REM (RISK OF ENDOMETRIAL MALIGNANCY): A PROPOSAL FOR A NEW SCORING SYSTEM TO EVALUATE RISK OF ENDOMETRIAL MALIGNANCY.


*Department of Obstetrics and Gynaecology Campus Bio Medico University of Rome, Italy

Corresponding author:

Angioli Roberto, MD
Department of Obstetrics and Gynecology
University of Rome “Campus Bio-Medico”
Via Alvaro del Portillo, 200 - 00128 Rome
Telephone: 0039 3452572851
Fax: 0039 06 22541456
E-mail address: r.angioli@unicampus.it

Conflict of interest
The authors have declared no conflicts of interest
STATEMENT OF TRANSLATIONAL RELEVANCE

For the first time in literature, we assessed a new risk stratification tool of EC, called REM, combining, in a prospective collection, serum markers, clinical and ultrasound features. Basis on our results, REM may be used to triage women at high risk of EC to centres of excellence for their treatment, optimizing health care costs and reducing the time needed to obtain a diagnosis, considering that EC patients are most likely to perform a complete surgical staging or optimal oncologic surgery, if correctly referred to a gynecologic oncologist surgeon.
ABSTRACT

PURPOSE: It is often difficult to distinguish a benign endometrial disease from a malignancy and tools to help physician are needed, in order to triage patients into high and low risk of endometrial cancer (EC). 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 45 and 80 years, diagnosed with ultrasound endometrial abnormalities and scheduled to have surgery were enrolled on a prospective study at Department of Gynaecologic Oncology of Campus Bio-Medico of Rome. Preoperative clinical, ultrasound and laboratory features were taken into account. Logistic regression algorithm was utilized to categorize patients into low and high risk groups for EC.

RESULTS: A total of 675 patients were considered for the analysis: 88 with EC 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 multivariate logistic regression model in order to determine the probability to have EC. In TS, REM reported 93.3% of sensitivity and 97.1% of specificity (PPV= 0.83, NPV= 0.98, AUC=0.957, 95%CI, 0.908 to 0.984). In VS REM reported 89.3% of sensitivity and 95.4% of specificity (PPV= 0.73, NPV= 0.98, AUC=0.919, 95%CI, 0.829 to 0.970).

CONCLUSIONS: Our data support the use of REM to triage patients into low and high risk of EC, even if an external validation of model is needed.

Keywords: endometrial cancer, score, nomogram, ultrasound, HE4
INTRODUCTION

In developed countries, endometrial cancer (EC) represents the most common gynaecologic 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 EC, increases worldwide (2).

In the United States approximately 42,160 cases are diagnosed annually, 7780 deaths occur and over 4000 new cases are diagnosed in Italy yearly (2). The diagnosis is usually performed in an early stage and approximately 70% of ECs are diagnosed as stage I; this results in better prognosis, with a 5-year overall survival rate of 90%-95% (3,4).

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

Thus, an earlier diagnosis represent an imperative goal, in order to improve survival and prognosis of EC patients. Actually, there are not certified screening tools for EC.

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

The role of tumour markers in EC is still debated, though some serum tumour markers have been studied during recent years. CA125 has been found to be elevated in only 10–20% of women with early-stage EC (5) and Human epididymis protein 4 (HE4) demonstrated 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 a 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 under-treatment of EC patients.

Actually there are no nomograms for predicting risk of EC; therefore our primary goal is to develop a new scoring system, called REM (Risk of Endometrial Malignancy) score, able to classify patients with endometrial abnormalities into high risk or low risk groups for having EC using clinical,
ultrasound and serum markers features. This could help identifying the appropriate timing of imaging and surgery in a more personalized, thus contributing to improvement of overall patient care, eventually triaging patients to centres 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 Department of Obstetrics & Gynecology and the Department of Gynaecologic Oncology of Campus Bio-Medico of Rome, were screened for enrollment and prospectively included in the study (N° NCT01830192 – www.clinicaltrialgov.org), after obtaining the approval of the institutional review board.

Inclusion criteria for enrollment were as follows: 1) aged between 45 and 80 years; 2) Eastern Cooperative Oncology Group performance status 0-2 according to World Health Organization (WHO) criteria; 3) informed consent obtained from the patients. Exclusion criteria included: 1) abnormal cardiac, haematological, renal, respiratory and/or hepatic functions; 2) presence of a previous malignancy; 3) presence of concomitant adnexal masses, 4) 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), co-morbidities, previous medical history. Presence of abnormal uterine bleeding (AUB), according to ACOG definition (10), was accurately recorded. The symptom value has to be considered 1 if the patient reports AUB or 0 if not. The day before surgery, ultrasonography was performed to all patients with a GE Voluson E8, using a 7.5 MHz vaginal probe. All examinations were performed 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, 1 when at least minimal flow could be detected, in order to have a clinical reproducibility. The day before surgery, blood samples were obtained for
HE4 and CA125 dosage, as previously reported (11). At all enrolled patients, a diagnostic hysteroscopy was performed under general anesthesia and an accurate inspection of endometrial cavity was conducted. In case of endometrial polyps, 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's syndrome and cervical stenosis were excluded from the study. All clinical, ultrasound, operative and histological features have been recorded in a dataset using Microsoft Office Excel 2010. After pathological examinations, patients with atypical endometrial hyperplasia and endocervical malignancy were excluded from the study. All patients with EC underwent complete surgical staging, according to 2011 NCCN guidelines.

Statistical analysis

Based on histological specimens, patients were allocated into benign (BG) and malignant groups (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 the 2/3 of the benign and malignant patients to the TS and the remaining 1/3 to VS, in order to follow the guideline for a minimum number of cases to include in a study for logistic regression analysis (12). The number of patients needed is \( N = 10k/p \), where \( N \) is the number of the patients needed for calculation, \( k \) is the number of covariates to be analyzed, \( p \) is the smallest of the proportions of negative or positive cases in the population. To compare patients characteristics of classified by type of pathology, Student’s or Mann–Whitney’s t-tests were used for quantitative variables, according to their pattern of distribution. Qualitative variables were compared by means of the chi-square test or Fisher’s 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 performed a logistic regression analysis to choose the sonographic and clinical characteristics of the endometrial disease that are best fitted to predict the pathological outcome of
benignity or malignity. 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 use software Medcalc © statistical software ver. 12.4.0.0 in stepwise mood. For the development of the nomogram, all patients in the TS set were included. The resulting nomogram is able to estimate outcome probability of EC. Finally in VS, we applied the REM to EC patients and age-matched patients with benign endometrial disease, defining the accuracy and reliability of the score.

RESULTS

Starting January 2010 to December 2012, 741 patients have been enrolled for hysteroscopy. 37 patients were excluded for Asherman's syndrome or cervical stenosis. After pathological examination, 88 patients had EC, 587 benign endometrial disease, 26 patients were diagnosed with atypical endometrial hyperplasia and 3 patients were affected from endocervical cancer. Therefore, a total of 675 patients were considered for the analysis and randomly divided into TS and VS, in the proportion previously reported (Figure 1). Clinical, ultrasound and operative patient’s characteristics are illustrated in Table 1. Evaluating the homogeneity of the two sets, there were no significant differences, regarding age (51.5 in TS versus 50.8 in VS), CA125 levels (33.4 in TS versus 34.5 in VS), HE4 levels (77.4 in TS versus 78.2 in VS) and endometrial thickness (13.3 in TS versus 12.7 in VS).

Training set (TS)

TS consisted of 451 patients: 60 (13.3%) affected by EC and 391 (86.7%) with benign endometrial disease. A statistically significant difference (p<0.001) was observed as regarding the presence of symptom in the MG (51 patients, 85%) compared to BG (156 patients, 40%). Similarly, mean endometrial thickness (18.3 mm in the MG and 8.4 mm in the BG) and mean HE4 levels (119.4 pmol/L in the MG and 35.6 pmol/L in the BG), resulted statistically significant.

Instead there was not a statistically significance as concerning mean CA125 levels (33.7 UI/mL in the MG and 33.3 UI/mL in the BG) and endometrial vascularization (83% of patients in the MG
versus 81.8% of patients in the BG).

On the basis of these statistical evaluations, all the significant variables were included into multivariate logistic regression model in order 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 & Lemeshow test.

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

$$REM = \frac{1}{1 + e^{-(7.4545 + 0.062343 \cdot \text{Age} + 0.039501 \cdot \text{HE4} + 0.25017 \cdot \text{Thickness} - 2.66275 \cdot (1 - \text{Symptom})}}$$

This equation allows to calculate the REM value, but for a practical, hands on, use a nomogram has been created (Figure 2). This equation can be easily reported in a computer program.

Using the Youden index in the ROC curve and 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 low and high risk of malignancy groups. Using 0.3185 as cut-off, in the TS we correctly classified 56/60 patients with EC and 380/391 patients with benign endometrial disease, reporting 93.3% of sensitivity and 97.1% of specificity (PPV = 0.83, NPV = 0.98) (Table 2). The ROC curve of REM in TS is reported in Figure 3 (AUC=0.957, 95%CI, 0.908 to 0.984).

**Verification Set (VS)**

VS consisted of 224 patients: 28 (12.5%) patients affected by EC and 196 (87.5%) presented benign endometrial disease. The symptom was present in 23/28 (82.1%) patients of the MG and 55/196 (27.9%) patients of the BG (p<0.001). Mean endometrial thickness was 16.3 mm in the MG and 9.2 mm in the BG (p<0.001). Mean HE4 was 150 pmol/L in the MG and 40.6 pmol/L in the BG (p<0.001). Using 0.3185 as cut-off, in the VS we correctly classified 25/28 patients with EC and 187/196 patients with benign endometrial disease, reporting 89.3% of sensitivity and 95.4% of specificity (PPV = 0.73, NPV = 0.98) (Table 2). The ROC curve of REM in VS (AUC=0.919, 95%CI, 0.829 to 0.970) is reported in Figure 3.
DISCUSSION

Cancer researchers and clinicians are increasingly interested in alternative tools such as nomograms to improve the management of cancer patients. In fact, there are many studies in literature in which the development of nomograms leads to a successful application for oncology prognostics. Concerning EC, promising nomograms have been developed to predict the risk of lymph node metastasis, EC recurrence and overall survival (13,14,15), but not for the EC detection. Therefore, we assessed, for the first time in literature, a new risk stratification tool of EC, called REM, combining, in a prospective collection, serum markers, clinical and ultrasound features. As no screening tests or scoring systems for EC detection have been already developed, our data are difficult to compare to 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 endometrial echo is >5 mm (6). Unfortunately this techniques present several pitfalls; in particular a low specificity (61%) and sensibility dramatically decreases of a 20% in postmenopausal asymptomatic women (7). Moreover about 23% of women who use hormone replacement therapy with normal histology may present abnormal ultrasound feature and in premenopausal women the cut-off to consider the endometrium thickened has not been standardized yet (6).

About tumour markers, CA125 showed a low sensitivity and specificity of 19.8% and 62.1% respectively in detecting EC (8) and resulted to be elevated in only 10–20% of women with early-stage disease (5). The best results in EC detection were observed with HE4, demonstrating a sensitivity of 59.4% and a specificity of 100% (8).

Our score showed an overall sensitivity of 92% and a specificity of 96% outperforming every single item. Furthermore, REM presented a ROC-AUC of 0.957 (95% CI, 0.908 to 0.984) in the training set and 0.919 (95% CI, 0.829 to 0.970) in the verification set, showing a better performance compared to the ROC-AUC of HE4 alone (0.864), presented in our recently published paper (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 world-widely used predictive scores, 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 (16,17,18). In fact, the ROMA score, based on a combination of age, CA125 and HE4 and considered the most accurate, showed lower sensitivity (88.1%), specificity (74.9%), PPV (38.1%) and NPV (97.3%), compared to our results (16).

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

Basis on our results, REM may be used to triage women at high risk of EC to centres of excellence for their treatment, optimizing health care costs and reducing the time needed to obtain a diagnosis, considering that EC patients are most likely to perform 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-effectiveness management and an improvement of survival and prognosis of these patients. Because we did not perform any central review, the most significant bias might be the single centre 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 7/88 false-negative patients in MG. Therefore, the false negatives might have resulted from measurement errors. Furthermore, we considered a specific set of women, aged between 45-80 years. It could be interesting to evaluate REM performance into all aged and/or high risk population (Lynch syndrome, Cowden syndrome, obesity, diabetes or breast cancer patients 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 counselling and in informed decision making. It could help to make
an early diagnosis even in those 30% of women, in which is usually much harder to make an early diagnosis.

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

Of course our REM outcomes should be considered promising but preliminary. In fact, it will need to be tested and validated in completely independent data sets from other institutions, in order to confirm its potential risk stratification role in EC and its spread in clinical practice.

Conflict of interest

The authors have declared no conflicts of interest

No Acknowledgements
References


(9) Valentini V, van Stiphout RG, Lammering G, Gambacorta MA, Barba MC, Bebenek M et al. Nomograms for predicting local recurrence, distant metastases, and overall survival for patients
with locally advanced rectal cancer on the basis of European randomized clinical trials. J Clin Oncol. 2011;29:3163-72

(10) www.acog.org/~/media/for%20patients/faq095.ashx


Table 1: The clinical characteristics of the patients enrolled

<table>
<thead>
<tr>
<th></th>
<th>TRAINING SET</th>
<th>VERIFICATION SET</th>
<th>OVERALL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MALIGNANT</td>
<td>BENIGN</td>
<td>MALIGNANT</td>
</tr>
<tr>
<td>n</td>
<td>60</td>
<td>391</td>
<td>28</td>
</tr>
<tr>
<td>AGE MEDIAN (range)</td>
<td>55 (45-75)</td>
<td>48 (48-69)</td>
<td>53 (47-78)</td>
</tr>
<tr>
<td>CA125 (U/ML) MEAN±SD</td>
<td>33.7±4.5</td>
<td>33.3±7.3</td>
<td>43.5±5.1</td>
</tr>
<tr>
<td>HE4 (pMOL/L) MEAN±SD</td>
<td>119.4±21.3</td>
<td>35.6±30.5</td>
<td>150±23.7</td>
</tr>
<tr>
<td>ENDOMETRIAL THICKNESS (mm) MEAN±SD</td>
<td>18.3±6.8</td>
<td>8.4±7.1</td>
<td>16.3±7.2</td>
</tr>
<tr>
<td>FIGO STAGE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>46 (77%)</td>
<td>-</td>
<td>23 (82%)</td>
</tr>
<tr>
<td>II</td>
<td>7 (12%)</td>
<td>-</td>
<td>2 (7%)</td>
</tr>
<tr>
<td>III</td>
<td>5 (8%)</td>
<td>-</td>
<td>2 (7%)</td>
</tr>
<tr>
<td>IV</td>
<td>2 (3%)</td>
<td>-</td>
<td>1 (4%)</td>
</tr>
<tr>
<td>HISTOLOGY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENDOMETRIOID ADENOCARCINOMA</td>
<td>58 (97%)</td>
<td>-</td>
<td>27 (96%)</td>
</tr>
<tr>
<td>NON-ENDOMETRIOID CARCINOMA</td>
<td>2 (3%)</td>
<td>-</td>
<td>1 (4%)</td>
</tr>
<tr>
<td>ENDOMETRIAL POLYP</td>
<td>-</td>
<td>184 (47%)</td>
<td>-</td>
</tr>
<tr>
<td>MYOMA</td>
<td>-</td>
<td>39 (10%)</td>
<td>-</td>
</tr>
<tr>
<td>ENDOMETRIUM WITHOUT ATYPIA</td>
<td>-</td>
<td>168 (43%)</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 2: REM accuracy in training, verification and overall set

<table>
<thead>
<tr>
<th></th>
<th>TRAINING</th>
<th>VERIFICATION</th>
<th>OVERALL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MALIGNANT</td>
<td>BENIGN</td>
<td>MALIGNANT</td>
</tr>
<tr>
<td>HIGH RISK</td>
<td>56</td>
<td>11</td>
<td>25</td>
</tr>
<tr>
<td>LOW RISK</td>
<td>4</td>
<td>380</td>
<td>3</td>
</tr>
<tr>
<td>SENSITIVITY</td>
<td>93.3%</td>
<td></td>
<td>89.3%</td>
</tr>
<tr>
<td>SPECIFICITY</td>
<td>97.1%</td>
<td></td>
<td>95.4%</td>
</tr>
<tr>
<td>PPV</td>
<td>0.83</td>
<td></td>
<td>0.73</td>
</tr>
<tr>
<td>NPP</td>
<td>0.98</td>
<td></td>
<td>0.98</td>
</tr>
</tbody>
</table>


Figure Legends

Figure 1: Consort trial flow diagram for patients who were accrued into the study.

Figure 2: The nomogram to estimate the percentage (%) of risk of endometrial cancer.

Tutorial: starting from the left, draw a line intercepting all the data you have available for the patient. Once you reach a reference line stop drawing and restart from it. Find the value of the Risk of Endometrial Malignancy (REM) on the right scale. The given example is for a patient having the following features: presence of symptom (bleeding), $HE4 = 70$ pmol/L, endometrial thickness = 14 mm, age = 45 years old. The calculated REM is 74%.

Figure 3: ROC curve of training and verification dataset.
FIGURE 1: CONSORT TRIAL FLOW DIAGRAM FOR PATIENTS WHO WERE ACCRUED INTO THE STUDY

PATIENTS SUITABLE FOR HYSTEROSCOPY
N=741

EXCLUDED: N=37
ASHERMAN'S SYNDROME OR CERVICAL STENOSIS
EXCLUDED: N=29
ENDOMETRIAL HYPERPLASIA OR ENDOCERVICAL CANCE

ASSESSED FOR ANALYSIS
N=675

TRAINING SET
N=451
MALIGN
N=60
BENIGN
N=391

VERIFICATION SET
N=224
MALIGN
N=28
BENIGN
N=196
FIGURE 2: THE NOMOGRAM TO ESTIMATE THE PERCENTAGE (%) OF RISK OF ENDOMETRIAL CANCER.
Figure 3: ROC curve of training and verification dataset.
Clinical Cancer Research

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

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.