Prediction of disease-free and overall survival in endometrial cancer using linear regressions and machine learning algorithms

Salvatore **Insinga** [1,*], Vito **Chiantera** [1], Francesco **Piva** [2], Luca **Giannella** [3], Andrea **Ciavattini** [3]

[1] Obstetrics and Gynecology Clinic, Policlinico P. Giaccone, Palermo, Italy.
[2] Odontostomatologic and Specialized Clinical Sciences, Università delle Marche, Ancona, Italy.
[3] Obstetrics and Gynecology Clinic, Università delle Marche, Ancona, Italy.

DOI: 10.36129/jog.2023.S05

Objective. To assess the value of prediction models in forecasting disease recurrence and death by analysing patient characteristics, imaging, histological findings, surgery, and disease stage in endometrial cancer (EC).

Materials and Methods. Retrospective study including women with EC and 3-5 years of follow-up, between January 2016 and December 2020. Overall and disease-free survival were the primary outcomes. Linear and quadratic regression analysis were used, followed by machine learning algorithms. The quality of the regressions was assessed based on adjusted R squared (R2) and root mean squared error (RMSE). Model I measured the recurrence rate, while Model II assessed the mortality rate and overall survival in patients with endometrial recurrence.

Results. The final analysis included 205 women with EC. In both prediction models, R2 and RMSE did not improve by increasing the number of regression variables. The best prediction of the measured outcomes was obtained with five variables (DFS - R2: 0.54; RMSE: 20.3; OS - R2: 0.83; RMSE 9.9). The variables most associated with DFS were "age", "personal history of cancer" and "residual disease after surgery". The OS model showed a significant association with the following variables: "preoperative imaging: tumour size in cm", "preoperative FIGO staging", "preoperative imaging: number of enlarged iliac LNs", "Lymphovascular space invasion" and in particular with "site of recurrence".

Conclusions. The studied variables were weakly correlated with DFS, while the analysed variables strongly influenced the OS. The latter predictive model could discriminate the timing of death in women with relapse with a fair approximation.