

Comment on 'External multicentre validation of a nomogram predicting the risk of relapse in patients with borderline ovarian tumours'

A Obermair^{*,1}

¹Research Gynaecological Oncology, Queensland Centre for Gynaecological Cancer, Royal Brisbane & Women's Hospital, 6th Floor Ned Hanlon Building, Brisbane, Queensland 4029, Australia

Sir,

We read with interest the paper by Bendifallah *et al*, 2013 on the validation of our previous nomogram (Obermair *et al*, 2013) to estimate the risk of recurrence after surgery for Borderline Ovarian Tumours (BOT) and we congratulate the authors on their work.

A diagnosis of BOT is typically established only postoperatively and many women diagnosed with BOT are in their childbearing years. Although most BOT patients will expect excellent outcomes, a small proportion of women will recur. Prediction of relapse is critical. Patients with remotely low risk of relapse can be discharged from regular follow-up. By contrast, patients at high risk of relapse may benefit from extended surgery or regular, lifelong follow-up because recurrences may develop late after surgery (Silva *et al*, 2006).

Our nomogram was the first attempt to quantify a patient's individual risk of relapse and included covariates from readily available clinical, biological and pathological characteristics. We made every attempt to create a representative sample and therefore included all consecutive patients from six gynaecological cancer centres. Hence, almost 80% of patients in our group were classified stage I.

The French group of clinicians abstracted information from 314 patients from two French institutions between 1980 and 2008. To validate our nomogram, they repeated our study using identical covariates. However, their patient sample was distinctly different to ours. Stage 2, 3 and 4 was almost five times as common in the French study than in ours. The pre-operative median serum CA125 was more than double as high in the French paper than in ours (77.6 U ml⁻¹ vs 36 U ml⁻¹). Expectedly, relapses developed in 5.5% vs 29.9% (Bendifallah *et al*, 2013).

After discussions with the French authors, it became clear that pathologists at those two French institutions regularly review high-risk BOT cases referred from other institutions. It seems that those cases were included in the reporting of this series, thus resulting in a very significant over-representation of high-risk cases.

While the Australian series reported the outcomes of a representative sample of all BOT, the French cohort was not representative of all BOT cases but provided an over-representation of high-risk patients.

Although both samples overlap to a degree, the French cohort was not a comparable patient cohort and therefore was not suited to validate the Australian cohort. Both samples were profoundly different in regards to patients' characteristics and outcomes. A comparison of those two samples should have excluded samples from external review to level the field.

REFERENCES

- Bendifallah S, Uzan C, Fauvet R, Morice P, Darai E (2013) External multicentre validation of a nomogram predicting the risk of relapse in patients with borderline ovarian tumours. *Br J Cancer* **109**: 2774–2777.
- Obermair A, Tang A, Kondalsamy-Chennakesavan S, Ngan H, Zusterzeel P, Quinn M, Carter J, Leung Y, Janda M (2013) Nomogram to predict the probability of relapse in patients diagnosed with borderline ovarian tumors. *Int J Gynecol Cancer* **23**: 264–267.
- Silva EG, Gershenson DM, Malpica A, Deavers M (2006) The recurrence and the overall survival rates of ovarian serous borderline neoplasms with noninvasive implants is time dependent. *Am J Surg Pathol* **30**: 1367–1371.

*Correspondence: Professor A Obermair; Email: Obermair@powerup.com.au

Published online 21 January 2014

© 2014 Cancer Research UK. All rights reserved 0007–0920/14



<http://creativecommons.org/licenses/by-nc-sa/3.0/>



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 Darai^{1,3}

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.