

RESEARCH

Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study

 OPEN ACCESS

Ben Van Calster *professor*¹, Kirsten Van Hoorde *doctoral researcher*^{2,3}, Lil Valentin *professor*⁴, Antonia C Testa *professor*⁵, Daniela Fischerova *consultant gynaecologist*⁶, Caroline Van Holsbeke *consultant gynaecologist*⁷, Luca Savelli *consultant gynaecologist*⁸, Dorella Franchi *consultant gynaecologist*⁹, Elisabeth Epstein *professor*¹⁰, Jeroen Kaijser *research fellow*^{1,11}, Vanya Van Belle *postdoctoral researcher*^{2,3}, Artur Czekierdowski *professor*¹², Stefano Guerriero *professor*¹³, Robert Fruscio *consultant gynaecologist*¹⁴, Chiara Lanzani *consultant gynaecologist*¹⁵, Felice Scala *consultant gynaecologist*¹⁶, Tom Bourne *professor*^{1,11,17}, Dirk Timmerman *professor*^{1,11}, International Ovarian Tumour Analysis (IOTA) group

¹Department of Development and Regeneration, KU Leuven, Herestraat 49 box 7003, 3000 Leuven, Belgium; ²Department of Electrical Engineering, KU Leuven, Leuven, Belgium; ³iMinds Medical Information Technologies, KU Leuven, Leuven, Belgium; ⁴Department of Obstetrics and Gynaecology, Skåne University Hospital Malmö, Lund University, Malmö, Sweden; ⁵Department of Oncology, Catholic University of the Sacred Heart, Rome, Italy; ⁶Gynaecological Oncology Center, Department of Obstetrics and Gynaecology, Charles University, Prague, Czech Republic; ⁷Department of Obstetrics and Gynaecology, Ziekenhuis Oost Limburg, Genk, Belgium; ⁸Gynaecology and Reproductive Medicine Unit, S Orsola-Malpighi Hospital, University of Bologna, Bologna, Italy; ⁹Preventive Gynaecology Unit, Division of Gynaecology, European Institute of Oncology, Milan, Italy; ¹⁰Department of Obstetrics and Gynaecology, Karolinska University Hospital, Stockholm, Sweden; ¹¹Department of Obstetrics and Gynaecology, University Hospitals Leuven, Leuven, Belgium; ¹²1st Department of Gynaecological Oncology and Gynaecology, Medical University in Lublin, Lublin, Poland; ¹³Department of Obstetrics and Gynaecology, Azienda Ospedaliero Universitaria di Cagliari, Cagliari, Italy; ¹⁴Clinic of Obstetrics and Gynaecology, University of Milan-Bicocca, San Gerardo Hospital, Monza, Italy; ¹⁵Department of Woman, Mother and Neonate, Buzzi Children's Hospital, Biological and Clinical School of Medicine, University of Milan, Milan, Italy; ¹⁶Department of Gynaecologic Oncology, Istituto Nazionale Tumori, Naples, Italy; ¹⁷Queen Charlotte's and Chelsea Hospital, Imperial College, London, UK

Abstract

Objectives To develop a risk prediction model to preoperatively discriminate between benign, borderline, stage I invasive, stage II-IV invasive, and secondary metastatic ovarian tumours.

Design Observational diagnostic study using prospectively collected clinical and ultrasound data.

Setting 24 ultrasound centres in 10 countries.

Participants Women with an ovarian (including para-ovarian and tubal) mass and who underwent a standardised ultrasound examination before

surgery. The model was developed on 3506 patients recruited between 1999 and 2007, temporally validated on 2403 patients recruited between 2009 and 2012, and then updated on all 5909 patients.

Main outcome measures Histological classification and surgical staging of the mass.

Results The Assessment of Different NEoplasias in the adneXa (ADNEX) model contains three clinical and six ultrasound predictors: age, serum CA-125 level, type of centre (oncology centres v other hospitals), maximum diameter of lesion, proportion of solid tissue, more than 10 cyst locules, number of papillary projections, acoustic shadows, and

Correspondence to: B Van Calster ben.vanecalster@med.kuleuven.be

Extra material supplied by the author (see <http://www.bmj.com/content/349/bmj.g5920?tab=related#datasupp>)

Supplementary information, tables, and figures

ascites. The area under the receiver operating characteristic curve (AUC) for the classic discrimination between benign and malignant tumours was 0.94 (0.93 to 0.95) on temporal validation. The AUC was 0.85 for benign versus borderline, 0.92 for benign versus stage I cancer, 0.99 for benign versus stage II-IV cancer, and 0.95 for benign versus secondary metastatic. AUCs between malignant subtypes varied between 0.71 and 0.95, with an AUC of 0.75 for borderline versus stage I cancer and 0.82 for stage II-IV versus secondary metastatic. Calibration curves showed that the estimated risks were accurate.

Conclusions The ADNEX model discriminates well between benign and malignant tumours and offers fair to excellent discrimination between four types of ovarian malignancy. The use of ADNEX has the potential to improve triage and management decisions and so reduce morbidity and mortality associated with adnexal pathology.

Introduction

Ovarian cancer is the most aggressive gynaecological malignancy. The five year survival rate of patients is around 40% and the disease accounts for approximately half of all deaths related to gynaecological cancer.^{1,2} The most important factor for survival is stage at diagnosis.³ Therefore attempts have been made to develop a screening method, which by detecting ovarian cancer at an early stage has the potential to decrease deaths from ovarian cancer. No such screening method is currently available.^{4,5} However, we are still awaiting the results of the United Kingdom Collaborative Trial on Ovarian Cancer Screening.⁶

An important factor that influences prognosis other than stage at diagnosis is referral to a gynaecology oncology centre for further diagnosis or staging, debulking surgery, and evaluation by an interdisciplinary tumour board.⁷⁻¹⁰ Although such centralised care is recommended because it results in improved prognosis, a large proportion of women with ovarian cancer remain treated by general surgeons,¹¹⁻¹³ possibly because the true nature of the disease is unknown before surgery. Optimal treatment of ovarian malignancies depends on the type of tumour. Treatment of borderline tumours can be less aggressive than treatment of invasive tumours, especially if the preservation of fertility is important.¹⁴ In selected cases, stage I ovarian cancer may be managed more conservatively than late stage disease, whereas for cancers metastasised to the ovary management depends on the origin of the primary tumour.¹⁵ An accurate specific diagnosis of adnexal tumours before surgery will almost certainly improve the triage of patients and so increase the likelihood that patients will receive appropriate treatment.

Recently, the International Ovarian Tumour Analysis (IOTA) group showed that polytomous risk prediction for the diagnosis of ovarian cancer is feasible.¹⁶ Mathematical models were developed to predict four tumour categories: benign, borderline, primary ovarian cancer, and secondary metastatic cancer. This work focused on comparing mathematical algorithms. From a clinical point of view it was preliminary for several reasons. Firstly, the model was built using information from only 754 patients with 40 borderline, 121 primary invasive, and 30 secondary metastatic cancers. Secondly, despite that more than 30 clinical and ultrasound candidate predictors were statistically evaluated, the tumour marker serum CA-125 was not considered. Although we have shown that serum CA-125 may not be needed in models with a binary outcome (benign v malignant),¹⁷ CA-125 is likely to be important for distinguishing between different types of malignant tumour.¹⁸ Thirdly, the models did not distinguish between stage I and stage II-IV primary cancer, which is clinically important.¹⁹

We developed a polytomous risk prediction model that can reliably distinguish between benign, borderline, stage I invasive, stage II-IV invasive, and secondary metastatic adnexal tumours.

Methods

Design and setting

We carried out an international multicentre prospective cohort study of women with at least one adnexal mass that required surgery, as judged by a clinician. The IOTA study group collected data between 1999 and 2012. IOTA was established to develop and validate diagnostic models for adnexal masses based on large multicentre datasets using a standardised ultrasound examination protocol, terms, and definitions.²⁰⁻²⁶ Patients were recruited from 24 centres in 10 countries. Twelve centres were labelled oncology centres, that is, tertiary referral centres with a specific gynaecology oncology unit. The remaining centres included general hospitals and gynaecology ultrasound units not linked to an oncology centre. Data collection was carried out in phases: phase 1 between 1999 and 2002, phase 1b between 2002 and 2005, phase 2 between 2005 and 2007, and phase 3 between 2009 and 2012.²¹⁻²⁴

Patients

Patients referred to one of the participating centres for an ultrasound examination because of a known or suspected adnexal mass were eligible for inclusion. We included consecutive patients with at least one adnexal mass judged not to be a physiological cyst, who were examined with transvaginal ultrasound by a principal investigator and later selected for surgical intervention. The decision to operate was made by the managing clinician on the basis of the full clinical picture, including the ultrasound report, the latter being based on the ultrasound examiner's subjective assessment of the ultrasound image. Following the requirements of the local ethics committees, we obtained oral or written informed consent from the women before their ultrasound scan and surgery. Exclusion criteria were refusal for transvaginal ultrasonography, pregnancy at the time of presentation, and surgical removal of the mass more than 120 days after the ultrasound examination. If more than one mass was detected, we used the mass with the most complex morphology on the ultrasound scan. When we observed masses with similar morphology, we used the largest or the one most easily accessible by ultrasound.²¹⁻²³

Data collection and reference standard

To collect clinical information we took a standardised history from each patient. All patients underwent a standardised transvaginal ultrasound examination.²⁰ Transabdominal sonography was added for women with large masses that could not be visualised in full by a transvaginal probe. We collected gray scale and Doppler ultrasound information in line with the research protocols. More information can be found in previous reports.²¹⁻²³ Participating centres were encouraged to measure serum CA-125. We used second generation immunoradiometric assay kits for CA-125 II from Roche Diagnostics, Centocor, Cis-Bio, Abbott Laboratories, Bayer Diagnostics, bioMérieux, DiaSorin, Siemens, and Beckman Coulter. All kits used the OC125 antibody.

The reference standard was the histopathological diagnosis of the mass after surgical removal by laparotomy or laparoscopy as considered appropriate by the surgeon, and the stage of malignant tumours using the classification of the International Federation of Gynecology and Obstetrics (FIGO).²⁷ The excised

tissues underwent histological examination at the local centre. Histological classification was performed without knowledge of the ultrasound results. The final diagnosis was divided into five tumour types: benign, borderline, stage I invasive, stage II-IV invasive, and secondary metastatic cancer.

Data were entered through dedicated and secure data collection systems, web based for phase 1, and through a local study screen (Astraia software, Munich, Germany) for later phases.²¹⁻²³ To ensure data integrity, several clinicians and statisticians used built-in automatic checks and manual review and cleaning of data.

Statistical analysis

We developed a prediction model using data from the women included in IOTA phases 1, 1b, and 2 (n=3506) and validated the model on data from the women included in phase 3 (n=2403).

The serum CA-125 tumour marker was not a mandatory variable, and measurements were missing in 31% of the patients. As described in detail in supplementary appendix A, we used multiple imputation to deal with missing values for CA-125.²⁸ We created 100 imputations, resulting in 100 completed datasets.

We selected variables in two stages (see supplementary appendix B for details). Firstly, to avoid over-fitting we reduced the number of potential predictors to 10 based on subject matter knowledge²⁹⁻³⁰ and the stability of the predictors over centres.³¹ We selected four clinical variables—age (years), serum CA-125 level (U/mL), family history of ovarian cancer (yes/no), and type of centre (oncology centre v other hospitals), and six ultrasound variables—the maximum diameter of the lesion (mm), proportion of solid tissue (that is, the maximum diameter of the largest solid component divided by the maximum diameter of the lesion), presence of more than 10 cyst locules (yes/no), number of papillary projections (0, 1, 2, 3, >3), presence of acoustic shadows (yes/no), and presence of ascites (yes/no). Oncology centres were defined as tertiary referral centres with a specific gynaecology oncology unit. We included the variable “type of centre” because the risk of a malignant tumour is likely to be higher in oncology centres than in other centres, even after adjustment for the characteristics of patients and tumours. Secondly, we carried out further data driven selection using a method based on multivariable fractional polynomials.³² This method simultaneously selects variables and determines the optimal transformation of numerical variables using fractional polynomials. We forced age and type of centre into the model by default.

To acknowledge variability between centres we used multinomial logistic regression with random centre intercepts to construct the polytomous model.³³ We multiplied the predictor coefficients with uniform “shrinkage factors” to avoid exaggerated model coefficients (see supplementary appendix C for details).³⁰⁻³⁴ We trained the model on each of the 100 completed datasets following multiple imputation. Probabilities were derived by averaging linear predictors (without the random effects) and odds ratios by averaging model coefficients.

We evaluated the model for discrimination and calibration performance.³⁵ To assess discrimination we first obtained the area under the receiver operating characteristic curve (AUC) for the basic discrimination between benign and malignant tumours. We calculated sensitivity and specificity for the cut-offs 3%, 5%, 10%, and 15% total risk of malignancy (that is, the sum of the estimated risks of the four malignant subtypes). We then also computed AUCs for each pair of tumour types using the conditional risk method.³⁶ For the five tumour types,

there are 10 pairwise AUCs. Finally, we calculated the polytomous discrimination index, a polytomous version of the AUC.³⁷ This index estimates the average proportion of patients who are correctly identified by the model when presented with five patients, one with each tumour type. For five groups, the polytomous discrimination index ranges between 0.20 (worthless) and 1 (perfect). A discrimination plot was used to visualise discrimination performance.³⁶

To assess calibration of the predicted probabilities we produced calibration plots showing the relation between predicted and observed probabilities for each type of tumour. The plots were based on a parametric multinomial logistic recalibration analysis,³⁸ using random centre intercepts. We used the probabilistic results of this analysis, including the random effects, as observed probabilities, which were plotted against the predicted probabilities.

Because model validation was successful, we updated the model on the pooled data (n=5909) to make full use of all available information. Predicted probabilities based on this model can then be compared with baseline probabilities for each type of tumour. The baseline probabilities were estimated through a random intercepts multinomial logistic regression model containing only intercept terms. All analyses were performed with SAS 9.3 (SAS Institute, Cary, USA).

Results

In total, data on 6169 patients were recorded in the databases for phases 1, 1b, 2, and 3. We excluded 255 patients (4.1%): 163 (2.6%) based on exclusion criteria (51 pregnant women, 112 women received surgery >120 days after the ultrasound examination), 91 (1.5%) because of data errors or uncertain or missing final histology, and one due to protocol violation. Based on logistic regression influence diagnostics³⁹ and further data review of the archived datasets, we omitted five additional cases. Thus data on 5909 women were used. Table 1⇓ gives an overview of participating centres, included patients, and the reference standard; supplementary table S1 the histological diagnoses and FIGO stages; and supplementary table S2 the personal and reproductive characteristics of the patients. The observed rate of malignancy varied between 22% and 66% in oncology centres and between 0% and 30% in other hospitals.

Model development, temporal validation, and updating

We included nine variables in the Assessment of Different NEoplasias in the adneXa (ADNEX) model: age, serum CA-125 level (log transformed), type of centre, maximum diameter of the lesion (log transformed), proportion of solid tissue (with quadratic term), number of papillary projections, more than 10 cyst locules, acoustic shadows, and ascites. Family history of ovarian cancer was dropped by the variable selection analysis. Table 2⇓ shows descriptive statistics for the 10 variables selected a priori. The AUC of the ADNEX model for the basic discrimination between benign and malignant tumours was 0.954 (95% confidence interval 0.947 to 0.961) on the development data and 0.943 (0.934 to 0.952) on the validation data (table 3⇓). The discrimination between benign and malignant was consistent over centres (see supplementary figure S1). Using a cut-off of 10% to predict malignancy, the sensitivity was 96.5% and specificity 71.3% on the validation data (table 3). The validation AUC was 0.85 for benign tumours compared with borderline tumours, 0.92 for benign tumours compared with stage I cancer, 0.99 for benign tumours compared with stage II-IV cancer, and 0.95 for benign tumours compared with

secondary metastatic cancer (table 4). Validation AUCs between malignant subtypes varied between 0.71 and 0.95. The model showed fair discrimination between stage I cancer and borderline tumours (validation AUC 0.75) and between stage I cancer and secondary metastatic cancer (validation AUC 0.71). It was well able to distinguish stage II-IV cancer from other malignancies (AUCs for stage II-IV cancer versus borderline tumours was 0.95, versus stage I cancer was 0.87, and versus secondary metastatic cancer was 0.82). The polytomous discrimination index was 0.56 (0.54 to 0.59) on the validation data. Supplementary table S3 presents separate results for oncology centres and other hospitals.

The calibration plots for all five tumour types showed acceptable calibration of the estimated risks (fig 1). High risks for secondary metastatic cancer were overestimated, but such high risks were uncommon. Calibration plots for oncology centres and other hospitals were similar (see supplementary figures S2 and S3).

Tables 3 and 4 and supplementary table S3 show the discrimination performance of the ADNEX model after it was updated on the pooled data. The discrimination plot shows that the predicted probability of a specific tumour type is highest for patients with a matching reference standard (fig 2)—for example, patients with histologically confirmed borderline tumours had the highest probabilities of a borderline malignancy. The ADNEX model formula is given in supplementary appendix D. The effects of the predictors are presented as odds ratios in table 5. Proportion of solid tissue and serum CA-125 level had the strongest independent relations with the outcome, as judged by the test statistic for the model coefficients (not shown). Type of centre was the weakest predictor, indicating that most of the differences in malignancy rates were captured by the other predictors.

Deriving a similar model without CA-125 level as a predictor mainly affected discrimination between stage II-IV cancer and other malignancies (see supplementary table S4): validation AUCs decreased from 0.82 to 0.59 (stage II-IV cancer *v* metastatic cancer), from 0.87 to 0.76 (stage II-IV cancer *v* stage I cancer), and from 0.95 to 0.91 (stage II-IV cancer *v* borderline tumours).

Implementation of ADNEX and illustrative example

The final ADNEX model is available online and in mobile applications (www.iotagroup.org/adnexmodel/). The applications allow risk calculation even without information on serum CA-125 level, despite the decrease in performance. As an example, we assess a 55 year old woman at a centre for gynaecological oncology. Her serum CA-125 level is 42 U/mL. Ultrasound examination reveals an adnexal mass with more than 10 cyst locules, no papillary projections, no acoustic shadows, ascites, a maximum lesion diameter of 120 mm, and a maximum diameter of the largest solid component of 20 mm (that is, proportion of solid tissue is 20/120). The ADNEX model gives the following probabilities: 37.4% for borderline tumour, 10.8% for stage I cancer, 8.4% for stage II-IV cancer, and 11.0% for secondary metastatic cancer. The total risk of malignancy is $37.4+10.8+8.4+11.0=67.6\%$. The tumour is most likely to be a borderline tumour as opposed to any other type of malignancy. If the CA-125 level was unavailable, predicted probabilities would be 25.2% (borderline), 8.3% (stage I), 35.8% (stage II-IV), and 11.5% (metastatic). Baseline probabilities for each type of tumour are 6.3% for borderline tumour, 7.5% for stage I, 14.1% for stage II-IV, and 4.0% for metastatic cancer.

Discussion

We developed and temporally validated a prediction model that is able to discriminate between five types of adnexal tumour (benign, borderline, stage I cancer, stage II-IV cancer, and secondary metastatic cancer), while still showing excellent overall discriminative capacity between benign and all malignant tumours. On the validation data, the previously proposed 10% risk cut-off for the total risk of malignancy²¹ resulted in 96.5% sensitivity and 71.3% specificity. The ADNEX model discriminated well between benign tumours and each of four types of malignancy (validation area under the receiver operating characteristic curves (AUCs) between 0.85 and 0.99). Moreover, the model was able to distinguish stage II-IV cancer from other malignancies (validation AUCs between 0.82 and 0.95) and showed fair discrimination between stage I cancer and borderline tumours (AUC 0.75) and stage I cancer and secondary metastatic cancer (AUC 0.71). The model uses three clinical predictors (age, serum CA-125 level, type of centre) and six ultrasound predictors (maximal diameter of lesion, proportion of solid tissue, more than 10 cyst locules, number of papillary projections, acoustic shadows, and ascites). Serum CA-125 level and proportion of solid tissue were the strongest predictors.

Results in relation to other studies

The polytomous approach to adnexal tumour diagnosis is novel. We do not know of multivariable polytomous models in this area outside the work of the International Ovarian Tumour Analysis (IOTA) group.¹⁶ In a recent meta-analysis evaluating the performance of prediction models and rules to characterise adnexal pathology, approaches by IOTA such as the logistic regression model LR2²¹ and the simple rules^{25 26} (a set of 10 ultrasound features) performed best for the overall discrimination between benign and all malignant masses.⁴⁰ The Royal College of Obstetricians and Gynaecologists has included the simple rules in their guidelines on management of adnexal tumours in premenopausal patients.⁴¹ The ADNEX model's performance is similar to, or even slightly better than, that of LR2 and simple rules. For example, the AUC of LR2 on the validation data (IOTA phase 3) was 0.92.⁴² In contrast with LR2 and simple rules, the ADNEX model also enables specific subtyping of malignancy using risk estimates.

Strengths and weaknesses of this study

Our study has several strengths and limitations. Firstly, the strengths of the present study are that we used a large number of patients that were prospectively examined at 24 centres in 10 countries using a standardised protocol, avoided strong data driven variable selection, and conducted a large temporal validation of the model. After validation, we used the pooled data from almost 6000 patients to update the model coefficients. We would therefore expect our results to be generalisable. Secondly, it may be seen as an advantage that a histological diagnosis was obtained for every included tumour. This could also be regarded as a limitation, because the model is based on patients who were selected for surgery. Hence we cannot be certain that the test performance of the ADNEX model would be maintained if applied to a population of tumours, of which some were selected for expectant management. However, this argument holds for all prediction models for the diagnosis of ovarian tumours. Thirdly, the centres used different assay kits for CA-125 assessment. This can also be interpreted as both a strength and a limitation: using different kits introduces variability in CA-125 levels (although this variability is minor⁴³), reflects clinical reality, and yields results that are less dependent

on assay. Fourthly, a potential limitation is that experienced operators examined all tumours in the study. However, other studies have shown that dichotomous models developed by the IOTA group using ultrasound variables similar to those in the current study, work well in the hands of non-expert level 2⁴⁴ ultrasound examiners.^{45 46} Fifthly, there was no central review of pathology. In phase 1 of the IOTA study, 10% of the patients were selected at random for central review of pathology.²¹ Because we found no clinically important differences in reported outcomes between local and central reports, such centralised review was not performed in later phases of the IOTA study. This may nevertheless have introduced bias. For example, distinguishing borderline tumours from benign tumours or stage I cancer may be difficult for pathologists, and confusion of these tumour types might have impacted on the ability of the ADNEX model to correctly distinguish between them.

Implications for clinical practice

The ADNEX model has clear potential to optimise management of women with an adnexal tumour. Currently the risk of malignancy index (RMI)⁴⁷ is often used to characterise adnexal masses as benign or malignant. However, the index had much poorer performance for discrimination between benign and malignant tumours (AUC 0.88, 67.1% sensitivity, and 90.6% specificity at the typical risk of malignancy index cut-off of 200) than the ADNEX model when tested on our validation data.⁴² In addition to offering excellent discrimination between benign and malignant tumours, the ADNEX model predicts type of malignancy. Knowledge of the specific type of adnexal pathology before surgery is highly likely to improve patient triage, and it also makes it possible to optimise treatment. This in turn may reduce morbidity and lead to enhanced survival from different types of ovarian malignancy. The correct identification of stage I cancer is particularly important.¹⁹ The ADNEX model can discriminate well between stage I cancers and benign tumours and between stage I cancers and advanced stage cancer. In addition, the ADNEX model can discriminate well between advanced primary cancer and secondary metastatic cancer. The latter result is largely achieved through the use of serum CA-125 level as a predictor. Although CA-125 level has little added value over ultrasound information when distinguishing benign from malignant tumours,¹⁷ the present study shows that serum CA-125 level is important for good discrimination between stage II-IV cancer and stage I and secondary metastatic cancer. An inconvenience that ADNEX shares with well known models to predict ovarian malignancy, such as the risk of malignancy index⁴⁷ and the risk of ovarian malignancy algorithm (ROMA),⁴⁸ is that predictions can only be made once the results of blood sample analyses are available. ADNEX implementations also allow risk calculation without a CA-125 level, but this will result in poorer discrimination between stage II-IV cancers and other types of malignancy.

We expect that the performance of the ADNEX model will be maintained in the hands of non-expert ultrasound examiners on condition that the examiners are familiar with the IOTA terms and definitions and use the IOTA examination and measurement techniques (see the IOTA consensus statement²⁰). How the predicted risks from ADNEX should be used clinically must be decided on an individual basis, because patient management depends on many factors. When deciding on treatment of an adnexal mass, the likelihood of a specific type of malignancy is pivotal, but age, symptoms, wish to preserve fertility, comorbidity, and operative risks are also important factors. However, the ADNEX predictions may form a solid and

objective base for optimal management of patients and could be incorporated in national and international clinical guidelines.

Key future research

Future work entails regular updating of ADNEX model coefficients using newly collected data, and monitoring of model performance. In addition, studies including patients who are managed conservatively are critically needed. This is the subject of phase 5 of the IOTA study, for which data collection started early in 2013. Finally, the ADNEX model could be optimised for use as a second stage test if screening for ovarian cancer is introduced into clinical practice.⁶

Conclusion

The ADNEX model has the potential to change management decisions for women with an adnexal tumour. This could impact considerably on the morbidity and mortality associated with adnexal pathology.

Contributors: BVC conceived and designed the study, with additional support from KVH, LV, TB, and DT. LV, ACT, DFi, CVH, LS, DFr, EE, JK, AC, SG, RF, CL, FS, and DT enrolled patients and acquired data. BVC, KVH, CVH, JK, and DT were involved in data cleaning. BVC analysed the data, with support from KVH and VVB. BVC, KVH, LV, ACT, JK, VVB, TB, and DT were involved in data interpretation. BVC, JK, TB, and DT wrote the first draft of the manuscript, which was then critically reviewed and revised by the other coauthors. All authors approved the final version of the manuscript for submission. All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. BVC, LV, TB, and DT are the guarantors

Funding: This study was supported by the Flemish government: Research Foundation–Flanders (FWO) project G049312N, Flanders' Agency for Innovation by Science and Technology (IWT) project IWT-TBM 070706-IOTA3, and iMinds 2013. BVC and VVB are postdoctoral fellows of FWO. KVH is a doctoral fellow of IWT. TB is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. The views expressed are those of the authors and not necessarily those of the NHS, NIHR or Department of Health. LV is supported by the Swedish Medical Research Council (grants K2001-72X-11605-06A, K2002-72X-11605-07B, K2004-73X-11605-09A, and K2006-73X-11605-11-3), funds administered by Malmö University Hospital and Skåne University Hospital, Allmänna Sjukhusets i Malmö Stiftelse för bekämpande av cancer (the Malmö General Hospital Foundation for fighting against cancer), and two Swedish governmental grants (ALF-medel and Landstingsfinansierad Regional Forskning). The sponsors had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the work for publication. The researchers performed this work independently of the funding sources.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: The research protocols were approved by the ethics committee of the University Hospitals KU Leuven and by each centre's local ethics committee.

Data sharing: No additional data available.

What is already known on this topic

- Referring patients with ovarian cancer to specialised gynaecology oncology centres impacts positively on survival
- Currently in Europe and the United States only a minority of women are triaged to receive specialist care in a gynaecology oncology centre
- Personalised management, including fertility sparing surgery, requires knowledge of the nature of an ovarian mass
- Prediction models exist that can discriminate between benign and malignant ovarian tumours but they do not subclassify malignant tumours

What this study adds

- The ADNEX model discriminated well between benign and malignant ovarian tumours
- The model was also able to discriminate between benign, borderline, stage I invasive, stage II-IV invasive, and secondary metastatic tumours
- The ADNEX model may improve patient triage and decisions about management, and so positively impact on the morbidity and mortality associated with adnexal pathology

Transparency: The manuscripts' guarantors (BVC, LV, TB, and DT) affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

- Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin* 2014;64:9-29.
- Berrino F, De Angelis R, Sant M, Rosso S, Lasota MB, Coebergh JW, et al. Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995-1999: results of the EURO-CARE-4 study. *Lancet Oncol* 2007;8:773-83.
- Howlander N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, et al. SEER cancer statistics review, 1975-2011. National Cancer Institute (Bethesda, MD), Apr 2014. http://seer.cancer.gov/csr/1975_2011/.
- Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C, et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA* 2011;305:2295-303.
- Kobayashi H, Yamada Y, Sado T, Sakata M, Yoshida S, Kawaguchi R, et al. A randomized study of screening for ovarian cancer: a multicenter study in Japan. *Int J Gynecol Cancer* 2008;18:414-20.
- Menon U, Gentry-Maharaj A, Hallett R, Ryan A, Burnell M, Sharma A, et al. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol* 2009;10:327-40.
- Earle CC, Schrag D, Neville BA, Yabroff KR, Topor M, Fahey A, et al. Effect of surgeon specialty on processes of care and outcomes for ovarian cancer patients. *J Natl Cancer Inst* 2006;98:172-80.
- Engelen MJA, Kos HE, Willems PHB, Alders JG, de Vries EGE, Schaapveld M, et al. Surgery by consultant gynecologic oncologists improves survival in patients with ovarian carcinoma. *Cancer* 2006;106:589-98.
- Bristow RE, Chang J, Ziogas A, Anton-Culver H. Adherence to treatment guidelines for ovarian cancer as a measure of quality care. *Obstet Gynecol* 2013;121:1226-34.
- Woo YL, Kyrgiou M, Bryant A, Everett T, Dickinson HO. Centralisation of services for gynaecological cancers—a Cochrane systematic review. *Gynecol Oncol* 2012;126:286-90.
- Vernooij F, Heintz APM, Witteveen PO, van der Heiden-van der Loo M, Coebergh JW, van der Graaf Y. Specialized care and survival of ovarian cancer patients in the Netherlands: nationwide cohort study. *J Natl Cancer Inst* 2008;100:399-406.
- Bristow RE, Chang J, Ziogas A, Randall LM, Anton-Culver H. High-volume ovarian cancer care: survival impact and disparities in access for advanced-stage disease. *Gynecol Oncol* 2014;132:403-10.
- Verleye L, Vergote I, van der Zee AGJ. Patterns of care in surgery for ovarian cancer in Europe. *Eur J Surg Oncol* 2010;36(Suppl 1):S108-14.
- Darai E, Fauvet R, Uzan C, Gouy S, Duvillard P, Morice P. Fertility and borderline ovarian tumor: a systematic review of conservative management, risk factors, risk of recurrence and alternative options. *Hum Reprod Update* 2013;19:151-66.
- Hennessy BT, Coleman RL, Markman M. Ovarian cancer. *Lancet* 2009;374:1371-82.
- Van Calster B, Valentin L, Van Holsbeke C, Testa AC, Bourne T, Van Huffel S, et al. Polytomous diagnosis of ovarian tumors as benign, borderline, primary invasive or metastatic: development and validation of standard and kernel-based risk prediction models. *BMC Med Res Methodol* 2010;10:96.
- Timmerman D, Van Calster B, Jurkovic D, Valentin L, Testa AC, Bernard JP, et al. Inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumors. *J Clin Oncol* 2007;25:4194-200.
- Van Calster B, Valentin L, Van Holsbeke C, Zhang J, Jurkovic D, Lissoni AA, et al. A novel approach to predict the likelihood of specific ovarian tumor pathology based on serum CA-125: a multicenter observational study. *Cancer Epidemiol Biomarkers Prev* 2011;20:2420-8.
- Vergote I, De Brabanter J, Fyles A, Bertelsen K, Einhorn N, Sevelid P, et al. Prognostic importance of degree of differentiation and cyst rupture in stage I invasive epithelial ovarian carcinoma. *Lancet* 2001;357:176-82.
- Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) group. *Ultrasound Obstet Gynecol* 2000;16:500-5.
- Timmerman D, Testa AC, Bourne T, Ferrazzi E, Amey L, Konstantinovic ML, et al. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis group. *J Clin Oncol* 2005;23:8794-801.
- Van Holsbeke C, Van Calster B, Testa AC, Domali E, Lu C, Van Huffel S, et al. Prospective internal validation of mathematical models to predict malignancy in adnexal masses: results from the International Ovarian Tumor Analysis study. *Clin Cancer Res* 2009;15:684-91.
- Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D, Lissoni AA, et al. Ovarian cancer prediction in adnexal masses using ultrasound-based logistic regression models: a temporal and external validation study by the IOTA group. *Ultrasound Obstet Gynecol* 2010;36:226-34.
- Kajiser J, Bourne T, Valentin L, Sayasneh A, Van Holsbeke C, Vergote I, et al. Improving strategies for diagnosing ovarian cancer: a summary of the International Ovarian Tumor Analysis (IOTA) studies. *Ultrasound Obstet Gynecol* 2013;41:9-20.
- Timmerman D, Testa AC, Bourne T, Amey L, Jurkovic D, Van Holsbeke C, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol* 2008;31:681-90.
- Timmerman D, Amey L, Fischerova D, Epstein E, Melis GB, Guerriero S, et al. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by the IOTA group. *BMJ* 2010;341:c6839.
- Heintz APM, Odicino F, Maisonneuve P, Quinn MA, Benedet JL, Creasman WT, et al. Carcinoma of the ovary. *Int J Gynecol Obstet* 2006;95(Suppl 1):S161-92.
- Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.
- Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer, 2001.
- Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Springer, 2009.
- Wynants L, Timmerman D, Bourne T, Van Huffel S, Van Calster B. Screening for data clustering in multicenter studies: the residual intraclass correlation. *BMC Med Res Methodol* 2013;13:128.
- Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Wiley, 2008.
- Kuss O, McLerran D. A note on the estimation of the multinomial logistic model with correlated responses in SAS. *Comput Methods Programs Biomed* 2007;87:262-9.
- Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. *Stat Med* 1990;9:1303-25.
- Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
- Van Calster B, Vergouwe Y, Looman CW, Van Belle V, Timmerman D, Steyerberg EW. Assessing the discriminative ability of risk models for more than two outcome categories. *Eur J Epidemiol* 2012;27:761-70.
- Van Calster B, Van Belle V, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW. Extending the c-statistic to nominal polytomous outcomes: the Polytomous Discrimination Index. *Stat Med* 2012;31:2610-26.
- Van Hoorde K, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW, Van Calster B. Assessing calibration of multinomial risk prediction models. *Stat Med* 2014;33:2585-96.
- Hosmer DW, Lemeshow S. Applied logistic regression. Wiley, 2000.
- Kajiser J, Sayasneh A, Van Hoorde K, Ghaem-Maghami S, Bourne T, Timmerman D, et al. Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: a systematic review and meta-analysis. *Hum Reprod Update* 2014;20:449-62.
- Royal College of Obstetricians and Gynaecologists. Management of suspected ovarian masses in premenopausal women. Green-top guideline No 62. RCOG, Nov 2011. www.rcog.org.uk/files/rcog-corp/GTG62_021211_OvarianMasses.pdf.
- Testa A, Kajiser J, Wynants L, Fischerova D, Van Holsbeke C, Franchi D, et al. Strategies to diagnose ovarian cancer: new evidence from phase 3 of the multicentre international IOTA study. *Br J Cancer* 2014;111:680-8.
- Davelaar EM, van Kamp GJ, Verstraeten RA, Kenemans P. Comparison of seven immunoassays for the quantification of CA 125 antigen in serum. *Clin Chem* 1998;44:1417-22.
- Education and Practical Standards Committee, European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB). Minimum training recommendations for the practice of medical ultrasound. *Ultraschall Med* 2006;27:79-105.
- Sayasneh A, Wynants L, Preisler J, Kajiser J, Johnson S, Stalder C, et al. Multicentre external validation of IOTA prediction models and RMI by operators with varied training. *Br J Cancer* 2013;108:2448-54.
- Nunes N, Ambler G, Hoo WL, Naftalin J, Foo X, Widschwendter M, et al. A prospective validation of the IOTA logistic regression models (LR1 and LR2) in comparison to subjective pattern recognition for the diagnosis of ovarian cancer. *Int J Gynecol Cancer* 2013;23:1583-9.
- Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *Br J Obstet Gynaecol* 1990;97:922-9.
- Moore RG, McMeekin DS, Brown AK, DiSilvestro P, Miller MC, Allard WJ, et al. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009;112:40-6.

Accepted: 5 September 2014

Cite this as: [BMJ 2014;349:g5920](#)

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>.

Tables

Table 1 | Number of patients in each centre, and type of centre

| Participating centres and data summaries | Dataset | Total | Benign* | Borderline | Stage I | Stage II-IV | Metastatic |
|---|---------|-------|-----------|------------|----------|-------------|------------|
| Oncology centres: | | | | | | | |
| University Hospitals Leuven, Belgium | D, V | 930 | 596 (64) | 64 | 48 | 171 | 51 |
| Università Cattolica del Sacro Cuore, Rome, Italy | D, V | 787 | 377 (48) | 44 | 79 | 213 | 74 |
| Ospedale San Gerardo, Monza, Italy | D, V | 401 | 308 (77) | 30 | 17 | 40 | 6 |
| General Faculty Hospital, Prague, Czech Republic | D, V | 354 | 120 (34) | 46 | 31 | 133 | 24 |
| Istituto Europeo di Oncologia, Milan, Italy | D, V | 311 | 135 (43) | 21 | 27 | 109 | 19 |
| Medical University Lublin, Poland | D, V | 285 | 183 (64) | 8 | 25 | 61 | 8 |
| University of Bologna, Italy† | V | 213 | 148 (69) | 19 | 10 | 31 | 5 |
| Karolinska University Hospital, Stockholm, Sweden | V | 120 | 67 (56) | 12 | 7 | 26 | 8 |
| King's College Hospital, London, UK | D | 119 | 78 (66) | 13 | 8 | 15 | 5 |
| Skåne University Hospital Lund, Sweden | D, V | 77 | 57 (74) | 2 | 4 | 11 | 3 |
| Chinese PLA General Hospital, Beijing, People's Republic of China | D | 73 | 57 (78) | 1 | 0 | 12 | 3 |
| Università degli Studi di Udine, Italy | D, V | 64 | 45 (70) | 1 | 10 | 6 | 2 |
| Istituto Nazionale dei Tumori, Naples, Italy | D, V | 15 | 7 (47) | 0 | 2 | 4 | 2 |
| Other hospitals: | | | | | | | |
| Skåne University Hospital Malmö, Sweden | D, V | 776 | 608 (78) | 35 | 38 | 77 | 18 |
| Ziekenhuis Oost-Limburg, Genk, Belgium | D, V | 428 | 367 (86) | 14 | 17 | 28 | 2 |
| Ospedale San Giovanni di Dio, Cagliari, Italy | D, V | 261 | 224 (86) | 8 | 8 | 13 | 8 |
| DCS Sacco University of Milan, Italy | D, V | 223 | 195 (87) | 4 | 8 | 13 | 3 |
| University of Bologna, Italy† | D | 135 | 124 (92) | 3 | 3 | 3 | 2 |
| Università degli Studi di Napoli, Naples, Italy | D, V | 103 | 82 (80) | 2 | 3 | 13 | 3 |
| Hôpital Boucicaud, Paris, France | D | 80 | 71 (89) | 2 | 2 | 5 | 0 |
| Centre Medical des Pyramides, Maurepas, France | D | 64 | 57 (89) | 1 | 4 | 2 | 0 |
| Institut Universitari Dexeus, Barcelona, Spain | V | 37 | 26 (70) | 8 | 2 | 1 | 0 |
| Macedonio Melloni Hospital, Italy | D | 21 | 17 (81) | 1 | 2 | 1 | 0 |
| Ospedale dei Bambini Vittore Buzzi, Milan, Italy | V | 21 | 21 (100) | 0 | 0 | 0 | 0 |
| St Joseph's Hospital, Hamilton, Canada | D | 11 | 10 (91) | 0 | 1 | 0 | 0 |
| Data summaries: | | | | | | | |
| Oncology centres only | D, V | 3749 | 2178 (58) | 261 (7)* | 268 (7)* | 832 (22)* | 210 (6)* |
| Other hospitals only | D, V | 2160 | 1802 (83) | 78 (4)* | 88 (4)* | 156 (7)* | 36 (2)* |
| Development data only | D | 3506 | 2557 (73) | 186 (5)* | 176 (5)* | 467 (13)* | 120 (3)* |
| Validation data only | V | 2403 | 1423 (59) | 153 (6)* | 180 (7)* | 521 (22)* | 126 (5)* |
| Total pooled dataset | D, V | 5909 | 3980 (67) | 339 (6)* | 356 (6)* | 988 (17)* | 246 (4)* |

D=contributed to development dataset; V=contributed to validation dataset.

*Number (percentage).

†Centre changed to an oncology referral centre after completion of IOTA phase 2 (that is, between patient recruitment for development and validation datasets).

Table 2| Descriptive statistics of the a priori considered predictors by tumour type in pooled dataset (n=5909). Values are numbers (percentages) unless stated otherwise

| Variables | Benign (n=3980) | Borderline (n=339) | Stage I (n=356) | Stage II-IV (n=988) | Metastatic (n=246) |
|---|-----------------|--------------------|-----------------|---------------------|--------------------|
| Median (interquartile range) age (years) | 42 (32-54) | 49 (36-62) | 54 (44-64) | 59 (50-67) | 57 (47-68) |
| Median (interquartile range) serum CA-125 (U/mL)* | 18 (11-39) | 30 (16-86) | 51 (20-195) | 442 (145-1238) | 91 (29-271) |
| Family history of ovarian cancer | 79 (2.0) | 10 (3.0) | 13 (3.7) | 57 (5.8) | 5 (2.0) |
| Median (interquartile range) maximal diameter of lesion (mm) | 63 (45-87) | 86 (51-150) | 106 (71-153) | 85 (56-123) | 86 (56-124) |
| Solid tissue: | | | | | |
| Presence of solid tissue | 1322 (33.2) | 267 (78.8) | 328 (92.1) | 968 (98.0) | 234 (95.1) |
| Median (interquartile range) proportion solid tissue if present (%) | 42 (20-100) | 37 (24-59) | 61 (38-100) | 100 (56-100) | 100 (64-100) |
| No of papillary projections: | | | | | |
| 0 | 3424 (86.0) | 135 (39.8) | 227 (63.8) | 772 (78.1) | 213 (86.6) |
| 1 | 333 (8.4) | 69 (20.4) | 25 (7.0) | 56 (5.7) | 12 (4.9) |
| 2 | 80 (2.0) | 21 (6.2) | 17 (4.8) | 30 (3.0) | 0 (0) |
| 3 | 66 (1.7) | 24 (7.1) | 17 (4.8) | 28 (2.8) | 2 (0.8) |
| >3 | 77 (1.9) | 90 (26.5) | 70 (19.7) | 102 (10.3) | 19 (7.7) |
| >10 cyst locules | 199 (5.0) | 74 (21.8) | 69 (19.4) | 93 (9.4) | 36 (14.6) |
| Acoustic shadows | 676 (17.0) | 8 (2.4) | 18 (5.1) | 30 (3.0) | 10 (4.1) |
| Ascites | 64 (1.6) | 28 (8.3) | 65 (18.3) | 473 (47.9) | 90 (36.6) |
| Missing values for CA-125 | 1447 (36.4) | 62 (18.3) | 71 (19.9) | 163 (16.5) | 62 (25.2) |

*Results based on multiple imputation of missing values.

Table 3| Diagnostic performance of ADNEX model when using different thresholds for total probability of malignancy (sum of probabilities of four subtypes of ovarian malignancy)

| Threshold for probability of malignancy* | Development data (n=3506) | | | | Validation data (n=2403) | | | | After updating on pooled data (n=5909) | | | |
|--|---------------------------|---------------------|---------------------|-----------------------|---------------------------|---------------------|---------------------|-----------------------|--|---------------------|---------------------|-----------------------|
| | AUC | Sensitivity | Specificity | Diagnostic odds ratio | AUC | Sensitivity | Specificity | Diagnostic odds ratio | AUC | Sensitivity | Specificity | Diagnostic odds ratio |
| | | | | | | | | | | | | |
| Not applicable | 0.954 (0.947 to 0.961) | — | — | — | 0.943 (0.934 to 0.952) | — | — | — | 0.950 (0.944 to 0.955) | — | — | — |
| 3% | — | 98.8 (97.9 to 99.4) | 52.3 (50.4 to 54.3) | 93.6 | — | 98.9 (98.0 to 99.4) | 46.6 (44.0 to 49.2) | 76.8 | — | 99.1 (98.6 to 99.5) | 43.4 (41.8 to 45.0) | 86.2 |
| 5% | — | 97.9 (96.8 to 98.7) | 65.4 (63.6 to 67.3) | 87.9 | — | 98.4 (97.4 to 99.1) | 59.4 (56.8 to 62.0) | 88.1 | — | 98.0 (97.3 to 98.6) | 61.1 (59.5 to 62.6) | 78.0 |
| 10% | — | 95.9 (94.4 to 97.1) | 75.5 (73.8 to 77.2) | 72.0 | — | 96.5 (95.2 to 97.6) | 71.3 (68.9 to 73.7) | 69.2 | — | 96.4 (95.4 to 97.2) | 73.2 (71.8 to 74.6) | 72.7 |
| 15% | — | 94.4 (92.8 to 95.8) | 81.0 (79.4 to 82.5) | 71.9 | — | 94.2 (92.5 to 95.6) | 77.2 (74.9 to 79.3) | 54.7 | — | 94.5 (93.4 to 95.5) | 78.7 (77.4 to 79.9) | 63.4 |

AUC=area under receiver operating characteristic curve.

Exact binomial 95% confidence intervals are reported in parentheses.

*Probability equal to or more than threshold indicates malignancy.

Table 4| Polytomous discrimination performance of ADNEX model on development data, validation data, and after updating on pooled data

| Performance measures | Development data (n=3506) | Validation data (n=2403) | After updating on pooled data (n=5909) |
|---------------------------------|---------------------------|--------------------------|--|
| AUC benign v borderline | 0.91 (0.88 to 0.93) | 0.85 (0.82 to 0.88) | 0.88 (0.87 to 0.90) |
| AUC benign v stage I | 0.94 (0.92 to 0.96) | 0.92 (0.90 to 0.93) | 0.93 (0.92 to 0.94) |
| AUC benign v stage II-IV | 0.99 (0.98 to 0.99) | 0.99 (0.98 to 0.99) | 0.99 (0.98 to 0.99) |
| AUC benign v metastatic | 0.96 (0.95 to 0.98) | 0.95 (0.93 to 0.97) | 0.96 (0.95 to 0.97) |
| AUC borderline v stage I | 0.71 (0.65 to 0.76) | 0.75 (0.69 to 0.79) | 0.75 (0.71 to 0.79) |
| AUC borderline v stage II-IV | 0.91 (0.88 to 0.93) | 0.95 (0.93 to 0.96) | 0.93 (0.91 to 0.95) |
| AUC borderline v metastatic | 0.86 (0.81 to 0.90) | 0.87 (0.82 to 0.91) | 0.88 (0.85 to 0.91) |
| AUC stage I v stage II-IV | 0.83 (0.79 to 0.86) | 0.87 (0.83 to 0.90) | 0.85 (0.82 to 0.87) |
| AUC stage I v metastatic | 0.77 (0.71 to 0.82) | 0.71 (0.65 to 0.76) | 0.75 (0.70 to 0.78) |
| AUC stage II-IV v metastatic | 0.76 (0.71 to 0.81) | 0.82 (0.78 to 0.86) | 0.80 (0.76 to 0.83) |
| Polytomous discrimination index | 0.554 (0.530 to 0.579) | 0.567 (0.540 to 0.591) | 0.569 (0.553 to 0.586) |

AUC=area under the receiver operating characteristic curve.

With five tumour types, the polytomous discrimination index for random prediction equals 0.2, hence its value cannot be directly compared with AUCs.

95% confidence intervals are shown in parentheses.

Table 5| Odds ratios for predictors in ADNEX model after it was updated on pooled dataset (n=5909)

| Predictor | Borderline v benign | Stage I v benign | Stage II-IV v benign | Metastatic v benign |
|---|---------------------|---------------------|----------------------|---------------------|
| Patient age, per 10 years | 1.05 (0.96 to 1.14) | 1.19 (1.09 to 1.30) | 1.67 (1.50 to 1.86) | 1.40 (1.24 to 1.57) |
| Serum CA-125, per doubling* | 1.12 (1.03 to 1.22) | 1.22 (1.12 to 1.32) | 2.15 (1.96 to 2.36) | 1.32 (1.19 to 1.46) |
| Maximal diameter of lesion, per doubling* | 1.45 (1.22 to 1.73) | 2.40 (1.97 to 2.91) | 1.54 (1.25 to 1.89) | 1.57 (1.23 to 1.99) |
| Proportion solid tissue (%)†: | | | | |
| 33 v 0 (no solid tissue) | 5.44 (3.88 to 7.64) | 12.8 (8.62 to 18.9) | 16.9 (10.5 to 27.0) | 7.09 (4.01 to 12.5) |
| 67 v 33 | 1.55 (1.32 to 1.81) | 3.49 (2.99 to 4.08) | 4.74 (3.92 to 5.73) | 4.25 (3.46 to 5.23) |
| 100 v 67 | 0.44 (0.29 to 0.67) | 0.95 (0.68 to 1.35) | 1.33 (0.92 to 1.94) | 2.55 (1.60 to 4.06) |
| >10 cyst locules | 3.96 (2.65 to 5.90) | 2.21 (1.42 to 3.43) | 1.31 (0.74 to 2.32) | 2.46 (1.33 to 4.56) |
| No of papillary projections | 1.83 (1.65 to 2.03) | 1.49 (1.33 to 1.68) | 1.48 (1.28 to 1.71) | 1.24 (1.01 to 1.52) |
| Acoustic shadows | 0.13 (0.06 to 0.28) | 0.15 (0.09 to 0.26) | 0.09 (0.05 to 0.17) | 0.08 (0.04 to 0.18) |
| Ascites | 2.64 (1.44 to 4.86) | 1.57 (0.93 to 2.67) | 3.85 (2.39 to 6.20) | 5.14 (3.00 to 8.79) |
| Oncology referral centre | 2.59 (1.32 to 5.11) | 1.57 (0.89 to 2.78) | 1.58 (0.78 to 3.21) | 2.25 (1.04 to 4.87) |

*This variable is log transformed (log with base 2) such that the odds ratio represents the effect for each doubling of the value.

†This variable represents the maximal diameter of the largest solid component divided by the maximal diameter of the lesion (range 0% to 100%), with 0% indicating that there is no solid tissue and 100% indicating that the maximal diameter of the largest solid component equals the maximal diameter of the lesion. The variable has a quadratic effect in the model, hence we report odds ratios for 33% v 0%, 67% v 33%, and 100% v 67%.

Figures

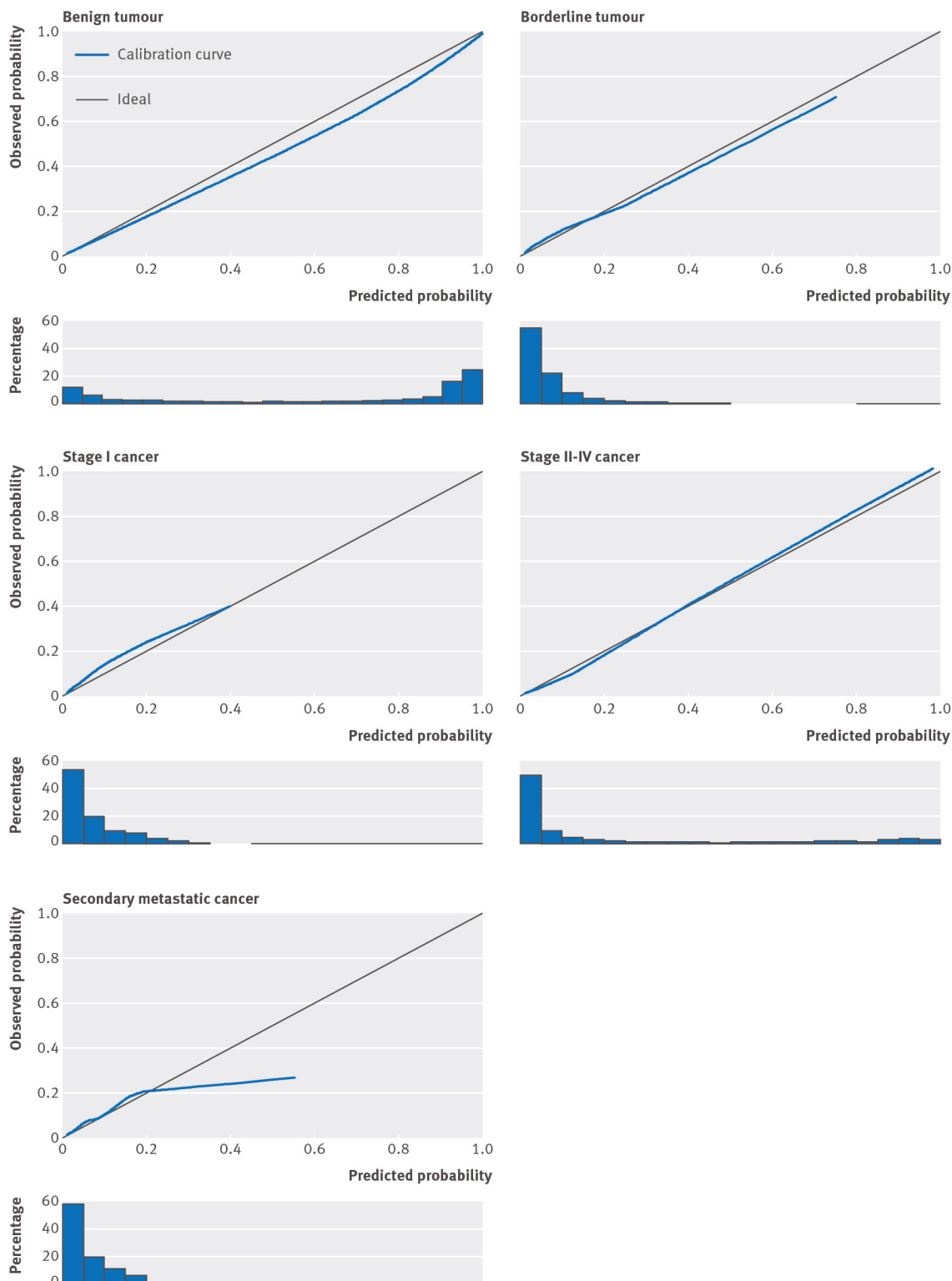


Fig 1 Calibration plots of predicted probabilities for each type of tumour. Data have been calculated using validation data (n=2403). Plots show how well the predicted probabilities (x axis) agree with observed probabilities (y axis). For perfect agreement, the calibration curve falls on the ideal diagonal line. Histograms below plots show distribution of predicted probabilities

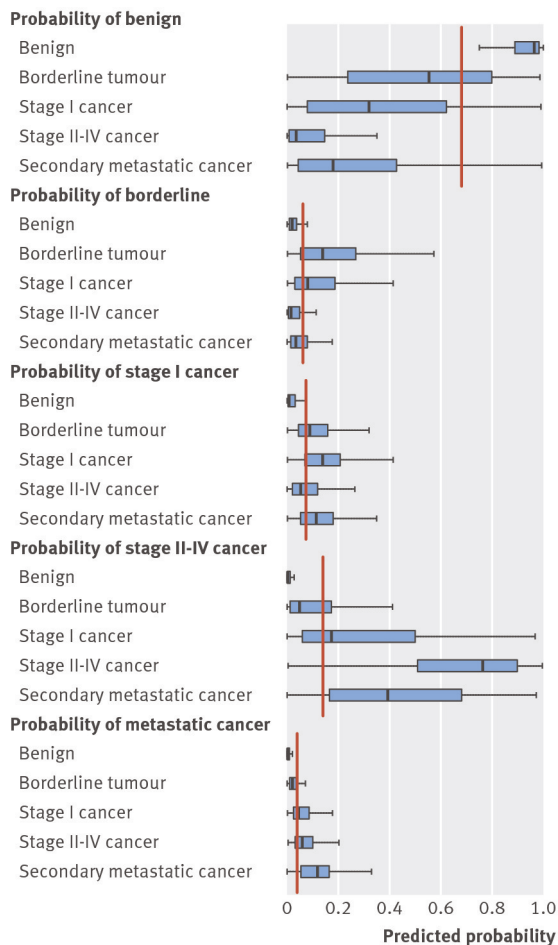


Fig 2 Discrimination plot of ADNEX model after it was updated on pooled dataset (n=5909). For each predicted tumour type, box plots of probabilities are presented for each confirmed tumour type (reference standard). Red vertical lines show baseline probabilities for each type of tumour. For example, the baseline probability of a benign tumour is 0.681; for most women with a benign tumour the predicted probability of a benign tumour was higher than 0.9, whereas most women with an ovarian malignancy (most notably stage II-IV cancer) had clearly lower predicted probabilities of a benign tumour