SWEDISH AGENCY FOR HEALTH TECHNOLOGY ASSESSMENT AND ASSESSMENT OF SOCIAL SERVICES

English Summary

Diagnostics of ovarian cancer; test accuracy of algorithm-based diagnostics in cases of suspected ovarian cancer.

A systematic review including ethical aspects and overview of health economic considerations.

2025. The full report in Swedish (www.sbu.se/395)

Main message

SBU assessed six algorithms (SR, PR, RMI, ROMA, ADNEX, LR2*) for the evaluation of pre-operative ovarian cancer risk. The most frequently used algorithms in Sweden are PR and SR. General finding: SR and PR demonstrated comparable diagnostic accuracy to most alternatives for the whole population.

Conclusions

- For premenopausal women, using indirect comparisons: SR, PR, and ADNEX algorithms exhibited superior test accuracy.
- For postmenopausal women, SR, PR, RMI, ROMA, and ADNEX algorithms showed high test accuracy, which was indistinguishable from each other.
- For LR2, there was insufficient data to make firm conclusions.

Aim

The aim of this systematic review was to assess the diagnostic test accuracy of six algorithms used for suspected ovarian cancer. It encompasses an ethical discourse on the findings.

Background

The vague symptoms of ovarian cancer complicate early diagnosis for patients and physicians, though timely detection strongly correlates with improved survival. Definitive diagnosis requires microscopic tissue analysis or structured clinical follow-up. Preoperative biopsy is contraindicated due to cancer dissemination risk. Diagnostic algorithms for preoperative evaluation integrate menopausal status, ultrasound findings, and serum tumor markers. In Sweden, ultrasound-based methods – Simple Rules and Pattern Recognition – are predominantly utilized.

* SR: Simple Rules, PR: Pattern Recognition, RMI: Risk of Malignancy Index, ROMA: Risk of Ovarian Malignancy Algorithm, ADNEX: Assessment of Different NEoplasias in the adneXa, LR2: Logistic Regression Model 2.

Method

We performed a systematic review following the PRISMA guidelines. The certainty of the evidence was evaluated using GRADE.

PIRO

Population:
Pre- and postmenopausal (reported separately) women with suspected ovarian, fallopian tube or peritoneal tumors or symptoms indicative of ovarian cancer.

Indextest:
The algorithms: RMI (Risk of Malignancy Index), ROMA (Risk of Ovarian Malignancy Algorithm), ADNEX (Assessment of Different NEoplasias in the adneXa model), LR2 (Logistic Regression 2), SR (Simple Rules), and PR (Pattern Recognition).

Reference test:
Surgery: a morphological microscopic diagnosis (by a pathologist) or structured follow-up with no development of a diagnosis requiring surgery for at least 12 months, the variants of were considered equivalent.

Outcome:
Principle outcome was sensitivity, specificity, and other derivatives from the 2x2 tables (number of true/false positive and true/false negative results)

Study design:
Diagnostic prospective or retrospective cross-sectional studies and randomized studies.

Language:
English, Swedish, Norwegian, and Danish.

Databases searched:
Cochrane Library, CDSR, Central (Wiley), Embase (Elsevier), Medline (Ovid), Clinicaltrials.gov (NLM), WHO ICTRP (WHO).

Result

A total of 59 primary studies were included in the analysis, with a total of close to 71 500 observations.

Figure 1 Flowchart.

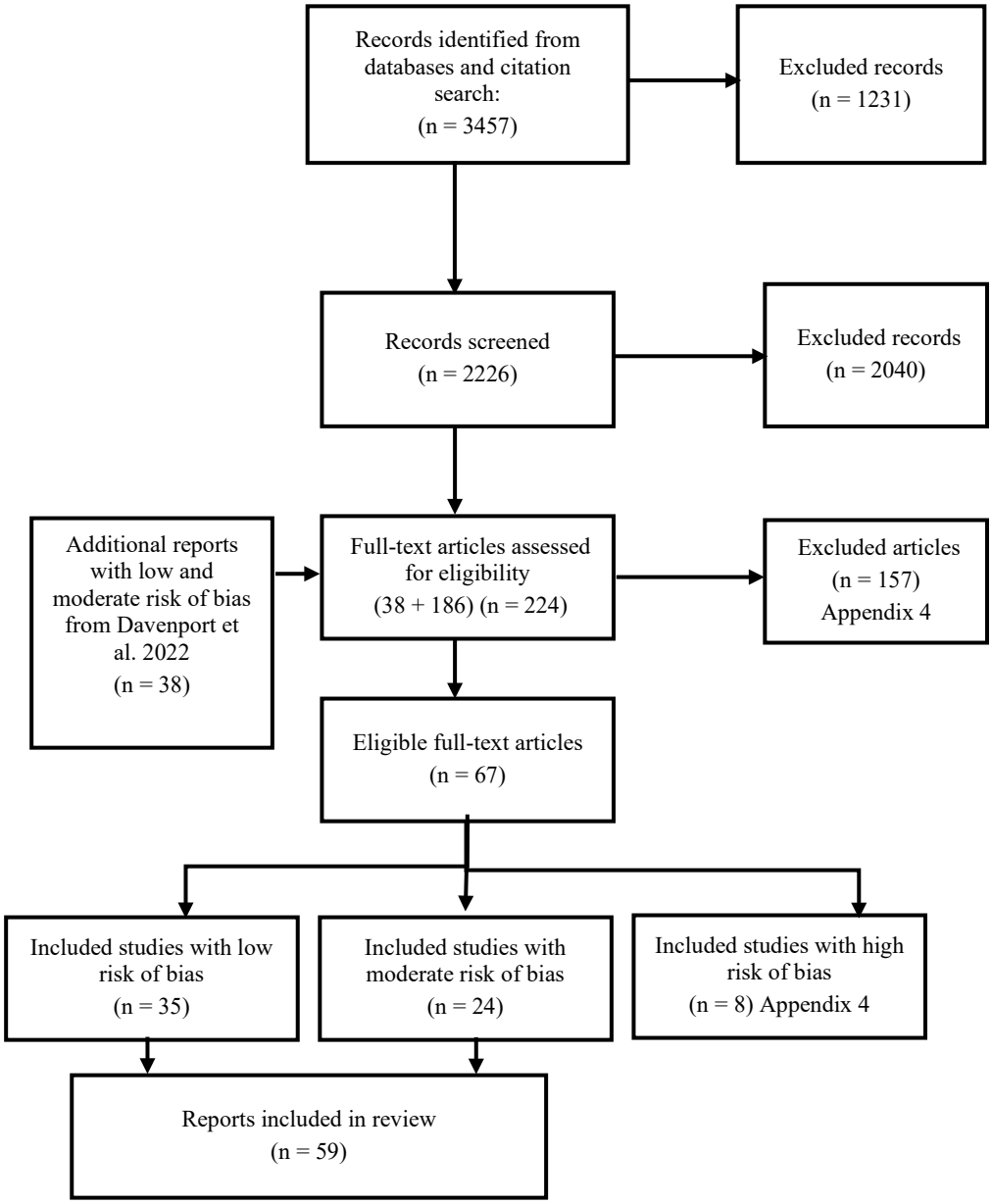
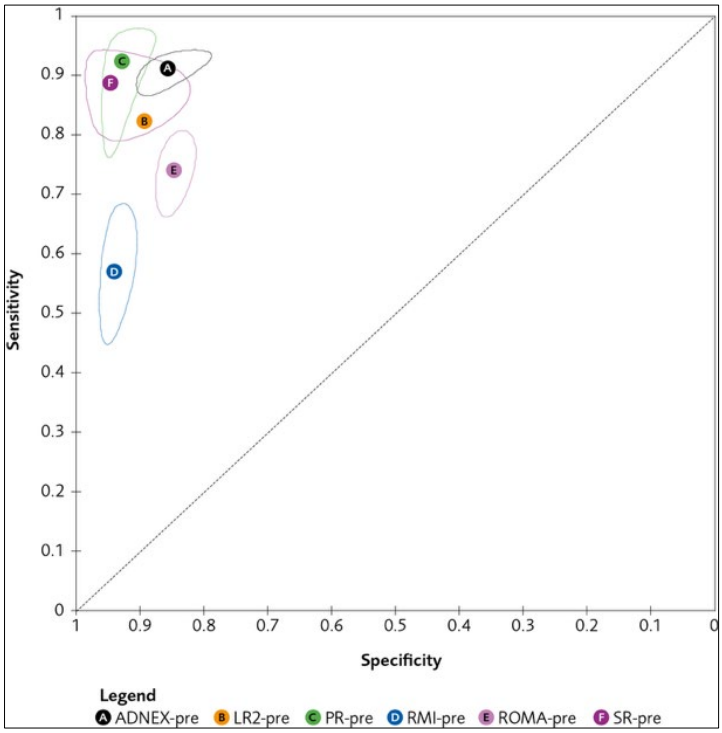


Table 1 Summary of main findings for premenopausal women.

Algorithm (threshold)	**Total number of observations** ----- **Number of included studies**	**Sensitivity (95 % CI)** ------ **GRADE**	**Specificity (95 % CI)** ------ **GRADE**	**Comments on GRADE evaluation**
RMI (200 IU/ml)	n=8229 ----- n=18	0.57 (0.48–0.66) ------ ⊕⊕⊙⊙	0.94 (0.92–0.96) ----- ⊕⊕⊕⊕	Sens: -1 heterogeneity, -1 precision Spec: /
ROMA (11,4–12,5%)	n=6150 ----- n=26	0.74 (0.68–0.79) ------ ⊕⊕⊙⊙	0.85 (0.82–0.87) ----- ⊕⊕⊕⊕	Sens: -2 precision Spec: /
LR2 (10%)	n=4578 ----- n=4	0.82 (0.82–0.83) ------ ⊕⊕⊕⊙	0.89 (0.89–0.90) ----- ⊕⊕⊕⊙	Sens and spec: -1 combination of heterogeneity and limited observations respectively
ADNEX (10% and with CA125)	n=6843 ----- n=13	0.91 (0.88–0.93) ------ ⊕⊕⊕⊙	0.86 (0.81–0.89) ----- ⊕⊕⊕⊙	Sens and spec: -1 heterogeneity respectively
SR (according to IOTA guidelines, varying handling of inconclusive results)	n=5453 ----- n=9	0.89 (0.83–0.93) ------ ⊕⊕⊕⊙	0.95 (0.88–0.98) ----- ⊕⊕⊕⊙	Sens and spec: -1 combination of heterogeneity and precision
PR (ultrasound performed by an experienced sonographer)	n=5280 ----- n=7	0.92 (0.85–0.96) ------ ⊕⊕⊕⊙	0.93 (0.90–0.95) ----- ⊕⊕⊕⊕	Sens: -1 combination of heterogeneity and precision Spec: /

Figure 2 HS-ROC (Hierarchical Summation Receiver Operating Curve) with point estimates and confidence regions for the indirect comparison between the studied algorithms in premenopausal women.

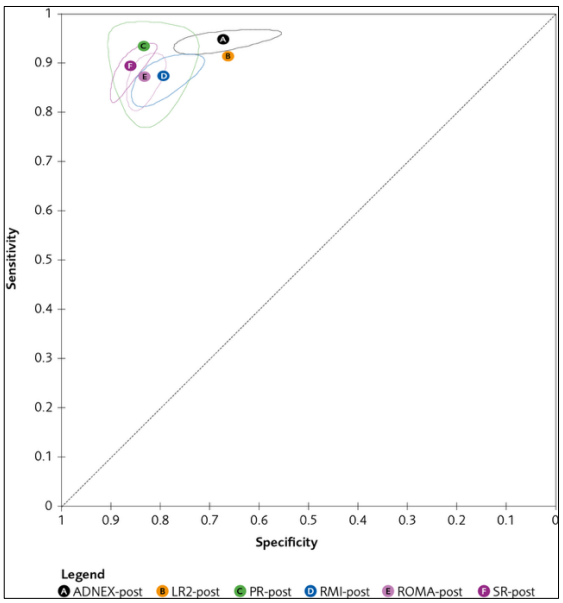


There was no significant difference in diagnostic accuracy among ADNEX, PR, and SR. These methods demonstrated superior accuracy relative to RMI and ROMA. The distinction between RMI and ROMA was significant. RMI exhibited lower sensitivity but greater specificity. Conversely, ROMA showed higher sensitivity with reduced specificity. The point estimate for LR2 aligned more closely with ADNEX, PR, and SR, nonetheless, uncertainty existed for LR2 because of insufficient observations to calculate a confidence region.

Table 2 Summary of main findings for postmenopausal women.

Algorithm (threshold)	**Total number of observations** ----- **Number of included studies**	**Sensitivity (95 % CI)** ------ **GRADE**	**Specificity (95 % CI)** ------ **GRADE**	**Comments on GRADE evaluation**
RMI (200 IU/ml)	n=8731 ----- n=22	0.87 (0,83–0,91) ------ ⊕⊕⊕⊙	0.79 (0,73–0,84) ----- ⊕⊕⊙⊙	Sens: -1 heterogeneity Spec: -1 heterogeneity, -1 precision
ROMA (14,4–29,9%)	n=7896 ----- n=32	0.87 (0,82–0,91) ------ ⊕⊕⊙⊙	0.83 (0,80–0,86) ----- ⊕⊕⊕⊙	Sens: -1 heterogeneity, -1 precision Spec: -1 heterogeneity
LR2 (10%)	n=3735 ----- n=5	0.91 (0.91–0.92) ------ ⊕⊕⊕⊙	0.66 (0.66–0.77) ----- ⊕⊕⊙⊙	Sens: -1 combination of heterogeneity and limited observations Spec: -1 heterogeneity and limited observations
ADNEX (10% and with CA125)	n=6412 ----- n=14	0.95 (0.93–0.96) ------ ⊕⊕⊕⊕	0.67 (0.59–0.75) ----- ⊕⊕⊙⊙	Sens: / Spec: -1 heterogeneity, -1 precision
SR (according to IOTA guidelines, varying handling of in-conclusive results)	n=4795 ----- n=10	0.89 (0.85–0.93) ------ ⊕⊕⊕⊙	0.86 (0.83–0.89) ----- ⊕⊕⊕⊙	Sens and Spec: -1 combination of heterogeneity and precision
PR (ultrasound performed by an experienced sonographer)	n=3187 ----- n=6	0.93 (0.87–0.97) ------ ⊕⊕⊕⊙	0.83 (0.78–0.88) ----- ⊕⊕⊕⊙	Sens and Spec: -1 combination of heterogeneity and precision

Figure 3 HS-ROC (Hierarchical Summation Receiver Operating Curve) with point estimates and confidence regions for the indirect comparison between the studied algorithms in postmenopausal women.



The point estimates for PR, SR, RMI, and ROMA were clustered around 0.9, for sensitivity and around one-tenth lower for the specificity, it was not possible to statistically separate them. Both ADNEX and LR2 had a point estimate for sensitivity that was comparable in relation to the others but with a lower specificity. ADNEX was indistinguishable PR and RMI. The number of observations for LR2 was insufficient calculate a confidence region.

Ethics

There are several ethical aspects to consider when assessing the preoperative risk of ovarian cancer, such as the risks of over/underdiagnosis, invasion of privacy and integrity, equal access to care and the associated costs.

The ethical considerations in this report should be understood in the context of Swedish legislation and the fact that Swedish healthcare is largely publicly funded.

Discussion

Our assessment indicated that no algorithms surpassed the diagnostic accuracy of SR and PR, currently the most used methods in Sweden. The findings were derived from extensive data from multiple studies characterized by high observations. The studies primarily focused on specialized care settings with elevated cancer prevalence. The RMI algorithm, the most established, provided high specificity and negative predictive value, allowing for confidence in negative results. Notably, both ultrasound methodologies, SR and PR, along with RMI, exhibited high specificity in the premenopausal group, reinforcing the notion that negative results likely indicate benign cause, especially in low-prevalence scenarios with a high negative predictive value. The results for SR and PR indicate that sonographers found it easier to discern benign changes in premenopausal women compared to cancer, influenced by the understanding of lower cancer risk in this population. Both PR and SR present limitations, with PR necessitating a skilled and experienced sonographer, and IOTA certification ensures competency in examinations. SR is challenged by a substantial number of uncertain findings, approximating 20%. Different studies adopt varied approaches to manage these ambiguous cases. Most institutions pursue further diagnostics for uncertain SR findings, typically utilizing PR or imaging modalities like CT or MRI, while some classify these cases as malignant. Other facilities opt for structured follow-up with repeat ultrasound and CA125 testing within 8 to 12 weeks until malignancy risk is assessed, or the patient is deemed healthy. Implementing structured follow-ups may mitigate the incidence of suspected false positives and ambiguous results.

Conflict of Interest

In accordance with SBU's requirements, the experts and scientific reviewers participating in this project have submitted statements about conflicts of interest. These documents are available at SBU's secretariat. SBU has determined that the conditions described in the submissions are compatible with SBU's requirements for objectivity and impartiality.

Appendices

- Appendix 3. Search strategies
- Appendix 4. Excluded references and references with high risk of bias