Response to ‘Comment on external multicentre validation of a nomogram predicting the risk of relapse in patients with borderline ovarian tumours’

S Bendifallah*,1,2 and E Darai1,3

1Department of Obstetrics and Gynaecology, Hôpital Tenon, Assistance Publique des Hôpitaux de Paris, Université Pierre et Marie Curie, Paris 6, Institut Universitaire de Cancérologie, Paris, France; 2INSERM UMR_S 707, ‘Epidemiology, Information Systems, Modeling’, University Pierre and Marie Curie, Paris, France and 3INSERM UMR_S 938, Université Pierre et Marie Curie, Paris, France

Sir,

We would like to thank Dr Obermair (Obermair, 2014) for the comments regarding our recent article (Bendifallah et al, 2013). Although the postoperative Obermair’s nomogram (Obermair et al, 2013) was based on the common evidence-based high-risk factors and constitute a valuable contribution for improving health care for women with borderline ovarian tumours (BOT), we demonstrated that the tool showed limitations in its generalisability to a new and independent French population.

Theoretically, the published nomogram offers the advantage of condensing the high heterogeneity of the disease into a simple and easily interpretable format to guide the decision-making process towards the most adapted treatment options or follow-up strategies.

The comments of Dr Obermair suggest that the proper question to ask is how to study the generalisability, clinical utility and level of complexity of the published tool. As previously reported, we were unable to confirm the validity of the nomogram due to differences in the epidemiological and surgical characteristics and histological patterns between the French cohort and the Australian series. We highlighted that the relatively low incidence of patients with stage II-IV in the Obermair et al cohort is a potential cause of underestimating the relapse rate, and therefore reciprocally a potential cause of overestimating that rate into the French cohort (Bendifallah et al, 2013).

Secondarily, we also underlined that the low rate of BOT stage I in our cohort (45% versus 80%) in contrast to the prevalence of classical BOTs could be explained by the fact that the two institutions that participated in the study are reference centres (Bendifallah et al, 2013). In comparison, both samples were profoundly different with regard to patients’ characteristics and outcomes. Nevertheless, this fact does not represent a limitation to validate the published nomogram. The French cohort was representative of all BOT cases treated at the two reference centres, which represents an illustration of the real practice scenario.

The predictive accuracy studied with our external validation set represents the gold standard technique. Indeed that external validation aims to address the accuracy of a model in patients from a different but plausibly related population, which may be defined as a selected study population representing the underlying disease domain (Iasonos et al, 2008).

The French physicians should ensure that the model is applicable both in terms of clinical relevancy and statistical accuracy before using it as a guide in the decision-making process. To achieve this level of evidence, the model should predict accurately which patients will and will not reach the end point (discrimination), demonstrate maximal correlation between actual and predicted values (calibration), should be accurate consistently when applied to different data sets (validation), be easy to use (level of complexity) and applicable to heterogeneous novel populations with the same accuracy (generalisability) (Iasonos et al, 2008).

To conclude, our intention is to promote individualised predictive approach with evidence-based results of its relevancy.