patients suspected for having ovarian cancer should be referred to a gynaecological oncologist. Bailey et al. examined whether RMI can be used for adequate referral of ovarian cancer to a tertiary center. Bailey found a positive predictive value of 86.8%.
In conclusion, there are several validated prediction models available for the pre-operative assessment of the adnexal mass. All validated prediction models have an acceptable sensitivity and specificity. The RMI I as described by Jacobs is a simple scoring system, appears to be very accurate and is therefore useful in clinical practice.\(^9,13\) Whether the RMI is adequate for referring patients suspected for ovarian malignancy, is subject of further research.

Chapter 3 compares the clinical judgment of gynaecologists of the predictive performance of existing prediction models in the pre-operative assessment of the adnexal mass. From a dataset of 170 cases, we randomly selected 34 cases to form five different sets. We offered each set to 15 gynaecological oncologists, 15 general gynaecologists and 15 senior residents. Each case consisted of data on female age and menopausal status, a written description of the transvaginal ultrasonography, transvaginal Doppler flow measurement and serum CA125 measurement. The observers were asked to express their perception on the chance that the adnexal mass was malignant. We assessed the accuracy of the judgment of the clinicians by Receiver Operating Characteristic (ROC) analyses, in which the test result, i.e. the risk of malignancy as estimated by the gynaecologists, was compared to the presence or absence of malignancy, as established at histological examination by a pathologist. We constructed a ROC-curve, and calculated an area under the ROC-curve. The areas under the ROC-curve were 0.82 (95% CI 0.78 - 0.86), 0.84 (95% CI 0.80 - 0.86) and 0.82 (95% CI 0.78 - 0.86) for gynaecological oncologists, general gynaecologists and residents, respectively. The clinical judgment of gynaecologists was compared to the predictive performance of two existing prediction models, i.e. a neural network and logistic regression model as described by Timmerman.\(^14\) At external validation, the areas under the ROC-curve of these models were 0.84 and 0.85 respectively.\(^15\) No statistically significant differences could be found between the accuracy of the risk assessments made by the clinicians and the accuracy of the risk assessments made by the prediction models.

We found that intuitive estimation of probabilities by experienced clinicians was as good as the prediction by mathematical models. This is hardly surprising, as after all this is what expertise is about. However, somewhat disturbingly, senior clinical residents performed as good as seasoned clinicians. One might hypothesize that the cases were so straightforward that no additional benefit could be accrued from either expertise or modelling. However, the case series were derived from a random selection of 170 consecutive patients in our hospital, and all cases were used three times by each group of observers. Moreover, the information that was provided to the clinicians was limited to that information that was designed to be important to prediction models.
Thus, additional data, for example results of the physical examination, could not improve the performance of the mathematical models, since this information has not been included in those models. Consequently, we think that the decision for referral of a patient with an adnexal mass can be made by a gynaecologist in a general hospital, without prior consultation of a gynaecological oncologist.

This study demonstrated that there is no need to introduce complicated predictive scoring systems like neural networks or logistic regression models for the pre-operative assessment of the adnexal masses. Using data from patients' history, ultrasonography and CA125 measurement the clinician is capable of making a decision with accuracy similar to the one obtained from predictive models. To establish the place of new diagnostic tools like three-dimensional ultrasonography in the assessment of the adnexal mass we formulated the second question of this thesis.

Is three-dimensional ultrasonography of additional value in the assessment of adnexal masses?

A potentially new diagnostic tool in the assessment of the adnexal mass is three-dimensional (3D) ultrasonography. This technique visualizes the adnexal mass in all three planes (longitudinal, transverse and coronal), thus allowing for 3D spatial evaluation as opposed to 2D ultrasonography, in which only two planes can be visualized. With 3D ultrasonography, a volume of ultrasonographic data is acquired and stored, while with 2D ultrasonography only slices can be assessed. The computer-assisted integration of the acquired information enables reconstruction of the organ being scanned. The addition of 3D power Doppler imaging allows visualization and quantitative assessment of vascularity in all three planes. Analogous to 2D pixels, the stored volume ultrasound information obtained using 3D ultrasonography is defined by voxels (smallest measurable volume). Gray scale voxels contain all 3D gray scale information grades from white to black. Voxels with a power Doppler signal are defined as colour voxels. The range of gray scale and colour voxels can vary between zero and 100.

Assessment by three-dimensional ultrasonography introduces new terms. The mean gray index is a dimension for the mean value of the gray scale voxels and is calculated by means of specially designed software. The mean gray index is a quantification of the echogenicity of an adnexal mass. The same software program is able to calculate three vascularization indices, based on colour voxels: vascularization index (VI), flow index (FI) and vascularization flow index (VFI). The vascularization index as dimension for percentage of vessels present in a defined object, is the ratio of the number of colour voxels to the number of all voxels in the defined contour. The VI is expressed as a percentage value. The flow index (scale 0 - 100) is the mean value of the colour voxels
and is thought to express the average intensity of flow in the vessels. The vascularization flow index measures the ratio of the mean value of the colour voxels and all the voxels in the defined contour (VI*FI/100) and is a feature of both vascularization and flow. It identifies the extremes between low vascularization and low blood flow on one side, and high vascularization and high blood flow, on the other.

Volume measurements of organs, scanned by three-dimensional ultrasonography, can be performed by the virtual organ computer-aided analysis system (VOCAL™).16

Chapters 4 and 5 focus on the question whether three-dimensional ultrasonography has additional value in the assessment of adnexal masses. To answer this question we performed a prospective multicenter study. Between January 2003 and March 2005, patients scheduled for surgery in Maxima Medical Centre because of an adnexal mass, were included consecutively. To enlarge the study population, also patients from four surrounding hospitals were included. In the week prior to surgery, women were scheduled for 2D and 3D gray scale ultrasonography and 2D and 3D power Doppler ultrasonography. The 2D and 3D ultrasound were performed using a 7.5-MHz transvaginal transducer (Voluson 730, Kretztechnik, Austria).

One hundred and eighty one women were included. At surgery, there appeared to be 144 (80%) benign adnexal masses were found, 26 (14%) malignant adnexal masses and 11 (6%) borderline malignant masses.

Chapter 4 evaluates whether data acquired by 3D ultrasonography and 3D power Doppler investigation could potentially contribute to the differentiation between benign and malignant ovarian masses. When central vessels were present the risk of a malignancy increased, as indicated by likelihood ratios of 3.3 (95% CI 1.3 - 6.8) and 4.9 (95% CI 2.1 - 12) for central vessels only and central and peripheral vessels in the same mass, respectively. The malignant masses also had a higher mean flow index (32.6 ± 7.1) compared to the benign (28.5 ± 7.3) and borderline (30.8 ± 6.6) masses (p-value 0.04, area under the ROC-curve 0.65). Moreover, the malignant masses had a higher mean gray index as compared to the benign and borderline masses (p-value 0.006, area under the ROC-curve 0.69). The mean vascularization index, the vascularization flow index and the volume of the mass, did not differ significantly between the three categories.

Malignant neoplasm's are highly vascular. The neovascularization in malignant neoplasm's is of poor quality with low compliance and low resistance. Vascular flow within an adnexal mass is often assessed by 2D power Doppler ultrasonography and described in terms of Resistance Index (RI), Pulsatility Index (PI) and Peak Systolic Velocity (PSV).17 Another possibility is to describe the localization of the vessels or the branching pattern. In malignant masses, central vascular flow with a chaotic branching
pattern is expected. By means of 3D power Doppler ultrasonography it is possible to visualize vessels in all three planes (longitudinal, transverse and coronal), thus facilitating spatial evaluation. Assessing the localization of vessels by 3D power Doppler ultrasonography is a simple procedure. In our study the risk of malignancy was increased in the presence of central vessels (likelihood ratio 4.9 (95% CI 2.1 - 11)). Our results are in accordance with the results of others. However Alcazar et al. assessed localization of blood flow by 2D and 3D power Doppler examination, and in their hands 3D power Doppler imaging did not have a better performance than 2D power Doppler imaging. We did not assess blood flow localization by 2D power Doppler examination, so we cannot confirm these results.

Because of the low resistance of neovascularization, the flow in malignant masses will be increased. This is in concordance with the increased flow index we found in the malignant masses as compared to the borderline masses and benign masses.

Adnexal masses which contain papillary formations and masses that are mixed or highly echogenic are suspicious for malignancy. These features result in a high mean gray index. We found the mean gray index to be significantly higher in malignant adnexal masses compared to benign and borderline masses. However also benign teratomas or hemorrhagic masses often show mixed echogenicity. This probably explains the low likelihood ratio of the mean gray index.

In conclusion, the central localization of vessels in an adnexal mass, the mean gray index and the flow index as observed by 3D ultrasonography, are potentially important parameters for distinguishing benign from malignant adnexal masses.

Chapter 5 assesses whether 3D ultrasonography had additional value in the assessment of an adnexal mass as compared to 2D ultrasonography. Multivariable logistic regression analysis with a stepwise backwards selection procedure was used to construct prediction models. Data were obtained from the prospective multicenter study described earlier. We developed two models. The first model ("2D model") was based on age and CA125 level as well as results from the 2D ultrasonography. In the second model ("3D model"), the results of the 3D ultrasonography were added. The capacity of both models to distinguish between benign and malignant adnexal masses was assessed and compared by calculating areas under the ROC-curve (AUC).

For the "2D model", age, CA125 level, presence of vessels suspected for malignancy, random echogenicity, presence of septations and ascites were selected as independent predictors. For the 3D model, age, CA125 level, ascites, uniform echogenicity, flow index, vascularity flow index, volume of the mass on 3D ultrasonography, mean gray index and the visualization of central vessels showed an association with the probability of malignancy. The AUC was 0.82 (95% CI 0.73 - 0.92) and 0.92 (95% CI 0.87 - 0.98) for the 2D and 3D model respectively (p = 0.02). We found that a model
that incorporated data of 3D ultrasonography predicted significantly better than a model that included data of 2D ultrasonography and patient characteristics only.

We want to stress that the findings must be considered in the context of a pilot study in light of our relatively small sample size. We included nine parameters in the 3D model. As a general rule, about ten cases of malignancy are needed for every factor that is included in the model, which implicates that about 90 malignant masses would be needed. Moreover, external validation, an essential and final step in the development of prediction models, is needed before these models can be used in clinical practice, because prediction models tend to be overoptimistic when applied in the same populations than the one in which they were developed.

Several studies on predicting malignancy in a patient with an adnexal mass by means of 3D ultrasonography have been published. Most of these studies conclude that 3D ultrasonography has a higher specificity in predicting malignancy of adnexal masses compared to 2D ultrasonography. However, such isolated test research does not reflect 'real life' clinical practice in which the findings of ultrasonography are always interpreted in the context of patient characteristics and laboratory results, and the additional value of 3D ultrasonography next to patients' characteristics and 2D ultrasonography can only be assessed with multivariable analyses.

Several authors have published models for the interpretation of data of 2D ultrasonography in clinical practice. A comparison of our newly developed 3D model with existing 2D models would be erroneous, since the 3D model could only be validated on the data set in which it was developed. Therefore, we limited our comparison to the 2D model and the 3D model that we developed in the present study.

As the addition of 3D ultrasonography improves the accuracy of the pre-operative diagnosis of malignancy of the adnexal mass, the next question is whether this improvement is clinically relevant. To answer this question we need a large prospective multicenter study with a higher proportion of malignancies. Such data are needed to establish whether the improvement in diagnostic accuracy can be translated in more laparoscopies, less frozen section diagnostic procedures at laparotomy and less relaporatomies after an unclear first diagnosis.

What is the accuracy of frozen section diagnosis and which factors have impact on the accuracy of frozen section diagnosis?

Chapter 6 presents a meta-analysis of studies reporting on accuracy of frozen section diagnosis. We performed a computerized Medline and EMBASE search to identify all registered articles published between January 1966 and June 2003 comparing the frozen section diagnosis of ovarian pathology with the final histopathological diagnosis. For each included study, we calculated the prevalence of
malignant and borderline tumours and the sensitivity, specificity and likelihood ratios of the frozen section diagnosis using the final histopathological diagnosis as reference. We performed the calculations in two ways. In the first analysis, tumours that were found to be borderline were considered as malignant in the two by two tables. In the second analysis, tumours that were found to be borderline at the final assessment were considered benign in the two by two tables.

All 18 included studies were retrospective cohort studies. We did not find a significant association between study characteristics and the diagnostic accuracy of frozen section diagnosis. When a diagnosis borderline malignancy was considered malignant, the sensitivity of frozen section diagnosis varied between 65% and 97%. The specificity varied between 97% and 100%. When a diagnosis borderline malignancy was considered benign, the sensitivity varied between 71% and 100%. The specificity varied between 98.3% and 100%.

In absence of clear evidence how to deal with borderline malignancy, there are wide variances in local practice. Our study offers the clinician the possibility to deal with this issue: in the first analysis by considering borderline tumours as malignant and in the second analysis by considering borderline tumours as benign. The clinician can choose which of these situations is appropriate for a specific patient. In case a borderline malignancy is treated with a bilateral salpingo-oophorectomy and hysterectomy with subsequent surgical staging, a borderline malignancy at frozen section diagnosis has to be considered as malignant disease. However, if borderline disease is considered as not requiring such extensive surgery, it might be preferable to classify the disease as benign.

In case a pathologist cannot judge an adnexal mass with frozen section diagnosis, he can choose to defer the diagnosis and wait until the definitive paraffin section diagnosis. Six studies reported explicitly that a diagnosis was deferred at frozen section diagnosis. Since it would generate bias if we would exclude these cases from our study, we included them, and considered the result of the frozen section diagnosis as benign. Such a choice is justified, if one realises that in case malignancy is not suggested at frozen section diagnosis, subsequent radical surgical treatment is not justified. The likelihood ratio for malignancy in case of deferred diagnosis appeared to be 1.9, a value much too small to justify radical surgery in case of a deferred diagnosis.

We found homogeneity among the reported specificity of frozen section diagnosis, whereas estimates of the sensitivity of the test varied between 65% and 100%. One explanation for the imperfect sensitivity of frozen section diagnosis might be the size of the tumour. It is not unlikely that malignant disease is not detected in case the adnexal mass is large. Moreover, in case of borderline malignancy and mucinous tumours, it is difficult for frozen section examination to meet the sampling requirements and hence frozen section diagnosis will be less sensitive. Some of the included studies implicated
this sort of explanation, however the data were insufficient to use in this meta-analysis.

Despite the limitations mentioned above, we found excellent test characteristics for frozen section diagnosis. The question arises whether these test characteristics justify the use of frozen section diagnosis in clinical practice. To answer this question we calculated expected regret reduction based on the consequences of misclassification due to false positive or false negative results of the frozen section diagnosis and based on the prevalence of malignant disease. It became apparent that frozen section diagnosis is expected to reduce regret in a majority of patients.

In conclusion, our meta-analysis has demonstrated that the accuracy of frozen section diagnosis for the assessment of the ovarian mass is good, with acceptable sensitivities for almost perfect specificities. Future studies on patients' preferences for the different outcomes as well as economic analysis are needed for establishing the definite position of this diagnostic technique.

Chapter 7 reports the results of a retrospective study of the impact of tumour size on the accuracy of frozen section diagnosis. Whereas malignancy is always detected in case the size of the adnexal mass is relatively small, frozen section diagnosis might miss malignancy in case the adnexal mass has a larger volume. Since this hypothesis could not be evaluated in our previous meta-analysis, we examined in this retrospective study whether size of the adnexal mass had impact on the accuracy of frozen section diagnosis. The study was performed at Máxima Medical Centre in Veldhoven, Tweesteden Hospital in Tilburg and Saint Elisabeth Hospital in Tilburg. We included all women who had undergone frozen section diagnosis for an adnexal mass between January 2000 and December 2003 retrospectively. For each case, the following data were recorded: pre-operative level of serum CA125, menopausal status, tumour size at surgery, and histological classification both at frozen section diagnosis and at final histology.

During the study period, 257 consecutive frozen section diagnoses of adnexal masses were performed. The ultimate paraffin diagnosis showed benign pathology in 142 cases (55%), malignancy in 87 cases (34%) and borderline malignancy in 28 cases (11%). In women with a tumour size < 10 cm, the likelihood ratios for benign frozen section diagnosis, borderline frozen section diagnosis, and malignant frozen section diagnosis were 0.03 (95% CI 0.01 - 0.23), 12.9 (95% CI 1.7 - 98.2) and infinite (95% CI 7.1 - ∞), respectively. In women with a tumour size ≥ 10 cm, the likelihood ratios for benign frozen section diagnosis, borderline frozen section diagnosis, and malignant frozen section diagnosis were 0.15 (95% CI 0.08 - 0.25), 4.9 (95% CI 2.1 - 11.4) and infinite (95% CI 9.7 - ∞), respectively.

These findings have consequences for clinical management. In case a tumour has a diameter of more than 10 cm and the frozen section diagnosis is benign, 11% (11 / 97)
of the women turned out to eventually have a malignant or borderline tumour. This means one out of nine patients needed further surgery despite of a negative frozen section diagnosis in case a tumour had a diameter of more than 10 cm.

Other studies confirm the influence of the size of adnexal masses on the accuracy of frozen section diagnosis.53,54 The influence of the size of the adnexal mass on the accuracy of frozen section diagnosis can be explained by the nature of the procedure of frozen section diagnosis. One of the major problems of frozen section diagnosis is the difficulty of sampling. Because frozen section diagnosis should be a rapid procedure during surgery and the diagnosis is expected as soon as possible there is inadequate time to make as much slices as can be performed in a paraffin procedure.

Besides the size of the tumour, the histologic type can be a problem for frozen section procedures. Mucinous tumours are known as the largest of all ovarian tumours. We calculated the risk of a false negative result and consequently a second operation is 23% in case of a mucinous tumours ≥ 10 cm. The histological characteristics of mucinous tumours make them less suitable for frozen section diagnosis.55 The problem of false negative frozen section diagnosis in case of mucinous tumours has been described in literature at several occasions.56-59 Our results are similar to those reported in the literature.

In conclusion, in adnexal masses ≥ 10 cm, a benign result of the frozen section diagnosis is less reliable than in masses < 10 cm. Diagnosing malignancy in mucinous tumours is fraught with problems in frozen section diagnosis because of their relative large size and their histological characteristics. It would be useful if the pathologist would be able to differentiate mucinous tumours from other tumours during frozen section diagnosis, as the non-malignant findings in mucinous tumours are less reliable.

Nowadays it is generally acknowledged that patients' preferences should be incorporated into medical decision making.60 Several studies have shown that patients' perspectives on the burden and benefits of therapy can differ from those of health professionals.61 This leads to the fourth question of the thesis.

What are patients' preferences towards the use of frozen section diagnosis and towards the mode of surgery?

Chapter 8 studies patients' preferences with regard to the mode of surgery in case of an adnexal mass. A false negative test result of frozen section diagnosis, i.e. the situation in which frozen section diagnosis is benign and paraffin diagnosis is malignant, implicates that the patient has to be operated a second time to perform an optimal surgical staging. Apart from the risks of a second surgical procedure, a false negative frozen section diagnosis can also cause disappointment for the patient who was reassured after initial surgery. An alternative could be to perform radical primary
staging in every patient operated for an adnexal mass without using frozen section diagnosis. This implicates that all patients with adnexal masses, including those whose eventual paraffin sections show a benign lesion, would undergo a median laparotomy with two sided adnectomy, hysterectomy and omentectomy, peritoneal biopsies and lymph node sampling.

Another issue is whether to perform laparoscopy or laparotomy in case of an adnexal mass. Although laparoscopy has advantages (short hospital admission, small scars and early resumption of daily activities), in case of malignancy there might be preference for laparotomy, because of uncertain risks of a laparoscopic approach to malignancy with possible negative consequences regarding the prognosis.

To study patients' preferences in mode of surgery in case of an adnexal mass, we designed a structured interview. Both dilemmas\(^1\), applying frozen section diagnosis versus radical primary surgery as well as laparoscopy versus laparotomy in case of an adnexal mass, were subject of the interview. We interviewed 43 women.

The mean threshold at which the women switched their preference from radical primary surgery to frozen section diagnosis was a risk of 50% on a false negative test result of frozen section diagnosis. Women scheduled for surgery for an adnexal mass, should be informed about the possibility of frozen section diagnosis if there is any (pre-operative) suspicion on malignancy, rather than to perform radical primary staging without any evidence of malignancy.

In case of benign adnexal masses, a laparoscopic approach is preferable to a laparotomic approach because of less postoperative complications, less postoperative pain and shorter admission time.\(^5\) In case of malignant masses, no explicit evidence exists regarding the question whether a laparoscopic or a laparotomic approach is to be preferred. Abdominal wall metastases at trocar sites following laparoscopy have been described, however its influence on survival is not clear.\(^62,63\) The negative effects of carbon dioxide on the peritoneum with risk on tumour dissemination have been described in animal studies, but until now there is no clinical evidence in case of ovarian malignancy.\(^64\) We explained the dilemma of laparoscopy versus laparotomy to patients scheduled for surgery because of an adnexal mass. These patients preferred laparoscopy to laparotomy up to a 50% risk on malignancy. We are aware of the fact that we interviewed a low risk population on ovarian malignancy. Three women suffered a borderline malignancy of the ovary and none of the interviewed women had an invasive ovarian malignancy. It is possible that a high risk population will have different preferences. It would be interesting to perform further research with regard to the preference of women for a surgical approach (laparoscopy versus laparotomy) and to find out whether preferences do change after the diagnosis of ovarian malignancy.

In conclusion, women prefer frozen section diagnosis to radical primary staging and
accept the risk of a negative result of frozen section diagnosis. Women with low risk of malignancy prefer laparoscopy to laparotomy.

Outcome of ovarian cancer is better when surgery is provided by a gynaecological oncologist than by a general gynaecologist. However, when all patients with an adnexal mass have to be operated by gynaecological oncologists, this requires a change in the organization of care, which generates additional costs.

What are the implications of centralized and regular care for women with an ovarian malignancy in The Netherlands?

In chapter 9 we assessed the implications of centralized and regular care for women with an ovarian malignancy in The Netherlands. We performed a cost-effectiveness analysis in which we integrated epidemiologic data, data on the survival of ovarian cancer patients in relation to surgical modality and estimates of diagnostic test accuracy to assess malignancy of the adnexal mass prior to surgery. We considered three strategies. In the first strategy, patients were operated by a general gynaecologist (general care strategy). In the second strategy, patients were operated by a gynaecological oncologist (specialized care strategy). In the third strategy, evaluation of the adnexal mass took place prior to surgery by means of the Risk of Malignancy Index (diagnostic strategy). Patients at high risk for malignancy were supposed to be operated in a specialized care setting, whereas low risk patients were supposed to be operated in a general care setting. All patients were supposed to receive adjuvant chemotherapy, except patients with stage I disease that had been operated by a gynaecological oncologist.

In The Netherlands, since 1971 a nationwide network of histo- and cytopathology (PALGA) has been active. In the PALGA database, we found all 1,096 reports of the epithelial ovarian malignancies diagnosed in 2003 in The Netherlands. We included 6,502 reports on benign pathology, of which the reports on simple cysts in premenopausal women were excluded (estimated on 5%). Consequently, we calculated a prevalence of epithelial ovarian malignancy of 14 % (1096 / 7598).

Based on studies reporting on the impact of the type of surgeon, i.e. a gynaecological oncologist or a general gynaecologist we also estimated survival in advanced stage disease.\textsuperscript{65-68} Mean survival for stage III / IV was estimated 15 months in the general care setting and 22 months in the specialized care setting. Mean survival for stage I / II was set on 60 months for both settings.

For each strategy, we calculated life expectancy and incremental costs per life year gained. Life expectancy of a patient with an ovarian malignancy in the general strategy was 2.7 years, in the diagnostic strategy 3.0 years and in the specialized strategy 3.1
years. The diagnostic strategy appeared to be dominant over the general strategy. The incremental costs to gain one additional life year with specialized surgery as compared to the diagnostic strategy were € 63,477 per LYG.

Because we did not know the real costs and survival data of this cohort, we had to make some assumptions. However, despite of the uncertainties of some assumptions included in this economic analysis, we illustrated clearly that costs in a general strategy (all patients with an adnexal mass operated by a general gynaecologist) were low compared to costs in a specialized strategy (all patients with an adnexal mass operated by a gynaecological oncologist) detrimental to survival. On the other side, treatment of all adnexal masses in specialized hospitals by gynaecological oncologists would be very expensive.

Greving et al. performed a cost-effectiveness analysis for centralization of ovarian cancer based on actual patient data. Greving found treatment of ovarian cancer patients (early stage and advanced stage disease) in semi-specialized hospital settings was the most cost-effective strategy. Treatment of ovarian cancer in semi-specialized hospitals yielded 0.15 QALY compared to treatment in general hospitals, resulting in an incremental cost-effectiveness ratio of € 5,700. Treatment of patients in tertiary referral hospitals resulted in an incremental cost-effectiveness ratio of € 93,700. If the optimal debulking rate in specialized hospitals would increase from 46% to 70%, ovarian care in these hospitals also would become a cost-effective strategy (incremental cost-effectiveness ratio of € 25,000). However, in this study, costs of unjustified referrals of patients with a benign mass to specialized hospitals (false positive pre-operative assessment of an adnexal mass), and loss of survival years because of unjustified not referred patients with malignant adnexal masses (false negative pre-operative assessment of an adnexal mass), were not included.

We assessed the implications of centralized and regular care for women with an ovarian malignancy in The Netherlands and showed centralized care is cost-effective on condition that pre-operative assessment of an adnexal mass has been performed.

**Clinical implications**

This thesis deals with the diagnosis of malignancy in adnexal masses, both pre-operatively as well as peri-operatively. For pre-operative assessment of an adnexal mass, several mathematical models have been described in literature and some of them have been validated externally. However, intuitive estimation of probabilities by experienced clinicians appeared to be as good as the prediction by mathematical models. In case the clinician prefers to use of a mathematical model, we advice to use the RMI I of Jacobs et al. This model appeared to have the best test characteristics. In
the economic analysis, the use of this model appeared to be cost-effective over strategies without diagnosis, be it immediate operation by a general gynaecologist or referral to a specialized oncological centre.

For peri-operative assessment of the adnexal mass, we advice to use frozen section diagnosis because of its accuracy. In case of large tumours with a diameter ≥ 10 cm, the diagnostic accuracy of frozen section is not perfect. This risk of a false negative test result of the frozen section diagnosis has to be taken into account when the patient is counselled prior to the operation. However, patients with an adnexal mass prefer frozen section diagnosis to primary radical surgery, even in case of an increased risk of a false negative test result.

In case malignancy is strongly suspected in an adnexal mass, a gynaecological oncologist should operate the patient. In the economic analysis in chapter 7, we found that surgery by a gynaecological oncologist is cost-effective over other strategies when the pre-operative probability of malignancy exceeds 58%. Referral of all patients with an adnexal mass to a gynaecological oncologist appeared to be very expensive, and therefore not cost-effective.

**Future research**

In chapter 3, we showed that at present clinical assessment is as good as the use of mathematical models. In this thesis, we reported that a new diagnostic instrument, 3D ultrasonography, improved diagnostic accuracy. We found that a model that incorporated data on 3D ultrasonography predicted significantly better than a model that included data on 2D ultrasonography and patient characteristics only. However, we did not perform external validation. Furthermore, our study included a small sample of women. We included nine parameters in the 3D model. As a general rule, about ten cases of malignancy are needed for every factor that is included in the model, which implicates that about 90 malignant masses would be needed. From our meta-analysis on existing prediction models, we know the Risk of Malignancy Index from Jacobs turned out to be a simple and accurate model. A prospective multicenter study with a large study population to validate our 3D model and compare its performance with Jacobs' Risk of Malignancy Index and with clinical estimates is the next step in research to the value of 3D ultrasonography.

Several studies implicate that the management of ovarian malignancy should be centralized to improve survival for Dutch patients with ovarian malignancy. Patients with early stage disease are staged adequately more often by gynaecological oncologists than by general gynaecologists. However, the procedure of adequately staging itself does not influence survival. Adequately staging differentiates between
patients with truly stage I disease and patients with occult stage III disease. As Vernooij reported in her thesis, in The Netherlands just 24% of patients with early stage disease were adequately staged in case surgery was performed in general hospitals compared to 60% adequately staged patients in (semi-) specialized hospitals.\textsuperscript{70} Heintz et al. underline in a recent article in Medisch Contact that adequately staging is just a matter of having surgical skills and executing a protocol.\textsuperscript{71} In a specialized hospital, the necessarily surgical skills will be available all the time, but not in a general hospital. In case ovarian malignancy is suspected, patients in general hospitals can be referred to a specialized hospital or a skilled gynaecological oncologist can be invited to operate the patient in the general hospital. However, in case malignancy is not suspected, the patient in a general hospital will have surgery by a general gynaecologist and consequently staging may be inadequately. So, for early stage disease, the skills of a gynaecological oncologist do not need to be improved, but the pre-operative assessment of an adnexal mass should be improved.

For advanced stage disease, optimally debulking (i.e. no residual tumour or resttumour with a diameter less than one cm) is an important prognostic factor for survival. Vernooij showed that patients operated by low-volume gynaecologists (i.e. one ovarian cancer operations per two months) were optimally debulked in 37%, while high volume gynaecologists (eg > two ovarian cancer operations per two months) achieved optimally debulking in 57%.\textsuperscript{70} Consequently, patients with high stage disease should be operated by high volume gynaecologists which implicates centralization of surgical treatment of patients with ovarian malignancy.

In chapter 9, we illustrated that a diagnostic strategy prior to the decision for surgery by a general gynaecologist or a gynaecological oncologist provides the best balance between costs and effects. The PROBAAT study (Radboud University Medical Centre, Nijmegen, The Netherlands) examined whether the Risk of Malignancy Index is an accurate test to decide whether consultation of a gynaecological oncologist is needed. Unfortunately, inclusion of patients in this study was not consecutive, but depended on the general gynaecologist who decided whether or not to calculate the RMI of a patient with an adnexal mass. In a future study on the implications of centralized care for ovarian malignancy in The Netherlands, selection of patients who would benefit from consulting a gynaecological oncologist should be based on an accurate diagnostic test, and, for a representative sample of study patients to be obtained, they should be included consecutively into the study.
Beschouwing en samenvatting

Een gynaecoloog wordt frequent geconsulteerd over de aanwezigheid van een vergroot ovarium. Meestal heeft een patiënt met een vergroot ovarium weinig klachten. Omdat er een kans op maligniteit bestaat, moet een vergroot ovarium zorgvuldig worden beoordeeld. De behandeling van een maligne vergroot ovarium is namelijk volledig anders dan de behandeling van een benigne vergroot ovarium.

Indien er sprake is van ovariumcarcinoom is optimale stadiëring zeer belangrijk, omdat op basis van stadiëring wordt bepaald of een patiënt wel of geen chemotherapie nodig heeft.1 Bij een gevorderd of laat stadium ovariumcarcinoom is primair chirurgische cyto-reductie de standaardbehandeling. Er wordt gestreefd om naar een bij voorkeur complete debulking, maar in ieder geval een optimale debulking waarbij de resttumor kleiner is dan één centimeter in doorsnede. Vervolgens wordt nagebehandeld met chemotherapie gebaseerd op Taxane en Platinum. Verschillende studies suggereren dat de overlevingskans zowel in geval van vroeg stadium als in geval van laat stadium ovariumcarcinoom beter is indien de chirurgie wordt uitgevoerd door een gynaecologisch oncoloog dan door een algemeen gynaecoloog.2-4

In geval van een benigne vergroot ovarium kan ovarium sparend geopereerd worden door een cystectomie te verrichten in plaats van een ovarietomie. Daarbij heeft een laparoscopische benadering de voorkeur boven een laparotomische benadering.5

Gezien de therapeutische consequenties is het belangrijk om het risico op de aanwezigheid van een ovariumcarcinoom zowel preoperatief als tijdens een operatie goed in te schatten. Er zijn vele diagnostische middelen beschikbaar om een benigne vergroot ovarium te onderscheiden van een maligne vergroot ovarium. In de preoperatieve setting zijn deze onder andere: patiëntkenmerken, serum tumor markers, echoscopie en beeldvormende technieken zoals CT en MRI. Om de nauwkeurigheid van het beoordelen van een vergroot ovarium te verbeteren, zijn de meeste van deze onderzoeken gecombineerd in diagnostische modellen. Deze diagnostische modellen blijken bij externe validatie minder goed te presteren dan beschreven door de oorspronkelijke auteurs van de modellen.

Een potentieel nieuw instrument om een vergroot ovarium te beoordelen is driedimensionale echoscopie (3D). Door deze techniek worden drie vlakken (longitudinaal, transversaal en coronaal) van een vergroot ovarium in beeld gebracht. Door een 3D reconstructie wordt ruimtelijk inzicht verkregen in tegenstelling tot 2D onderzoek waarbij slechts twee vlakken gevisualiseerd kunnen worden.

Om tijdens een operatie een vergroot ovarium te beoordelen kan gebruik worden gemaakt van vriescoupe-onderzoek. Dit proefschrift behandelt de diagnostiek van maligniteit in geval van een vergroot ovarium, zowel preoperatief als tijdens chirurgie.
Hoe accuraat zijn bestaande modellen in de preoperatieve beoordeling van een vergroot ovarium, en zijn deze modellen nauwkeuriger dan klinische beoordeling?

Hoofdstuk 2 presenteert een review van bestaande diagnostische modellen. Het risico op maligniteit in een vergroot ovarium kan preoperatief worden geschat met behulp van diagnostische modellen waarin verschillende factoren zijn opgenomen. In een systematisch literatuur onderzoek hebben we de accuraatheid van diagnostische modellen onderzocht. Met behulp van Medline en EMBASE databases werd gezocht naar literatuur vanaf januari 1966 tot maart 2008. We vonden 109 studies die voldeden aan de inclusiecriteria. In deze studies werd gerapporteerd over 21.750 vergrote ovaria (5.826 maligniteiten) en over 83 verschillende diagnostische modellen. Deze 83 modellen werden door ons ingedeeld in vier types. Op de eerste plaats beschrijvende numerieke diagnostische modellen die gebaseerd zijn op echoscopisch waargenomen morfologische kenmerken van een vergroot ovarium. Aan elk kenmerk wordt een numerieke waarde toegekend. De waarden worden bij elkaar opgeteld en de totale score geeft het risico op maligniteit weer. Het model van Sassone is het meest frequent gevalideerde voorbeeld van een dergelijk diagnostisch model. Bij een afkapwaarde van negen was de gecombineerde schatting voor de sensitiviteit 84% met een specificiteit van 80%. Een tweede type diagnostisch model is gebaseerd op logistische regressie. Een logistische regressie analyse is een mathematische procedure voor het systematisch ontwikkelen van een algoritme waarmee de kans op een bepaalde gebeurtenis kan worden bepaald. De formule van de Risk of Malignancy Index (RMI), oorspronkelijk beschreven door Jacobs (RMI I) en gemodificeerd door Tinguistad (RMI II en III), is een vereenvoudiging van een formule op basis van logistische regressie analyse. De RMI gebruikt het product van drie variabelen: echo score, menopauzale status en de absolute waarde van het serum CA125. De score van de RMI varieert tussen nul en oneindig in de RMI I en tussen één en oneindig in de RMI II en III. De meest gebruikte afkapwaarde is 200, zowel in de RMI I, II als III. In onze meta-analyse was bij een afkapwaarde van 200 de gecombineerde schatting voor sensitiviteit 78% (RMI I) en 79% (RMI II) met een specificiteit van 87% (RMI I) en 81% (RMI II). Een derde type diagnostisch model is het artificiële neurale netwerk. Het neurale netwerk onderscheidt zich van logistische regressie analyse door de mogelijkheid eventuele aanwezigheid van niet-lineaire interactie tussen variabelen te onderzoeken. Een vierde type modellen zijn eveneens rekenkundig, namelijk de least squares support vector machines (LS-SVM) en relevance vector machines (RVM). Deze zijn erg complex en de ervaring ermee is nog zeer beperkt.

Alle beschreven modellen hadden acceptabele sensitiviteit en specificiteit. Echter, de RMI I bleek de beste voorspeller.

Aangezien een gynaecologisch oncoloog het ovariumentum beter kan
behandelen dan een algemeen gynaecoloog. Zouden patiënten bij wie een ovariumcarcinoom wordt vermoed, verwezen moeten worden naar een gynaecologisch oncoloog.\textsuperscript{12} Bailey e.a. onderzochten of de RMI gebruikt zou kunnen worden voor adequate verwijzing van patiënten naar een derdelijns centrum.\textsuperscript{13} Zij vonden een positief voorspellende waarde van 86,8%.

Samenvattend zijn er verschillende types gevalideerde modellen beschikbaar voor de preoperatieve beoordeling van een vergroot ovarium. Alle diagnostische modellen hebben een acceptabele sensitiviteit en specifiched. De RMI I zoals beschreven door Jacobs is een eenvoudig diagnostisch model dat zeer accuraat is en daardoor bruikbaar in de klinische praktijk.\textsuperscript{7-9,13} Er zou verder moeten worden onderzocht of de RMI gebruikt kan worden om patiënten verdacht voor ovariumcarcinoom te verwijzen naar een derdelijns centrum.

**Hoofdstuk 3** vergelijkt de preoperatieve klinische beoordeling door gynaecologen met de waarde van diagnostische modellen. Uit een dataset van 170 casus selecteerden we at random 34 casus waarmee vijf verschillende sets werden gevormd. Elke set werd aangeboden aan 15 gynaecologisch oncoloog, 15 algemeen gynaecologen en 15 gynaecologen in opleiding. Elke casus bevatte gegevens over de leeftijd en menopauzale status van de patiënt, een beschrijving van de transvaginale echoscopie, transvaginale Doppler meting en de waarde van het serum CA125. Aan de beoordelaars werd gevraagd hoe groot zij het risico op maligniteit schatten. De accuraatheid van de beoordeling van de klinici werd beoordeeld met Receiver Operating Characteristic (ROC) analyse, waarin het test resultaat, dus het risico op maligniteit geschat door de gynaecologen, werd vergeleken met de aan- of afwezigheid van maligniteit zoals bepaald aan de hand van histologisch onderzoek door de patholoog. We construeerden een ROC-curve en berekenen het oppervlak onder de ROC-curve. De oppervlakken onder de ROC-curven waren 0,82 (95% BI 0,78 - 0,86), 0,84 (95% BI 0,80 - 0,86) en 0,82 (95% BI 0,78 - 0,86) voor respectievelijk gynaecologisch oncoloog, algemeen gynaecologen en gynaecologen in opleiding. De klinische beoordeling door de gynaecologen werd vergeleken met de voorspelling door twee bestaande diagnostische modellen, namelijk een neuraal netwerk en een logistisch regressiemodel beiden ontwikkeld door Timmerman.\textsuperscript{14} Bij externe validatie waren de oppervlakken onder de ROC-curve van deze modellen 0,84 en 0,85.\textsuperscript{15} Er werden geen statistisch significante verschillen gevonden tussen de accuraatheid van risico schatting door klinici en de risico inschatting door diagnostische modellen.

We vonden dat intuitieve inschatting van risico's door ervaren klinici net zo goed was als de voorspelling door mathematische modellen. Dit is niet verbazingwekkend aangezien het hier om expertise gaat. Het is wat verwarrend dat gynaecologen in
opleiding net zo goed presteerden als ervaren gynaecologen. Hypothetisch zou het zo kunnen zijn dat de casuïstiek zo evident was dat er geen additioneel voordeel te halen zou zijn uit expertise of modellen. Echter de casuïstiek was een at random selectie van 170 opeenvolgende patiënten uit ons ziekenhuis, en elke casus werd drie keer gebruikt door elke groep beoordelaars. Verder was de informatie voor de klinici beperkt tot de informatie die belangrijk is voor de modellen. Dus bijkomende gegevens zoals het resultaat van een lichamelijk onderzoek zouden de prestatie van de modellen niet verbeteren, omdat deze informatie geen onderdeel uitmaakt van de modellen. Concluderend denken wij dat de beslissing om een patiënt met een vergroot ovarium te verwijzen naar een derdelings ziekenhuis gemaakt kan worden door een gynaecoloog in een algemeen ziekenhuis zonder consultatie van een gynaecologisch oncoloog.

Deze studie toonde aan dat het niet nodig is om ingewikkelde voorspellende score systemen te introduceren zoals neurale netwerken of logistische regressie modellen voor de preoperatieve beoordeling van een vergroot ovarium. De klinicus is in staat om op basis van de anamnese, echoscopie en CA125 te beslissen met een zelfde betrouwbaarheid als een diagnostisch model. Wij wilden in dit proefschrift ook de plaats bepalen van de 3D echoscopie als nieuw diagnostisch instrument voor de beoordeling van een vergroot ovarium.

Een potentiële nieuw diagnostisch instrument in de beoordeling van een vergroot ovarium is driedimensionale (3D) echoscopie. Bij deze techniek wordt een object in drie vlakken gevisualiseerd (longitudinaal, transversaal en coronaal), waardoor een ruimtelijke 3D evaluatie kan plaatsvinden. Dit in tegenstelling tot de 2D echoscopie waarin slechts twee vlakken kunnen worden gevisualiseerd. Met de 3D echoscopie wordt een volume van echoscopische data verkregen en opgeslagen, terwijl bij 2D echoscopie alleen doorsneden beoordeeld kunnen worden. Computer gestuurde integratie van de verkregen data maakt het mogelijk om het gescande orgaan te reconstrueren. Door het toevoegen van 3D power Doppler is het mogelijk om de vascularisatie te visualiseren en te kwantificeren in alle drie de vlakken. Analoog aan de 2D pixel, wordt de informatie verkregen met 3D echo gedefinieerd als voxels (kleinste volume eenheden). Grijze voxels bevatten alle zwart wit informatie; kleuren voxels zijn voxels verkregen met het power Doppler signaal. De range van grijze en kleuren voxels varieert tussen nul en honderd.

Beoordeling met 3D echoscopie brengt een aantal nieuwe termen met zich mee. De mean gray index representeert de gemiddelde waarde van de grijze voxels en wordt berekend met behulp van speciaal ontworpen software. De mean gray index is een kwantificatie van de echogeniciteit van een vergroot ovarium. Dezelfde software kan drie vascularisatie indices berekenen: vascularisatie index (VI), flow index (FI) en
vascularisatie flow index (VFI). De vascularisatie index (VI) representeert het percentage vaten dat aanwezig is in een gedefinieerd object. De VI is de verhouding tussen het aantal kleuren voxels en het totale aantal voxels in een gedefinieerde contour. Het wordt uitgedrukt in een percentage. De flow index (schaal 0 - 100) is de gemiddelde waarde van de kleuren voxels en representeert de gemiddelde intensiteit van de bloeddoorstroming in de vaten. De vascularisatie flow index is de verhouding tussen de gemiddelde waarde van de kleuren voxels en het totale aantal voxels in een gedefinieerde contour (VI*FI/100). Het is een maat voor zowel vascularisatie als bloeddoorstroming. Het identificeert de extremen tussen lage vascularisatie met weinig bloeddoorstroming aan de ene zijde van het spectrum en hoge vascularisatie met veel bloeddoorstroming aan de andere zijde van het spectrum.

Volume metingen van organen, gescand met 3D echoscopie, kunnen worden bepaald met het zogenaamde virtual organ computer-aided analysis systeem (VOCAL™).16

Heeft driedimensionale echoscopie een toegevoegde waarde bij de beoordeling van een vergroot ovarium?

Hoofdstukken 4 en 5 behandelen de vraag of 3D echoscopie een toegevoegde waarde heeft bij de beoordeling van een vergroot ovarium. Om deze vraag te beantwoorden hebben we een prospectieve multicentrum studie opgezet. Tussen januari 2003 en maart 2005 werden patiënten die op de wachtlijst stonden voor een operatie in verband met een vergroot ovarium consecutief geïncludeerd. Om de studiepopulatie te vergroten werden ook patiënten uit omliggende ziekenhuizen in het onderzoek opgenomen. In de week voorafgaand aan de operatie werd bij elke geïncludeerd patiënt een 2D en 3D echo gemaakt en een 2D en 3D power Doppler. Het 2D and 3D echoscopisch onderzoek werd verricht met een 7,5-MHz transvaginale transducer (Voluson 730, Kretztechnik, Oosterrijk).

De onderzoeksgrup betrof 181 vrouwen. Na de operatie bleken 144 (80%) ovaria benigne te zijn, 26 (14%) ovaria maligne en 11 (6%) ovaria bevatten borderline maligniteit.

Hoofdstuk 4 evalueert of data verkregen met 3D echoscopie en 3D power Doppler onderzoek kunnen bijdragen in het onderscheid tussen benigne en maligne ovaria. Indien centrale vaten aanwezig waren was het risico op maligniteit verhoogd. Dit gold zowel voor aanwezigheid van alleen centrale vaten als voor aanwezigheid van centrale en perifere vaten in één vergroot ovarium met aannemelijkheidsverhoudingen (likelihood ratio’s) van respectievelijk 3,3 (95%BI 1,3 - 6,8) en 4,9 (95%BI 2,1 - 12). De maligne ovaria hadden een hogere gemiddelde flow index (32,6 ± 7,1) dan de benigne (28,5 ± 7,3) en borderline (30,8 ± 6,6) ovaria (p-waarde 0,04, oppervlakte onder de ROC-curve 0,65). Bovendien hadden de maligne ovaria een hoger gemiddelde
Hoofdstuk 10

mean gray index in vergelijking met de benigne and borderline ovaria (p-waarde 0,006, oppervlakte onder de ROC-curve 0,69). De gemiddelde vascularisatie index, de vascularisatie flow index en het volume van de ovaria verschillen niet significant tussen de drie categorieën ovaria.

Maligne neoplasma zijn rijkelijk gevasculariseerd. De neovascularisatie in maligne neoplasma is van slechte kwaliteit met een lage compliance en een lage vaatweerstand. De bloeddoorstroming in een vergroot ovarium wordt vaak beoordeeld met 2D power Doppler echoscopie en beschreven in termen als Resistance Index (RI), Pulsatility Index (PI) and Peak Systolic Velocity (PSV).\(^\text{12}\) Het is ook mogelijk om de lokalisatie of het vertakkingspatroon van de vaten te beschrijven. In maligne ovaria wordt centrale vascularisatie met een chaotisch vertakkingspatroon verwacht.\(^\text{18-20}\) Met 3D power Doppler echoscopie is het mogelijk om de vaten te visualiseren in alle drie de vlakken (longitudinaal, transversaal en coronaal), waardoor ruimtelijk inzicht kan worden verkregen. Het is een eenvoudige procedure om de lokalisatie van de vaten te beoordelen met behulp van 3D power Doppler echoscopie. In onze studie was het risico op maligniteit verhoogd indien centrale vaten werden gezien (aannemelijkheidsverhouding 4,9 (95% BI 2,1 - 11)). Deze resultaten komen overeen met die van anderen. Alcazar e.a. beoordeelden de lokalisatie van de vaten met 2D en 3D power Doppler onderzoek en in hun handen presteerde het 3D power Doppler onderzoek niet beter dan het 2D power Doppler onderzoek.\(^\text{18}\) Wij hebben de vaatlokalisatie niet beoordeeld met 2D power Doppler onderzoek en kunnen deze resultaten dus niet bevestigen.

De bloeddoorstroming in maligne ovaria is verhoogd ten gevolge van de lage vaatweerstand van de neovascularisatie. Dit komt overeen met de verhoogde flow index die we vonden in de maligne ovaria vergeleken met de borderline en benigne ovaria.

Vergrote ovaria met papillaire formaties en ovaria met wisselende echogeniciteit of dense afwijkingen zijn verdacht van maligniteit. Deze kenmerken leiden tot een hoge mean gray index. Ook wij vonden een verhoogde gemiddelde mean gray index in de maligne ovaria in vergelijking met de benigne en borderline ovaria. Echter ook benigne teratomen of hemorragische cysten laten vaak een wisselend echogeon aspect zien. Dit is waarschijnlijk de verklaring voor de lage aannemelijkheidsverhouding van de gemiddelde mean gray index.

Concluderend zijn de centrale lokalisatie van de vaten, de gemiddelde mean gray index en de flow index gemeten met 3D (power Doppler) echoscopie potentiële belangrijke parameters in het onderscheid tussen het benigne en maligne ovarium. Hoofdstuk 5 onderzocht of 3D echoscopie, in vergelijking met 2D echoscopie, een toegevoegde waarde heeft in de beoordeling van een vergroot ovarium. We gebruikten multivariabele logistische regressie analyse om diagnostische modellen te construeren.
De gegevens werden verkregen uit de eerder beschreven multicentrum studie. We ontwikkelden twee modellen. Het eerste model ("2D model") was gebaseerd op leeftijd, waarde van het serum CA125 en resultaten van het 2D echoscopie onderzoek. In het tweede model ("3D model"), werden de resultaten van het 3D echoscopische onderzoek toegevoegd. De capaciteit van beide modellen om onderscheid te maken tussen het benigne en maligne ovarium werd beoordeeld en vergeleken door het berekenen van de oppervlakten onder de ROC-curve (AUC).\textsuperscript{21}

Voor het "2D model" werden leeftijd, waarde van het serum CA125, aanwezigheid van voor maligniteit suspecte vaten, wisselende echogeniciteit, aanwezigheid van septa en de aanwezigheid van ascites geselecteerd als onafhankelijke predictoren. Voor het "3D model" dienen leeftijd, waarde van het serum CA125, ascites, uniforme echogeniciteit, flow index, vascularisatie flow index, volume van het vergrote ovarium met 3D echoscopie, gemiddelde mean gray index en de aanwezigheid van centrale vaten een associatie zien met een risico op maligniteit. De AUC was 0,82 (95% BI 0,73 - 0,92) en 0,92 (95% BI 0,87 - 0,98) voor respectievelijk het 2D en 3D model (p = 0,02). Concluderend gaf het model dat 3D data bevatte een significant betere voorspelling dan het 2D model.

Vanwege de relatief kleine onderzoeksopopulatie willen we benadrukken dat onze bevindingen moeten worden gezien in de context van een pilot studie. We gebruikten negen parameters in het 3D model. Als algemene regel zijn ongeveer 10 casus van maligniteit nodig voor elke factor die wordt geïncludeerd in een model, hetgeen impliceert dat ongeveer 90 maligne ovarium tumoren nodig zijn.\textsuperscript{22} Externe validatie is een essentiële en bestissende stap in de verdere ontwikkeling van diagnostische modellen. Externe validatie is dan ook nodig voordat modellen gebruikt kunnen worden in de klinische praktijk, omdat diagnostische modellen de neiging hebben overoptimistisch te zijn indien ze gebruikt worden in deelzijdse populatie als waarin ze ontwikkeld zijn.\textsuperscript{23-25}

Er zijn diverse studies gepubliceerd betreffende het voorspellen van maligniteit met 3D echoscopie bij patiënten met een vergroot ovarium.\textsuperscript{18-20,26-28} De meeste studies concluderen dat 3D echoscopie een hogere specifiekheid heeft in de voorspelling van maligniteit bij een vergroot ovarium in vergelijking met 2D echoscopie. Echter onderzoek naar een op zich zelf staande test weerspiegelt niet 'real life' in de klinische praktijk. In de klinische praktijk zullen de echoscopie bevindingen altijd geïnterpreteerd worden in het kader van patiëntkenmerken en laboratorium uitslagen. De toegevoegde waarde van 3D echoscopie naast patiënt kenmerken en 2D echoscopie kan alleen beoordeeld worden met multivariabele analyses.\textsuperscript{29}

Verschillende auteurs hebben modellen gepubliceerd betreffende de interpretatie van 2D echoscopie in de klinische praktijk. Ons nieuw ontwikkelde 3D model kon niet
worden vergeleken met bestaande 2D modellen omdat het alleen is gevalidateerd in een dataset waarin het ook is ontwikkeld. Daarom hebben we onze vergelijking beperkt tot het 2D model dat we ontwikkeld hebben in deze studie.

Het toevoegen van 3D echoscopie verhoogt de betrouwbaarheid van de preoperatieve diagnostiek naar maligniteit in een vergroot ovarium. De volgende vraag is of deze verbetering klinisch relevant is. Om deze vraag te kunnen beantwoorden zal een grote prospectieve multicentrum studie uitgevoerd moeten worden met daarin een groter aantal patiënten met maligne tumoren. Deze data zijn nodig om te onderzoeken of verbetering van de diagnostiek met 3D echoscopie leidt tot minder laparotomische procedures, minder vriescoupe-onderzoeken tijdens laparotomie en minder relaparotomieën vanwege een onduidelijke diagnose bij de eerste operatie.

Wat is de nauwkeurigheid van vriescoupe-onderzoek en welke factoren hebben invloed op de nauwkeurigheid van vriescoupe-onderzoek?

Hoofdstuk 6 presenteert een meta-analyse van studies die rapporteren over de betrouwbaarheid van vriescoupe-onderzoek. We verrichten een gecomputeriseerde zoekprocedure in Medline en EMBASE om alle artikelen te identificeren die gepubliceerd zijn tussen januari 1966 en juni 2003 waarin vriescoupe-onderzoek bij ovarium pathologie werd vergeleken met de definitieve histopathologische diagnose. We berekenden van elke geïncludeerde studie de prevalentie van maligne en borderline tumoren. Verder werden de sensitiviteit, specifiekheid en aannemelijkheidsverhoudingen (likelihood ratio's) van vriescoupe-onderzoek berekend, waarbij de definitieve histopathologische uitslag als referentie werd gebruikt. De berekeningen werden op twee manieren uitgevoerd. In de eerste analyse werden de borderline tumoren als maligne beschouwd, terwijl in de tweede analyse borderline tumoren als benigne werden beschouwd.

De 18 geïncludeerde artikelen betroffen retrospectieve cohrefstudies.\textsuperscript{30-46} We vonden geen significant verband tussen studie karakteristieken en de betrouwbaarheid van vriescoupe-onderzoek. Indien de diagnose borderline maligniteit werd beschouwd als maligne, varieerde de sensitiviteit van vriescoupe-onderzoek tussen 65% en 97%. De specifiekheid varieerde tussen 97% en 100%. Indien de diagnose borderline maligniteit werd beschouwd als benigne varieerde de sensitiviteit tussen 71% en 100%. De specificiteit varieerde tussen 98% en 100%.

Er is geen duidelijkheid over het optimale beleid bij borderline maligniteit en daardoor is er veel variatie in het beleid rondom borderline maligniteit in de klinische praktijk. De klinicus dient beleid te voeren afhankelijk van de klinische setting. Indien een borderline maligniteit wordt behandeld door middel van bilaterale salpingoovariectomie, hysterectomie en stadiëring, dient de borderline maligniteit beschouwd te
worden als een maligne proces. Echter indien een borderline maligniteit niet met dergelijke uitgebreide chirurgie behandeld hoeft te worden, heeft het wellicht de voorkeur om de borderline maligniteit te beschouwen als een benigne afwijking.

Indien een patholoog tijdens vriescoupe-onderzoek geen diagnose kan stellen, kan hij er voor kiezen geen uitspraak te doen en te wachten op de definitieve paraffine diagnose. In zes studies werd uitdrukkelijk gerapporteerd dat de diagnose werd uitgesteld na vriescoupe-onderzoek.47-52 Aangezien exclusie van deze casus zou leiden tot bias, werden ze geïncludeerd in de studie. Deze keuze is gerechtvaardigd, als men zich realiseert dat indien vriescoupe-onderzoek niet suggereert dat een tumor maligne is, uitgebreide radicale chirurgie niet te rechtvaardigen is. De aannemelijkheids-verhouding voor maligniteit in geval van een uitgestelde diagnose was 1,9, hetgeen een waarde is die veel te laag is om radicale chirurgie te rechtvaardigen.

We vonden homogeniteit bij de gerapporteerde specificiteit van vriescoupe-onderzoek, echter de schatting voor sensitiviteit voor deze test varieerde tussen 65% en 100%. Een verklaring voor een minder goede sensitiviteit zou de grootte van de tumor kunnen zijn. Het zou kunnen zijn dat een maligniteit niet wordt gedetecteerd in een grote ovariële tumor. Ook in geval van borderline maligniteit of mucineuze tumoren is het moeilijk om de juiste coupe te kiezen en daardoor zal de sensitiviteit van vriescoupe-onderzoek lager zijn. Enkele van de geïncludeerde studies impliceerden een dergelijke verklaring, echter er waren onvoldoende data om dit in de meta-analyse mee te nemen.

Ondanks de bovengenoemde beperkingen vonden we zeer goede testkarakteristieken van vriescoupe-onderzoek. De vraag is of deze test karakteristieken het gebruik van vriescoupe-onderzoek in de dagelijkse praktijk rechtvaardigen. Om deze vraag te beantwoorden, berekenden wij een te verwachten spijt vermindering gebaseerd op de consequenties van een fout positief of fout negatief testresultaat en de prevalentie van maligniteit. Het bleek dat vriescoupe-onderzoek zeer waarschijnlijk bij de meeste patiënten spijt zal verminderen.

Concluderend heeft onze meta-analyse gedemonstreerd dat het verrichten van vriescoupe-onderzoek voor de beoordeling van een ovarium tumor nuttig is, gezien de redelijke sensitiviteit en de bijna perfecte specificiteit. Toekomstige studies naar de voorkeur van patiënten evenals een economische analyse zijn nodig om de definitieve positie van deze diagnostische techniek te bepalen.

Hoofdstuk 7 rapporteert over de resultaten van een retrospectieve studie naar de invloed van tumor grootte op de betrouwbaarheid van vriescoupe-onderzoek. In een kleine ovariële tumor wordt maligniteit bijna altijd gedetecteerd, maar in geval van een grote tumor zou tijdens vriescoupe-onderzoek maligniteit gemist kunnen worden. Omdat deze hypothese niet onderzocht kon worden in onze vorige meta-analyse,

Gedurende de studieperiode werd bij 257 opeenvolgende patiënten met een vergroot ovariun vriescoupé-onderzoek verricht. De definitieve paraffine diagnose liet benigne pathologie zien in 142 casus (55%), maligniteit in 87 casus (34%) en borderline maligniteit in 28 casus (11%). Bij vrouwen met een tumorgrootte < 10 cm, waren de aannemelijkheidsverhoudingen (likelihood-ratio's) voor benigne vriescoupé, borderline vriescoupé, en maligne vriescoupé respectievelijk 0,03 (95% BI 0,01 - 0,23), 1,29 (95% BI 1,7 - 98,2) en oneindig (95% BI 7,1 - ∞). Bij vrouwen met een tumor grootte ≥ 10 cm, waren de aannemelijkheidsverhoudingen voor benigne vriescoupé diagnose, borderline vriescoupé diagnose, en maligne vriescoupé diagnose respectievelijk 0,15 (95% BI 0,08 - 0,25), 4,9 (95% BI 2,1 - 11,4) en oneindig (95% BI 9,7 - ∞).

Deze bevindingen hebben consequenties voor het klinische beleid. Indien een tumor een diameter heeft van meer dan 10 centimeter en het vriescoupé-onderzoek laat een benigne uitslag zien, blijkt 11% (11 / 97) van de vrouwen uiteindelijk een maligne of borderline tumor te hebben. Dit betekent dat, in geval van een tumor met een diameter van meer dan 10 centimeter, één op de negen patiënten aanvullende chirurgie nodig heeft ondanks een eerste negatieve vriescoupé uitslag.

De invloed van de grootte van een ovariëum tumor bij vriescoupé-onderzoek wordt bevestigd in andere studies.33-34 De verklaring hiervoor is te vinden in het basisprincipe van vriescoupé-onderzoek, meer in het bijzonder het probleem van sampling. Het onderzoek tijdens een operatie moet een snelle procedure zijn, waardoor er onvoldoende tijd is om evenveel coupes te beoordelen als tijdens een paraffine procedure.

Behalve de grootte van de tumor kan bij vriescoupé-onderzoek ook het histologische type een probleem zijn. Het is bekend dat mucineuze tumoren de grootste van alle ovariëum tumoren zijn. We berekenen dat het risico op een fout negatieve uitslag met als gevolg een tweede chirurgische ingreep 25% was in geval van mucineuze tumoren ≥ 10 cm. De histologische karakteristieken van mucineuze tumoren maken dit tumortype minder geschikt voor vriescoupé-onderzoek.35 Het probleem van fout negatief vriescoupé-onderzoek in geval van mucineuze tumoren werd al eerder in de
literatuur beschreven. Onze resultaten komen overeen met de literatuur.

Concluderend stellen wij dat een benigne uitslag van vriescoupe-onderzoek van een ovarium tumor ≥ 10 cm minder betrouwbaar is dan bij een ovarium tumor < 10 cm. De diagnose maligniteit in mucineuze tumoren kan lastig zijn bij vriescoupe-onderzoek door de relatief grote omvang en de histologische kenmerken van dit tumortype. Het zou nuttig zijn als de patholoog met vriescoupe-onderzoek zou kunnen differentiëren tussen mucineuze tumoren en niet-mucineuze tumoren, aangezien de niet maligne bevindingen in mucineuze tumoren minder betrouwbaar zijn.

Tegenwoordig is het algemeen geaccepteerd dat de voorkeur van patiënten betrokken moet worden in het medische beleidsplan. Verschillende studies hebben laten zien dat het perspectief van de patiënt voor wat betreft voor- en nadelen van therapie kan verschillen van die van de medische professional. Dit brengt ons tot de vierde vraag van dit proefschrift.

Wat zijn de preferenties van patiënten ten aanzien van vriescoupe-onderzoek en ten aanzien van de manier van opereren?

Hoofdstuk 8 bestudeert de preferenties van patiënten betreffende een fout negatief testresultaat van vriescoupe-onderzoek, dus de situatie waarin de uitslag van vriescoupe-onderzoek benigne is en de paraffine diagnose maligne. Deze situatie impliceert dat een patiënt een tweede maal geopereerd dient te worden om optimale chirurgische stadiëring te bereiken. Naast de risico's van een tweede chirurgische procedure veroorzaakt een fout negatieve uitslag van een vriescoupe ook teleurstelling bij de patiënt die in eerste instantie gerustgesteld was. Het zou een alternatief kunnen zijn om bij alle patiënten met een ovarium tumor radicale primaire chirurgie te verrichten zonder vriescoupe-onderzoek. Dit impliceert dat alle patiënten met een ovarium tumor, inclusief de patiënten bij wie de paraffine uitslag benigne is, een mediane laparotomie ondergaan met beiderzijds adnexextirpatie, hysterectomie en omentectomy, peritoneumsampling en lymfeklier sampling.

Een andere vraag is of een ovarium tumor per laparoscopie dan wel per laparotomie geopereerd dient te worden. Een laparoscopie heeft voordelen zoals korte optredtduur, kleine littekens en snelle werkervatting. In geval van maligniteit zou een laparotomie echter de voorkeur kunnen hebben. Een laparoscopische benadering van ovariumcarcinoom brengt mogelijk negatieve effecten op de prognose met zich mee.

Om de preferentie van patiënten naar de wijze van opereren te bestuderen, hebben we een gestructureerd interview ontwikkeld. Beide dilemma's, dus toepassing van vriescoupe-onderzoek versus primaire radicale chirurgie en laparoscopie versus laparotomie in geval van een ovarium tumor, waren onderwerp van het interview. We interviewden 43 vrouwen.
Hoofdstuk 10

De gemiddelde drempelwaarde waarop de vrouwen hun preferentie voor radicale primaire chirurgie switchten naar vriescoupe-onderzoek was een risico van 50% op een fout negatief testresultaat van het vriescoupe-onderzoek. Vrouwen die gepland werden voor een operatie in verband met een ovarium tumor dienen geïnformeerd te worden over de mogelijkheid van vriescoupe-onderzoek indien er enige verdenking is op maligniteit. Radicale primaire chirurgie dient niet plaats te vinden zonder bewijs van maligniteit.

In geval van een benigne ovarium tumor heeft de laparoscopische benadering de voorkeur boven een laparotomische benadering vanwege het lagere risico op postoperatieve complicaties, minder postoperatieve pijn en de kortere opnemduur.5 In geval van ovariumcarcinoom is er geen eenduidig bewijs of een laparoscopische benadering dan wel een laparotomische benadering de voorkeur heeft. Buikwandmetastasen ter plekke van de trocart-opening na laparosco pie zijn beschreven, echter de invloed op overleving is niet duidelijk.62-63 In dierstudies zijn de negatieve effecten van koolstofdioxide op het peritoneum beschreven met risico op tumordisseminatie, echter tot op heden is hier geen klinisch bewijs voor in geval van ovariumcarcinoom.64 We legden het dilemma van laparoscopie versus laparotomie voor aan patiënten die gepland waren voor chirurgie in verband met een ovarium tumor. Deze patiënten prefereerden laparoscopie boven laparotomie tot een risico van 50% op maligniteit. We zijn ons bewust van het feit dat we een laag risico populatie hebben geïnterviewd. Drie vrouwen hadden een borderline maligniteit van het ovarium en geen van de geïnterviewde vrouwen had een ovariumcarcinoom. Het is mogelijk dat een hoog risico populatie een andere preferentie heeft. Het zou interessant zijn om verder onderzoek te doen naar de preferentie van vrouwen betreffende de chirurgische benadering (laparoscopie versus laparotomie) en te onderzoeken of preferenties veranderen na de diagnose ovariumcarcinoom.

Concluderend hebben vrouwen voorkeur voor een strategie met vriescoupe-onderzoek boven radicale primaire chirurgie en accepteren ze het risico op een fout negatief testresultaat van vriescoupe-onderzoek. Vrouwen met een laag risico op maligniteit hebben een voorkeur voor laparoscopie boven laparotomie.

De afloop van ovariumcarcinoom is beter indien chirurgie wordt uitgevoerd door een gynaecologisch oncoloog dan door een algemeen gynaecoloog. Echter indien alle patiënten met een ovarium tumor geopereerd dienen te worden door een gynaecologisch oncoloog, vereist dit een verandering in de organisatie van de gezondheidszorg en dit genereert bovendien additionele kosten.
Wat zijn de implicaties van gecentraliseerde en reguliere zorg voor vrouwen met een ovariumcarcinoom in Nederland?

In hoofdstuk 9 onderzochten we de implicaties van gecentraliseerde en reguliere zorg voor vrouwen met een ovariumcarcinoom in Nederland. We verrichten een kosteneffectiviteitsanalyse waarin epidemiologische gegevens werden geïntegreerd met gegevens over overleving in relatie tot de chirurgische benadering en de preoperatieve inschatting van de betrouwbaarheid van diagnostiek naar de aanwezigheid van maligniteit in een ovarium tumor. We evaluateerden drie strategieën. In de eerste strategie werden de patiënten geopereerd door een algemeen gynaecoloog (algemene zorg strategie). In de tweede strategie werden de patiënten geopereerd door een gynaecologisch oncoloog (gespecialiseerde zorg strategie). In de derde strategie werd de ovarium tumor preoperatief geëvalueerd met de RMI (diagnostische strategie). Patiënten met een hoog risico op maligniteit werden verondersteld geopereerd te worden in de gespecialiseerde zorg setting, terwijl patiënten met een laag risico op maligniteit werden verondersteld geopereerd te worden in een algemene zorg setting. Alle patiënten werden verondersteld adjuvante chemotherapie te krijgen, behalve patiënten met een stadium I ovariumcarcinoom die geopereerd waren door een gynaecologisch oncoloog.

In Nederland is sinds 1971 een nationaal netwerk van histo- en cytopathologie (PALGA) actief. In de PALGA database vonden we alle 1096 rapporten over de in 2003 in Nederland gediagnosticeerde epitheliaal ovarium carcinomen terug. We includeerden 6502 rapporten over benigne ovarium pathologie. Rapporten over de simpele cysten in premenopauzale vrouwen werden geëxcludeerd (geschat op 5%). We berekenden een prevalentie van epitheliaal ovariumcarcinoom van 14% (1096 / 7598).

Gebaseerd op studies die de invloed van het type chirurg op overleving onderzochten, dus een operatie door een gynaecologisch oncoloog dan wel door een algemeen gynaecoloog, maakten we een schatting van de overleving in het laatst stadium ovariumcarcinoom.65-68 De mediane overleving voor stadium III / IV werd geschat op 15 maanden in de algemene zorg setting en op 22 maanden in de gespecialiseerde zorg setting. De mediane overleving voor stadium I / II werd vastgesteld op 60 maanden voor beide settings.

Voor elke strategie berekenden we de levensverwachting en de incrementele kosten per gewonnen levensjaar. De levensverwachting van een patiënt in de algemene strategie was 2,7 jaar, in de diagnostische strategie 3,0 jaar en in de gespecialiseerde strategie 3,1 jaar. De diagnostische strategie bleek dominant over de algemene strategie. De incrementele kosten om een additionele levensjaar te winnen met gespecialiseerde chirurgie vergeleken met de diagnostische strategie, waren € 63.477 per gewonnen levensjaar.
We moesten enkele aannameën doen, omdat we de werkelijke kosten en overlevingsdata van de cohort niet wisten. Echter ondanks enkele onzekerheden betreffende aannameën in deze economische analyse, illustreerden we duidelijk dat de kosten in de algemene strategie (alle patiënten met een ovarium tumor geopereerd door een algemeen gynaecoloog) duidelijk lager waren dan de kosten in de gespecialiseerde strategie (alle patiënten met een ovarium tumor geopereerd door een gynaecologisch oncoloog) ten koste van overleving. Aan de andere kant zou behandeling van alle ovariële tumoren door een gynaecologisch oncoloog erg duur zijn.

Grevings e.a. verrichtten een kosteneffectiviteitsanalyse betreffende centralisering van ovariumcarcinoom op basis van actuele patiëntendata. Grevings vond dat behandeling van ovariumcarcinoom patiënten (vroeg stadium en laat stadium) het meest kosten effectief was in een semigespecialiseerde ziekenhuis setting. Behandeling van ovariumcarcinoom in semigespecialiseerde ziekenhuizen leverde 0,15 QALY op, vergeleken met behandeling in algemene ziekenhuizen. Dit resulteerde in een incrementele kosten effectiviteitsratio van € 5700. Behandeling van patiënten in tertiaire verwijzingscentra resulteerde in een incrementele kosten effectiviteitsratio van € 93.700. Zodra het aandeel optimale debulking procedures in gespecialiseerde ziekenhuizen zou toenemen van 46% naar 70%, zou de zorg voor het ovariumcarcinoom in gespecialiseerde ziekenhuizen kosten effectief worden (incrementele kosten effectiviteits ratio van € 25.000). In deze studie werden de kosten van ontrecht verwezen patiënten met een benigne ovarium tumor naar gespecialiseerde ziekenhuizen (fout positieve beoordeling van een ovarium tumor) en het verlies in overleving ten gevolge van ontrecht niet verwezen patiënten met een maligne ovarium tumor (fout negatieve beoordeling van een ovariumtumor) niet geïncludeerd.

We beoordeelden de implicaties van gecentraliseerde en reguliere zorg voor patiënten met een ovariumcarcinoom in Nederland en lieten zien dat gecentraliseerde zorg kosten effectief is op voorwaarde dat de ovarium tumor preoperatief beoordeeld is op maligniteitsrisico.

**Klinische implicaties**

Dit proefschrift behandelt het diagnostisch proces van maligniteit in een ovarium tumor, zowel preoperatief als perioperatief. Voor de preoperatieve beoordeling van een ovarium tumor zijn verschillende mathematische modellen beschreven in de literatuur en een aantal daarvan zijn extern gevalideerd. Intuïtieve inschatting van risico's door ervaren klinici bleek echter even goed als voorspelling door mathematische modellen. Indien de voorkeur uitgaat naar een mathematisch model, adviseren we om de RMI I van Jacobs te gebruiken omdat dit model de beste testkarakteristieken heeft.
Voor perioperatieve beoordeling adviseren we gebruik te maken van vriescoupé-onderzoek vanwege de betrouwbaarheid van dit onderzoek. In geval van grote tumoren (≥ 10 cm) en speciaal in geval van mucineuze tumoren met een diameter ≥ 10 cm, is de diagnostische betrouwbaarheid lager en moet de patiënt preoperatief geïnformeerd worden over een risico op een fout negatief testresultaat van het vriescoupé-onderzoek. Patiënten met een ovariële tumor hebben voorkeur voor vriescoupé-onderzoek boven primaire radicale chirurgie, zelfs in geval van een verhoogd risico op een fout negatief testresultaat.

Indien er de verdenking is op maligniteit in een ovarium tumor, dient de patiënt geopereerd te worden door een gynaecologisch oncoloog. Het bleek te duur te zijn om alle patiënten met een ovarium tumor te verwijzen naar een gynaecologisch oncoloog. Echter, indien een algemeen gynaecoloog een ovarium carcinoom opereert, is de overleving slechter. Gecentraliseerde zorg is kosten effectief op voorwaarde dat preoperatieve beoordeling van een ovarium tumor plaatsvindt.

Toekomstig onderzoek

In hoofdstuk 3 hebben we laten zien dat momenteel klinische beoordeling even goed is als het gebruik van mathematische modellen. In dit proefschrift hebben we ook gerapporteerd dat een nieuw diagnostisch instrument, namelijk de 3D echoscopie, de diagnostische nauwkeurigheid verbeterde. We vonden dat een model met daarin opgenomen gegevens verkregen met de 3D echoscopie, beter maligniteit voorspelde dan een model bestaande uit gegevens verkregen met 2D echoscopie. Echter dit model is niet extern gevalideerd, en bovendien betrof onze studie slechts een kleine groep patiënten. We beoordeelden negen parameters in ons 3D model. Als algemene regel wordt aangehouden dat ongeveer tien casus van maligniteit nodig zijn voor elke parameter die wordt onderzocht. Dit impliceert dat ongeveer 90 maligne tumoren nodig zijn. Uit onze meta-analyse over bestaande diagnostische modellen hebben we geleerd dat de RMI van Jacobs een simpel en accuraat diagnostisch model is. Een volgende stap in het onderzoek naar de waarde van de 3D echoscopie is een prospectieve multicentrum studie met een grote studie populatie om ons 3D model te valideren en te vergelijken met de prestatie van Jacobs' RMI.

Verschillende studies concluderen dat de behandeling van ovariumcarcinoom zou moeten worden gecentraliseerd om de overleving van Nederlandse patiënten met ovariumcarcinoom te verbeteren. Patiënten met een vroeg stadium ovariumcarcinoom worden vaker adequaat gestadiëerd door een gynaecologisch oncoloog dan door een algemeen gynaecoloog. Echter de procedure van adequate stadiëring op zich heeft geen invloed op de overleving. Adequate stadiëring differentieert tussen patiënten met een
werkelijk stadium I ovariumcarcinoom en patiënten met een occult stadium III ovariumcarcinoom. Vernooij rapporteerde in haar proefschrift dat in Nederland slechts 24% van de patiënten met een vroeg stadium ovariumcarcinoom adequaat gestadiëerd werd indien de operatie was uitgevoerd in een algemeen ziekenhuis, vergeleken met 60% adequaat gestadiëerde patiënten in (semi) gespecialiseerde ziekenhuizen.\textsuperscript{70} Heintz onderstreepte in een recent artikel in Medisch Contact dat adequate stadiëring een kwestie is van het beheersen van chirurgische vaardigheden en het uitvoeren van een protocol.\textsuperscript{71} In een gespecialiseerd ziekenhuis zijn de benodigde chirurgische vaardigheden continu beschikbaar, maar dat is niet het geval in een algemeen ziekenhuis. Indien er preoperatief een verdenking is op ovariumcarcinoom kunnen patiënten uit een algemeen ziekenhuis verwezen worden naar een gespecialiseerd ziekenhuis. Ook zou een gynaecologisch oncoloog kunnen worden uitgenodigd om de patiënt in een algemeen ziekenhuis te opereren. Echter indien er (ten onrechte) geen verdenking is op maligniteit, zal de patiënt uit een algemeen ziekenhuis door een algemeen gynaecoloog worden geopereerd met als gevolg een niet adequate stadiëring. In geval van een vroeg stadium ovariumcarcinoom is het niet nodig om de vaardigheden van een gynaecologisch oncoloog te verbeteren, maar moet de preoperatieve beoordeling van een ovarium tumor worden geoptimaliseerd.

In geval van een laat stadium ovarium carcinoom is complete of optimale debulking (een rest tumor of resttumor met een doorsnede kleiner dan één centimeter) een belangrijke prognostische factor voor overleving. Vernooij liet zien dat patiënten die werden geopereerd door laag-volume gynaecologen (> één ovariumcarcinoom operatie per twee maanden), in 37% optimaal werden gedebulked, terwijl hoog volume gynaecologen (> twee ovarium carcinoom operaties per twee maanden) een optimale debulking bereikten in 57%. Patiënten met een laat stadium ovariumcarcinoom dienen daarom geopereerd te worden door hoog volume gynaecologen. Dit impliceert dan weer dat de chirurgische behandeling van patiënten met een ovariumcarcinoom gecentraliseerd moet worden.

In hoofdstuk 9 hebben we geïllustreerd dat een diagnostische strategie voorafgaande aan de beslissing tot chirurgie door een algemeen gynaecoloog of een gynaecologisch oncoloog, de beste balans biedt tussen kosten en effectiviteit. De PROBAAT studie (Radboud Universitair Medisch Centrum, Nijmegen, Nederland) onderzocht of de RMI een accurate test is om te beslissen of de consultatie van een gynaecologisch oncoloog nodig is. Helaas werden de patiënten niet consecutief geïncludeerd, maar was inclusie afhankelijk van de algemeen gynaecoloog die de beslissing nam om al dan niet de RMI van een patiënt te een ovariumtumor te berekenen. In toekomstig onderzoek naar de consequenties van gecentraliseerde zorg voor ovariumcarcinoom, zou selectie van patiënten die voordeel zouden kunnen hebben van consultatie van een gynaecologisch oncoloog, gebaseerd moeten zijn op een accurate diagnostische test. De patiënten
zouden consecutief moeten worden geïncludeerd om tot een representatieve onderzoekspopulatie te komen.

References


62. Kruitwagen RF, Swinkels BM, Keyser KG, Doesburg WH, Schijff CP. Incidence and effect on survival of abdominal wall metastases at trocar or puncture sites following laparoscopy or paracentesis in women with ovarian cancer. Gynecol Oncol 1996; 60:233-7.
68. Vermeij F, Heintz P, Witteveen E, van der GY. The outcomes of ovarian cancer treatment are better when provided by gynaecologic oncologists and in specialized hospitals: a systematic review. Gynecol Oncol 2007; 103:801-12.
Chapter 11

Dankwoord
Dankwoord

Als eerste wil ik alle patiënten bedanken, die deel hebben genomen aan het onderzoek naar de waarde van de 3D echo. Patiënten met een ovariumtumor werden de avond voor hun operatie nog meegenomen naar een andere afdeling biometrie voor een 3D echo. Daarnaast kwamen er patiënten, vaak vergezeld door hun partner, met de taxi vanuit Helmond en Tilburg voor extra echo-onderzoek naar Veldhoven. Deze vrouwen waren op dat moment nog onwetend over hun diagnose en uiteraard erg ongerust. Goede preoperatieve counseling gebaseerd op een zorgvuldige inschatting op het risico op ovariumcarcinoom is voor deze patiëntengroep van groot belang.

Prof. dr. B.W.J. Mol, lieve Ben Willem, ik leerde je kennen toen jij nog in opleiding was en ik co-assistent. Vervolgens kwam ik in opleiding en werd jij gynaecoloog en heel even zaten we samen in de maatschap. Nu ben jij hoogleraar en mag ik als eerste bij je promoveren. Daar ben ik bijzonder trots op. Jouw mailtjes en telefoontjes op de meest onverwachte tijdstippen hebben ervoor gezorgd dat ook dit boekje af kwam. Ben Willem, wat jij voor elkaar krijgt is onvoorstellbaar. Dank voor alles wat jij voor mij hebt gedaan.

Prof. dr. J.L.H. Evers, dank voor de ruimte die ik in Maastricht kreeg voor het afronden van mijn proefschrift.

Prof. dr. R.F.P.M. Kruitwagen, lieve Roy, dank dat je altijd reageerde op elke mail die ik rond stuurde. Je kwam altijd met een kritische vraag en suggesties voor nieuwe literatuur. Gefeliciteerd met je gloednieuwe hoogleraarschap in Maastricht. Ik verheug me op de verdere onderzoekssamenwerking.

Gérard Bremer, lieve Gérard, jij was vanaf het allereerste begin betrokken bij mijn wetenschappelijke carrière. De evaluatie van het Halfway Grading System leidde tot grote hilariteit op de poli, de sacropexie werd geëvalueerd, hormonale suppletie therapie kwam even om de hoek en nu ligt er een boekje over de diagnostiek van adenopathologie. Dank voor al je hulp.

Prof. dr. G.G.M. Essed, Prof. dr. R.G.H. Beets- Tan, Prof. dr. M.P.M. Burger, Dr. H.J. Hoogland en Prof. dr. D. Timmerman dank voor het beoordelen en goedkeuren van mijn manuscript. Prof. dr. M.J.E. Mourits en Prof. dr. H.A.M. Bröllmann, beste Marian en Hans, super dat jullie helemaal naar Maastricht komen de 19e.
Dankwoord


Kirsten Kluivers, opleidingsmaatje en onderzoeksmaatje. We hebben samen heel wat verzucht. Maar nu hebben we het allebei voor elkaar! Dank voor alle echo's die je voor mij maakte als ik er weer eens niet was. Elk jaar samen naar een congres is de afspraak!

Evelien Moret, als ik dacht dat we een adnex vergeten waren, bleek jij hem al te hebben opgeslagen. Ik waardeer jouw warme hart voor al onze patiënten enorm. Ik stel voor om de afdeling biometrie om te dopen in biomoretrie.

Maatschap gynaecologie en verloskunde van Máxima Medisch Centrum. Ik noem ons altijd de leukste maatschap van Nederland en dat is ook zo. Het bijzondere is dat we elkaar respecteren en waarderen in wat we doen en laten en elkaar aanvullen waar kan. Dank voor de ruimte die jullie me gaven om mijn boekje af te ronden.

Assistenten gynaecologie van MMC. Dank voor al jullie belangstelling, mijn boekje is af, dat van jullie gaat ook af komen. Het motto van Ben Willem "een dag niet gerandomiseerd is een dag niet geleefd" baart mij geen zorgen, mede dankzij jullie leven we elke dag!

Sjors, ook toen je in Amsterdam zat was je betrokken bij mijn onderzoek en bood je hulp bij ingewikkeld statistiek en het maken van figuren. Het was heerlijk om samen te zuchten over hoe vreselijk het is om een artikel te submitten, maar ook hoe super het gevoel is als je artikel wordt gecentreerd! Dankjewel.


Marjine Bremer, lieve Marjine, dank voor het afnemen van alle interviews. Je nam de tijd om de patiënten op hun gemak te stellen en had een luisterend oor voor al hun vragen, een aanvulling op het consult bij de gynaecoloog. Jammer dat de interviews zijn afgelopen.
Hoofdstuk 11

Dierk Hendriks, lieve Dierk, als herinnering aan mijn vader wilde ik graag dat jij als fotograaf mijn kaftje zou maken. Het duurde even voordat we goede ideeën kregen, maar uiteindelijk is het super geworden! Dank je.

Nol van Montfort, dank voor je kritische blik op de Nederlandse samenvatting. Je opmerkingen heb ik zeer gewaardeerd, ik ben blij dat je de 19e geen vragen mag stellen!

Henk Dinnissen, dank voor het omtoveren van mijn "world bestand" tot dit prachtige boekje.

Lieve mama, ik hoef jou maar te bellen en je komt me helpen. Mede namens Patrick, Youp en Teun, dank je dat je er altijd bent voor ons. Wat hadden we graag gehad dat we dit met papa konden delen.

Lieve Youp en Teun, liefste vriendjes. Jullie kwamen het liefste bij me op schoot zitten als ik achter de computer kroop. Het is super om jullie mama te zijn. Lieve Patrick, jij bent een kanjer in relativeren, maar als er iemand blij is dat dit boekje af is ben jij het wel! Dank je voor wie je bent en dat je er altijd bent voor mij.
Curriculum Vitae


Peggy Geomini was born on the 27th of June in Blerick, The Netherlands. After finishing secondary school in 1991, she was not admitted to the study of Medicine and studied Health Science for one year at the University of Maastricht, The Netherlands. Fortunately, she was allowed to start Medicine in 1992 at the University of Maastricht. After her medical finals in 1998, she worked as a resident at the department of Obstetrics and Gynaecology in Máxima Medical Centre in Veldhoven, The Netherlands. She started her specialist training in 2001, firstly at the department of Obstetrics and Gynaecology of Máxima Medical Centre in Veldhoven (head of the department Prof. H.A.M. Bröllmann MD PhD and Prof. S.G. Oei MD PhD) and afterwards at the University Hospital of Maastricht, The Netherlands (head of the department Prof. J.L.H. Evers MD PhD and Prof. G.G.M. Essed MD PhD). During her specialist training, she started research work, which resulted in this thesis. Since April 2006, she has been working as a gynaecologist in Máxima Medical Centre in Veldhoven. Peggy Geomini has special interest in diagnostic and therapeutic endoscopy and oncology.