

The ADNEX risk prediction model for ovarian cancer diagnosis:

A systematic review and meta-analysis of external validation studies

Lasai BARREÑADA¹, Ashleigh LEDGER¹, Paula DHIMAN², Gary S. COLLINS², Laure WYNANTS^{1,3},

Jan Y. VERBAKEL^{4,5}, Dirk TIMMERMAN^{1,6}, Lil VALENTIN^{7,8}, Ben VAN CALSTER^{1,9}

¹ Department of Development and Regeneration, KU Leuven, Belgium

² Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK

³ Department of Epidemiology, CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands.

⁴ Department of Public Health and Primary Care, KU Leuven, Belgium

⁵ NIHR Community Medtech and IVD cooperative, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

⁶ Department of Obstetrics and Gynaecology, University Hospitals Leuven, Leuven, Belgium

⁷ Department of Obstetrics and Gynaecology, Skåne University Hospital, Malmö, Sweden

⁸ Department of Clinical Sciences Malmö, Lund University, Lund, Sweden

⁹ Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, the Netherlands

Corresponding author

Ben Van Calster

KU Leuven, Department of Development and Regeneration

Herestraat 49 box 805

3000 Leuven

Belgium

Ben.vancalster@kuleuven.be

INTRODUCTION

The optimal management of patients with an ovarian mass depends on the histology of the mass. Patients with a benign mass can be managed non-surgically with clinical and ultrasound follow-up or using conservative surgical techniques.[1,2] Malignant tumours benefit from management in specialised oncology centres, but borderline malignancies, stage I primary invasive tumours, and advanced primary invasive tumours may require different surgical approaches.[3,4] To optimise patient triage without operating on all masses, diagnostic models can be used to estimate the likelihood of malignancy and so be used to plan treatment for patients.

Given the potential advantages of accurately predicting risk of malignancy, the International Ovarian Tumour Analysis (IOTA) group developed the Assessment of Different NEoplasias in the adneXa (ADNEX) risk prediction model, using three clinical and six ultrasound predictor variables.[5] The clinical variables are age, serum CA-125 level, and type of centre (oncology centre vs other). An oncology centre is defined as a tertiary referral centre with a specific gynaecology oncology unit. The ultrasound variables are the maximal diameter of the lesion, proportion of solid tissue (defined as the largest diameter of the largest solid component divided by the largest diameter of the lesion), number of papillary projections, presence of more than 10 cyst locules, presence of acoustic shadows, and ascites. ADNEX is a multinomial logistic regression model that estimates the risk of five tumour types: benign, borderline, stage I primary invasive, stage II-IV primary invasive, and secondary metastatic. The total risk of malignancy calculated by ADNEX is the sum of the risks for each malignant subtype. ADNEX has two versions: one with and one without CA125 as a predictor.[5,6] The model was developed on data from 5909 patients with an adnexal mass that subsequently underwent surgery and that were recruited at 24 centres across 10 countries (Belgium, Italy, Czech Republic, Poland, Sweden, China, France, Spain, UK and Canada). Although developed on data from patients that underwent surgery, ADNEX has been validated on cohorts that also include non-surgically managed patients.[7–10]

ADNEX is widely used and is included in national guidelines, e.g. in Belgium, The Netherlands and Sweden.[11–13] It is recommended by scientific societies, such as International Society of Ultrasound in Obstetrics and Gynecology (ISUOG), European Society of Gynaecological Oncology (ESGO), European Society for Gynaecological Endoscopy (ESGE), and the American College of Radiology.[4,14] In addition, manufacturers of ultrasound machines have begun to incorporate ADNEX directly into their machines.[15–17]

Several external validation studies of ADNEX have been carried out.[8–10,18–61] To date, five published systematic reviews and meta-analyses of ADNEX have summarised between 3 and 22 external validation studies.[62–67] All systematic reviews evaluated ADNEX solely as a diagnostic test, reporting a summary sensitivity and specificity at a threshold for the estimated risk of malignancy of 10% or 15%.[62,64,65,67,68] However, the ADNEX model is not only a diagnostic test. It is also a risk prediction model to provide probability estimates of five different tumour types at the individual patient level. Using classification thresholds reduces the information given by the model.[69] Further, when analysing ADNEX merely as a diagnostic test with a 10% threshold we assume that there is no difference between patients with an estimated risk of 11% versus 99% of having a malignancy. This means that these meta-analyses have not fully validated the diagnostic performance of ADNEX. For example, pooling discrimination performance (area under the receiver operating characteristic curve, AUC) allows us to determine the ability of the model to differentiate between patients with and without the outcome across thresholds. There are now guidelines on how to evaluate the quality and risk of bias of external validation studies of risk prediction models. [70–72] These should be used in meta-analyses of validation studies.

The objective of this study is to conduct a systematic review of studies that externally validate ADNEX, to describe reporting completeness and risk of bias of the validation studies, and to conduct meta-analyses of model performance measures.

METHODS

Protocol registration

We report this study according to the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analysis) and TRIPOD-SRMA (Transparent Reporting of multivariable prediction models for Individual Prognosis Or Diagnosis: checklist for Systematic Reviews and Meta-Analyses) checklists.[72,73] The study protocol was prospectively registered in the International prospective register of systematic reviews (PROSPERO; ID CRD42022373182).[74]

Eligibility criteria

Any study that carried out an external validation to evaluate the performance of the ADNEX model using any study design and any study population was eligible for inclusion in the systematic review.

The exclusion criteria were (1) studies that did not evaluate the model performance of ADNEX in any way; (2) studies that evaluated the predictive performance only of updated versions of ADNEX; (3) studies for which only an abstract is available, or the full text cannot be obtained; (4) case studies presenting ADNEX performance for individual patients (this criterion was not pre-specified in the protocol but was added post hoc upon review of the search results). Updating can refer to recalibration, refitting, or extension with additional predictors.[75]

Information sources and search strategy

We created a search string and overall search strategy with the help of biomedical reference librarians of the KU Leuven Libraries. We searched the electronic databases Medline (via PubMed), EMBASE, Web of Science, and Scopus for published articles, and EuropePMC for preprints. The search dates were from the publication of the first ADNEX paper (15/10/2014) until 15/05/2023 (date the final search was run). We also screened all articles citing the original ADNEX paper.[5] The reference lists of relevant review and opinion articles retrieved by the search strings were checked for other potentially eligible articles. Forward and backward snowballing (forward/back cross-reference checking) of the included articles was performed to identify additional publications.[76] There was no language restriction, but for papers in languages other than English, Spanish, Dutch, French or Swedish an automatic translation tool (deepl.com) was used to decide whether to include or exclude a paper and to extract information. The full search strategy is provided in **Supplementary Material S1**.

Study selection

The studies we identified in our search were imported into Zotero reference manager, where they were automatically de-duplicated. The de-duplicated records were then imported into Rayyan web application for manual de-duplication (LB) and subsequent screening of title and abstract by two independent authors with any conflicts being resolved by discussions between them (LB, AL).[77]

As three authors (BVC, LV, DT) are members of the IOTA group that developed ADNEX, we separated the studies as linked or not linked to IOTA. A study was IOTA-linked if it was co-authored by a member of the IOTA steering committee (**Supplementary Material S2**). IOTA-linked papers, as well as a few others with a potential conflict of interest (i.e., including authors that are or were IOTA collaborators), were independently assessed by authors PD and GSC, two medical statisticians with expertise in prediction modelling and unrelated to IOTA. All other studies were independently assessed by authors LB and AL.

Conflicts were resolved through discussion between reviewers and for the non-IOTA papers also through discussion with authors BVC, LV and JYV.

Data extraction and data items

Data was extracted and entered into a standardised data extraction form in Microsoft Excel. The data extraction focused on general and design characteristics of studies, target population, reference standard, sample size, performance results, reporting quality, methodological quality, and risk of bias (**Table S1**). The extraction form was based on and adapted from the CHARMS (Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies), TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis), and PROBAST (Prediction model Risk Of Bias ASsessment Tool) tools.[71,78,79]

To describe model performance, we extracted information on any reported measure related to discrimination, calibration, diagnostic accuracy, or clinical utility. The reference standard could be binary (e.g., benign vs malignant) or multinomial (e.g., the five tumour types predicted by ADNEX). Performance data was extracted for all reported validations (i.e., for both ADNEX versions), subgroup analyses, sensitivity analyses, and centre-specific results in multicentre studies.

For each study we assessed the reporting of all TRIPOD items that are applicable to external validation studies (**Table S2**). We also checked PROBAST's 'signalling questions' and evaluated risk of bias for each subdomain (participants, predictors, outcome, and analysis) and overall. We included rationale for the risk of bias classification.

We contacted study authors to obtain further information or results in the following situations: (1) when centre-specific results were not reported in multicentre studies, (2) the type of centre was not explicitly reported (in case of nonresponse, the clinical co-authors, JYV, DT, and LV, classified the centre), (3) overall performance was reported but not performance by menopausal status, or (4) performance was not reported for operated patients separately in studies including both surgically and non-surgically managed patients.

Details on all extracted items are available in **Supplementary Material (Table S1)** and **Open Science Framework repository (Extraction sheet)**. [80]

Statistical analysis and quantitative data synthesis

Data were summarised with descriptive statistics and data visualisations. Meta-analysis of performance, using centre-specific results for multicentre studies where possible, was performed using random effects meta-analysis methods. Meta-analysis was done separately for the two versions of ADNEX. In addition to 95% confidence intervals (CI) for the summary performance, we assessed heterogeneity using tau squared and 95% prediction intervals (PI). **Supplementary Material S3-5** provides details on the meta-analysis methodology.

Meta-analysis of the AUC for benign versus malignant tumours was done on the logit scale. Meta-analysis for sensitivity and specificity was also performed on the logit scale using random effects meta-analysis.[81] As the meta-analysis was only conducted for the 10% threshold for the risk of malignancy, we did not use the bivariate random effects model as specified in the protocol.[74] Meta-analysis of Net Benefit (NB) [82,83] and Relative Utility (RU) [84,85] at the 10% risk of malignancy threshold was performed using Bayesian trivariate random effects meta-analysis of sensitivity, specificity and prevalence of malignancy.[86] For Bayesian methods, 95% credible intervals (CrI) are reported instead of 95% CIs. The Bayesian approach allowed to estimate the probability that the model is useful in a new centre, i.e. the probability that RU is >0.

To address multinomial discrimination performance, meta-analysis of AUCs between pairs of tumour outcomes ('pairwise AUCs') was conducted on logit scale. We only included studies that used the "conditional risk method" to calculate pairwise AUCs.[87]

Subgroups are defined based on geographical location, type of centre, and menopausal status. Sensitivity analysis are based on the risk of bias judgment and on whether the study was IOTA linked or not. As pre-specified in the protocol, we only meta-analysed performance if at least three estimates in a specific analysis could be retrieved from the included studies.[74] To assess the association of prevalence of malignancy with the AUC and sensitivity and specificity at the 10% risk of malignancy threshold, we used meta-regression.[88]

Reporting bias and small study effects were visually explored using funnel plots adapted for the AUC. The body of evidence was assessed using an adapted version of GRADE.[89]

All analyses were performed in R version 4.2.2 using package "metamisc" for the AUC, "meta" and "mada" for sensitivity and specificity, and rjags for NB and RU. [89–95] Bayesian methods were computed using JAGS version 4.3.1.[95]

RESULTS

We identified 1843 records and screened 490 after de-duplication. Forty-seven studies met our inclusion criteria and were included in this systematic review (**Figure 1, Table S3**). [8–10,18–61] Three studies were excluded because they used the same data as in another included study, and one study was excluded because a preliminary version of ADNEX was used. [96–99] The data of three studies that were IOTA-linked and three other studies with a potential conflict of interest were extracted by authors PD and GSC. [7,27,40,49,51,52]

Key study characteristics are summarised in **Table 1** (**Table S3** shows study-specific details). The unit of analysis was the patient in 42 (89%) studies and the tumour in five (11%) studies. When tumour was the unit of analysis, multiple tumours for the same patient could be included. The 47 studies reported on 17007 tumours, with a median study sample size of 261 tumours (range 24 to 4905). The validations were conducted in 28 countries with most studies being conducted in Asia (51%) and Europe (38%). **Supplementary Material S6** presents a list of reporting inconsistencies.

Table 1. Characteristics of the 47 included studies.

Study characteristic	N (%)	Comments
Unit of analysis		
Patient	42 (89)	1 tumour per patient
Tumour	5 (11)	>1 tumour per patient possible
Region		
Asia	24 (51)	Common countries: China (12), South Korea (3)
Europe	18 (38)	Common countries: Poland (7), Italy (4), Spain (4), UK (3), Sweden (3)
South America	3 (6)	
North America	2 (4)	
Number of centres		
1	37 (79)	
2-5	7 (15)	
>5	3 (6)	Range: 8-17
Type of centre		
Oncology centre(s)	35 (74)	
Non-oncology centre(s)	2 (4)	
Both types of centres	4 (9)	
Unclear	6 (13)	
ADNEX version		
ADNEX with CA125	37 (79)	21 only used ADNEX with CA125, 16 used both
ADNEX without CA125	19 (40)	3 only used ADNEX without CA125, 16 used both
Mixed	3 (6)	ADNEX with CA125 used if CA125 was available
Unclear	4 (9)	
Selection based on histology		
No	39 (83)	
Yes	8 (17)	e.g., borderlines excluded, only invasive tumours
Focus only on clinical subgroup		
No	42 (89)	
Yes	5 (11)	e.g., only pregnant patients
Target population		
Surgically managed patients	43 (91)	
Surgically and non-surgically managed patients	4 (9)	

ADNEX with CA125 was validated in 37 (79%) studies, and ADNEX without CA125 in 19 (40%) studies (16 studies evaluated both versions). Three (6%) studies conducted a mixed validation, using ADNEX with CA125 when CA125 was available. For four (9%) studies, the ADNEX version was unclear. In total, 63 validations of ADNEX were performed after distinguishing between the ADNEX versions used (**Table S4**). When also including reported results for subgroups (e.g., by menopausal status, or by centre in multicentre studies), the total number of validations reported across the 47 studies was 159.

Five (11%) studies focused only on a specific clinical subgroup, such as pregnant women or tumours for which the clinician's subjective assessment of the outcome was uncertain.[10,24,38,42,52] Eight (17%) studies selected patients based on histology (**Table S3** for details). Thirty-six (77%) studies did not focus on a specific clinical subgroup and did not select tumours based on histology. In these 36 studies that were eligible for the meta-analysis, the median sample size was 284 tumours (range 50-4905), the median number of malignant tumours was 68 (7-1041) and the median prevalence of malignancy 28% (3%-57%). Fourteen of the 36 (39%) studies had at least 100 benign and at least 100 malignant tumours.

The target population of the studies could be surgically managed patients or surgically and non surgically managed patients. The reference standard for determining the tumour type in surgically managed patients was always histopathology. Four (9%) studies included surgically and non-surgically managed patients. In non-surgically managed patients, the outcome determination was the clinician's subjective assessment of the tumour as benign or malignant or spontaneous resolution of the tumour during follow-up. The required follow-up time to determine the outcome was 3-4 months, 1 year, or 2 years, depending on the study.[8-10,56]

The most commonly reported performance measure was the AUC for benign vs malignant tumours (72%) (**Table S5**). About two-thirds of studies (66%) presented a receiver operating characteristic curve, 31 (66%) reported sensitivity and specificity performance at the 10% threshold for the risk of malignancy, 12 (26%) reported measures for multinomial discrimination and four (9%) studies reported calibration performance.

Critical appraisal: reporting and risk of bias

Completeness of reporting the TRIPOD items was assessed for the 63 validations. Adherence to TRIPOD items was on average 61%: studies reported on average 16.5 out of 27 items (**Figures 2 and S1**). The least commonly reported items were 'comparison of demographics, predictors and outcome between the model development and external validation data' (item 13c; 5%), 'the reporting of performance measures

with confidence intervals' (item 16; 11%), 'specification of all performance measures' (item 10d; 11%), 'rationale for study sample size' (item 8; 13%), and 'description of how missing data were handled' (item 9; 22%).

Fifty-seven (90%) of 63 validations were rated at high risk of bias, two (3%) at uncertain risk of bias, and four (6%) at low risk of bias (**Figures 3 , S2-5, OSF Extraction sheet**). 43 (68%) validations were at high risk of bias for the participant domain, mostly by having incomplete data as an exclusion criterion. Fifty-seven (90%) validations were at high risk of bias for the analysis domain, mostly due to small sample size (69%), not including all participants in the analysis (85%), inappropriate handling of missing data (82%), and incomplete evaluation of model performance - in most instances by not reporting an assessment of calibration (89%).

TRIPOD adherence in 36 studies without focus on selected histologies or clinical subgroups was on average 65% (17.47 items out of 27). On these studies, two were at low risk of bias, one at unclear risk of bias and 33 at high risk of bias.

Meta-analysis

The meta-analysis included studies without post hoc selection based on histology and without focus on a clinical subgroup only (n = 36) that reported the meta-analysed metrics.

The AUC for benign vs malignant tumours in operated patients was reported in 12 studies (6309 tumours, 31 centres, 13 countries) for ADNEX without CA125 (**Table S4 and S6**), and the summary AUC was 0.93 (CI 95% 0.91-0.94, 95% PI 0.85-0.98) (**Table 2, Figure 4**). [22,26,30,31,37,43-46,49,56,61] Twenty-one studies (9202 tumours, 43 centres, 18 countries) reported the AUC for benign vs malignant tumours in operated patients for ADNEX with CA125, [18,19,23,26,30,35,37,40,41,44-46,49,51,52,55,56,58,60,61,100] (**Table S4 and S6**) and the summary AUC was 0.93 (CI 95% 0.92-0.94, 95% PI 0.85-0.98) (**Table 2, Figure 5**).

Table 2. Meta-analysis of the AUC to discriminate between benign and malignant tumours, including sensitivity and subgroup analyses.

Meta-analysis	Number of studies (centres)	Summary estimate (95% CI)	95% PI	τ^2
<i>Main analysis</i>				
Operated patients, with CA125	21 (43)	0.93 (0.92-0.94)	0.85 to 0.98	0.25
Operated patients, without CA125	12 (31)	0.93 (0.91-0.94)	0.85 to 0.98	0.21
<i>Sensitivity analyses (operated patients)</i>				
High or Unclear risk of bias studies, with CA125	19 (23)	0.93 (0.91-0.95)	0.84 to 0.99	0.33
Low risk of bias studies, with CA125	2 (20)	0.93 (0.91-0.94)	0.87 to 0.97	0.12
IOTA studies, with CA125	4 (23)	0.92 (0.91-0.94)	0.88 to 0.97	0.09
Non-IOTA studies, with CA125	17 (20)	0.93 (0.91-0.95)	0.83 to 0.99	0.38
High or Unclear risk of bias studies, without CA125	10 (11)	0.94 (0.92-0.96)	0.86 to 0.99	0.26
Low risk of bias studies, without CA125	2 (20)	0.91 (0.90-0.93)	0.85 to 0.96	0.11
IOTA studies, without CA125	2 (20)	0.91 (0.89-0.93)	0.85 to 0.97	0.14
Non-IOTA studies, without CA125	10 (11)	0.94 (0.92-0.96)	0.86 to 0.99	0.26
<i>Subgroup analyses (operated patients)</i>				
Asian centres, with CA125	11 (13)	0.94 (0.91-0.96)	0.81 to 1	0.54
Asian centres, without CA125	7 (8)	0.95 (0.93-0.97)	0.89 to 0.99	0.19
Chinese centres, with CA125	5 (5)	0.94 (0.91-0.97)	0.87 to 0.99	0.25
Chinese centres, without CA125	4 (4)	0.94 (0.90-0.97)	0.85 to 0.99	0.33
European centres, with CA125	8 (28)	0.92 (0.91-0.94)	0.88 to 0.96	0.07
European centres, without CA125	3 (21)	0.91 (0.89-0.93)	0.84 to 0.97	0.16
Non-oncology centres, with CA125	2 (9)	0.91 (0.88-0.94)	0.85 to 0.97	0.11
Non-oncology centres, without CA125	1 (8)	0.90 (0.86-0.95)	0.80 to 0.99	0.24
Oncology centres, with CA125	18 (31)	0.93 (0.92-0.95)	0.84 to 0.98	0.29
Oncology centres, without CA125	12 (23)	0.93 (0.91-0.94)	0.85 to 0.98	0.22
Postmenopausal patients, with CA125	11 (32)	0.92 (0.90-0.94)	0.83 to 0.98	0.25
Postmenopausal patients, without CA125	4 (22)	0.90 (0.87-0.93)	0.81 to 0.97	0.19
Premenopausal patients, with CA125	11 (32)	0.91 (0.89-0.94)	0.81 to 0.98	0.28
Premenopausal patients, without CA125	4 (22)	0.91 (0.89-0.93)	0.85 to 0.96	0.08
<i>Target population</i>				
Operated and non-surgically managed patients, with CA125	2 (18)	0.94 (0.93-0.96)	0.88 to 0.99	0.22
Operated and non-surgically managed patients, without CA125	1 (17)	0.94 (0.91-0.95)	0.82 to 0.98	0.27

Sensitivity and specificity at the 10% risk of malignancy threshold in operated patients were reported in ten studies for ADNEX without CA125 (**Tables S4 and S7**). [22,26,31,37,43,45,46,49,56,61] The summary sensitivity and specificity were 0.93 (95% CI 0.90-0.95, 95% PI 0.73-0.99) and 0.75 (95% CI 0.70-0.79, 95% PI 0.46-0.91), respectively (**Table S8, Figure 6**). For ADNEX with CA125, sensitivity and specificity at the 10% risk of malignancy threshold in operated patients were reported in 24 studies (**Tables S4 and S7**). [9,18,22,26,30,35–37,40,41,43–46,48,49,51,55,56,58–61,100] The summary sensitivity and specificity were 0.94 (CI 95% 0.92-0.95; 95% PI 0.80-0.98) and 0.77 (CI 95% 0.73-0.81; 95% PI 0.47-0.93, respectively (**Table S8, Figure 7**).

NB and RU were calculated using studies that presented sensitivity and specificity at the 10% risk of malignancy threshold in operated patients (**Table S7**). For ADNEX without CA125, the summary NB was 0.28 (95% CrI 0.22-0.35, 95% PI 0.05-0.67) and the summary RU 0.50 (95% CrI 0.36-0.62, 95% PI -0.41-0.81). The probability that the model is clinically useful in a random new centre was estimated to be 91% (**Table S9, Figures S6-7**). For ADNEX with CA125, the summary NB was 0.27 (95% CrI 0.23-0.33, 95% PI 0.06-0.64) and the summary RU 0.54 (95% CrI 0.45-0.61, 95% PI -0.12-0.78). The probability that the model is clinically useful in a random new centre was estimated to be 95% (**Table S9, Figures S8-9**).

Pairwise AUCs in operated patients were reported in 4 studies for ADNEX without CA125 and in 5 studies for ADNEX with CA125 (**Table S10**). [58,101–104] The summary pairwise AUCs for ADNEX without CA125 ranged from 0.66 (stage II-IV primary invasive vs metastatic) to 0.97 (benign vs stage II-IV primary invasive) (**Table S11**). For ADNEX with CA125, the summary pairwise AUCs ranged from 0.72 (borderline vs stage I primary invasive) to 0.98 (benign vs stage II-IV primary invasive).

The AUC for benign vs malignant tumours in operated and non-surgically managed patients combined was reported in 2 studies (5167 tumours, 18 centres, 8 countries) with a summary estimate of 0.94 (95% CI 0.93-0.96, 95% PI 0.88-0.99) for ADNEX with CA125 (**Table 2**). [8,104] ADNEX without CA125 was assessed in only 1 study (4905 tumours, 17 centres, 7 countries). [104] This study reported summary AUC 0.94 (95% CI 0.91-0.95, 95% PI 0.82-0.98).

Sensitivity and subgroup results for AUC, specificity, sensitivity, NB and RU are shown in **Tables 2 and S6-9**. These results showed that findings were robust (AUCs ranged from 0.90 to 0.95 across all analyses) and clinical utility was suggested in all subgroups. When limiting the analyses to low RoB studies, summary AUCs were 0.93 (with CA125) and 0.91 (without CA125). Sensitivity was higher and specificity lower in oncology versus non-oncology centres and in postmenopausal versus premenopausal patients. In line with

this, meta-regression suggested that prevalence of malignancy was not related to AUC but positively to sensitivity and negatively to specificity (**Figures S10-12**).

Meta-analysis of calibration in only operated patients was not feasible as only one study reported calibration slope and intercept.[104] Four studies presented a calibration plot: in three studies [102,103,105] the estimated risks were close to the observed ones, in one study [104] the risk of malignancy was slightly underestimated.

Based on the subdomains of GRADE assessment, we have identified the risk of bias in the studies included in this meta-analysis as a significant limitation affecting the certainty of our meta-analysis results. Two studies (representing 5511 tumours or 32% of total tumours included in this review) did not fall under the classification of high risk of bias. [103,104] However, the sensitivity analysis according to risk of bias showed consistent findings (Table 2, Table S8 and S9). Funnel plots for the AUC did not suggest publication bias (**Figure S13**).

DISCUSSION

The ADNEX model exhibited robust discrimination and classification performance between benign and malignant tumours across various settings and populations. In addition, the results indicate that ADNEX has clinical utility at the 10% risk of malignancy threshold, for example to decide whether a patient should be referred for assessment in a gynaecological oncology centre.

We found deficiencies in study reporting, and most studies were judged to be at high risk of bias. However, our sensitivity analyses indicated that performance was almost identical in low risk and high risk of bias studies. For ADNEX with CA125, the AUC was 0.93 based on all studies vs 0.93 when based on only low risk of bias studies. For ADNEX without CA125, the AUCs were 0.93 (all studies) and 0.91 (low risk of bias only). High risk of bias was mainly caused by low sample size, absence of calibration performance, and unjustified use of complete case analysis. Low sample size makes the estimated AUC less precise but does not systematically affect the AUC, unless there is publication bias. The funnel plots do not suggest publication bias. Absence of calibration does not affect the AUC. Using complete cases regarding CA125 or other predictors may lead to underestimation of the AUC, because missing values tend to be associated with the examiner's subjective impression that the tumour is benign.[106,107] Complete case analysis would then tend to exclude clearly benign tumours, which would make the sample more homogeneous and reduce

the AUC. However, our results suggest that the impact of complete case analysis on performance may have been minimal. The impact on calibration could not be assessed. Taken together, the results presented in our meta-analysis are reliable.

Strengths of our systematic review include the meta-analysis of ADNEX both as a risk model and as a diagnostic test, and the thorough critical appraisal regarding risk of bias and reporting quality using recommended checklists.[79,108] Our study also has limitations. First, some study authors have a conflict of interest because they were involved in developing ADNEX or in some of the included external validation studies. To address this, independent researchers with expertise in study methodology and prediction modelling evaluated the IOTA-related studies. Second, calibration performance was reported in only four studies, and meta-analysis of calibration was not possible.

Previous systematic reviews have conducted meta-analyses of the diagnostic performance of ADNEX with CA125.[62–67] They used the QUADAS-2 tool [109] to assess risk of bias, and found between 0 and 64% of studies to be at high risk of bias in at least one domain. We identified 45/47 studies as high risk of bias using the PROBAST tool designed for the appraisal of risk prediction models. None of the previous meta-analyses of ADNEX included clinical utility, calibration or AUC. Our results align with those of other systematic reviews of risk prediction modelling studies in various domains. These consistently indicate that reporting in the original studies was poor and that many studies were at high risk of bias.[110–116]. Still, the results in terms of sensitivity and specificity of ADNEX in our meta-analysis were similar to those in the other meta-analyses of ADNEX performance.

Our findings support the use of ADNEX (1) to decide where, by whom and how to operate when a decision to operate a patient has already been taken (meta-analysis in operated patients), or (2) to choose between surgery and conservative follow-up (meta-analysis in operated and non-surgically managed patients). Because the AUC of ADNEX with and without CA125 are similar, and because adding CA125 mainly helps to distinguish between different types of malignant tumours, we argue that the main use of ADNEX without CA125 is to support the decision whether to operate or follow non-surgically, whereas the main use of ADNEX with CA125 is to support decisions regarding type of surgery. Therefore, we believe that measuring CA125 is often not needed.

While our findings suggest that ADNEX has clinical utility, well conducted validations of any model are always of value to monitor its performance in diverse regions and clinical settings, and over time.[117] In line with this, to improve ADNEX performance even further, efforts to update the ADNEX formula are of

interest.[117,118] If further validation studies are conducted, we recommend to include a validation of ADNEX without CA125, to obtain a sufficiently large sample to assess calibration and multinomial discrimination, and to include patients irrespective of whether they are managed surgically or non-surgically, despite the challenges regarding reference standard for non-surgically managed patients.[87,119] Methodological recommendations for validation studies include using available tools to guarantee adequate sample size, describe missing data in detail and use methods such as imputation when needed, and assess calibration performance.[118,120–124] Finally, adherence to the TRIPOD reporting checklists is important to maximise the value of the validation study (www.tripod-statement.org).

CONCLUSION

ADNEX has been widely validated with AUC values >0.90 to discriminate between benign and malignant tumours across various settings, and with strong results regarding clinical utility at the 10% risk of malignancy threshold.

ACKNOWLEDGMENT

The authors wish to thank Thomas Vandendriessche, Chayenne Van Meel and Krizia Tuand, the biomedical reference librarians of the KU Leuven Libraries – 2Bergen – learning Centre Désiré Collen (Leuven, Belgium), for their help in conducting the systematic literature search.

We gratefully acknowledge the authors of the external validation of ADNEX for providing their valuable data.

This research will be presented at ISUOG World Congress 2023 in Seoul (October 2023).

OTHER INFORMATION

Funding

This research was supported by the Research Foundation – Flanders (FWO) under grant G097322N with BVC and DT as supervisors. LW and BVC were supported by Internal Funds KU Leuven (grant C24M/20/064). JYV was supported by the National Institute for Health and Care Research (NIHR) Community Healthcare MedTech and In Vitro Diagnostics Co-operative at Oxford Health NHS Foundation

Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. GSC is supported by Cancer Research UK (programme grant: C49297/A27294). PD is supported by CRUK (project grant: PRCPJT-Nov21\100021)

Conflicts of interest

BVC, LV and DT are members of the steering committee of the International Ovarian Tumour Analysis (IOTA) consortium and were involved in the development of the ADNEX model. BVC and DT report consultancy work done by KU Leuven to help implementing and testing the ADNEX model in ultrasound machines by Samsung Medison, GE Healthcare, Canon Medical Systems Europe, and Shenzhen Mindray Bio-medical Electronics, outside the submitted work. GSC is a statistics editor for the BMJ.

Transparency declaration

The lead authors affirm that this 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 (and, if relevant, registered) have been explained.

Availability of data and code

The data and code used for the analysis and the generation of the figures is publicly available in Open Science Framework (<https://osf.io/jtsvd/>). [80] This includes the extraction sheet, one script for running the meta-analysis and one script for generating all the figures included in the paper (except PRISMA Flowchart). Note that for some rows in the extraction sheet we deleted the performance measures for the menopausal subgroups because this information was provided by authors upon request therefore it is not publicly available.

REFERENCES

- 1 Carley ME, Klingele CJ, Gebhart JB, *et al.* Laparoscopy versus laparotomy in the management of benign unilateral adnexal masses. *J Am Assoc Gynecol Laparosc* 2002;**9**:321–6. doi:10.1016/s1074-3804(05)60411-2
- 2 Froyman W, Landolfo C, De Cock B, *et al.* Risk of complications in patients with conservatively managed ovarian tumours (IOTA5): a 2-year interim analysis of a multicentre, prospective, cohort study. *Lancet Oncol* Published Online First: 2019. doi:10.1016/S1470-2045(18)30837-4
- 3 Woo YL, Kyrgiou M, Bryant A, *et al.* Centralisation of services for gynaecological cancers - a Cochrane systematic review. *Gynecol Oncol* 2012;**126**:286–90. doi:10.1016/j.ygyno.2012.04.012
- 4 Timmerman D, Planchamp F, Bourne T, *et al.* ESGO/ISUOG/IOTA/ESGE Consensus Statement on preoperative diagnosis of ovarian tumors. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol* 2021;**58**:148–68. doi:10.1002/uog.23635
- 5 Van Calster B, Hoorde KV, Valentin L, *et al.* 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. *BMJ* 2014;**349**:g5920. doi:10.1136/bmj.g5920
- 6 Van Calster B, Wynants L, Timmerman D, *et al.* Predictive analytics in health care: how can we know it works? *J Am Med Inform Assoc* 2019;**26**:1651–4. doi:10.1093/jamia/ocz130
- 7 Van Calster B, Valentin L, Froyman W, *et al.* Validation of models to diagnose ovarian cancer in patients managed surgically or conservatively: multicentre cohort study. *BMJ* 2020;**370**:m2614. doi:10.1136/bmj.m2614
- 8 Hack K, Gandhi N, Bouchard-Fortier G, *et al.* External Validation of O-RADS US Risk Stratification and Management System. *Radiology* Published Online First: 2022. doi:10.1148/radiol.211868
- 9 Jeong SY, Park BK, Lee YY, *et al.* Validation of IOTA-ADNEX Model in Discriminating Characteristics of Adnexal Masses: A Comparison with Subjective Assessment. *J Clin Med* 2020;**9**. doi:10.3390/jcm9062010
- 10 Quaranta M, Nath R, Mehra G, *et al.* Surgery of Benign Ovarian Masses by a Gynecological Cancer Surgeon: A Cohort Study in a Tertiary Cancer Centre. *Cureus* 2020;**12**:e9201. doi:10.7759/cureus.9201
- 11 Diagnostik - RCC Kunskapsbanken. <https://kunskapsbanken.cancercentrum.se/diagnoser/aggstockscancer-epitelial/vardprogram/diagnostik/> (accessed 14 Jun 2023).
- 12 Vergote Ignace, Vlayen Joan, Heus Pauline, *et al.* Eierstokkanker: diagnose, behandeling en follow-up. Brussel: : Federaal Kenniscentrum voor de Gezondheidszorg (KCE) 2016.
- 13 Geomini P. ACCEPT (ACcuracy, Cost Effectiveness of Prediction models for ovarian Tumors): a study on the cost-effectiveness of risk scoring models for the discrimination between benign or malignant

- ovarian tumors. ZonMw Proj. 2021.<https://projecten.zonmw.nl/nl/project/accept-accuracy-cost-effectivenss-prediction-models-ovarian-tumors-study-cost-effectiveness> (accessed 3 Jul 2023).
- 14 Andreotti RF, Timmerman D, Strachowski LM, *et al.* O-RADS US risk stratification and management system: A consensus guideline from the ACR ovarian-Adnexal Reporting and Data System committee. *Radiology* Published Online First: 2020. doi:10.1148/radiol.2019191150
 - 15 Wereldprimeur: integratie van diagnosemodel ADNEX in echoapparatuur van Samsung. <https://news.samsung.com/nl/wereldprimeur-integratie-van-diagnosemodel-adnex-in-echoapparatuur-van-samsung> (accessed 20 Apr 2023).
 - 16 GE Voluson™ Club - Ultrasound Practice Guidelines. <https://www.volusonclub.net/emea/guidelines/IOTA-Risk-Assessment> (accessed 20 Apr 2023).
 - 17 Supporting Advances in Gynecologic Oncology | Ultrasound | Canon Medical Systems. https://ca.medical.canon/publication/ultrasound/V37_ULEU210081 (accessed 20 Apr 2023).
 - 18 Araujo KG, Jales RM, Pereira PN, *et al.* Performance of the IOTA ADNEX model in preoperative discrimination of adnexal masses in a gynecological oncology center. *Ultrasound Obstet Gynecol* Published Online First: 2017. doi:10.1002/uog.15963
 - 19 Behnamfar F, Esmaeilian F, Adibi A, *et al.* Comparison of Ultrasound and Tumor Marker CA125 in Diagnosis of Adnexal Mass Malignancies. *Adv Biomed Res* 2022;**11**:18. doi:10.4103/abr.abr_164_20
 - 20 Budiana ING, Suwiyoga K, Suwardewa TGA, *et al.* Skor assessment of different neoplasias in the adnexa (ADNEX) untuk memprediksi keganasan ovarium di RSUP Sanglah Denpasar. *Intisari Sains Medis* 2022;**13**:197–201. doi:10.15562/ism.v13i1.1328
 - 21 Butureanu T, Socolov D, Matasariu DR, *et al.* Ovarian masses-applicable iota adnex model versus morphological findings for accurate diagnosis and treatment. *Appl Sci Switz* Published Online First: 2021. doi:10.3390/app112210789
 - 22 Chen G-Y, Hsu T-F, Chan I-S, *et al.* Comparison of the O-RADS and ADNEX models regarding malignancy rate and validity in evaluating adnexal lesions. *Eur Radiol* 2022;**32**:7854–64. doi:10.1007/s00330-022-08803-6
 - 23 Chen H, Qian L, Jiang M, *et al.* Performance of IOTA ADNEX model in evaluating adnexal masses in a gynecological oncology center in China. *Ultrasound Obstet Gynecol* Published Online First: 2019. doi:10.1002/uog.20363
 - 24 Czekierdowski A, Stachowicz N, Smoleń A, *et al.* Sonographic Assessment of Complex Ultrasound Morphology Adnexal Tumors in Pregnant Women with the Use of IOTA Simple Rules Risk and ADNEX Scoring Systems. *Diagn Basel Switz* 2021;**11**. doi:10.3390/diagnostics11030414
 - 25 Czekierdowski A, Stachowicz N, Smolen A, *et al.* Performance of IOTA Simple Rules Risks, ADNEX Model, Subjective Assessment Compared to CA125 and HE4 with ROMA Algorithm in Discriminating between Benign, Borderline and Stage I Malignant Adnexal Lesions. *Diagnostics* 2023;**13**:885. doi:10.3390/diagnostics13050885

- 26 Díaz L, Santos M, Zambrano B, *et al.* Ovarian tumors: Risk of malignancy and IOTA ADNEX model indexes. No technology Doppler diagnostic options [Tumores de ovario: Índices de riesgo de malignidad y modelo ADNEX de IOTA. Opciones diagnósticas sin tecnología doppler]. *Rev Obstet Ginecol Venez* Published Online First: 2017. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85041059604&partnerID=40&md5=e4865970adeccaac5c4f5c384ee90747>
- 27 Epstein E, Van Calster B, Timmerman D, *et al.* Subjective ultrasound assessment, the ADNEX model and ultrasound-guided tru-cut biopsy to differentiate disseminated primary ovarian cancer from metastatic non-ovarian cancer. *Ultrasound Obstet Gynecol* Published Online First: 2016. doi:10.1002/uog.14892
- 28 Esquivel Villabona AL, Rodríguez JN, Ayala N, *et al.* Two-Step Strategy for Optimizing the Preoperative Classification of Adnexal Masses in a University Hospital, Using International Ovarian Tumor Analysis Models: Simple Rules and Assessment of Different NEoplasias in the adneXa Model. *J Ultrasound Med* Published Online First: 2022. doi:10.1002/jum.15728
- 29 Gaurilcikas A, Gedgaudaite M, Cizauskas A, *et al.* Performance of the IOTA ADNEX Model on Selected Group of Patients with Borderline Ovarian Tumours. *Med Kaunas Lith* 2020;**56**. doi:10.3390/medicina56120690
- 30 He P, Wang J-J, Duan W, *et al.* Estimating the risk of malignancy of adnexal masses: validation of the ADNEX model in the hands of nonexpert ultrasonographers in a gynaecological oncology centre in China. *J Ovarian Res* 2021;**14**:169. doi:10.1186/s13048-021-00922-w
- 31 Hiett AK, Sonek JD, Guy M, *et al.* Performance of IOTA Simple Rules, Simple Rules risk assessment, ADNEX model and O-RADS in differentiating between benign and malignant adnexal lesions in North American women. *Ultrasound Obstet Gynecol* Published Online First: 2022. doi:10.1002/uog.24777
- 32 Hu Y, Chen B, Dong H, *et al.* Comparison of ultrasound-based ADNEX model with magnetic resonance imaging for discriminating adnexal masses: a multi-center study. *Front Oncol* 2023;**13**:1101297. doi:10.3389/fonc.2023.1101297
- 33 Jiang M-J, Le Q, Yang B-W, *et al.* Ovarian sex cord stromal tumours: analysis of the clinical and sonographic characteristics of different histopathologic subtypes. *J Ovarian Res* 2021;**14**:53. doi:10.1186/s13048-021-00805-0
- 34 JIANHONG S, Lei T, Wu L, *et al.* Comparison of performance between O-RADS, IOTA Simple Rules Risk assessment and ADNEX model in the discrimination of ovarian Brenner tumors. Published Online First: 2022. doi:10.21203/rs.3.rs-2160740/v1
- 35 Joyeux E, Miras T, Masquin I, *et al.* Before surgery predictability of malignant ovarian tumors based on ADNEX model and its use in clinical practice [Prédictibilité préopératoire de la malignité des tumeurs ovariennes à partir du score ADNEX et utilisation en pratique clinique]. *Gynecol Obstet Fertil* Published Online First: 2016. doi:10.1016/j.gyobfe.2016.07.007
- 36 Lai H-W, Lyu G-R, Kang Z, *et al.* Comparison of O-RADS, GI-RADS, and ADNEX for Diagnosis of Adnexal Masses: An External Validation Study Conducted by Junior Sonologists. *J Ultrasound Med* Published Online First: 2022. doi:10.1002/jum.15834

- 37 Lam Huong L, Thi Phuong Dung N, Hoang Lam V, *et al.* The Optimal Cut-Off Point of the Andex Model for the Prediction of the Ovarian Cancer Risk. *Asian Pac J Cancer Prev APJCP* 2022;**23**:2713–8. doi:10.31557/APJCP.2022.23.8.2713
- 38 Lee SJ, Kim Y-H, Lee M-Y, *et al.* Ultrasonographic evaluation of ovarian mass for predicting malignancy in pregnant women. *Gynecol Oncol* Published Online First: 2021. doi:10.1016/j.ygyno.2021.09.007
- 39 Liu B, Liao J, Gu W, *et al.* ADNEX model-based diagnosis of ovarian cancer using MRI images. *Contrast Media Mol Imaging* 2021;**2021**. doi:10.1155/2021/2146578
- 40 Meys EMJ, Jeelof LS, Achten NMJ, *et al.* Estimating risk of malignancy in adnexal masses: external validation of the ADNEX model and comparison with other frequently used ultrasound methods. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol* 2017;**49**:784–92. doi:10.1002/uog.17225
- 41 Nam G, Lee SR, Jeong K, *et al.* Assessment of different NEoplasias in the adneXa model for differentiation of benign and malignant adnexal masses in Korean women. *Obstet Gynecol Sci* 2021;**64**:293–9. doi:10.5468/ogs.21012
- 42 Nohuz E, De Simone L, Chêne G. Reliability of IOTA score and ADNEX model in the screening of ovarian malignancy in postmenopausal women. *J Gynecol Obstet Hum Reprod* Published Online First: 2019. doi:10.1016/j.jogoh.2018.04.012
- 43 Pelayo M, Pelayo-Delgado I, Sancho-Saucó J, *et al.* Comparison of Ultrasound Scores in Differentiating between Benign and Malignant Adnexal Masses. *Diagn Basel Switz* 2023;**13**. doi:10.3390/diagnostics13071307
- 44 Peng X-S, Ma Y, Wang L-L, *et al.* Evaluation of the Diagnostic Value of the Ultrasound ADNEX Model for Benign and Malignant Ovarian Tumors. *Int J Gen Med* 2021;**14**:5665–73. doi:10.2147/IJGM.S328010
- 45 Poonyakanok V, Tanmahasamut P, Jaishuen A, *et al.* Preoperative Evaluation of the ADNEX Model for the Prediction of the Ovarian Cancer Risk of Adnexal Masses at Siriraj Hospital. *Gynecol Obstet Invest* Published Online First: 2021. doi:10.1159/000513517
- 46 Qian L, Du Q, Jiang M, *et al.* Comparison of the Diagnostic Performances of Ultrasound-Based Models for Predicting Malignancy in Patients With Adnexal Masses. *Front Oncol* 2021;**11**:673722. doi:10.3389/fonc.2021.673722
- 47 Rashmi N, Singh S, Begum J, *et al.* Diagnostic Performance of Ultrasound-Based International Ovarian Tumor Analysis Simple Rules and Assessment of Different NEoplasias in the adneXa Model for Predicting Malignancy in Women with Ovarian Tumors: A Prospective Cohort Study. *Womens Health Rep New Rochelle N* 2023;**4**:202–10. doi:10.1089/whr.2022.0072
- 48 Sandal K, Polat M, Yassa M, *et al.* Comparison of ‘risk of malignancy indices’ and ‘assessment of different neoplasia in the adnexa’ (ADNEX) model as preoperative malignancy evaluation methods for adnexal masses. *Zeynep Kamil Tip Bul* 2018;**49**:324–9. doi:10.16948/zkctipb.445755

- 49 Sayasneh A, Ferrara L, De Cock B, *et al.* Evaluating the risk of ovarian cancer before surgery using the ADNEX model: a multicentre external validation study. *Br J Cancer* 2016;**115**:542–8. doi:10.1038/bjc.2016.227
- 50 Stukan M, Alcazar JL, Gębicki J, *et al.* Ultrasound and Clinical Preoperative Characteristics for Discrimination Between Ovarian Metastatic Colorectal Cancer and Primary Ovarian Cancer: A Case-Control Study. *Diagn Basel Switz* 2019;**9**. doi:10.3390/diagnostics9040210
- 51 Stukan M, Badocha M, Ratajczak K. Development and validation of a model that includes two ultrasound parameters and the plasma D-dimer level for predicting malignancy in adnexal masses: an observational study. *BMC Cancer* 2019;**19**:564. doi:10.1186/s12885-019-5629-x
- 52 Szubert S, Szperek D, Wójtowicz A, *et al.* Performance of Selected Models for Predicting Malignancy in Ovarian Tumors in Relation to the Degree of Diagnostic Uncertainty by Subjective Assessment With Ultrasound. *J Ultrasound Med Off J Am Inst Ultrasound Med* 2020;**39**:939–47. doi:10.1002/jum.15178
- 53 Szubert S, Wojtowicz A, Moszynski R, *et al.* External validation of the IOTA ADNEX model performed by two independent gynecologic centers. *Gynecol Oncol* Published Online First: 2016. doi:10.1016/j.ygyno.2016.06.020
- 54 Tavoraitė I, Kronlachner L, Opolskienė G, *et al.* Ultrasound Assessment of Adnexal Pathology: Standardized Methods and Different Levels of Experience. *Med Kaunas Lith* 2021;**57**. doi:10.3390/medicina57070708
- 55 Tug N, Yassa M, Sargin MA, *et al.* Preoperative discriminating performance of the IOTA-ADNEX model and comparison with risk of malignancy index: An external validation in a non-gynecologic oncology tertiary center. *Eur J Gynaecol Oncol* Published Online First: 2020. doi:10.31083/j.ejgo.2020.02.4971
- 56 Van Calster B, Valentin L, Froyman W, *et al.* Validation of models to diagnose ovarian cancer in patients managed surgically or conservatively: multicentre cohort study. *BMJ* 2020;:m2614. doi:10.1136/bmj.m2614
- 57 Velayo CL, Reforma KN, Sicam RVG, *et al.* Diagnostic Performances of Ultrasound-Based Models for Predicting Malignancy in Patients with Adnexal Masses. *Healthcare* 2022;**11**:8. doi:10.3390/healthcare11010008
- 58 Viora E, Piovano E, Baima Poma C, *et al.* The ADNEX model to triage adnexal masses: An external validation study and comparison with the IOTA two-step strategy and subjective assessment by an experienced ultrasound operator. *Eur J Obstet Gynecol Reprod Biol* Published Online First: 2020. doi:10.1016/j.ejogrb.2020.02.022
- 59 Wang R, Yang Z. Evaluating the risk of malignancy in adnexal masses: validation of O-RADS and comparison with ADNEX model, SA, and RMI. *Ginekol Pol* Published Online First: 2023. doi:10.5603/GP.a2023.0019
- 60 Yang S, Tang J, Rong Y, *et al.* Performance of the IOTA ADNEX model combined with HE4 for identifying early-stage ovarian cancer. *Front Oncol* 2022;**12**:949766. doi:10.3389/fonc.2022.949766

- 61 Zhang Y, Zhao Y, Feng L. External Validation of the Assessment of Different NEoplasias in the adneXa Model Performance in Evaluating the Risk of Ovarian Carcinoma Before Surgery in China: A Tertiary Center Study. *J Ultrasound Med* Published Online First: 2022. doi:10.1002/jum.15920
- 62 Cui L, Xu H, Zhang Y. Diagnostic Accuracies of the Ultrasound and Magnetic Resonance Imaging ADNEX Scoring Systems For Ovarian Adnexal Mass: Systematic Review and Meta-Analysis. *Acad Radiol* 2022;**29**:897–908. doi:10.1016/j.acra.2021.05.029
- 63 Davenport CF, Rai N, Sharma P, *et al.* Diagnostic Models Combining Clinical Information, Ultrasound and Biochemical Markers for Ovarian Cancer: Cochrane Systematic Review and Meta-Analysis. *Cancers* 2022;**14**. doi:10.3390/cancers14153621
- 64 Davenport C, Rai N, Sharma P, *et al.* Menopausal status, ultrasound and biomarker tests in combination for the diagnosis of ovarian cancer in symptomatic women. *Cochrane Database Syst Rev* 2022;**2022**. doi:10.1002/14651858.CD011964.pub2
- 65 Huang X, Wang Z, Zhang M, *et al.* Diagnostic Accuracy of the ADNEX Model for Ovarian Cancer at the 15% Cut-Off Value: A Systematic Review and Meta-Analysis. *Front Oncol* 2021;**11**:684257. doi:10.3389/fonc.2021.684257
- 66 Westwood M, Ramaekers B, Lang S, *et al.* Risk scores to guide referral decisions for people with suspected ovarian cancer in secondary care: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2018;**22**:1–264. doi:10.3310/hta22440
- 67 Yue X, Zhong L, Wang Y, *et al.* Value of Assessment of Different Neoplasias in the Adnexa in the Differential Diagnosis of Malignant Ovarian Tumor and Benign Ovarian Tumor: A Meta-analysis. *Ultrasound Med Biol* Published Online First: 2022. doi:10.1016/j.ultrasmedbio.2022.02.001
- 68 Westwood M, Ramaekers B, Lang S, *et al.* Risk scores to guide referral decisions for people with suspected ovarian cancer in secondary care: A systematic review and cost-effectiveness analysis. *Health Technol Assess* Published Online First: 2018. doi:10.3310/hta22440
- 69 on behalf of the Topic Group ‘Evaluating diagnostic tests and prediction models’ of the STRATOS initiative, Wynants L, van Smeden M, *et al.* Three myths about risk thresholds for prediction models. *BMC Med* 2019;**17**:192. doi:10.1186/s12916-019-1425-3
- 70 Debray TPA, Damen JAAG, Snell KIE, *et al.* A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;**356**:i6460. doi:10.1136/bmj.i6460
- 71 Moons KGM, Wolff RF, Riley RD, *et al.* PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* 2019;**170**:W1. doi:10.7326/M18-1377
- 72 Snell KIE, Levis B, Damen JAA, *et al.* Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). *BMJ* 2023;**381**:e073538. doi:10.1136/bmj-2022-073538
- 73 Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;**n71**. doi:10.1136/bmj.n71

- 74 Barreñada L, Ledger A, Verbakel JY, *et al.* The ADNEX model for ovarian cancer diagnosis: A systematic review and meta-analysis of external validation studies. Published Online First: 2022. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=373182 (accessed 21 Apr 2023).
- 75 Steyerberg EW. Updating for a New Setting. In: Steyerberg EW, ed. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Cham: : Springer International Publishing 2019. 399–429. doi:10.1007/978-3-030-16399-0_20
- 76 Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*. London, England, United Kingdom: : ACM Press 2014. 1–10. doi:10.1145/2601248.2601268
- 77 Ouzzani M, Hammady H, Fedorowicz Z, *et al.* Rayyan—a web and mobile app for systematic reviews. *Syst Rev* 2016;**5**. doi:10.1186/s13643-016-0384-4
- 78 Moons KGM, de Groot JAH, Bouwmeester W, *et al.* Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Med* 2014;**11**:e1001744. doi:10.1371/journal.pmed.1001744
- 79 Gary S Collins, Johannes B Reitsma, Douglas G Altman, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis tripod the tripod statement. *BMJ* Published Online First: 2015. doi:10.1136/bmj.g7594
- 80 Barreñada L, Ledger A, Collins G, *et al.* The ADNEX risk model for ovarian cancer diagnosis: A systematic review and meta-analysis of external validation studies. Published Online First: 3 July 2023. doi:10.17605/OSF.IO/JTSVD
- 81 Riley RD, Abrams KR, Sutton AJ, *et al.* Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol* 2007;**7**:3. doi:10.1186/1471-2288-7-3
- 82 Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *The BMJ* 2016;**352**:i6. doi:10.1136/bmj.i6
- 83 Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med Decis Making* 2006;**26**:565–74. doi:10.1177/0272989X06295361
- 84 Baker SG. Putting Risk Prediction in Perspective: Relative Utility Curves. *JNCI J Natl Cancer Inst* 2009;**101**:1538–42. doi:10.1093/jnci/djp353
- 85 Baker SG, Cook NR, Vickers A, *et al.* Using Relative Utility Curves to Evaluate Risk Prediction. *J R Stat Soc Ser A Stat Soc* 2009;**172**:729–48. doi:10.1111/j.1467-985X.2009.00592.x
- 86 Wynants L, Riley RD, Timmerman D, *et al.* Random-effects meta-analysis of the clinical utility of tests and prediction models. *Stat Med* 2018;**37**:2034–52. doi:10.1002/sim.7653

- 87 Van Calster B, Vergouwe Y, Looman CWN, *et al.* Assessing the discriminative ability of risk models for more than two outcome categories. *Eur J Epidemiol* 2012;**27**:761–70. doi:10.1007/s10654-012-9733-3
- 88 Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;**21**:1559–73. doi:10.1002/sim.1187
- 89 Guyatt GH, Oxman AD, Vist GE, *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;**336**:924–6. doi:10.1136/bmj.39489.470347.AD
- 90 Debray TP, Damen JA, Riley RD, *et al.* A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat Methods Med Res* 2019;**28**:2768–86. doi:10.1177/0962280218785504
- 91 Schwarzer G, Carpenter JR, Rücker G. *Meta-Analysis with R*. Cham: : Springer International Publishing 2015. doi:10.1007/978-3-319-21416-0
- 92 Doebler P, Holling H. Meta-Analysis of Diagnostic Accuracy with mada.
- 93 Reitsma JB, Glas AS, Rutjes AWS, *et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;**58**:982–90. doi:10.1016/j.jclinepi.2005.02.022
- 94 Plummer M. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. 2003.
- 95 Plummer M. *rjags: Bayesian Graphical Models using MCMC*. 2022. <https://CRAN.R-project.org/package=rjags>
- 96 Landolfo C, Bourne T, Froyman W, *et al.* Benign descriptors and ADNEX in two-step strategy to estimate risk of malignancy in ovarian tumors: retrospective validation on IOTA 5 multicenter cohort. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol* Published Online First: 30 September 2022. doi:10.1002/uog.26080
- 97 Poonyakanok V, Tanmahasamut P, Jaishuen A. Prospective comparative trial comparing O-RADS, IOTA ADNEX model, and RMI score for preoperative evaluation of adnexal masses for prediction of ovarian cancer. *J Obstet Gynaecol Res* 2023.
- 98 Froyman W, Wynants L, Landolfo C, *et al.* Validation of the Performance of International Ovarian Tumor Analysis (IOTA) Methods in the Diagnosis of Early Stage Ovarian Cancer in a Non-Screening Population. *Diagn Basel Switz* 2017;**7**. doi:10.3390/diagnostics7020032
- 99 Wynants L, Timmerman D, Verbakel JY, *et al.* Clinical utility of risk models to refer patients with adnexal masses to specialized oncology care: Multicenter external validation using decision curve analysis. *Clin Cancer Res* Published Online First: 2017. doi:10.1158/1078-0432.CCR-16-3248
- 100 Szubert S, Wojtowicz A, Zywicka P. Response to letter to the editor concerning validation of IOTA ADNEX model. *Gynecol Oncol Rep* 2016;**18**:51–2. doi:10.1016/j.gore.2016.10.009

- 101 Poonyakanok V, Tanmahasamut P, Jaishuen A, *et al.* Preoperative Evaluation of the ADNEX Model for the Prediction of the Ovarian Cancer Risk of Adnexal Masses at Siriraj Hospital. *Gynecol Obstet Invest* Published Online First: 2021. doi:10.1159/000513517
- 102 Zhang W, Jia S, Xiang Y, *et al.* Factors associated with misdiagnosis of frozen section of mucinous borderline ovarian tumor. *J Int Med Res* 2019;**47**:96–104. doi:10.1177/0300060518795582
- 103 Sayasneh A, Ferrara L, De Cock B, *et al.* Evaluating the risk of ovarian cancer before surgery using the ADNEX model: a multicentre external validation study. *Br J Cancer* 2016;**115**:542–8. doi:10.1038/bjc.2016.227
- 104 Van Calster B, Valentin L, Froyman W, *et al.* Validation of models to diagnose ovarian cancer in patients managed surgically or conservatively: multicentre cohort study. *BMJ* 2020;**370**:m2614. doi:10.1136/bmj.m2614
- 105 Viora E, Piovano E, Baima Poma C, *et al.* The ADNEX model to triage adnexal masses: An external validation study and comparison with the IOTA two-step strategy and subjective assessment by an experienced ultrasound operator. *Eur J Obstet Gynecol Reprod Biol* Published Online First: 2020. doi:10.1016/j.ejogrb.2020.02.022
- 106 Rusanov A, Weiskopf NG, Wang S, *et al.* Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research. *BMC Med Inform Decis Mak* 2014;**14**:51. doi:10.1186/1472-6947-14-51
- 107 Landolfo C, Bourne T, Froyman W, *et al.* Benign descriptors and ADNEX in two-step strategy to estimate risk of malignancy in ovarian tumors: retrospective validation in IOTA5 multicenter cohort. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol* 2023;**61**:231–42. doi:10.1002/uog.26080
- 108 Wolff RF, Moons KGM, Riley RD, *et al.* PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019;**170**:51. doi:10.7326/M18-1376
- 109 Whiting PF, Rutjes AWS, Westwood ME, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36. doi:10.7326/0003-4819-155-8-201110180-00009
- 110 Collins GS, de Groot JA, Dutton S, *et al.* External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014;**14**:40. doi:10.1186/1471-2288-14-40
- 111 Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020;**369**:m1328. doi:10.1136/bmj.m1328
- 112 Helmrich IRAR, Mikolić A, Kent DM, *et al.* Does poor methodological quality of prediction modeling studies translate to poor model performance? An illustration in traumatic brain injury. *Diagn Progn Res* 2022;**6**:8. doi:10.1186/s41512-022-00122-0

- 113 Dhiman P, Ma J, Navarro CA, *et al.* Reporting of prognostic clinical prediction models based on machine learning methods in oncology needs to be improved. *J Clin Epidemiol* 2021;**138**:60–72. doi:10.1016/j.jclinepi.2021.06.024
- 114 Andaur Navarro CL, Damen JAA, Takada T, *et al.* Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ* 2021;**375**:n2281. doi:10.1136/bmj.n2281
- 115 Allotey J, Whittle R, Snell KIE, *et al.* External validation of prognostic models to predict stillbirth using International Prediction of Pregnancy Complications (IPPIC) Network database: individual participant data meta-analysis. *Ultrasound Obstet Gynecol* 2022;**59**:209–19. doi:10.1002/uog.23757
- 116 Bouwmeester W, Zuithoff NPA, Mallett S, *et al.* Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* 2012;**9**:1–12. doi:10.1371/journal.pmed.1001221
- 117 Van Calster B, Steyerberg EW, Wynants L, *et al.* There is no such thing as a validated prediction model. *BMC Med* 2023;**21**:70. doi:10.1186/s12916-023-02779-w
- 118 On behalf of Topic Group ‘Evaluating diagnostic tests and prediction models’ of the STRATOS initiative, Van Calster B, McLernon DJ, *et al.* Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;**17**:230. doi:10.1186/s12916-019-1466-7
- 119 Van Hoorde K, Vergouwe Y, Timmerman D, *et al.* Assessing calibration of multinomial risk prediction models: K. VAN HOORDE *ET AL.* *Stat Med* 2014;**33**:2585–96. doi:10.1002/sim.6114
- 120 Pate A, Riley RD, Collins GS, *et al.* Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. *Stat Methods Med Res* Published Online First: 2023. doi:10.1177/09622802231151220
- 121 Snell KIE, Archer L, Ensor J, *et al.* External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. *J Clin Epidemiol* 2021;**135**:79–89. doi:10.1016/j.jclinepi.2021.02.011
- 122 Riley RD, Debray TPA, Collins GS, *et al.* Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med* 2021;**40**:4230–51. doi:10.1002/sim.9025
- 123 E.W. Steyerberg. Clinical Prediction Models. *Stat Health Sci* Published Online First: 2009. doi:10.1007/978-0-387-77244-8
- 124 Sisk R, Sperrin M, Peek N, *et al.* Imputation and missing indicators for handling missing data in the development and deployment of clinical prediction models: A simulation study. *Stat Methods Med Res* 2023;**9622802231165000**. doi:10.1177/09622802231165001

Figure 1. PRISMA (preferred reporting items for systematic reviews and meta-analyses) flowchart of study inclusions and exclusions

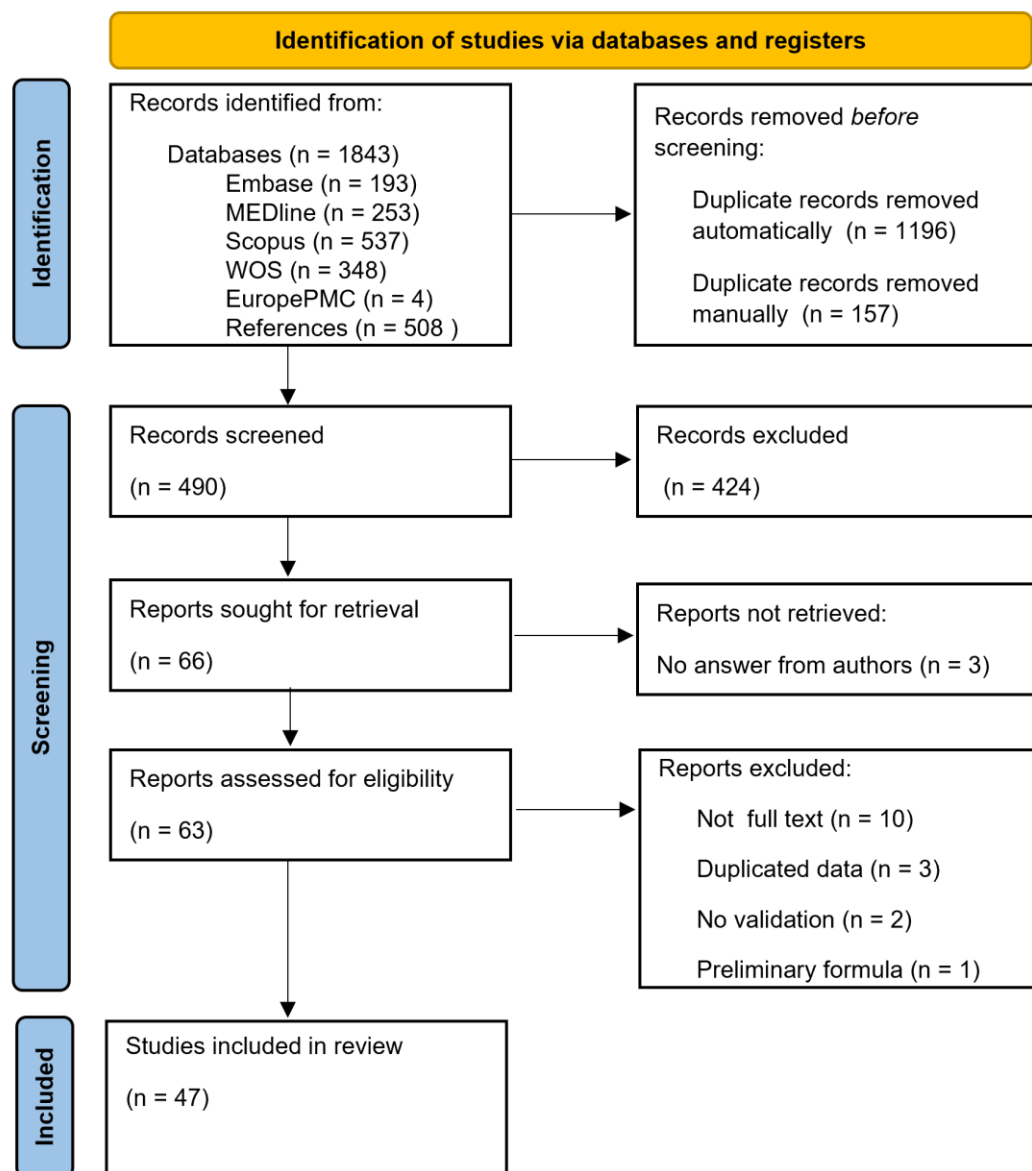


Figure 2. TRIPOD adherence per item in 63 validations

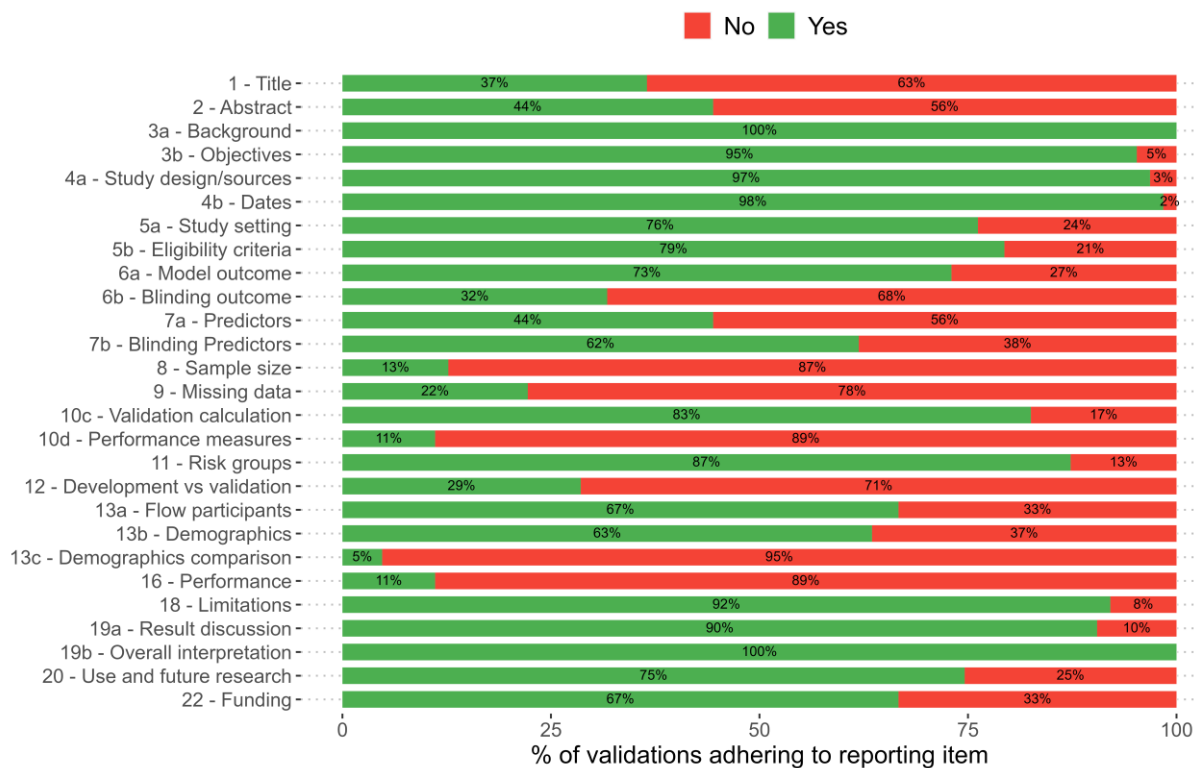


Figure 3. PROBAST risk of bias by subdomain and overall in 63 validations

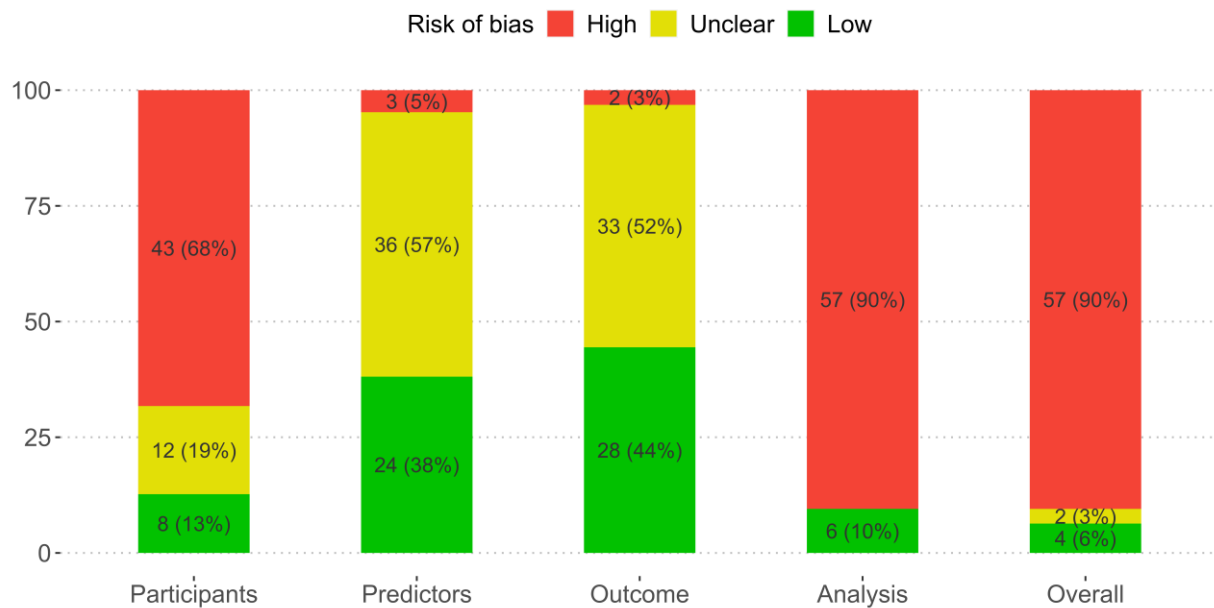


Figure 4. Forest plot of area under the receiver operating curve (AUC) for ADNEX without CA125. CI, confidence interval; Prev, prevalence of malignancy

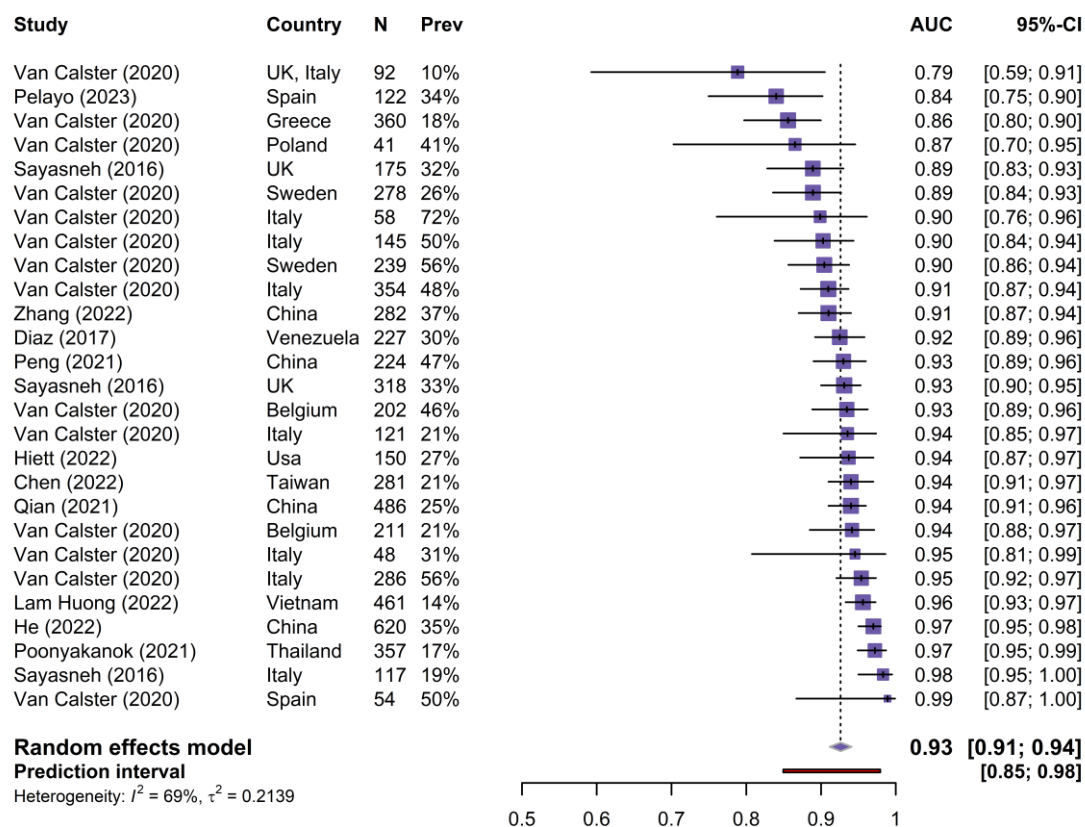


Figure 5. Forest plot of area under the receiver operating characteristic curve (AUC) for ADNEX with CA125. CI, confidence interval; Prev, prevalence of malignancy

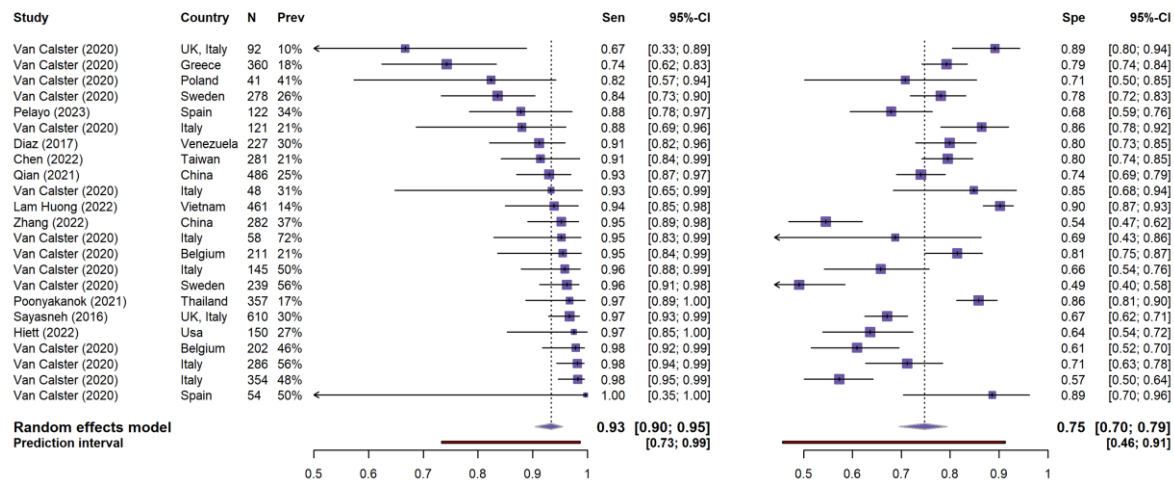


Figure 6. Forest plot of sensitivity (left) and specificity (right) at the 10% risk of malignancy threshold for ADNEX without CA125. CI, confidence interval Prev, prevalence of malignancy; Sen, sensitivity; Spe, specificity

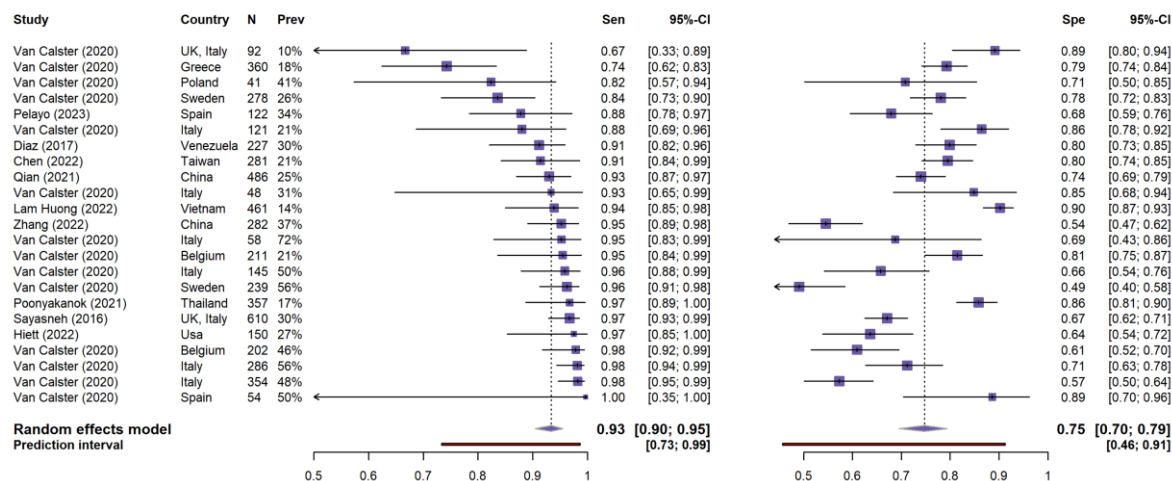


Figure 7. Forest plot of sensitivity (left) and specificity (right) at the 10% risk of malignancy threshold for ADNEX with CA125. CI, confidence interval Prev, prevalence of malignancy; Sen, sensitivity; Spe, specificity

