European Journal of Cancer 58 (2016) 17–29

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Review

Subjective assessment versus ultrasound models to diagnose ovarian cancer: A systematic review and meta-analysis

E.M.J. Meys [a], J. Kaijser [b,1], R.F.P.M. Kruitwagen [a], B.F.M. Slangen [a], B. Van Calster [c], B. Aertgeerts [d], J.Y. Verbakel [d,e], D. Timmerman [b], T. Van Gorp [a,*]

[a] *Department of Obstetrics and Gynaecology, Maastricht University Medical Centre+ (MUMC+), GROW – School for Oncology and Developmental Biology, P.Debyelaan 25, 6202 AZ, Maastricht, The Netherlands*
[b] *Department of Obstetrics and Gynaecology and Leuven Cancer Institute, University Hospital KU Leuven, Herestraat 49, Leuven 3000, Belgium*
[c] *Department of Development and Regeneration, KU Leuven, Herestraat 49, Leuven 3000, Belgium*
[d] *Department of General Practice, KU Leuven, Kapucijnenvoer 33, Leuven 3000, Belgium*
[e] *Nuffield Department of Primary Care Health Sciences, University of Oxford, Woodstock Road, Oxford OX2 6GG, UK*

Received 30 September 2015; received in revised form 8 January 2016; accepted 14 January 2016
Available online 27 February 2016

KEYWORDS
Ovarian neoplasms;
Ovarian cancer;
Ultrasonography;
Sensitivity and specificity;
Systematic review;
Meta-analysis

Abstract ***Introduction:*** Many national guidelines concerning the management of ovarian cancer currently advocate the risk of malignancy index (RMI) to characterise ovarian pathology. However, other methods, such as subjective assessment, International Ovarian Tumour Analysis (IOTA) simple ultrasound-based rules (simple rules) and IOTA logistic regression model 2 (LR2) seem to be superior to the RMI.

Our objective was to compare the diagnostic accuracy of subjective assessment, simple rules, LR2 and RMI for differentiating benign from malignant adnexal masses prior to surgery.
Materials and methods: MEDLINE, EMBASE and CENTRAL were searched (January 1990 –August 2015). Eligibility criteria were prospective diagnostic studies designed to preoperatively predict ovarian cancer in women with an adnexal mass.
Results: We analysed 47 articles, enrolling 19,674 adnexal tumours; 13,953 (70.9%) benign and 5721 (29.1%) malignant. Subjective assessment by experts performed best with a pooled

* *Corresponding author:* Tel.: +31 43 387 4767.
E-mail addresses: evelyne.meys@mumc.nl (E.M.J. Meys), j.kaijser@ikazia.nl (J. Kaijser), r.kruitwagen@mumc.nl (R.F.P.M. Kruitwagen), brigitte.slangen@mumc.nl (B.F.M. Slangen), ben.vancalster@med.kuleuven.be (B. Van Calster), bert.aertgeerts@med.kuleuven.be (B. Aertgeerts), jan.verbakel@med.kuleuven.be (J.Y. Verbakel), dirk.timmerman@uzleuven.be (D. Timmerman), Toon.van.gorp@mumc.nl (T. Van Gorp).
[1] Present address: Montessoriweg 1, 3083 AN, Rotterdam, The Netherlands.

http://dx.doi.org/10.1016/j.ejca.2016.01.007

sensitivity of 0.93 (95% confidence interval [CI] 0.92–0.95) and specificity of 0.89 (95% CI 0.86–0.92). Simple rules (classifying inconclusives as malignant) (sensitivity 0.93 [95% CI 0.91–0.95] and specificity 0.80 [95% CI 0.77–0.82]) and LR2 (sensitivity 0.93 [95% CI 0.89–0.95] and specificity 0.84 [95% CI 0.78–0.89]) outperformed RMI (sensitivity 0.75 [95% CI 0.72–0.79], specificity 0.92 [95% CI 0.88–0.94]). A two-step strategy using simple rules, when inconclusive added by subjective assessment, matched test performance of subjective assessment by expert examiners (sensitivity 0.91 [95% CI 0.89–0.93] and specificity 0.91 [95% CI 0.87–0.94]).

Conclusions: A two-step strategy of simple rules with subjective assessment for inconclusive tumours yielded best results and matched test performance of expert ultrasound examiners. The LR2 model can be used as an alternative if an expert is not available.

1. Introduction

1.1. Rationale and objectives

In order to ensure that ovarian cancer patients receive appropriate treatment, an accurate characterisation of any adnexal mass that needs surgery is pivotal to improve the outcome of this disease. Subjective assessment by experienced examiners, also called 'pattern recognition', is generally accepted to be the best way to classify adnexal masses prior to surgery. Several individual reports have demonstrated that subjective assessment is superior to the use of scoring systems and mathematical models, such as International Ovarian Tumour Analysis (IOTA) simple ultrasound-based rules (simple rules), IOTA logistic regression model 2 (LR2) or the risk of malignancy index (RMI) [1–4]. However, both LR2 and simple rules closely approximate the performance of subjective assessment by expert examiners [5,6]. An advantage of these models over subjective assessment is their objectivity and simplicity which facilitates their use by ultrasonographers with different backgrounds and various levels of experience [7–10]. Despite accumulating and compelling evidence in favour of both subjective assessment and the ultrasound-based models such as simple rules and LR2, many national guidelines concerning the management of ovarian masses still advocate the use of RMI in the classification of adnexal masses. Consequently, the RMI is still the most commonly used model in clinical practice.

Several reviews have critically appraised the evidence relating to this subject [5,6,11–16]. However, none of these has provided a meta-analysis on the test performance of subjective assessment of adnexal tumours, while in general this method is considered the most accurate way to distinguish benign from malignant adnexal tumours. The aim of this meta-analysis was to compare the diagnostic accuracy of subjective assessment, simple rules, LR2 and RMI for the pre-operative differentiation of benign and malignant adnexal masses.

2. Methods

2.1. Protocol and registration

All methods described in this manuscript were determined in advance and recorded in a study protocol (Prospero CRD42013004334, http://www.crd.york.ac.uk/PROSPERO). The conduct of this systematic review and meta-analysis was done in accordance with prevailing guidelines (http://www.prisma-statement.org and http://srdta.cochrane.org/handbook-dta-reviews).-

2.2. Eligibility criteria

Eligible studies had to evaluate diagnostic accuracy of subjective assessment, simple rules, LR2 and/or RMI for the characterisation of adnexal tumours in women scheduled for surgery (in order to obtain a final histological diagnosis). Regarding subjective assessment, studies were only eligible when the diagnosis of the tumour was based purely on the ultrasonographic interpretation of the examiner (whether or not complemented with clinical information, such as medical history).

The simple rules comprise two strategies; simple rules supplemented with subjective assessment in case the simple rules could not be applied, or classification of all masses in which simple rules could not be applied as malignant [17]. Studies evaluating either of these strategies or both were eligible.

Three principal variants of the RMI have been described (RMI-I, II and III) which differ according to points attributed to the different ultrasound variables and the menopausal status of the patient [18–20]. All studies regarding one or more of these three versions were eligible.

Furthermore, eligible studies had to contain sufficient data to extract 2 × 2 contingency tables of diagnostic test performance.

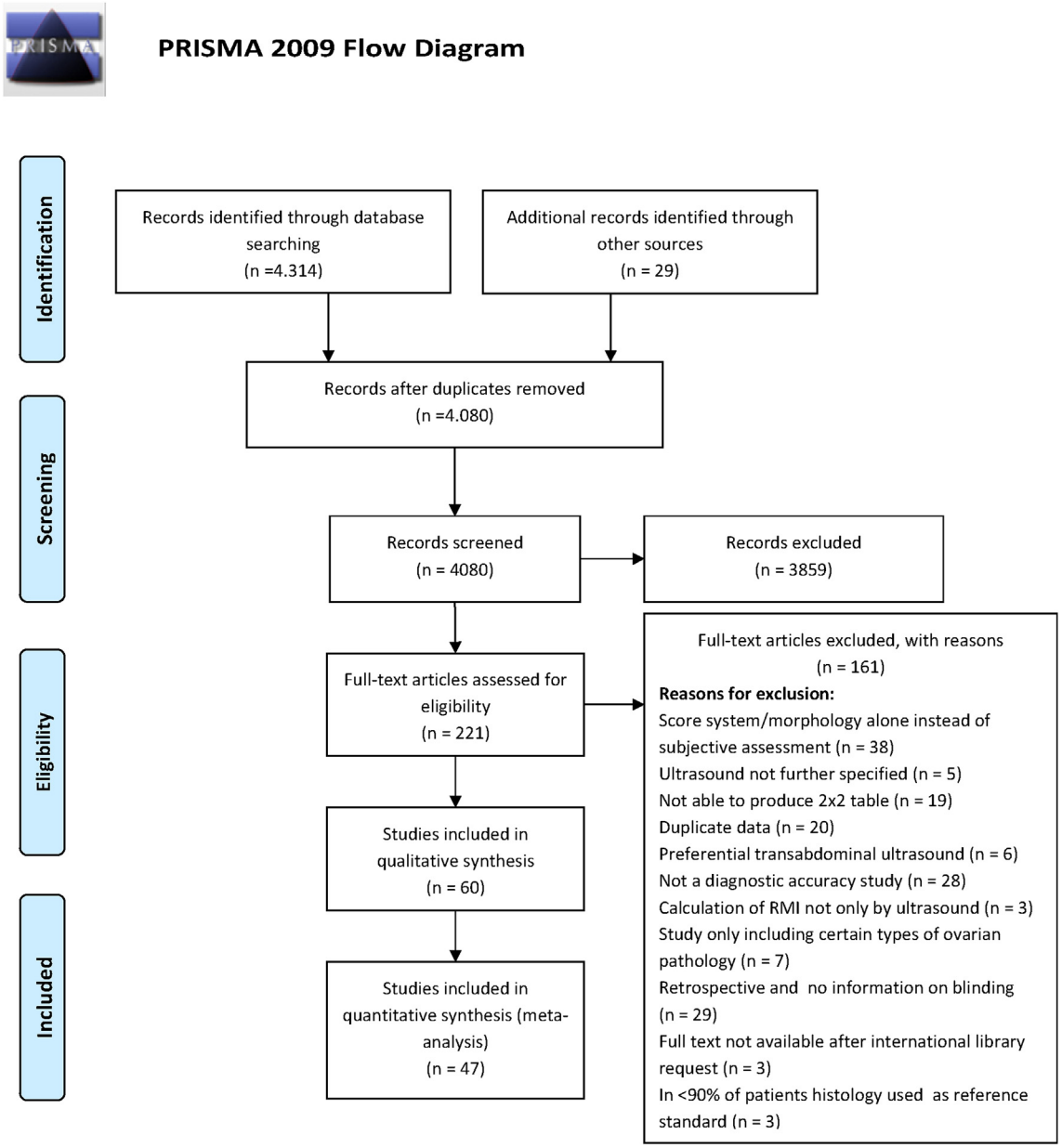

Fig. 1. Flow diagram of the study selection process (based on PRISMA 2009). Abbreviation: RMI, risk of malignancy index.

Studies meeting the following criteria were excluded: studies that evaluated model performance in only a specific histological subgroup of ovarian cancer, studies that used only transabdominal ultrasound, studies containing duplicate data, studies using another method than histology as a reference standard in >10% of patients, studies of which the full text was unavailable even after an international library request, case-control studies and 'cohort-type' studies with a retrospective design not providing any information on blinding of test results to the final study outcome. The latter are excluded because knowledge of the reference test will bias relatively subjective index tests like those based on ultrasonography, and tend to overestimate diagnostic test performance.

Studies excluding certain histological subtypes of ovarian cancer, those containing data inconsistency or errors, and those classifying borderline tumours as benign for statistical analysis, were also excluded from meta-analysis.

2.3. Information sources

Studies were identified by searching the following electronic bibliographic databases; MEDLINE(OvidSP), EMBASE(OvidSP) and the Cochrane Central Register of Controlled Trials (CENTRAL). Additional eligible articles were identified by manually searching cross-references of articles retrieved from the computerised databases and relevant reviews.

2.4. Search

The MEDLINE, EMBASE and CENTRAL search was performed on July 31, 2015 using key words consisting of terminology for the different index tests and target condition, completed with methodological terms. No language restrictions were made. The search was limited from 1990 onwards, since RMI -the 'oldest' prediction model- was first described in 1990. A detailed description of the search strategy is presented in Appendix A.

2.5. Study selection

Two authors (EM/JK) independently reviewed all records retrieved by the search. Disagreements were solved by discussion and – if necessary – by consulting a third party (TVG). Reasons for exclusion after full text evaluation (n = 161) are summarised in Fig. 1.

2.6. Data extraction

Extracted data from selected studies were entered independently by the same two authors in a pre-defined data collection form (Appendix B). If necessary, authors were contacted for additional information. Expertise in gynaecological ultrasonography is defined by guidelines of the European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) [21]. In articles not defining the level of expertise by EFSUMB guidelines, we considered ultrasonographers with at least 10 years of experience, who work in a tertiary referral centre, as level-III experts for the purpose of further statistical analysis regarding subjective assessment.

2.7. Risk of bias assessment

All original studies were assessed by two authors (EM/JK) independently, using the QUADAS-2 checklist in order to evaluate the risk of bias and applicability (Appendix C) [22].

2.8. Statistical analysis and synthesis of results

Only studies in which the models (simple rules, LR2 and RMI) were externally validated, were included in the meta-analysis [17–20,23]. All individual study results were grouped according to type of index test used and presented graphically by plotting estimates of sensitivity and specificity (and corresponding 95%-confidence intervals [CIs]) in forest plots. Levels of heterogeneity were primarily deducted from interpretation of these CIs, although subgroup analyses were performed if a sufficient number of studies (≥3) were available in the subgroup. Specific subgroup analyses were performed for potential sources of heterogeneity between the different index tests under review (e.g. prevalence, expert versus non-expert and QUADAS items) by fitting bivariate models with covariates using multi-level mixed effects logistic regression (xtmelogit-function in Stata). The same meta-regression method was used to compare the accuracy between tests and between pre- and post-menopausal populations at a p-value of 0.05, using a chi square test.

Allowing a variation in cut-off values to define a positive test (e.g. RMI I–II–III), a hierarchical summary receiver operating curve (HSROC) model was fitted using the metandi-command in STATA 13.1 (Stata Corp., College Station, TX, USA). Due to the low number of studies on LR2 (n = 3) we had to use a mixed-methods function (GLLAMM) in order to calculate the pooled summary estimates for sensitivity and specificity. All graphs were designed using RevMan (version 5.3. Copenhagen: Nordic Cochrane Centre, Cochrane Collaboration, 2014).

3. Results

3.1. Study selection and characteristics

In total 60 studies were included in the qualitative data synthesis (Fig. 1) [1,6,7,9,18–20,24–75]. Of these, 47 were valid for quantitative data synthesis (meta-analysis), containing 19,674 masses; 13,953 (70.9%) benign and 5721 (29.1%) malignant (Appendix D) [1,3,6–10,24–58,70–74]. Out of a total of 4343 first hits, there was only a 2% difference in studies selected for full text evaluation by both authors (EM/JK). Overall, only ten studies were included in the meta-analysis that simultaneously validated more than one index test [1,6–8,35,41,42,49,56,71].

3.1.1. Subjective assessment

Subjective assessment of grey scale and colour Doppler ultrasound findings was the most frequently validated method in our review (Fig. 2) [1,6–8,24–28,30–40,43,44,72,73,76]. Only three studies investigated the performance of subjective assessment in 'non-experts' (i.e. operators with level-I/II experience by EFSUMB-criteria) [6,36,72], one study investigated both experts and 'non-experts' [73], and in one study the expertise of the performing ultrasonographer was unclear [28].

3.1.2. Simple rules

In five out of ten studies validating simple rules, simple rules were performed by 'non-experts' [6,8–10,71], in two studies by experts [3,57], in one study by both, and in two studies level of expertise of the performing

Study	TP	FP	FN	TN	7. prevalence	Sensitivity (95% CI)	Specificity (95% CI)
van Trappen 2007	40	44	0	58	high prevalence (>25%)	1.00 [0.91, 1.00]	0.57 [0.47, 0.67]
Mancuso 2004	14	23	0	88	low prevalence (<25%)	1.00 [0.77, 1.00]	0.79 [0.71, 0.86]
Sohaib 2005	29	24	0	19	high prevalence (>25%)	1.00 [0.88, 1.00]	0.44 [0.29, 0.60]
Utrilla-Layna 2015	84	60	2	221	low prevalence (<25%)	0.98 [0.92, 1.00]	0.79 [0.73, 0.83]
Knafel 2015	80	1	2	143	high prevalence (>25%)	0.98 [0.91, 1.00]	0.99 [0.96, 1.00]
Van Gorp 2012	145	22	5	202	high prevalence (>25%)	0.97 [0.92, 0.99]	0.90 [0.86, 0.94]
Timmerman 1999	80	22	3	195	high prevalence (>25%)	0.96 [0.90, 0.99]	0.90 [0.85, 0.94]
Moszynski 2013	87	31	4	196	high prevalence (>25%)	0.96 [0.89, 0.99]	0.86 [0.81, 0.91]
Nunes 2013	126	11	6	149	high prevalence (>25%)	0.95 [0.90, 0.98]	0.93 [0.88, 0.97]
Romagnolo 2006	20	18	1	182	low prevalence (<25%)	0.95 [0.76, 1.00]	0.91 [0.86, 0.95]
Nunes 2014	128	11	7	157	high prevalence (>25%)	0.95 [0.90, 0.98]	0.93 [0.89, 0.97]
Guerriero 2010	441	101	27	1579	low prevalence (<25%)	0.94 [0.92, 0.96]	0.94 [0.93, 0.95]
Lucidarme 2010	128	68	8	51	high prevalence (>25%)	0.94 [0.89, 0.97]	0.43 [0.34, 0.52]
Testa 2014	913	189	67	1234	high prevalence (>25%)	0.93 [0.91, 0.95]	0.87 [0.85, 0.88]
Daemen 2011	490	102	52	1294	high prevalence (>25%)	0.90 [0.88, 0.93]	0.93 [0.91, 0.94]
Van Holsbeke 2009	129	26	14	338	high prevalence (>25%)	0.90 [0.84, 0.95]	0.93 [0.90, 0.95]
Sayasneh 2015	86	17	10	200	high prevalence (>25%)	0.90 [0.82, 0.95]	0.92 [0.88, 0.95]
Roman 1997	38	24	5	159	low prevalence (<25%)	0.88 [0.75, 0.96]	0.87 [0.81, 0.91]
Van Calster 2007	234	38	32	762	high prevalence (>25%)	0.88 [0.83, 0.92]	0.95 [0.94, 0.97]
Valentin 1999	21	6	3	143	low prevalence (<25%)	0.88 [0.68, 0.97]	0.96 [0.91, 0.99]
Strigini 1996	16	4	3	86	high prevalence (>25%)	0.84 [0.60, 0.97]	0.96 [0.89, 0.99]
Valentin 2001	30	9	6	91	high prevalence (>25%)	0.83 [0.67, 0.94]	0.91 [0.84, 0.96]
Granberg 1990	32	11	7	130	low prevalence (<25%)	0.82 [0.66, 0.92]	0.92 [0.86, 0.96]
Salle 1995	8	6	3	74	low prevalence (<25%)	0.73 [0.39, 0.94]	0.93 [0.84, 0.97]

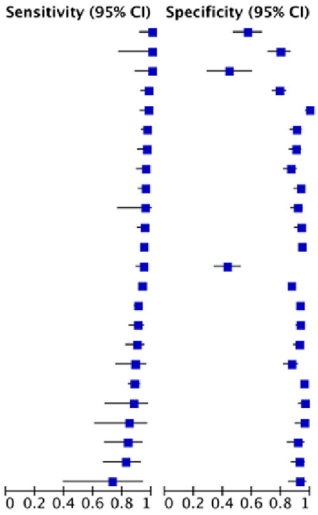

Fig. 2. Forest plots of studies investigating accuracy of subjective assessment included in this meta-analysis. Abbreviations: TP, true positives; FP, false positives; FN, false negatives; TN, true negatives; CI, confidence interval.

ultrasonographer was unclear (Fig. 3) [58,74]. Six studies investigated subjective assessment as a second stage test for inconclusive test results [3,6,8–10,71].

3.1.3. LR2

Our search identified eight articles on LR2 [3,7,24,40,42,62,63,77]. Only three studies on LR2 were included in the meta-analysis (Fig. 4) [7,8,42].

3.1.4. RMI

Eighteen RMI validation studies were included in the final meta-analysis (Fig. 5) [1,8,29,35,41,42,45–56], four of which addressed more than one version of the RMI [41,42,49,56]. RMI-I was most reviewed (n = 14).

3.2. Risk of bias within studies

Fig. 6 shows a summary of the quality assessment of all studies included in the meta-analysis, by using the QUADAS-2 tool. For more information on the risk of bias within each study we refer to Appendix E.

All studies included in the meta-analysis were prospective, used histology as a reference standard, and avoided a case-control design. One third of the studies yielded a risk of inappropriate exclusions, i.e. when no information about the number of patients before exclusion or no reason of exclusion of patients was provided.

It was unclear whether blinding of the pathologist to the outcome of the index test took place in 41 studies. Moreover, only in 59.6% it was clearly stated how the index test was performed. For example, expertise of the performing clinician was not recorded in studies examining subjective assessment (n = 17), or it was not clear which version of the RMI was used (n = 3).

In 48.9% an appropriate interval between conducting the index test (i.e. ultrasound) and getting the histology results was applied (Appendix C). In 76.6% the study took place in only second or only third line hospitals. Furthermore, 43 studies used real time imaging.

3.3. Primary analysis

Pooled summary point estimates (and SROC curves if appropriate) of subjective assessment, simple rules, RMI (at various cut-off points) and LR2 (at original cut-off point of ≥10%) were calculated and are shown in Fig. 7 and Table 1.

Subjective assessment performed best of all methods under investigation (Table 1). For studies investigating 'non-expert' examiners pooled estimates for sensitivity and specificity were similar (chi square: p = 0.12). Simple rules were applicable in 2490 of 3073 masses (81.0%). Simple rules in combination with subjective assessment for inconclusive results gave only small differences when compared to subjective assessment, which were not significant according to meta-regression (Appendix F). When inconclusive findings of simple rules were classified as malignant specificity dropped significantly.

LR2 had comparable sensitivity and specificity to the strategy using simple rules and classifying inconclusives as malignant.

RMI performed worst of the methods under investigation. Although diagnostic test accuracy is highest in RMI-I, differences are not statistically significant between the three versions (Appendix F). There was however, a statistically significant difference in sensitivity of RMI (I, II and III) compared to all other methods (all p-values <0.0004).

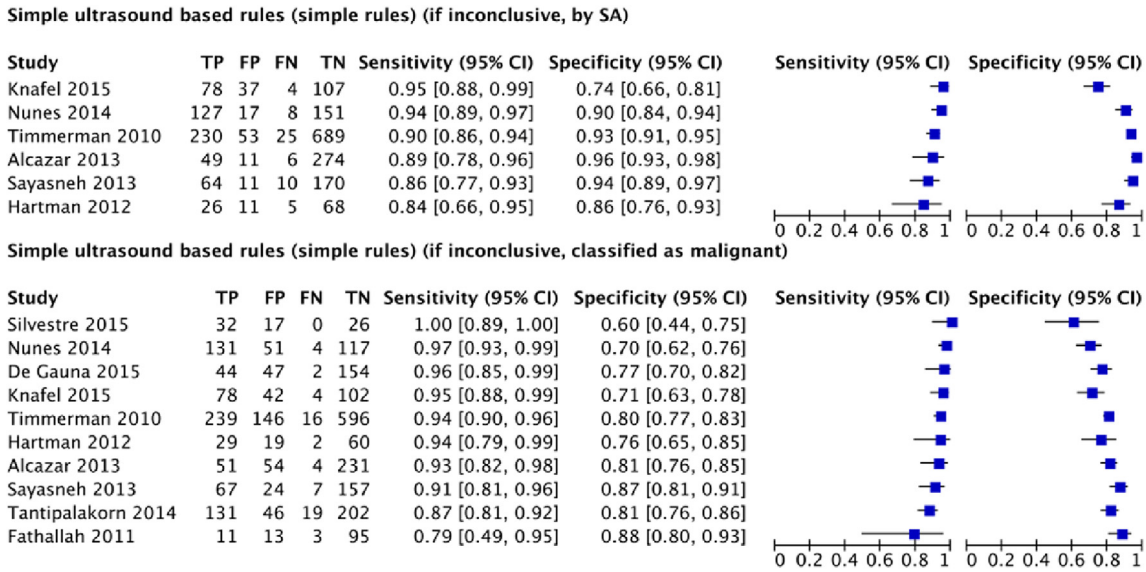

Fig. 3. Forest plots of studies investigating accuracy of simple rules supplemented with either subjective assessment (SA) or with the strategy that inconclusive results were classified as malignant. Abbreviations: TP, true positives; FP, false positives; FN, false negatives; TN, true negatives; CI, confidence interval.

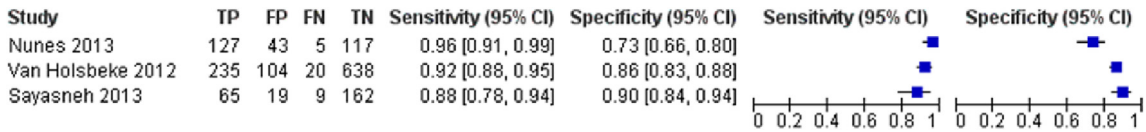

Fig. 4. Forest plots of studies investigating accuracy of LR2 included in meta-analysis. Abbreviations: TP, true positives; FP, false positives; FN, false negatives; TN, true negatives; CI, confidence interval; LR2, logistic regression model 2.

3.4. Secondary analysis

3.4.1. Investigating heterogeneity

The prevalence of ovarian carcinoma in this review ranged from 9.5% to 57% (28.4% overall prevalence across studies). Subgroup analyses through meta-regression dividing all studies in two distinct prevalence settings (above/below 25%) revealed no significant influence of prevalence on test performance for those methods with three or more studies available in each subgroup (i.e. subjective assessment, simple rules and RMI). Neither did subgroup analyses performed for site of recruitment (second/third line hospital or both) or the QUADAS-items for these methods (with p-values > 0.10 for all heterogeneity analyses).

3.4.2. Pre- versus postmenopausal

We retrieved menopausal status of patients in 38 studies with a total number of 6444 postmenopausal patients out of 17060 subjects (37.7%) (Appendix D). Subjective assessment had the highest diagnostic accuracy for differentiation between benign and malignant adnexal masses in premenopausal women (Table 2 and Appendix G). Nonetheless, meta-regression analysis showed no significant differences in accuracy between subjective assessment and simple rules, supplemented with subjective assessment when simple rules yielded inconclusive results (Appendix F). LR2 had the highest diagnostic accuracy among postmenopausal women, with a statistically significant difference in accuracy between LR2 and subjective assessment.

4. Discussion

4.1. Main results

This review and meta-analysis summarises the evidence currently available on the diagnostic accuracy of different pre-operative ultrasound methods for differentiating benign from malignant adnexal masses. According to this systematic review and meta-analysis, we believe an evidence-based approach should incorporate either simple rules with referral for subjective assessment of ultrasound findings by expert examiners if the rules are not applicable, or alternatively the LR2 model if such expertise is not available.

4.2. Strengths and weaknesses

To the best of our knowledge, this is to date the most comprehensive review on diagnostic test accuracy for

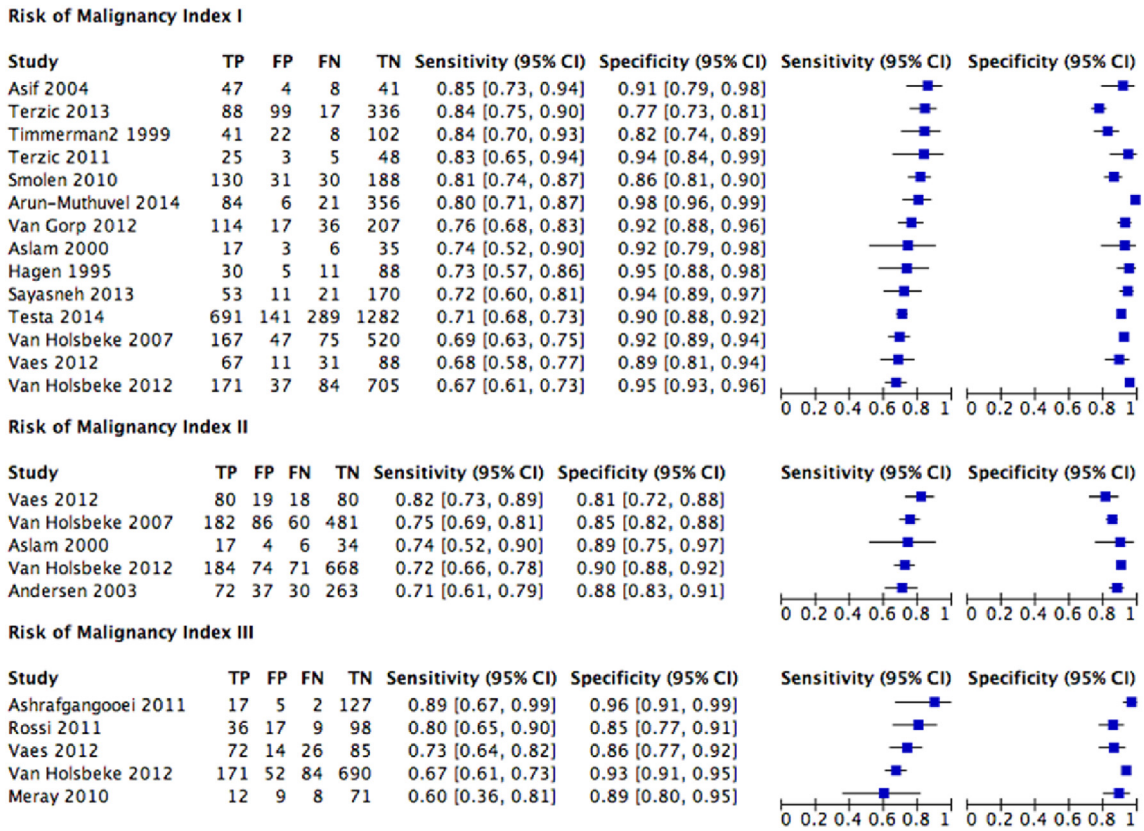

Fig. 5. Forest plots of studies investigating different versions of RMI; 14 studies on RMI-I, five studies on RMI-II and five studies on RMI-III were included in this meta-analysis. Abbreviations: TP, true positives; FP, false positives; FN, false negatives; TN, true negatives. RMI, risk of malignancy index; CI, confidence interval.

differentiating benign and malignant adnexal masses. Previous meta-analysis has not included studies considering subjective assessment. We have worked by a clearly defined protocol to provide transparency in the review process and avoid reporting bias. The fact that there was no substantial disagreement in inclusion of articles by the two authors (EM/JK) can be regarded as a strong point in the review process. Furthermore, this review comprises a large number of studies (n = 47) and adnexal masses (n = 19674), thereby increasing the power of the analysis. Finally, we successfully collected additional information by contacting authors of included studies.

For inclusion it was mandatory that histology was used as a reference standard. Although interpretation of histology is a relatively subjective method, it can be deduced from previous studies that different pathologists report no relevant differences in terms of discriminating between malignant and benign adnexal tumours [78]. We have classified borderline tumours as malignant for statistical analysis, since previous research has shown that comprehensive surgery influences prognosis in borderline tumours [79]. The overall prevalence of malignancy across included studies was 28.4%. Because we excluded studies that also included benign-looking masses that were managed expectantly, this prevalence is higher than in daily practice. However, this prevalence is comparable to previous reviews on prediction models in the preoperative assessment of adnexal masses (27%,[5] 23.8%[6] and 28.8%[13], respectively).

A limitation of reviews in general is the possible heterogeneity between studies. More specifically for this meta-analysis, only low levels of heterogeneity were found in most of the subsets allocated to the various methods. Therefore, in general, differences in included studies could be explained by sampling variation. For RMI these differences could also be explained by the variation in cut-offs used. Most studies have used 200 as a cut-off since this is suggested in the original articles. However four of 18 studies (also) used other cut-offs [47,48,55,56]. In studies where more than one cut-off was available a cut-off different from 200 was used to avoid the analyses to be based on a single cut-off value for RMI, resulting in a point estimate in the ROC space, rather than a ROC curve.

On the other hand, the fact that we did not find a significant influence of prevalence and site of recruitment on test performance could be explained by the small number of studies in some of the subgroups. For example, only four of the included studies were conducted exclusively in second line hospitals [44,50,53,54].

We have focused on sensitivity and specificity instead of the area under the curve (AUC) as a primary

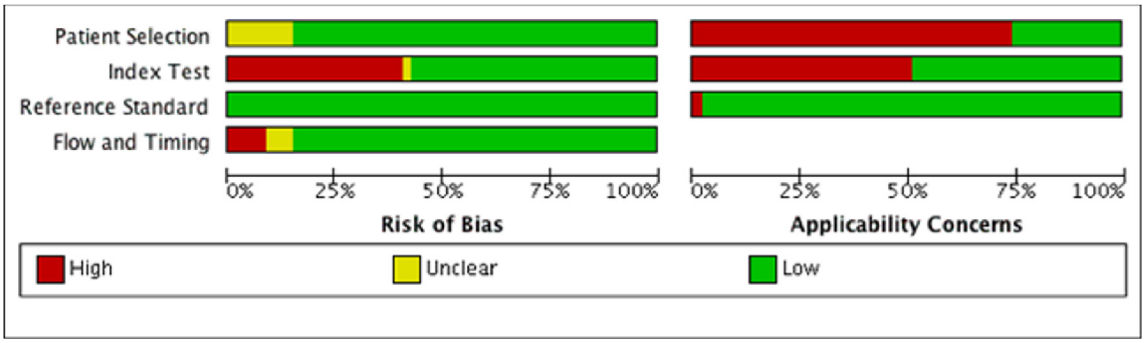

Fig. 6. Summary of the quality assessment of studies included in the meta-analysis (n = 47) by using the QUADAS-2 tool.

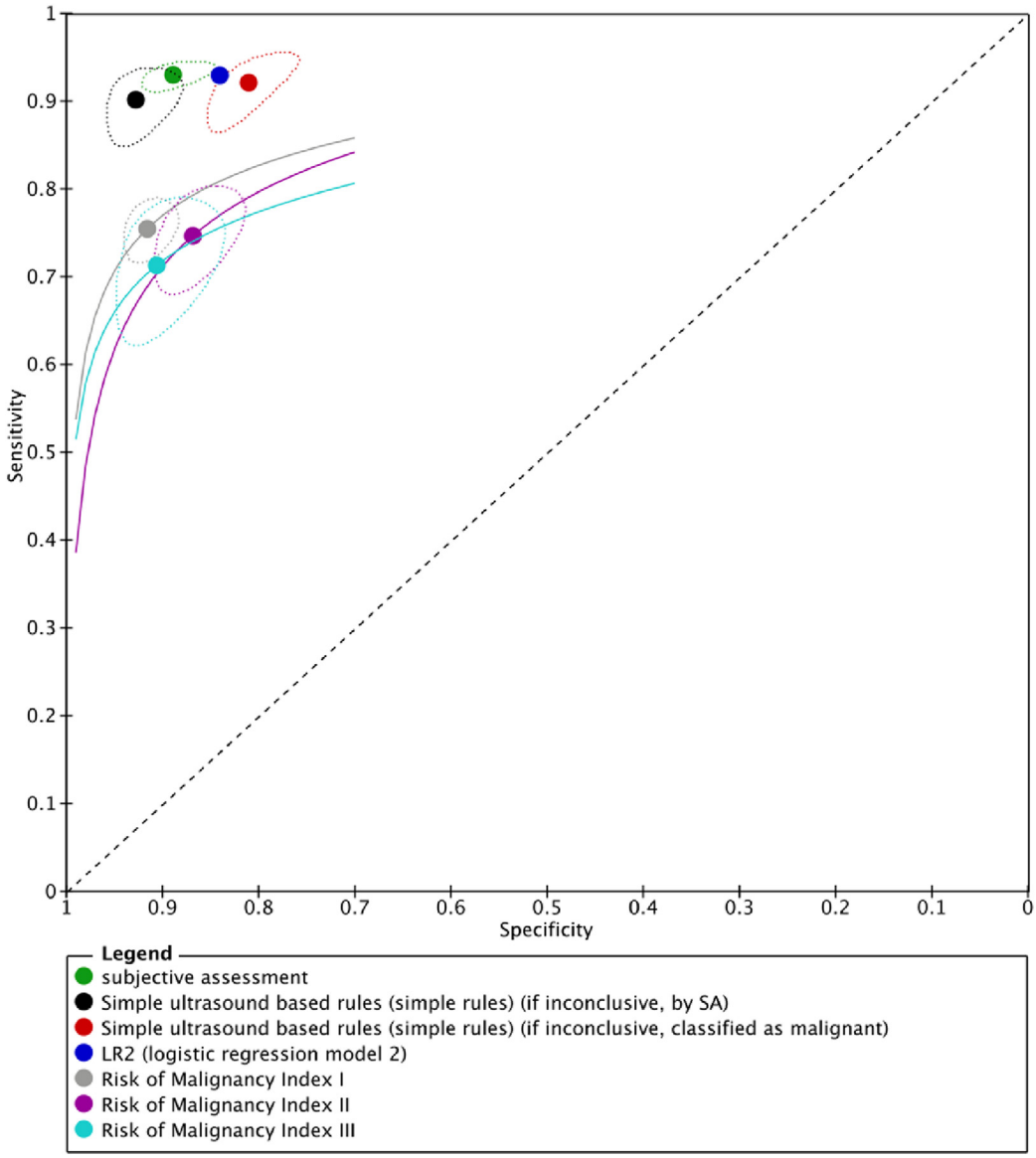

Fig. 7. Overview of summary point estimates of all study types included in the meta-analysis. All summary points are displayed as sensitivity/specificity pairs with their 95% confidence regions (except for LR2, due to insufficient studies). For RMI-I, RMI-II and RMI-III additional HSROC curves were fitted. SA; subjective assessment. RMI, risk of malignancy index; HSROC, hierarchical summary receiver operating curve; LR2, logistic regression model 2.

Table 1
Pooled summary point estimates of all methods included in this review (LR2 at original cut-off point and RMI at various cut-off points) together with their 95% confidence interval (95%CI).

	Sens. (95%CI)	Spec. (95%CI)	DOR (95%CI)	LR+ (95%CI)	LR− (95%CI)
SA	0.93 (0.92–0.95)	0.89 (0.86–0.92)	120 (91–157)	8.78 (6.81–11.32)	0.07 (0.06–0.09)
SR+SA	0.91 (0.89–0.93)	0.91 (0.87–0.94)	102 (74–139)	9.87 (6.97–13.99)	0.10 (0.08–0.12)
SR+Mal	0.93 (0.91–0.95)	0.80 (0.77–0.82)	52 (41–65)	4.56 (4.06–5.12)	0.09 (0.07–0.12)
LR2	0.93 (0.89–0.95)	0.84 (0.78–0.89)	69 (52–92)	5.80 (4.30–7.80)	0.08 (0.06–0.12)
RMI-I	0.75 (0.72–0.79)	0.92 (0.88–0.94)	33 (23–48)	8.96 (6.50–12.36)	0.27 (0.23–0.31)
RMI-II	0.75 (0.72–0.77)	0.87 (0.85–0.89)	19 (16–23)	5.67 (4.90–6.56)	0.29 (0.26–0.32)
RMI-III	0.71 (0.67–0.75)	0.91 (0.88–0.93)	24 (18–31)	7.58 (5.94–9.67)	0.32 (0.28–0.36)

Abbreviations: CI, confidence interval; SA, subjective assessment; SR+SA, simple rules, if inconclusive classified by subjective assessment; SR+Mal, simple rules, if inconclusive classified as malignant; LR2, logistic regression model 2; RMI, risk of malignancy index; sens., sensitivity; spec., specificity; DOR, diagnostic odds ratio; LR+, positive likelihood ratio; LR−, negative likelihood ratio.

outcome. AUC can be used to summarise diagnostic accuracy, by combining the accuracy of a test across a range of thresholds [80]. A disadvantage of the AUC is it has little clinical consequences if thresholds that are clinically relevant are combined with those that are clinically nonsensical [81]. Furthermore, AUC does not offer any solution to which cut-off should be used. We have used a HSROC model to summarise the results and compare RMI at different cut-off values.

The quality assessment did reveal a relatively high concern on the applicability of the selected patients. However, the meta-regression analysis showed no significant differences in sensitivity and specificity of all index tests for second and third line centres. Therefore, we believe it is safe to assume that the included patients do match the review question.

Furthermore, there were only three studies comparing all four methods within the same population and therefore only between-study comparisons of the index tests were possible [3,8,35]. This can possibly lead to confounding due to differences in patient characteristics and other inter-study variation. However, due to strict inclusion criteria hardly any confounding was possible for reference standard and study design.

Other limitations include subjective assessment being at greater risk of review bias due to prior knowledge of the ultrasonographer, and the limited number of studies concerning LR2.

Furthermore, we were unable to perform subgroup analysis for all subgroups, since these were only deemed suitable if a sufficient number of studies (three or more as permitted by our meta-analyses techniques) were available in the subgroup at hand. Therefore, some caution regarding interpretation of these results is recommended. Given that the subgroup analysis performed on other heterogeneity-items (i.e. prevalence, site of recruitment and QUADAS-items) of the same articles all showed similar trends regarding heterogeneity, we are confident that this applies to the overall heterogeneity of these articles. Some uncertainty however remains.

Subjective assessment by experienced examiners is generally considered the best method to classify adnexal masses prior to surgery. Our meta-analysis confirms this presumption. The downside of this method is it requires availability of an expert in ultrasonography. However, a subgroup analysis within this meta-analysis, comparing level-III experts with 'non-experts', revealed only small differences in sensitivity and specificity.

Simple rules were proposed as a tool for less experienced examiners, in order to achieve the same results as expert ultrasonographers. When comparing experts [3,57] with non-experts [6,8–10], diagnostic test accuracy does not differ much between both groups. This was also confirmed by a recent review [6].

Sensitivity was particularly low for RMI in premenopausal women. The outcome of RMI depends mainly on the level of serum-CA125, while this biomarker is of limited value for the diagnosis of ovarian cancer, especially when it concerns premenopausal women [2,39,63,82]. Another recent review on this subject by Kaijser et al. revealed a sensitivity of only 44% for RMI-I in premenopausal women, while we found a sensitivity of 63% [5]. This difference could be explained by the fact that Kaijser et al. only used two multicentre cohort studies, while in the current meta-analysis datasets of seven different studies were included. Furthermore, there is a difference in quality appraisal of the included studies, since − contrary to Kaijser et al.- we did not include retrospective studies and we have included various cut-offs for RMI instead of only 200.

Other strategies that were not included in this meta-analysis might give similar or better diagnostic accuracy results, but have not been investigated or validated yet. For example, LR2 could be used as a first line test and when the result of the LR2 is between certain percentages (e.g. 5–25%) subjective assessment could be added as a second line test to increase diagnostic accuracy. This combination of methods could also be applied for RMI.

Furthermore, new models are introduced frequently. Some of these have not been validated extensively but seem promising, such as the ADNEX-model [83]. This model predicts the risk for a specific type of adnexal pathology, which could optimise a patients' triage and thereby improve treatment and outcome. After all, referral of patients to a gynaecologic oncology centre is associated

Table 2
Pooled summary point estimates for the subgroups of pre- and postmenopausal patients for all methods included in this review at their original cut-off point together with their 95% confidence interval (95%CI).

	Premenopausal					Postmenopausal				
	Sens. (95%CI)	Spec. (95%CI)	DOR (95%CI)	LR+ (95%CI)	LR− (95%CI)	Sens. (95%CI)	Spec. (95%CI)	DOR (95%CI)	LR+ (95%CI)	LR− (95%CI)
RMI-I	0.63 (0.51–0.73)	0.93 (0.89–0.95)	21 (16–28)	8.51 (6.38–11.37)	0.40 (0.30–0.54)	0.79 (0.77–0.82)	0.86 (0.79–0.91)	23 (15–36)	5.60 (3.80–8.30)	0.24 (0.21–0.27)
RMI-II	0.58 (0.46–0.68)	0.91 (0.87–0.94)	13 (8–22)	6.3 (4.4–9.0)	0.47 (0.36–0.60)	0.84 (0.80–0.87)	0.80 (0.76–0.83)	20 (14–27)	4.10 (3.40–4.90)	0.21 (0.17–0.25)
RMI-III	0.57 (0.45–0.68)	0.90 (0.85–0.93)	11 (7–18)	5.44 (3.95–7.49)	0.48 (0.37–0.62)	0.79 (0.76–0.82)	0.89 (0.87–0.92)	32 (23–45)	7.50 (5.80–9.60)	0.23 (0.20–0.27)
SA	0.90 (0.88–0.91)	0.94 (0.93–0.95)	151 (123–185)	15.89 (12.86–19.62)	0.11 (0.09–0.12)	0.94 (0.93–0.95)	0.85 (0.82–0.88)	96 (66–139)	6.43 (5.14–8.05)	0.07 (0.05–0.08)
SR+SA	0.89 (0.86–0.92)	0.91 (0.85–0.95)	86 (43–172)	10.27 (5.86–18.01)	0.12 (0.09–0.16)	0.92 (0.90–0.94)	0.87 (0.84–0.90)	83 (60–115)	7.37 (5.80–9.36)	0.09 (0.07–0.12)
SR+Mal	0.93 (0.90–0.95)	0.82 (0.79–0.85)	61 (40–92)	5.23 (4.39–6.24)	0.09 (0.06–0.12)	0.95 (0.92–0.96)	0.77 (0.71–0.81)	57 (40–80)	4.04 (3.28–4.97)	0.07 (0.05–0.10)
LR2	0.87 (0.81–0.91)	0.91 (0.86–0.94)	65 (42–100)	9.60 (6.20–14.70)	0.15 (0.11–0.21)	0.94 (0.92–0.96)	0.86 (0.82–0.89)	92 (55–154)	6.59 (5.23–8.29)	0.07 (0.05–0.10)

Abbreviations: CI, confidence interval; RMI, risk of malignancy index; SA, subjective assessment; SR+SA, simple rules, if inconclusive classified by subjective assessment; SR+Mal, simple rules, if inconclusive classified as malignant; sens., sensitivity; spec., specificity; DOR, diagnostic odds ratio; LR+, positive likelihood ratio; LR−, negative likelihood ratio; LR2, logistic regression model 2.

with better median survival [84]. In addition, a new study shows how simple rules can be linked to an estimate of the risk that the tumour is malignant [85]. A classification with five levels is introduced (very low risk, low risk, intermediate risk, elevated risk, very high risk), allowing for a more fine-tuned assessment of adnexal masses.

5. Conclusions

5.1. Implications for research

Side-by-side comparison studies in which all methods are validated in the same population should be performed in order to prove which method demonstrates the best diagnostic performance. Furthermore, studies should employ strict blinding of the ultrasonographer to the outcome (histology) and vice versa.

More research should be performed on new methods, such as the ADNEX-model, or combinations of current methods.

5.2. Implications for practice

Evidence of this meta-analysis shows that, although RMI is used most commonly, both subjective assessment and simple rules, added with subjective assessment in case simple rules are inconclusive, show a superior test accuracy. Since these tests are equally invasive, they are just as acceptable for the patient. Furthermore, these methods provide immediate results (with the exception of about one out of five patients where simple rules yield an inconclusive result) since no blood test is needed, while RMI depends on the level of serum-CA125. The choice between using subjective assessment or simple rules can be made depending on the expertise present in the hospital.

Conflicts of interest statement

None declared.

Role of the funding source

EM, RK and TVG received grants from CZ Fund and the Academic Fund of Maastricht University Medical Centre+. This study is supported by the Flemish Government: FWO project G049312N, Flanders'. DT is Senior Clinical Investigator of the Research Foundation - Flanders (Belgium) (FWO).

The sponsors had no role in design of the review and meta-analysis; 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.

Acknowledgements

The authors acknowledge the authors of included articles who provided us with additional data, whenever possible; Dr. S. Granberg [25], Dr. B. Hagen [50], Dr. F Strigini [34], Dr. L. Roman [32], Dr. O. Lucidarme [27], Dr. A. Rossi and Dr. L. Forzano [51], Dr. S. Derchain [9], Dr. N. Nunes and Dr. D. Jurkovic [6,7,77], Dr. E Vaes and Dr. R. Manchanda [56], Dr. J.L. Alcazar [10,70,73], Dr. R. Moszynski [30], Dr. V. Arun-Muthuvel [46], Dr. T. Tongsong [58] and the members of the IOTA group [1,3,39,40,42].

This research was supported by CZ Fund, Academic Fund of Maastricht University Medical Centre+, Flemish Government: FWO project G049312N, Flanders'.

Appendices. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ejca.2016.01.007.

References

[1] Van Gorp T, Veldman J, Van Calster B, Cadron I, Leunen K, Amant F, et al. Subjective assessment by ultrasound is superior to the risk of malignancy index (RMI) or the risk of ovarian malignancy algorithm (ROMA) in discriminating benign from malignant adnexal masses. Eur J Cancer 2012;48(11):1649–56.

[2] Valentin L, Jurkovic D, Van Calster B, Testa A, Van Holsbeke C, Bourne T, et al. Adding a single CA 125 measurement to ultrasound imaging performed by an experienced examiner does not improve preoperative discrimination between benign and malignant adnexal masses. Ultrasound Obstet Gynecol 2009;34(3): 345–54.

[3] Timmerman D, Ameye 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 IOTA group. BMJ 2010;341:c6839.

[4] Timmerman D. The use of mathematical models to evaluate pelvic masses; can they beat an expert operator. Best Prac Res Clin Obstet Gynaecol 2004;18(1):91–104.

[5] Kaijser 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 2013; 20(3):449–62.

[6] Nunes N, Ambler G, Foo X, Naftalin J, Widschwendter M, Jurkovic D. Use of the IOTA simple rules for the diagnosis of ovarian cancer: a meta-analysis. Ultrasound Obstet Gynecol 2014;44(5):503–14.

[7] 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(9):1583–9.

[8] Sayasneh A, Wynants L, Preisler J, Kaijser 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.

[9] Hartman CA, Juliato CR, Sarian LO, Toledo MC, Jales RM, Morais SS, et al. Ultrasound criteria and CA 125 as predictive variables of ovarian cancer in women with adnexal tumors. Ultrasound Obstet Gynecol 2012;40(3):360–6.

[10] Alcazar JL, Pascual MA, Olartecoechea B, Graupera B, Auba M, Ajossa S, et al. IOTA simple rules for discriminating between benign and malignant adnexal masses: a prospective external validation. Ultrasound Obstet Gynecol 2013.

[11] Alcazar JL, Jurado M. Three-dimensional ultrasound for assessing women with gynecological cancer: a systematic review. Gynecol Oncol 2011;120(3):340–6.

[12] Dodge JE, Covens AL, Lacchetti C, Elit LM, Le T, Devries-Aboud M, et al. Preoperative identification of a suspicious adnexal mass: a systematic review and meta-analysis. Gynecol Oncol 2012;126(1):157–66.

[13] Geomini P, Kruitwagen R, Bremer GL, Cnossen J, Mol BW. The accuracy of risk scores in predicting ovarian malignancy: a systematic review. Obstet Gynecol 2009;113(2 Pt 1):384–94.

[14] Kinkel K, Lu Y, Mehdizade A, Pelte MF, Hricak H. Indeterminate ovarian mass at US: incremental value of second imaging test for characterization–meta-analysis and Bayesian analysis. Radiology 2005;236(1):85–94.

[15] Liu J, Xu Y, Wang J. Ultrasonography, computed tomography and magnetic resonance imaging for diagnosis of ovarian carcinoma. Eur J Radiol 2007;62(3):328–34.

[16] Stukan M, Dudziak M, Ratajczak K, Grabowski JP. Usefulness of diagnostic indices comprising clinical, sonographic, and biomarker data for discriminating benign from malignant ovarian masses. J Ultrasound Med 2015;34(2):207–17.

[17] Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. Ultrasound Obstet Gynecol 2008;31(6): 681–90.

[18] 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(10):922–9.

[19] Tingulstad S, Hagen B, Skjeldestad FE, Onsrud M, Kiserud T, Halvorsen T, et al. Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses. Br J Obstet Gynaecol 1996;103(8):826–31.

[20] Tingulstad S, Hagen B, Skjeldestad FE, Halvorsen T, Nustad K, Onsrud M. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals. Obstet Gynecol 1999;93(3): 448–52.

[21] EFSUMB. Minimum training recommendations for the practice of medical ultrasound. Ultraschall Med 2006;27(1):79–105.

[22] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011; 155(8):529–36.

[23] Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye 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(34):8794–801.

[24] Daemen A, Valentin L, Fruscio R, Van Holsbeke C, Melis GB, Guerriero S, et al. Improving the preoperative classification of adnexal masses as benign or malignant by second-stage tests. Ultrasound Obstet Gynecol 2011;37(1):100–6.

[25] Granberg S, Norstrom A, Wikland M. Tumors in the lower pelvis as imaged by vaginal sonography. Gynecol Oncol 1990;37(2): 224–9.

[26] Guerriero S, Alcazar JL, Ajossa S, Galvan R, Laparte C, Garcia-Manero M, et al. Transvaginal color Doppler imaging in the

detection of ovarian cancer in a large study population. Int J Gynecol Cancer 2010;20(5):781–6.
[27] Lucidarme O, Akakpo JP, Granberg S, Sideri M, Levavi H, Schneider A, et al. A new computer-aided diagnostic tool for non-invasive characterisation of malignant ovarian masses: results of a multicentre validation study. Eur Radiol 2010;20(8):1822–30.
[28] Mancuso A, De Vivo A, Triolo O, Irato S. The role of transvaginal ultrasonography and serum CA 125 assay combined with age and hormonal state in the differential diagnosis of pelvic masses. Eur J Gynaecol Oncol 2004;25(2):207–10.
[29] Meray O, Turkcuoglu I, Meydanli MM, Kafkasli A. Risk of malignancy index is not sensitive in detecting non-epithelial ovarian cancer and borderline ovarian tumor. J Turk Ger Gynecol Assoc 2010;11(1):22–6.
[30] Moszynski R, Szpurek D, Szubert S, Sajdak S. Analysis of false negative results of subjective ultrasonography assessment of adnexal masses. Ginekol Pol 2013;84(2):102–7.
[31] Romagnolo C, Trivella G, Bonacina M, Fornale M, Maggino T, Ferrazzi E. Preoperative diagnosis of 221 consecutive ovarian masses: scoring system and expert evaluation. Eur J Gynaecol Oncol 2006;27(5):487–9.
[32] Roman LD, Muderspach LI, Stein SM, Laifer-Narin S, Groshen S, Morrow CP. Pelvic examination, tumor marker level, and gray-scale and Doppler sonography in the prediction of pelvic cancer. Obstet Gynecol 1997;89(4):493–500.
[33] Salle B, Gaucherand P, Ecochard R, Rudigoz RC. Contribution of colour pulsed Doppler in pre-operative work-up for pelvic masses. J de Gynecol Obstet et Biol de la Reprod 1995;24(3):234–40.
[34] Strigini FA, Gadducci A, Del Bravo B, Ferdeghini M, Genazzani AR. Differential diagnosis of adnexal masses with transvaginal sonography, color flow imaging, and serum CA 125 assay in pre- and postmenopausal women. Gynecol Oncol 1996;61(1):68–72.
[35] Testa A, Kaijser 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(4):680–8.
[36] Timmerman D, Schwarzler P, Collins WP, Claerhout F, Coenen M, Amant F, et al. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. Ultrasound Obstet Gynecol 1999;13(1):11–6.
[37] Valentin L. Prospective cross-validation of Doppler ultrasound examination and gray-scale ultrasound imaging for discrimination of benign and malignant pelvic masses. Ultrasound Obstet Gynecol 1999;14(4):273–83.
[38] Valentin L, Hagen B, Tingulstad S, Eik-Nes S. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. Ultrasound Obstet Gynecol 2001;18(4):357–65.
[39] Van Calster B, Timmerman D, Bourne T, Testa AC, Van Holsbeke C, Domali E, et al. Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125. J Natl Cancer Inst 2007;99(22):1706–14.
[40] 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(2):684–91.
[41] Van Holsbeke C, Van Calster B, Valentin L, Testa AC, Ferrazzi E, Dimou I, et al. External validation of mathematical models to distinguish between benign and malignant adnexal tumors: a multicenter study by the International Ovarian Tumor Analysis Group. Clin Cancer Res 2007;13(15 Pt 1):4440–7.
[42] Van Holsbeke C, Van Calster B, Bourne T, Ajossa S, Testa AC, Guerriero S, et al. External validation of diagnostic models to estimate the risk of malignancy in adnexal masses. Clin Cancer Res 2012;18(3):815–25.
[43] Sohaib SA, Mills TD, Sahdev A, Webb JA, Vantrappen PO, Jacobs IJ, et al. The role of magnetic resonance imaging and ultrasound in patients with adnexal masses. Clin Radiol 2005;60(3):340–8.
[44] van Trappen PO, Rufford BD, Mills TD, Sohaib SA, Webb JA, Sahdev A, et al. Differential diagnosis of adnexal masses: risk of malignancy index, ultrasonography, magnetic resonance imaging, and radioimmunoscintigraphy. Int J Gynecol Cancer 2007;17(1):61–7.
[45] Andersen ES, Knudsen A, Rix P, Johansen B. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. Gynecol Oncol 2003;90(1):109–12.
[46] Arun-Muthuvel V, Jaya V. Pre-operative evaluation of ovarian tumors by risk of malignancy index, CA125 and ultrasound. Asian Pac J Cancer Prev 2014;15(6):2929–32.
[47] Ashrafgangooei T, Rezaeezadeh M. Risk of malignancy index in preoperative evaluation of pelvic masses. Asian Pac J Cancer Prev 2011;12(7):1727–30.
[48] Asif N, Sattar A, Dawood MM, Rafi T, Aamir M, Anwar M. Preoperative evaluation of ovarian mass: risk of malignancy index. J Coll Physicians Surg Pak 2004;14(3):128–31.
[49] Aslam N, Tailor A, Lawton F, Carr J, Savvas M, Jurkovic D. Prospective evaluation of three different models for the preoperative diagnosis of ovarian cancer. BJOG 2000;107(11):1347–53.
[50] Hagen B, Tingulstad S, Onsrud M, Moen M, Kiserud T, Eik-Nes S, et al. Preoperative identification of malignancy among women with a pelvic mass. Evaluation of a risk index based on ultrasound findings. CA 125 in serum and menopausal status. Tidsskr Nor Laegeforen 1995;115(7):820–2.
[51] Rossi A, Braghin C, Soldano F, Isola M, Capodicasa V, Londero AP, et al. A proposal for a new scoring system to evaluate pelvic masses: pelvic masses score (PMS). Eur J Obstet Gynecol Reprod Biol 2011;157(1):84–8.
[52] Smolen A, Stachowicz N, Czekierowski A, Kotarski J. The estimation of the probability of tumor malignancy on the basis of test combination in the primary diagnosis of adnexal tumors. Ginekol Pol 2010;81(4):254–61.
[53] Terzic M, Dotlic J, Ladjevic IL, Atanackovic J, Ladjevic N. Evaluation of the risk malignancy index diagnostic value in patients with adnexal masses. Vojnosanit Pregl 2011;68(7):589–93.
[54] Terzic M, Dotlic J, Likic I, Brndusic N, Pilic I, Ladjevic N, et al. Risk of malignancy index validity assessment in premenopausal and postmenopausal women with adnexal tumors. Taiwan J Obstet Gynecol 2013;52(2):253–7.
[55] Timmerman D, Verrelst H, Bourne TH, De Moor B, Collins WP, Vergote I, et al. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound Obstet Gynecol 1999;13(1):17–25.
[56] Vaes E, Manchanda R, Autier P, Nir R, Nir D, Bleiberg H, et al. Differential diagnosis of adnexal masses: sequential use of the risk of malignancy index and HistoScanning, a novel computer-aided diagnostic tool. Ultrasound Obstet Gynecol 2012;39(1):91–8.
[57] Fathallah K, Huchon C, Bats AS, Metzger U, Lefrre-Belda MA, Bensaid C, et al. External validation of simple ultrasound rules of Timmerman on 122 ovarian tumors. Gynecologie Obstetrique Fertilite 2011;39(9):477–81.
[58] Tantipalakorn C, Wanapirak C, Khunamornpong S, Sukpan K, Tongsong T. IOTA simple rules in differentiating between benign and malignant ovarian tumors. Asian Pac J Cancer Prev 2014;15(13):5123–6.
[59] Alcazar JL, Pascual MA, Olartecoechea B, Graupera B, Auba M, Ajossa S, et al. IOTA simple rules for discriminating between benign and malignant adnexal masses: a prospective external validation. Ultrasound Obstet Gynecol 2013;42(4):467–71.
[60] Engelen MJ, Bongaerts AH, Sluiter WJ, de Haan HH, Bogchelman DH, Tenvergert EM, et al. Distinguishing benign

and malignant pelvic masses: the value of different diagnostic methods in everyday clinical practice. Eur J Obstet Gynecol Reprod Biol 2008;136(1):94–101.
[61] Radosa MP, Camara O, Vorwergk J, Diebolder H, Winzer H, Mothes A, et al. Preoperative multimodal strategies for risk assessment of adnexal masses: analysis of 1362 cases in a gynecologic cancer center. Int J Gynecol Cancer 2011;21(6):1056–62.
[62] Sayasneh A, Wynants L, Preisler J, Kaijser 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(12):2448–54.
[63] Kaijser J, Van Gorp T, Van Hoorde K, Van Holsbeke C, Sayasneh A, Vergote I, et al. A comparison between an ultrasound based prediction model (LR2) and the risk of ovarian malignancy algorithm (ROMA) to assess the risk of malignancy in women with an adnexal mass. Gynecol Oncol 2013;129(2):377–83.
[64] van den Akker PA, Zusterzeel PL, Aalders AL, Snijders MP, Samlal RA, Vollebergh JH, et al. External validation of the adapted risk of malignancy index incorporating tumor size in the preoperative evaluation of adnexal masses. Eur J Obstet Gynecol Reprod Biol 2011;159(2):422–5.
[65] Mansour GM, El-Lamie IK, El-Sayed HM, Ibrahim AM, Laban M, Abou-Louz SK, et al. Adnexal mass vascularity assessed by 3-dimensional power Doppler: does it add to the risk of malignancy index in prediction of ovarian malignancy?: four hundred-case study. Int J Gynecol Cancer 2009;19(5):867–72.
[66] Hakansson F, Hogdall EV, Nedergaard L, Lundvall L, Engelholm SA, Pedersen AT, et al. Risk of malignancy index used as a diagnostic tool in a tertiary centre for patients with a pelvic mass. Acta Obstet Gynecol Scand 2012;91(4):496–502.
[67] Bensaid C, Belda MALF, Metzger U, Larousserie F, Clement D, Chatellier G, et al. Performance of laparoscopy in identifying malignant ovarian cysts. Surgical Endoscopy and Other Interventional Techniques 2006;20(9):1410–4.
[68] Akturk E, Karaca RE, Alanbay I, Dede M, Karasahin E, Yenen MC, et al. Comparison of four malignancy risk indices in the detection of malignant ovarian masses. J Gynecol Oncol 2011;22(3):177–82.
[69] Javdekar R, Maitra N. Risk of malignancy index (RMI) in evaluation of adnexal mass. J Obstet Gynaecol India 2015;65(2):117–21.
[70] Ruiz de Gauna B, Rodriguez D, Olartecoechea B, Auba M, Jurado M, Gomez Roig MD, et al. Diagnostic performance of IOTA simple rules for adnexal masses classification: a comparison between two centers with different ovarian cancer prevalence. Eur J Obstet Gynecol Reprod Biol 2015;191:10–4.
[71] Knafel A, Nocun A, Banas T, Wiechec M, Jach R, Pietrus M, et al. Iota simple ultrasound-based rules: why do we have inconclusive results? Int J Gynecol Cancer 2013;23(8 SUPPL. 1):155–6.
[72] Sayasneh A, Kaijser J, Preisler J, Smith AA, Raslan F, Johnson S, et al. Accuracy of ultrasonography performed by examiners with varied training and experience in predicting specific pathology of adnexal masses. Ultrasound Obstet Gynecol 2015;45(5):605–12.
[73] Utrilla-Layna J, Alcazar JL, Auba M, Laparte C, Olartecoechea B, Errasti T, et al. Performance of three-dimensional power Doppler angiography as third-step assessment in differential diagnosis of adnexal masses. Ultrasound Obstet Gynecol 2015;45(5):613–7.
[74] Silvestre L, Martins WP, Candido-Dos-Reis FJ. Limitations of three-dimensional power Doppler angiography in preoperative evaluation of ovarian tumors. J Ovarian Res 2015;8:47.
[75] Tinnangwattana D, Vichak-Ururote L, Tontivuthikul P, Charoenratana C, Lerthiranwong T, Tongsong T. IOTA simple rules in differentiating between benign and malignant adnexal masses by non-expert examiners. Asian Pac J Cancer Prev 2015;16(9):3835–8.
[76] Knafel A, Banas T, Nocun A, Wiechec M, Jach R, Ludwin A, et al. The prospective external validation of international ovarian tumor analysis (IOTA) simple rules in the hands of level I and II examiners. Ultraschall Med 2015.
[77] Nunes N, Yazbek J, Ambler G, Hoo W, Naftalin J, Jurkovic D. Prospective evaluation of the IOTA logistic regression model LR2 for the diagnosis of ovarian cancer. Ultrasound Obstet Gynecol 2012;40(3):355–9.
[78] Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye 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(34):8794–801.
[79] du Bois A, Ewald-Riegler N, de Gregorio N, Reuss A, Mahner S, Fotopoulou C, et al. Borderline tumours of the ovary: a cohort study of the Arbeitsgmeinschaft Gynakologische Onkologie (AGO) Study Group. Eur J Cancer 2013;49(8):1905–14.
[80] Mallett S, Halligan S, Thompson M, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. BMJ 2012;345:e3999.
[81] Greenland S. The need for reorientation toward cost-effective prediction: comments on 'Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). Stat Med 2008;27(2):199–206.
[82] 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(27):4194–200.
[83] Van Calster B, Van Hoorde K, Valentin L, Testa AC, Fischerova D, Van Holsbeke C, 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.
[84] American College of Obstetricians and Gynecologists Committee on Gynecologic Practice. Committee Opinion No. 477: the role of the obstetrician-gynecologist in the early detection of epithelial ovarian cancer. Obstet Gynecol 2011;117(3):742–6.
[85] Timmerman D, Van Calster B, Testa A, Savelli L, Fischerova D, Froyman W, et al. Predicting the risk of malignancy in adnexal masses based on the Simple Rules from the International Ovarian Tumor Analysis group. Am J Obstet Gynecol 2016. http://dx.doi.org/10.1016/j.ajog.2016.01.007 [Epub ahead of print].