

Adjusting prediction models for ovarian tumor classification to new clinical settings

L. De Clercq¹, B. Van Calster¹, C. Van Holsbeke^{2,12}, A. Czekierdowski³, S. Guerriero⁴, L. Savelli⁵, D. Franchi⁶, D. Fischerova⁷, J. Zhang⁸, E. Epstein⁹, L.I. Valentin¹⁰, A. Testa¹¹, S. Van Huffel¹, D. Timmerman¹²

¹ESAT-SCD-SISTA, Katholieke Universiteit Leuven, Heverlee, Belgium; ²Department of Obstetrics and Gynaecology, Ziekenhuis Oost-Limburg, Genk, Belgium; ³1st Department of Gynecology, Medical University in Lublin, Lublin, Poland; ⁴Department of Obstetrics and Gynaecology, Ospedale San Giovanni di Dio, University of Cagliari, Cagliari, Italy; ⁵Reproductive Medicine Unit, Department of Obstetrics and Gynecology, University of Bologna, Bologna, Italy; ⁶Dt.ssa Ginecologic Oncology Unit, Division of Gynecology, IEO, Milano, Italy; ⁷Oncogynecological Center Department of Obstetrics and Gynecology, General Faculty Hospital of Charles University, Prague, Czech Republic; ⁸Ultrasound Department, Chinese PLA General Hospital, Beijing, China; ⁹Department of Obstetrics and Gynaecology, Lund University, Lund, Sweden; ¹⁰Department of Obstetrics and Gynaecology, Malmö University Hospital, Malmö, Sweden; ¹¹Istituto di Clinica Ostetrica e Ginecologica, Università Cattolica di Sacro Cuore, Roma, Italy; ¹²Department of Obstetrics and Gynaecology, University Hospitals KU Leuven, Leuven, Belgium

Objectives: An existing prediction model may suffer from errors when applied to new data, leading to wrong decisions. This can happen if the new data are collected in a center that was not involved in developing the model, or if a different data collection protocol was used. A useful solution is to adjust the model coefficients. Additionally, one may want to include an additional predictor because data is available or to align the model to one's clinical practices. In this study, these methods are illustrated on logistic regression model LR2 from phase 1 of the International Ovarian Tumor Analysis (IOTA) project.

Methods: LR2 is modified based on 699 patients from centers that did not contribute to IOTA phase 1. In the first method all model coefficients are recalculated using the new data. In the second method the coefficients are recalculated and the color Doppler score of tumoral flow is added as a new predictor. The adjusted models were evaluated by the area under the ROC curve (AUC) and by odds ratios (OR). Given that IOTA phase 1 was a large multi-center study, and that IOTA protocol was followed when collecting the new data, we did not expect that model adjustment would greatly influence AUC in our study.

Results: The original LR2 model had an AUC of 0.947 on the new data. Recalculating coefficients (method 1) gave an AUC of 0.948, and changed some ORs. For example, the OR for papillary blood flow changed from 3.2 to 14.2. Method 2 resulted in an AUC of 0.952.

Conclusions: Depending on the situation, medical centers could adjust an existing model to tailor it to their setting. This could be considered when suboptimal performance is suspected, if the value of an extra predictor is to be checked, or if there is specific interest in the ORs in the new setting. Due to the nature of the IOTA project, suboptimal performance of LR2 or another IOTA model is not very likely, conditional on protocol adherence. This is supported by our results.