A REGRESSION TREE APPROACH TO PREDICT EXTRAUTERINE DISEASE IN CLINICALLY STAGE I ENDOMETRIAL CARCINOMA PATIENTS

Miscioscia RA[1], Ravaggi A[1], Bignotti E[1], Zanotti L[1], Vezzoli M[2], Sartori E[1], Odicino FE[1]

1. A. Nocivelli Inst. of Molecular Medicine, Division of Gynecologic Oncology, University of Brescia;
2. Department of Molecular and Translational Medicine, University of Brescia, Italy.

INTRODUCTION

The standard care of endometrial carcinoma (EC) at an early stage consists of extrafascial hysterectomy and bilateral salpingo-oophorectomy, however the optimal surgical management for the advanced stage type I EC is not yet univocally defined (1). Correctly assessing the preoperative stage of EC guarantees a tailored surgical treatment.

AIM

The aim of this study is to predict EC extrauterine diffusion before surgery through a new statistical approach that examines preoperative CA125 and HE4 serum levels and clinico-pathological variables .

MATERIAL & METHODS

CA125 and HE4 levels were analyzed on 293 preoperative serum samples from EC patients diagnosed at the Obstetric & Gynecology Department, University of Brescia, between January 2003 and April 2015. Biomarker concentrations were measured using the CMIA assay on the fully automated Architect instrument (Abbott Diagnostics). Patients who underwent preoperative neoadiuvant therapy or with past cancer history were excluded from the study. Using a Random Forest approach (2), relative variable importance was assessed in order to understand which covariates have a major impact on the FIGO stage (namely response variable Y). Moreover, Regression Trees (3) were grown in order to classify patients in homogeneous groups with respect to Y. In doing so, a step-by-step procedure was applied: (i) 1000 bootstrap samples were obtained from the dataset; (ii) 1000 trees were grown on each bootstrap sample; (iii) amongst the 1000 trees, the best one, in terms of Area Under the Curve (AUC) was selected. The result obtained and the related performance measures [AUC, Specificity, Sensitivity, Accuracy, Positive Predictive Value (PPV), Negative Predictive Value (NPV) computed in correspondence of the Youden's Index] were compared to the performance of the (a) Logistic regression analysis, (b) cross-validated Regression Tree, and (c) single biomarkers (HE4 and CA125). All the analysis where performed with R 3.1.2.

RESULTS

Figure 1: Best Regression Tree considering the entire EC cohort (N° 293)
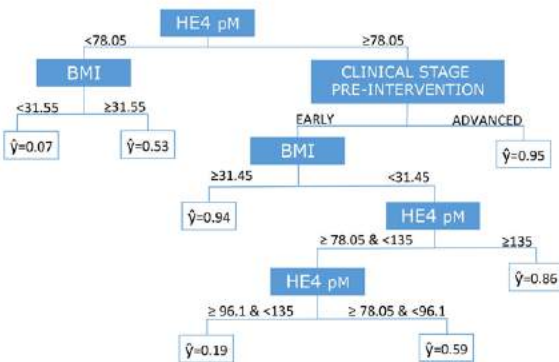
In each terminal node of the Regression Tree, the prediction corresponding to the mean value of the dependent variable (Surgical FIGO stage) for the patients clustered within the same final node is represented. Since the response variable is dummy, it ranges in 0-1 (0=FIGO stage I, or 1=FIGO stage>I). Low values of $\hat{y}$ can be interpreted as low probability of having an advanced FIGO stage (>I), and high values as high probability of having FIGO stage>I.

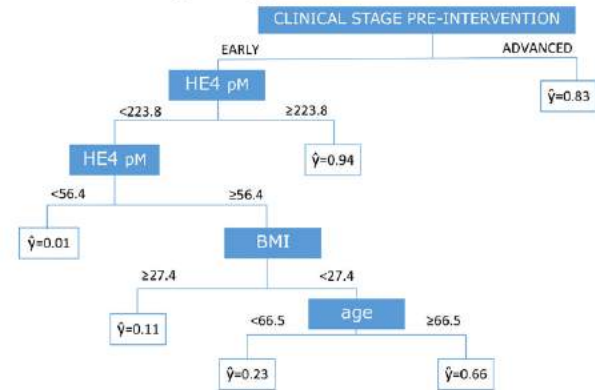Figure 2: Best Regression Tree considering endometrioid histotype only (N° 254)

Table 1: Metrics for evaluating the performance of the proposed methods considering the entire EC cohort

Metrics	Clinical stage	CA125 (U/ml)	HE4 (pM)	Best Regression Tree extracted from 1000 trees with CV	Logistic regression	Single Regression Tree with CV
AUC	-	0.66	0.74	**0.88**	0.52	0.74
Threshold (Youden Index)	-	33.25	81.80	**0.28**	0.28	0.64
Specificity	0.95	0.82	0.66	**0.77**	0.66	0.90
Sensitivity	0.36	0.44	0.69	**0.90**	0.50	0.54
Accuracy	0.75	0.69	0.67	**0.81**	0.61	0.78
PPV	0.78	0.56	0.51	**0.66**	0.41	0.74
NPV	0.74	0.74	0.81	**0.94**	0.73	0.79

Table 2: Metrics for evaluating the performance of the proposed methods considering endometrioid histotype only

Metrics	Clinical stage	CA125 (U/ml)	HE4 (pM)	Best Regression Tree extracted from 1000 trees with CV	Logistic regression	Single Regression Tree with CV
AUC	-	0.61	0.76	**0.87**	0.55	0.77
Threshold (Youden Index)	-	36.05	81.55	**0.45**	0.24	0.70
Specificity	0.95	0.87	0.65	**0.89**	0.36	0.96
Sensitivity	0.32	0.41	0.74	**0.77**	0.77	0.43
Accuracy	0.76	0.72	0.68	**0.85**	0.50	0.78
PPV	0.73	0.62	0.52	**0.78**	0.38	0.84
NPV	0.77	0.74	0.83	**0.88**	0.75	0.77

SUMMARY / CONCLUSION

Analyzing the entire cohort of 293 EC patients, the best cross-validated regression tree (built on the most statistically significant variables: HE4, CA125, BMI, and clinical presurgical stage) provided an AUC=0.88. With the threshold corresponding to Youden's Index, the ability of discrimination between early and advanced stage EC demonstrated a specificity=77%, a sensitivity =90% and an accuracy=81% (Fig.1 and Tab.1). Interestingly, considering uniquely the endometrioid histotypes (N°254), despite an identical AUC, setting the threshold in corrispondence to Youden's Index, the best regression tree showed an improved specificity= 89% and accuracy=85% (Fig.2 and Tab.2).

In conclusion, we developed a strong statistical model potentially able of preoperatively predict the presence of extrauterine disease in EC patients. This offers us a valid decisional process regarding the therapeutic options to be performed, since it better identifies the true early stage patients.

Spedali Civili Brescia

Angelo Nocivelli Institute of Molecular Medicine

ACKNOWLEDGEMENTS

REFERENCES

1. Benedetti Panici P, Basile S, Maneschi F. et al., Systematic pelvic lymphadenectomy vs. no lymphadenectomyin early-stage endometrial carcinoma: Natl Cancer Inst. 2008 Dec 3;100(23):1707-16. randomizedclinical trial. J

2. Breiman, L. (2001). Random forests. Machine Learning, 45(1), pp. 5–32.

3. Breiman L, Friedman J, Olshen R, Stone C (1984). *Classification and Regression Trees*. Wadsworth Inc., California

Contact information

marika.vezzoli@unibs.it
antonella.ravaggi@unibs.it
franco.odicino@unibs.it