



## Correspondence

**Response to letter to the editor  
concerning validation of IOTA  
ADNEX model**


Dear Editor,

We carefully read the letter by Van Calster et al. (Van Calster et al., 2016a) concerning our article (Van Calster et al., 2016a; Szubert et al., 2016). We want to thank the authors for the valuable feedback on our work, and we do appreciate it. However, some of the guidelines mentioned by the authors were recently published, hence, we could not conform to them at the time of an arrangement of our research. We had chosen a different methodology, which still remains mathematically valid. Nevertheless, we carefully considered each comment, and certainly we agree with most of them. Thus, we extended our results to gain a deeper insight into performance of the ADNEX model. We hope, this new approach will facilitate drawing reliable conclusions. We are aware that with a larger study group the results would be more valuable, and according to Van Calster et al., suggestion (Van Calster et al., 2016a), we combined the datasets from two Centers, and we performed subsequent analysis.

### 1. Study group

The combined dataset consists of 327 patients, including 104 malignant and 223 benign ovarian tumors. Comprehensive patients' characteristics are presented in Supplementary materials (Tables 1 and 2). In order to evaluate both versions of ADNEX model (with and without CA-125) on the same group of patients, we imputed 9 missing values of CA-125. This was done by the predictive mean matching in multiple imputation with 100 datasets. We used age, menopause, morphology, locules number, as cites, shadowing, number of papillary projections, lesion max diameter and solid part max diameter as predictors. Statistical evaluation was performed with use of software R version 3.3.1 (2016-06-21).

### 2. Calibration

We investigated calibration for both ADNEX models. For ADNEX with CA-125, the intercept and slope equal to  $-0.451$  ( $-0.843$ ,  $-0.065$ ) and  $0.843$  ( $0.684$ ,  $1.028$ ), respectively. This may suggest that predicted risks are on average overestimated (intercept  $<0$ ). For ADNEX without CA-125, the intercept and slope equal to  $-0.274$  ( $-0.641$ ,  $0.089$ ) and  $0.805$  ( $0.652$ ,  $0.979$ ), respectively. Such results may imply overfitting of the model (slope  $<1$ ). The aforementioned intercept and slope narrowly deviate from the ideal values. For this reason, the calibration plots should be investigated, however, according

to Van Calster et al., a graphical assessment should be done with at least 200 events and 200 nonevents (Van Calster et al., 2016b). This condition is not fulfilled within the study group; hence, we omit a detailed analysis of the plots (Fig. 1).

### 3. Results and discussion

In the current analysis, The ADNEX model with CA-125 had significantly higher AUC when compared to ADNEX without CA-125 (0.927; 95%CI 0.888–0.958 vs 0.907; 95%CI 0.868, 0.941; DeLong's test  $p = 0.009$ ; see Table 3). This may suggest that lack of CA125 evaluation leads to significant drop of the performance. AUCs for both models remain high, however, the confidence intervals are wide, so there is a need to investigate this problem in further research.

In the primary study we investigated only 10% cutoff, as it was proposed in the original report (Van Calster et al., 2014). However, ADNEX model may work differently depending on ovarian cancer prevalence, thus according to the guidelines by Van Calster et al., some clinical centers may prefer different cutoffs (Van Calster et al., 2015). For this reason, we have extended our original results and we have performed analysis of ADNEX at cutoffs from 3% to 30%. We have provided detailed results of ADNEX performance according to various cut-offs in supplementary materials (Tables 4 and 5). These are important findings of our study, showing the highest accuracy of the model at 30% threshold. This is in agreement with the guidelines by Van Calster et al. showing that for tertiary centers 30% cut-off may be more appropriate (Van Calster et al., 2015).

When tumor groups were analyzed in pairs, the achieved AUCs indicated good overall performance. In general, the results confirm good polytomous discrimination from the first study by Van Calster et al., and were comparable to later ADNEX model evaluation studies (Table 6) (Van Calster et al., 2014; Araujo et al., 2016; Meys et al., 2016; Sayasneh et al., 2016). Both polytomous discrimination indexes (PDI) were  $>0.2$ , which in this case is a threshold for a random performance. The models had problems in discrimination between some groups: borderline vs stage I, stage I vs stage II-IV, stage II-IV vs metastatic and in particular stage I vs metastatic. However, low AUCs for these pairs might be caused by the low prevalence of some malignant subclasses in the study group.

We gave a special attention to 'multiclass' performance evaluation. Van Calster et al., provided guidelines for applying ADNEX model in clinical practice. The authors "believe that the aim should not be to classify tumors into a single subgroup of malignancy" and the risk should be assessed per type of malignancy (Van Calster et al., 2015). However, in our setting of external validation we were curious to know how ADNEX models behave when they have to differentiate between five categories. This can be achieved by the assessment of relative change in risk, as it was described by the authors in the aforementioned article (Van Calster et al., 2015). We have used prevalence of the total pooled dataset described in the article by Van Calster et al. (Van Calster et al., 2014). We have investigated the performance of the ADNEX models with 10% cutoff (see Tables 7–10). Indeed, with the use of relative risk calculation, such an approach improves the performance comparing to

DOIs of original article: <http://dx.doi.org/10.1016/j.gore.2016.09.003>,  
<http://dx.doi.org/10.1016/j.ygyno.2016.06.020>.

absolute risk. However, many II-IV stage tumors are misclassified as borderline tumors. There is also problem with metastatic tumors, which are commonly misclassified as other malignant types. The latter case might be investigated with a larger study group, since in this validation there are only 10 such cases. Nevertheless, the results confirm that the ADNEX models should not be used in that manner. Thus, we speculate that the best method for evaluation of ADNEX performance would be prospective analysis of clinical decision making upon ADNEX results.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at [doi:10.1016/j.gore.2016.10.009](https://doi.org/10.1016/j.gore.2016.10.009).

#### References

- Araujo, K.G., Jales, R.M., Pereira, P.N., Yoshida, A., de Angelo, A.L., Sarian, L.O., et al., 2016. Performance of the IOTA ADNEX model in the preoperative discrimination of adnexal masses in a gynecologic oncology center. *Ultrasound Obstet. Gynecol.* <http://dx.doi.org/10.1002/uog.15963>.
- Meys, E.M.J., Jeelof, L.S., Achten, N.M.J., Slangen, B.F.M., Lambrechts, S., Kruitwagen, R.F.P.M., et al., 2016. Estimating the risk of malignancy in adnexal masses: an external validation of the ADNEX model and comparison with other frequently used ultrasound methods. *Ultrasound Obstet. Gynecol.* <http://dx.doi.org/10.1002/uog.17225>.
- Sayasneh, A., Ferrara, L., De Cock, B., Saso, S., Al-Memar, M., Johnson, S., et al., 2016. Evaluating the risk of ovarian cancer before surgery using the ADNEX model: a multicentre external validation study. *Br. J. Cancer* 115:542–548. <http://dx.doi.org/10.1038/bjc.2016.227>.
- Szubert, S., Wojtowicz, A., Moszynski, R., Zywica, P., Dyczkowski, K., Stachowiak, A., et al., 2016. External validation of the IOTA ADNEX model performed by two independent gynecologic centers. *Gynecol. Oncol.* 142:490–495. <http://dx.doi.org/10.1016/j.ygyno.2016.06.020>.
- Van Calster, B., Van Hoorde, K., Valentin, L., Testa, A.C., Fischerova, D., Van Holsbeke, C., et al., 2014. Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. *BMJ* 349:g5920. <http://dx.doi.org/10.1136/bmj.g5920>.
- Van Calster, B., Van Hoorde, K., Froyman, W., Kaijser, J., Wynants, L., Landolfo, C., et al., 2015. Practical guidance for applying the ADNEX model from the IOTA group to discriminate between different subtypes of adnexal tumors. *Facts, Views Vis ObGyn* 7, 32–41.
- Van Calster, B., Steyerberg, E., Bourne, T., Timmerman, D., Collins, G., 2016a. Flawed external validation study of the ADNEX model to diagnose ovarian cancer. *Gynecol Oncol Rep.* <http://dx.doi.org/10.1016/j.gore.2016.09.003>.
- Van Calster, B., Nieboer, D., Vergouwe, Y., De Cock, B., Pencina, M.J., Steyerberg, E.W., 2016b. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J. Clin. Epidemiol.* 74:167–176. <http://dx.doi.org/10.1016/j.jclinepi.2015.12.005>.

Sebastian Szubert

*Division of Gynecological Surgery, Poznan University of Medical Sciences,  
33 Polna St., 60-535 Poznan, Poland*

Corresponding author at: *Division of Gynecological Surgery, 33. Polna  
St., 60-535 Poznan, Poland.*

*E-mail address: szuberts@o2.pl.*

Andrzej Wojtowicz

Patryk Zywica

*Faculty of Mathematics and Computer Science, Adam Mickiewicz  
University, 87 Umultowska St., 61-614 Poznan, Poland*

27 October 2016