European Journal of Obstetrics & Gynecology and Reproductive Biology 159 (2011) 422–425

Contents lists available at ScienceDirect

European Journal of Obstetrics & Gynecology and Reproductive Biology

journal homepage: www.elsevier.com/locate/ejogrb

External validation of the adapted Risk of Malignancy Index incorporating tumor size in the preoperative evaluation of adnexal masses

Petronella A.J. van den Akker [a,*], Petra L.M. Zusterzeel [a], Anette L. Aalders [b], Marc P.L.M. Snijders [c], Rahul A.K. Samlal [d], Jos H.A. Vollebergh [e], Kirsten B. Kluivers [a], Leon F.A.G. Massuger [a]

[a] Radboud University Nijmegen Medical Centre, Department of Obstetrics and Gynecology, Nijmegen, The Netherlands
[b] Rijnstate Hospital, Department of Obstetrics and Gynecology, Arnhem, The Netherlands
[c] Canisius-Wilhelmina Hospital, Department of Obstetrics and Gynecology, Nijmegen, The Netherlands
[d] Gelderse Vallei Hospital, Department of Obstetrics and Gynecology, Ede, The Netherlands
[e] Bernhoven Hospital, Department of Obstetrics and Gynecology, Oss, The Netherlands

ARTICLE INFO

Article history:
Received 29 December 2010
Received in revised form 7 July 2011
Accepted 14 July 2011

Keywords:
Adnexal masses
Preoperative evaluation
Risk of Malignancy Index
Ultrasound

ABSTRACT

Objective: The Risk of Malignancy Index (RMI) is a simple scoring system to standardize and improve the preoperative evaluation of adnexal masses. Since 1990, three versions of the RMI have been validated in different clinical studies. Recently, a fourth version of the RMI (RMI-4) was introduced that includes tumor size as an additional parameter. The aim of this study was to validate the ability of RMI-4 to discriminate between non-invasive lesions and invasive malignant adnexal masses, and to compare its performance with RMI-3.

Study design: Women scheduled for surgery for an adnexal mass between 2005 and 2009 in 11 hospitals were included. Ultrasonographic characteristics, menopausal status and serum CA 125 level were registered preoperatively, and combined into the RMI. The performances of RMI-3 and RMI-4 were assessed and statistically tested for differences.

Results: A total of 643 patients were included: 469 benign, 73 borderline and 101 malignant tumors. The RMI-3 had a sensitivity of 76%, specificity of 82%, positive and negative predictive values (PPV and NPV) of 45% and 95%, and an accuracy of 81%. The RMI-4 had a sensitivity of 74%, specificity of 79%, PPV of 40%, NPV of 94%, and an accuracy of 78%. The accuracy of RMI-3 was significantly higher than the accuracy of RMI-4 ($p = .001$). Both models had an area under the curve of 0.86.

Conclusion: Both RMI-3 and RMI-4 were able to discriminate between non-invasive lesions and invasive malignant adnexal masses, with similar performances. Including tumor size in the RMI does not improve its performance.

1. Introduction

The discriminative preoperative evaluation of adnexal masses is rather complicated. A variety of diagnostic procedures has been used, leading to a wide range of variables which can result in an inaccurate interpretation of the nature of the adnexal mass. In view of treatment of ovarian cysts, the assessment between benign and malignant needs to be performed as accurate as possible. To standardize and improve the preoperative evaluation, Jacobs et al. [1] developed the Risk of Malignancy Index (RMI), which is a simplification of a formula found by logistic regression analysis. The RMI was the first diagnostic model that combined demographic, sonographic and biochemical data in the assessment of patients with adnexal masses. The main advantage of this method compared with other diagnostic models is that the RMI is a simple scoring system that can be applied directly into clinical practice without the introduction of expensive or difficult tools. The original RMI is known as RMI-1. The RMI has been modified by Tingulstad et al. in 1996 (RMI-2) [2] and again in 1999 (RMI-3) [3]. The difference between these three measurement tools lies in the different scoring of ultrasound characteristics and menopausal status. The three versions of the RMI have been validated retrospectively and prospectively in various clinical studies [1–15] where a cutoff value of 200 showed the best discrimination between benign and malignant adnexal masses, with high sensitivity and specificity levels (sensitivity 51–90%, specificity 51–97%).

Recently, a fourth RMI was introduced by Yamamoto et al. [16] which includes tumor size as an additional parameter. They found

* Corresponding author at: Radboud University Nijmegen Medical Centre, Department of Obstetrics and Gynecology (791), P.O. Box 9101, 6500 HB Nijmegen, The Netherlands. Tel.: +31 24 361 64 53; fax: +31 24 366 85 97.
E-mail address: S.vandenAkker@obgyn.umcn.nl (Petronella A.J. van den Akker).

0301-2115 © 2011 Elsevier Ireland Ltd. Open access under the Elsevier OA license.
doi:10.1016/j.ejogrb.2011.07.035

that a cutoff level of 450 in RMI-4 is comparable with a cutoff level of 200 in the three previous RMIs. When they compared the RMI-4 in a retrospective study with 253 cases with the previous RMIs, an improved performance at a cutoff level of 450 was found, with an accuracy of 90%. The RMI-4 still needs to be validated prospectively and in different institutions, to assess external validity.

The aim of the present study was to validate the ability of RMI-4 to discriminate between non-invasive lesions and invasive malignant adnexal masses in daily clinical practice, and to compare its performance with RMI-3.

2. Materials and methods

This study was conducted between January 2005 and September 2009 in the Radboud University Nijmegen Medical Centre (RUNMC), a third line regional referral hospital, and in 10 cooperating hospitals in the east of The Netherlands. The study was approved by the medical ethics committee of the Radboud University Nijmegen Medical Centre. The study included 643 women admitted for surgical procedure for an adnexal mass. We have previously published on the RMI-3 in a subgroup of the present study population [15]. Ultrasound was performed transvaginally combined with abdominal ultrasound when needed, by experienced gynecological oncologists, general gynecologists, or registrars in gynecology. Serum samples were analysed for CA 125 as part of routine preoperative assessment, and menopausal status was registered. Based on the data obtained, the RMI-3 was calculated prospectively as the multiplied value of the ultrasound score (U), menopausal status (M) and serum CA 125 level as follows:

RMI-3 [3] = $U \times M \times$ CA 125. Multilocularity, solid areas, bilaterality, ascites and intra-abdominal metastases score one point each. A total of 2 or more points was recalculated into $U = 3$, fewer than 2 points into $U = 1$. Postmenopausal status is defined as more than 1 year of amenorrhea, or age 50 years or older among women who had prior hysterectomies, and scores $M = 3$; premenopausal status scores $M = 1$. Serum CA 125 (U/mL) was entered directly into the equation.

Based on the obtained data, RMI-4 was calculated retrospectively as follows:

RMI-4 [16] = $U \times M \times S \times$ CA 125. A total ultrasound score of 0 or 1 was recalculated into $U = 1$, and a score of ≥ 2 into $U = 4$. Premenopausal status scores $M = 1$ and postmenopausal status scores $M = 4$. The tumor size was obtained from the pathology report. A tumor size (single greatest diameter) of <7 cm was recalculated into $S = 1$, and ≥ 7 cm into $S = 2$, as introduced by Yamamoto et al. [16]. The serum CA 125 (U/mL) was applied directly to the calculation.

Final diagnoses of included patients were based on the histopathological examination of surgical specimens. Patients that were diagnosed with non-gynecological malignancies were excluded from the study.

The RMI was merely registered and not applied in a standardized manner in the further planning of care. Based on the clinical impression by the gynecologist in the local hospital it was decided whether a gynecological oncologist should be involved in the surgical treatment. This clinical impression was based on routine preoperative assessment, consisting of physical examination, testing of serum samples, and ultrasound examination. The local gynecologists varied in levels of expertise, from gynecologists specialised in oncology to general gynecologists.

Statistical analyses were performed using the Statistical Packages for the Social Sciences Version 16.0.1 (SPSS Inc., Chicago, IL). The sensitivity, specificity, positive and negative predictive values, and accuracy of RMI-3 and RMI-4 were calculated. Borderline malignancies were allocated to the non-invasive group in all analyses. Comparison between patients with non-invasive (benign and borderline) lesions and invasive malignancies was performed using the Mann–Whitney U test for age and serum CA 125 level, the Pearson χ^2 test for menopausal status and tumor size and the Kruskal–Wallis test for ultrasound score. A receiver operating characteristic (ROC) curve was created to show the relation between sensitivity and specificity of both RMI-3 and RMI-4 in the discrimination between non-invasive lesions and invasive malignancies, and an area under the curve (AUC) was calculated for both models. The McNemar's test was used to test the difference in performances between RMI-3 and RMI-4. The cutoff level was set at 200 for RMI-3 and 450 for RMI-4 to be able to compare the results with the study of Yamamoto et al. [16]. A p-value $\leq .05$ was considered as statistically significant.

3. Results

A total of 643 patients was included in the study, of whom 469 (73%) were diagnosed with benign ovarian cysts, 73 patients (11%) with borderline malignancies, and 101 patients (16%) with malignant diseases. The distribution of age, menopausal status, ultrasound score, tumor size and serum CA 125 level in the non-invasive and invasive groups was as shown in Table 1. Statistically significant differences between the two groups were observed for all these variables. The histopathological diagnoses are listed in Table 2. The majority of non-invasive gynecological conditions included mucinous cystadenomas ($n = 118$) and serous cystadenomas ($n = 86$). Histopathological diagnoses in invasive malignant diseases were mainly serous cystadenocarcinomas ($n = 41$).

The performances of the RMI-3 and RMI-4 at different cutoff levels are presented in Table 3. At a cutoff level of 200, the RMI-3 gave a sensitivity of 76% and a specificity of 82%. Positive and negative predictive values at that cutoff level were 45% and 95%, respectively. The accuracy was 81%. The RMI-4 gave, at a cutoff level of 450, a sensitivity of 74% and a specificity of 79%. Positive and negative predictive values at that cutoff level were 40% and 94%, respectively. The accuracy was 78%. The diagnostic performances of both RMI-3 and RMI-4 are illustrated in Fig. 1. A comparison of the accuracy levels of the two indices showed that RMI-3 at a cutoff level of 200 was significantly better in predicting invasive malignancy than RMI-4 at a cutoff level of 450 ($p = .001$). Both models had an area under the curve of 0.86.

4. Comment

This study has confirmed that both RMI-3 and RMI-4 were able to discriminate between non-invasive lesions and invasive malignant masses. The RMI-4 tested on a new population of women with adnexal masses showed lower sensitivity and specificity levels compared with the original report [16].

External validation of proposed models often results in a decreased performance compared to the performance that is reported initially [8]. Therefore, external validation of a prediction model is essential before introduction into clinical practice. In this new population both RMI-3 and RMI-4 were able to discriminate between non-invasive lesions and invasive malignant adnexal masses, with similar performances. Although the accuracy was higher in RMI-4, the similar AUC and overlappig ROC curves indicate that the differences in performances are not statistically significant.

We have chosen to use the RMI-3 [3] over the original RMI [1] or RMI-2 [2]. The reason for eliminating RMI-1 is that it gives an ultrasound score (U) of 0 when none of the ultrasound features were present. This results in an RMI of 0 regardless of the CA 125 level, whereas we consider the CA 125 level as an important parameter of the RMI. CA 125 level does contribute in both RMI-2