REM (Risk of Endometrial Malignancy): A Proposal for a New Scoring System to Evaluate Risk of Endometrial Malignancy

Roberto Angioli, Stella Capriglione, Alessia Aloisi, Daniela Luvero, Ester Valentina Cafà, Nella Dugo, Roberto Monteria, Carlo De Cicco Nardone, Corrado Terranova, and Francesco Plotti

Abstract

Purpose: It is often difficult to distinguish a benign endometrial disease from a malignancy and tools to help the physician are needed to triage patients into high and low risk of endometrial cancer. The purpose of this study was to obtain a predictive model to assess the risk of endometrial malignancy (REM) in women with ultrasound endometrial abnormalities.

Experimental Design: Women, between ages 45 to 80 years, diagnosed through ultrasound with endometrial abnormalities and scheduled to have surgery were enrolled on a prospective study at the Department of Gynaecologic Oncology of Campus Bio-Medico, University of Rome. Preoperative clinical, ultrasound and laboratory characteristics were taken into account. Logistic regression algorithm was used to categorize patients into low- and high-risk groups for endometrial cancer.

Results: A total of 675 patients were considered for the analysis: 88 with endometrial cancer and 587 with benign endometrial disease. We divided the patients into two groups: training set (TS) and verification set (VS). Preoperative age, symptom, HE4 levels, and ultrasound endometrial thickness were found statistically significant, and were included into a multivariate logistic regression model to determine the probability to have endometrial cancer. In the TS, REM reported 93.3% sensitivity and 97.1% specificity [positive predictive value (PPV), 0.83; negative predictive value (NPV), 0.98; AUC, 0.957; 95% confidence interval (CI), 0.908–0.984]. In the VS, REM reported 89.3% sensitivity and 95.4% specificity (PPV, 0.73; NPV, 0.98; AUC, 0.919; 95% CI, 0.829–0.970).

Conclusions: Our data support the use of REM to triage patients into low- and high-risk groups for endometrial cancer, even if an external validation of the model is needed.

Introduction

In developed countries, endometrial cancer represents the most common gynecologic cancer (1), and it is expected to become an even greater public health concern as the prevalence of obesity, one of the most common risk factors for endometrial cancer, increases worldwide (2).

In the United States, approximately 42,160 cases are diagnosed annually, 7,780 deaths occur, and more than 4,000 new cases are diagnosed in Italy yearly (2). The diagnosis is usually done at an early stage, and approximately 70% of endometrial cancers are diagnosed as stage I; this results in better prognosis, with a 5-year overall survival rate of 90% to 95% (3, 4).

However, almost 20% of patients with endometrial cancer are in the premenopausal state and 10% are asymptomatic. In such a case, it is much harder to make an early diagnosis and usually they are probably diagnosed at advanced stages (5).

Thus, an earlier diagnosis represents an imperative goal to improve survival and prognosis of patients of endometrial cancer. Actually, there are no certified screening tools for endometrial cancer.

Pelvic ultrasound as screening for endometrial cancer reaches 80.5% of sensitivity, when endometrial echo is >5 mm (6), but it dramatically decreases to 20% in asymptomatic women; moreover, specificity is low (61%; ref. 7).

The role of tumor markers in endometrial cancer is still debated, although some serum tumor markers have been studied during recent years. CA125 has been found to be elevated in only 10% to 20% of women with early-stage endometrial cancer (5), and the human epididymis protein 4 (HE4) has been shown to be a useful preoperatively marker with a sensitivity of 59.4% and a specificity of 100% (8).

There are many studies in literature in which the development of nomograms leads to successful applications for oncology (9), and developing of an accurate predictive nomogram of malignancy risk would, therefore, be greatly beneficial, helping to avoid over- or undertreatment of patients with endometrial cancer. However, there are no
Translational Relevance
For the first time in the literature, we assessed a new risk stratification tool of endometrial cancer, called risk of endometrial malignancy (REM), combining in a prospective collection serum markers as well as clinical and ultrasound characteristics. Based on our results, REM may be used to triage women at high risk of endometrial cancer to centres of excellence for their treatment, optimizing health care costs and reducing the time needed to obtain a diagnosis, considering that patients with endometrial cancer are most likely to undergo a complete surgical staging or optimal oncologic surgery, if correctly referred to a gynecologic oncologist surgeon.

Nomograms for predicting risk of endometrial cancer; therefore, our primary goal is to develop a new scoring system, called risk of endometrial malignancy (REM) score, capable to classify patients with endometrial abnormalities into high- or low-risk groups for endometrial cancer using clinical, ultrasound, and serum markers features. This could help identify the appropriate timing of imaging and surgery in a more personalized manner, thus contributing to improvement of overall patient care, and eventually triaging patients to centers of excellence.

Materials and Methods
From January 2010 to December 2012, all patients with ultrasound endometrial abnormalities (endometrial thickness, polyp and submucous myoma) and a planned surgical intervention (Operative Hysteroscopy) drawn from two distinct institutions, the Departments of Obstetrics and Gynecology and Gynaecologic Oncology, Campus Bio-Medico of Rome, were screened for enrolment, and prospectively included in the study (N° NCT01830192; www.clinicaltrialgov.org) after obtaining the approval of the Institutional Review Board.

Inclusion criteria for enrolment were: (i) age between 45 and 80 years; (ii) Eastern Cooperative Oncology Group performance status 0 to 2 according to the World Health Organization (WHO) criteria; and (iii) informed consent obtained from the patients. Exclusion criteria included: (i) abnormal cardiac, hematologic, renal, respiratory, and/or hepatic functions; (ii) presence of a previous malignancy; (iii) presence of concomitant adnexal masses; and (iv) abnormal pap smear, within last 12 months. According to our protocol, a detailed anamnesis was taken for each patient, recording: age, parity, menarche, body mass index (BMI), comorbidities and previous medical history. Presence of abnormal uterine bleeding (AUB), according to definition of the American Congress of Obstetricians and Gynecologists (ACOG; ref. 10), was accurately recorded. The symptom value was considered 1 if the patient reported AUB, or 0 if not. The day before surgery, ultrasonography was carried out for all patients with a GE Voluson E8, using a 7.5 MHz vaginal probe. All examinations were done by the same operator and were carried out in a systematic and predetermined manner. The maximum endometrial thickness in the longitudinal plane was initially measured; in addition, a subjective semiquantitative assessment of the amount of blood flow within the examined lesion was made (color score): a score of 0 was recorded when no blood flow could be found, and 1 when at least minimal flow could be detected, to have a clinical reproducibility. The day before surgery, blood samples were obtained for HE4 and CA125 dosage, as previously reported (11). For all enrolled patients, a diagnostic hysteroscopy was done under general anesthesia, and an accurate inspection of the endometrial cavity was conducted. In case of endometrial polyps and submucous myoma, hysteroscopic resection and multiple endometrial and endocervical biopsies were performed. In case of focal or diffuse endometrial thickness, a directed biopsy was carried out with subsequent multiple endometrial and endocervical biopsies. Patients with Asherman syndrome and cervical stenosis were excluded from the study. All clinical, ultrasound, operative, and histologic features have been recorded in a dataset using Microsoft Office Excel 2010. After pathologic examinations, patients with atypical endometrial hyperplasia and endocervical malignancy were excluded from the study. All patients with endometrial cancer underwent complete surgical staging, according to the 2011 National Comprehensive Cancer Network (NCCN) guidelines.

Statistical analysis
On the basis of the examination of histologic specimens, patients were allocated into benign group (BG) and malignant group (MG). Using a computer-based random procedure, all enrolled patients were divided into two groups, creating one training set (TS) and one verification set (VS), assigning two-thirds of the benign and malignant patients to the TS and the remaining one-third to VS, to follow the guideline for a minimum number of cases to be included in a study for logistic regression analysis (12). The number of patients needed was $N = 10k/p$, where $N$ is the number of the patients needed for calculation, $k$ is the number of covariates to be analyzed, and $p$ is the smallest of the proportions of negative or positive cases in the population. To compare patients characteristics as classified by type of pathology, Student or Mann–Whitney $t$ tests were used for quantitative variables, according to their pattern of distribution. Qualitative variables were compared by means of the $\chi^2$ or Fisher test, according to the assumptions to be verified. Individual factors, suitably transformed where necessary, were assessed by means of the logistic regression model to evaluate their predictive ability. For all statistical comparisons, a level of $P < 0.05$ was accepted as statistically significant. In the TS, we conducted a logistic regression analysis to select the sonographic and clinical characteristics of the endometrial disease that are best fitted to predict the pathologic outcome of benignity or malignancy. Using the statistically significant features, a predicted probability (PP) was calculated for each
patient using the appropriate logistic regression formula, with the resulting PP values ranging from 0% to 100%. For the analysis of logistic regression, we used Medcalc statistical software ver. 12.4.0.0 in a stepwise manner. For the development of the nomogram, all patients in the TS were included. The resulting nomogram was able to estimate outcome probability of endometrial cancer. Finally in VS, we applied the REM to patients with endometrial cancer and age-matched patients with benign endometrial disease, thus defining the accuracy and reliability of the score.

Results
Starting January 2010 to December 2012, 741 patients were enrolled for hysteroscopy. Thirty-seven patients were excluded due to Asherman syndrome or cervical stenosis. After pathologic examination, 88 patients were found to have endometrial cancer, 587 had benign endometrial disease, 26 patients were diagnosed with atypical endometrial hyperplasia, and 3 patients had endocervical cancer. Therefore, a total of 675 patients were considered for the analysis and randomly divided into TS and VS, in the proportions previously reported (Fig. 1). Clinical, ultrasound, and operative patient characteristics are illustrated in Table 1. Evaluating the homogeneity of the two sets, there were no significant differences, regarding age (51.5 in TS vs. 50.8 in VS), CA125 levels (33.4 in TS vs. 34.5 in VS), HE4 levels (77.4 in TS vs. 78.2 in VS), and endometrial thickness (13.3 in TS vs. 12.7 in VS).

Table 1. The clinical characteristics of the patients enrolled

<table>
<thead>
<tr>
<th></th>
<th>Training set</th>
<th></th>
<th>Verification set</th>
<th></th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Malignant</td>
<td>Benign</td>
<td>Malignant</td>
<td>Benign</td>
<td>Malignant</td>
</tr>
<tr>
<td><strong>n</strong></td>
<td>60</td>
<td>391</td>
<td>28</td>
<td>196</td>
<td>88</td>
</tr>
<tr>
<td><strong>Age median (range)</strong></td>
<td>55 (45–75)</td>
<td>48 (48–69)</td>
<td>53 (47–78)</td>
<td>47 (45–75)</td>
<td>54 (46–78)</td>
</tr>
<tr>
<td><strong>CA125 (U/mL) mean ± SD</strong></td>
<td>33.7 ± 4.5</td>
<td>33.3 ± 7.3</td>
<td>43.5 ± 5.1</td>
<td>26.4 ± 7.9</td>
<td>36.8 ± 4.9</td>
</tr>
<tr>
<td><strong>HE4 (pmol/L) mean ± SD</strong></td>
<td>119.4 ± 21.3</td>
<td>35.6 ± 30.5</td>
<td>150 ± 23.7</td>
<td>40.6 ± 34.6</td>
<td>129.2 ± 22.4</td>
</tr>
<tr>
<td><strong>Endometrial thickness (mm) mean ± SD</strong></td>
<td>18.3 ± 6.8</td>
<td>8.4 ± 7.1</td>
<td>16.3 ± 7.2</td>
<td>9.2 ± 7.4</td>
<td>17.7 ± 7.1</td>
</tr>
<tr>
<td><strong>FIGO Stage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>46 (77%)</td>
<td>—</td>
<td>23 (82%)</td>
<td>—</td>
<td>69 (79%)</td>
</tr>
<tr>
<td>II</td>
<td>7 (12%)</td>
<td>—</td>
<td>2 (7%)</td>
<td>—</td>
<td>9 (10%)</td>
</tr>
<tr>
<td>III</td>
<td>5 (8%)</td>
<td>—</td>
<td>2 (7%)</td>
<td>—</td>
<td>7 (8%)</td>
</tr>
<tr>
<td>IV</td>
<td>2 (3%)</td>
<td>—</td>
<td>1 (4%)</td>
<td>—</td>
<td>3 (3%)</td>
</tr>
<tr>
<td><strong>Histology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endometrioid adenocarcinoma</td>
<td>58 (97%)</td>
<td>—</td>
<td>27 (96%)</td>
<td>—</td>
<td>85 (97%)</td>
</tr>
<tr>
<td>Non-endometrioid carcinoma</td>
<td>2 (3%)</td>
<td>—</td>
<td>1 (4%)</td>
<td>—</td>
<td>3 (3%)</td>
</tr>
<tr>
<td>Endometrial polyp</td>
<td>—</td>
<td>184 (47%)</td>
<td>—</td>
<td>106 (54%)</td>
<td>—</td>
</tr>
<tr>
<td>Myoma</td>
<td>—</td>
<td>39 (10%)</td>
<td>—</td>
<td>17 (9%)</td>
<td>—</td>
</tr>
<tr>
<td>Endometrium without atypia</td>
<td>—</td>
<td>168 (43%)</td>
<td>—</td>
<td>73 (37%)</td>
<td>—</td>
</tr>
</tbody>
</table>
Training set

The training set comprised 451 patients: 60 (13.3%) with endometrial cancer and 391 (86.7%) with benign endometrial disease. A statistically significant difference ($P < 0.001$) was observed with regard to the presence of symptom in the malignant group (51 patients, 85%) compared with the benign group (156 patients, 40%). Similarly, mean endometrial thickness (18.3 and 8.4 mm in the malignant group and benign group, respectively) and mean HE4 levels (119.4 and 35.6 pmol/L in the malignant group and benign group, respectively) showed statistically significant results.

However, there was no statistical significance concerning mean CA125 levels (33.7 and 33.3 UI/mL in the MG and BG, respectively) showed statistically significant results. (119.4 and 35.6 pmol/L in the malignant group and benign group, respectively) and mean HE4 levels (119.4 and 35.6 pmol/L in the malignant group and benign group, respectively) showed statistically significant results.

On the basis of these statistical evaluations, all the significant variables were included in the multivariate logistic regression model to identify the statistically relevance of each feature in the prediction model. The overall model significance is very high with $P < 0.0001$, and the goodness of fit of the model is high as well with the Hosmer and Lemeshow test.

Therefore, we obtained the equation of logistic regression to determine the probability to have endometrial cancer:

$$REM = \frac{1}{1 + e^{- \left[ -7.455 + 0.062343 \times \text{Age} + 0.039591 \times \text{HE4} + 0.26017 \times \text{Thickness} - 2.66275 \times \text{Symptom} \right]}}$$

This equation allows calculation of the REM value; however, for a practical hands-on use, a nomogram has been created (Fig. 2). This equation can be easily reported in a computer program.

Using the Youden index for the receiver operating characteristic (ROC) and area under the curve (AUC) values in the TS, we obtained that the high risk of malignancy was set as $> 0.3185$.

The selected threshold is able to stratify patients into groups with low and high risk of malignancy. Using 0.3185 as the cut-off, in the TS we correctly classified 56 of 60 patients with endometrial cancer and 380 of 391 patients with benign endometrial disease, reporting 93.3% of sensitivity and 97.1% of specificity [positive predictive value (PPV) = 0.83; negative predictive value (NPV) = 0.98; Table 2]. The ROC curve of REM in the TS is reported in Fig. 3 (AUC, 0.957; 95% CI: 0.908–0.984).

Verification set

The VS consisted of 224 patients: 28 (12.5%) patients were affected by endometrial cancer and 196 (87.5%) presented with benign endometrial disease. The symptom was present in 23 of 28 (82.1%) patients in the MG and 55 of 196 (27.9%) patients of the BG ($P < 0.001$). The mean endometrial thickness was 16.3 mm in the MG and 9.2 mm in the BG ($P < 0.001$). The mean HE4 was 150 pmol/L in the MG and 40.6 pmol/L in the BG ($P < 0.001$). Using 0.3185 as the cut-off, in the VS we correctly classified 250 of 28 patients with endometrial cancer and 187 of 196 patients with benign endometrial disease, reporting 89.3% sensitivity and 95.4% specificity (PPV = 0.73; NPV = 0.98; Table 2).

The ROC curve of REM in VS (AUC = 0.919; 95% CI, 0.829–0.970) is shown in Fig. 3.

Discussion

Cancer researchers and clinicians are increasingly interested in alternative tools such as nomograms to improve the management of patients with cancer. In fact, there are many studies in literature wherein the development of nomograms leads to a successful application for oncology prognostics. With regard to endometrial cancer, nomograms for endometrial cancer have been developed to predict the risk of lymph node metastasis, endometrial cancer recurrence, and overall survival (13–15), but not for the detection of endometrial cancer.

Therefore, we assessed, for the first time in literature, a new risk stratification tool for endometrial cancer, called REM, which combines, in a prospective collection, serum markers as well as clinical and ultrasound characteristics.
As no screening tests or scoring systems for endometrial cancer detection have been developed to date, our data are difficult to compare with those reported in literature; for this reason, we compared our results with the most used and effective diagnostic tools actually reported in literature, such as ultrasound, CA125, and HE4, each taken separately. Pelvic ultrasound as screening for endometrial cancer reaches 80.5% of sensitivity when the endometrial echo is more than 5 mm (6). Unfortunately, this technique presents several pitfalls; in particular, a low specificity (61%) and dramatic decreases in sensitivity of 20% in postmenopausal asymptomatic women (7). Moreover, approximately 23% of women who use hormone replacement therapy with normal histology may present abnormal ultrasound feature and, in premenopausal women, the cutoff to consider the endometrium as being thickened has not been standardized as yet (6).

With regard to tumor markers, CA125 showed a low sensitivity and specificity of 19.8% and 62.1%, respectively, in detecting endometrial cancer (8), and tumor marker results were elevated in only 10% to 20% of women with early-stage disease (5). The best results in endometrial cancer detection were observed with HE4, showing a sensitivity of 59.4% and a specificity of 100% (8).

Our REM score showed an overall sensitivity of 92% and a specificity of 96%, outperforming every other single item tested in this model. Furthermore, REM presented a ROC–AUC of 0.957 (95% CI, 0.908–0.984) in the training set and 0.919 (95% CI, 0.829–0.970) in the verification set, showing a better performance compared with the ROC–AUC of

| Table 2. REM accuracy in training, verification, and overall sets |
|-----------------------|-----------------------|-----------------------|
|                      | Training              | Verification          | Overall       |
|                      | Malignant | Benign | Malignant | Benign | Malignant | Benign |
| High risk            | 56        | 11     | 25        | 9      | 81        | 20     |
| Low risk             | 4         | 380    | 3         | 187    | 7         | 567    |
| Sensitivity          | 93.3%     | 89.3%  | 92%       |
| Specificity          | 97.1%     | 95.4%  | 96%       |
| PPV                  | 0.83      | 0.73   | 0.80      |
| NPV                  | 0.98      | 0.98   | 0.98      |

Figure 3. ROC curve of training and verification dataset.
HE4 alone (0.864), presented in our recently published article (8).

Furthermore, our score system reaches high sensitivity and specificity rates (92% and 96% respectively) and high PPV and NPV (80% and 98% respectively) that completely outperforms, in terms of diagnostic power, those reported by other predictive scores used worldwide such as Risk of Ovarian Malignancy Algorithm (ROMA) or Risk Malignancy Index (RMI) that have been developed in ovarian cancer to improve the accuracy of diagnosis (16–18). In fact, the ROMA score, based on a combination of age, CA125, and HE4 levels and considered the most accurate, showed lower sensitivity (88.1%), specificity (74.9%), PPV (38.1%), and NPV (97.3%) compared with our results (16).

Although REM score is still experimental as a diagnostic tool, it is interesting to see how well it performs even in early-stage disease; in fact, in our study group, most patients (79%) had stage I disease.

On the basis of these results, REM may be used to triage women at high risk of endometrial cancer to centers of excellence for their treatment, optimizing healthcare costs and reducing the time needed to obtain a diagnosis, considering that patients with endometrial cancer are most likely to undergo a complete surgical staging or optimal oncologic surgery, if correctly referred to a gynecologic oncologist surgeon.

In fact, a preoperative tool such as REM may be translated into an earlier diagnosis, a more cost-effective management, and an improvement of survival and prognosis of these patients. Because we did not conduct any central reviews, the most significant bias might be the single-center data and measurement error. Usually, measurement error in a covariate tends to attenuate the coefficient estimate for the covariate, that is, a bias toward the null occurs. Indeed, in our data, there were 7 of 88 as false-negative patients in the MG. Therefore, the false negatives might have resulted from measurement errors. Furthermore, we considered a specific set of women, aged between 45 and 80 years. It could be interesting to evaluate REM performance into all aged and/or high-risk population (patients of Lynch syndrome, Cowden syndrome, obesity, diabetes, or breast cancer on tamoxifen).

The most important benefit of REM is that the risk can be determined before surgery. For this reason, it can be used in patient counseling and in informed decision making. It could help to make an early diagnosis even in those 30% of women in which it is usually much harder to make an early diagnosis.

Furthermore, REM can be used in hospitals with limited resources in which a skilled gynecologic oncologist is not always present. Finally, the most appealing of our score is represented by its simplicity and reproducibility.

Nevertheless, our REM outcomes should be considered promising but preliminary. In fact, it will need to be tested and validated in completely independent datasets from other institutions, to confirm its potential risk stratification role in endometrial cancer and its application in clinical practice.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: R. Angioli, S. Capriglione, A. Aloisi, D. Luvero, E.V. Cafà, R. Monteria, C. De Cicco Nardone, C. Terranova, F. Plotti Development of methodology: R. Angioli, S. Capriglione, D. Luvero, E.V. Cafà, C. De Cicco Nardone, C. Terranova, F. Plotti Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R. Angioli, S. Capriglione, D. Luvero, E.V. Cafà, N. Dugo, R. Monteria, C. De Cicco Nardone, C. Terranova, F. Plotti Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R. Angioli, S. Capriglione, D. Luvero, E.V. Cafà, C. De Cicco Nardone, C. Terranova, F. Plotti Writing, review, and/or revision of the manuscript: R. Angioli, S. Capriglione, A. Aloisi, D. Luvero, E.V. Cafà, N. Dugo, R. Monteria, C. De Cicco Nardone, C. Terranova, F. Plotti Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): R. Angioli, S. Capriglione, D. Luvero, E.V. Cafà, N. Dugo, R. Monteria, F. Plotti Study supervision: R. Angioli, F. Plotti

Received May 18, 2013; revised August 9, 2013; accepted August 15, 2013; published OnlineFirst August 27, 2013.

References

REM (Risk of Endometrial Malignancy): A Proposal for a New Scoring System to Evaluate Risk of Endometrial Malignancy

Roberto Angioli, Stella Capriglione, Alessia Aloisi, et al.

*Clin Cancer Res*  Published OnlineFirst August 27, 2013.

**Updated version**
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-13-1376