FULL PAPER

BJC

British Journal of Cancer (2013) 109, 2774–2777 | doi: 10.1038/bjc.2013.678

Keywords: borderline ovarian tumours; nomogram; relapse risk; validation

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

S Bendifallah*[,1,2], C Uzan[3,4], R Fauvet[5,6], P Morice[3,4] and E Darai[1,7]

[1]*Department of Obstetrics and Gynaecology, Hôpital Tenon, Paris, France;* [2]*UMR S 707, Epidemiology, Information Systems, Modeling, Université Pierre et Marie Curie, Paris, France;* [3]*Department of Gynaecological Surgery, Institut Gustave Roussy, Villejuif, France;* [4]*INSERM U 10-30, Institut Gustave Roussy, Villejuif, France;* [5]*Department of Obstetrics and Gynaecology, CHU Amiens, Amiens, France;* [6]*INSERM ERI-12, Université de Picardie Jules Vernes, Amiens, France and* [7]*INSERM UMR S 938, Université Pierre et Marie Curie, Paris, France*

Background: The Obermair nomogram was recently developed to predict the risk of relapse in patients with borderline ovarian tumours (BOTs) based on five readily available clinical, biological, and pathological characteristics. We set out to externally validate and assess its robustness using a multi-institutional BOT database.

Methods: All consecutive patients treated for BOTs in the two participating centres between January 1980 and December 2008 and who had all the nomogram variables documented were identified for analysis.

Results: Three hundred and fourteen eligible patients were identified and used for external validation analysis. The median follow-up and initial relapse time were 46.43 (range: 0.1–360) and 66.64 (range: 8–77) months, respectively. The nomogram concordance index was 0.54 (95% CI, 0.52–0.56). The correspondence between the actual relapse and the nomogram predictions suggests a limited calibration of the nomogram in the validation cohort.

Conclusion: This external validation study of the Obermair nomogram showed limitations in its generalisability to a new and independent patient population.

Borderline ovarian tumours (BOTs), which have been recognised as an intermediate entity between benign and invasive tumours, account for 10–15% of epithelial ovarian tumours (Hart, 2005). Borderline ovarian tumour is characterised clinically by better overall survival with 5- and 10-year survival rates for stage I, II, and III disease of 99% and 97%, 98% and 90%, and 96% and 88%, respectively (Trimble *et al*, 2002). Despite this favourable prognosis, up to 25% of patients relapse or succumb to disease (Morice *et al*, 2012). Therefore, relapse prediction after primary surgical treatment is a cornerstone of patient management. In particular, predicting individualised outcome based on prognostic factors may help to identify patients at risk and hence to decide on the most adapted treatment options, which follow-up strategies to adopt and how to best counsel the patient. The past decade has been marked by several important advances in therapeutic options such as the advent of fertility-sparing surgery, ovarian cryopreservation, and the routine use of laparoscopy (Fauvet *et al*, 2005; du Bois *et al*, 2013). More recently, as for most types of cancer, a complementary approach based on prediction models has been developed (Bendifallah *et al*, 2012; Isariyawongse and Kattan, 2012). Cancer researchers, clinicians, and patients are increasingly interested in nomograms which are defined as a graphical representation of a statistical model to predict a particular end point according to the individual characteristics of a patient (Isariyawongse and Kattan, 2012). By providing predictions that are both evidence-based and individualised, these tools may

*Correspondence: Dr S Bendifallah; E-mail: sofiane.bendifallah@yahoo.fr

Received 2 August 2013; revised 15 September 2013; accepted 1 October 2013; published online 29 October 2013

improve medical management and guide the decision-making process (Isariyawongse and Kattan, 2012). Obermair *et al* (2013) recently presented a nomogram to predict the probability of relapse in individual patients who underwent surgery for BOT. The published nomogram included age at diagnosis, FIGO stage, CA125, the conservative surgical procedure, and the histologic subtype. Although the tool was internally validated, external validation on an independent set of women is required to ensure applicability to patients from different institutions. The aim of this study was to externally validate this recently introduced nomogram in a population of women with BOTs using large databases of academic cancer centres.

MATERIALS AND METHODS

Patients. Data of all women with BOTs, who had received primary surgical treatment between January 1980 and December 2008, were abstracted from two institutions with prospectively maintained databases (Tenon Hospital and Institut Gustave Roussy). Extracted data included patient demographics, histologic type, FIGO stage, details on treatment and relapse, and follow-up and survival data. To be included for validation analysis, the women had to have all the nomogram variables documented. Women were treated with upfront surgery according to international guidelines. Surgical treatment was considered conservative when one ovary and the uterus were respected. Conservative ovarian treatment consisted of unilateral cystectomy, unilateral salpingo-oophorectomy, unilateral salpingo-oophorectomy plus contralateral cystectomy, or bilateral cystectomy. Surgical treatment was considered non-conservative when bilateral salpingo-oophorectomy was performed. Histologic staging was performed according to the WHO classification system on the basis of the final evaluation of the pathological specimen (Tavassoli and Devilee, 2003). A central pathology review was performed for all tumours. No adjuvant therapy was performed except when invasive implants were detected according to international guidelines. All women were followed in the institutions' outpatient department. The protocol was approved by the Ethics Committee of the 'Collège National des Gynécologues et Obstétriciens Français (CNGOF)'.

Outcomes. As in Obermair *et al* (2013), disease-free survival was calculated from the date of surgery to either the last follow-up or the date of relapse to compare the nomogram predictions with the actual outcome. Relapse (invasive and BOT) disease was diagnosed by pathology or imaging studies.

Validation. The discrimination and calibration accuracy of the nomogram were assessed (Hanley and McNeil, 1982). Discrimination is the ability to differentiate between women with relapse and those without. It is measured using the receiver operating characteristic curve and summarised by the area under the curve (AUC). An AUC of 1.0 indicates perfect concordance, whereas an AUC of 0.5 indicates no relationship. Calibration is the agreement between the frequency of observed outcome and the predicted probabilities and was studied using graphical representations of the relationship between the two calibration curves. In addition, women were clustered according to their FIGO stage to evaluate the nomogram performance within each risk group.

Other statistical tests. The categorical variables were analysed using the χ^2 squared-test. Differences were considered significant at a level of $P<0.05$. All analyses were performed using the R software with the rms PresenceAbsence packages (http://lib.stat.cmu.edu/R/CRAN).

RESULTS

During the study period, 314 women were documented as having both received primary surgical treatment for BOTs and having all the nomogram variables documented. The demographics and clinical characteristics of both cohorts (Obermair *et al* (2013) and validation) are outlined in Table 1. Both cohorts were significantly different with a significantly higher rate of women with advanced FIGO stage tumour and a higher rate of serous BOT in the validation cohort. Women in the validation cohort were also slightly younger at the time of surgery. The median follow-up and initial relapse times were 46.43 (range: 0.1–360) and 66.64 (range: 8–77) months, respectively. Among the 314 patients included in these validation study, 13.7% (43 out of 314) received adjuvant therapy due to the presence of invasive implants at the time of surgery.

At the time of the last follow-up, the overall relapse rate was 29.9% (94 out of 314) with non-invasive and invasive rates of 25.2% (79 out of 314) and 4.7% (15 out of 314), respectively. Recurrences were diagnosed in 26.4% (23 out of 87) and 31.3% (71 out of 227), in the conservative and non-conservative treatment groups, respectively. Median follow-up time were 45.43 (range: 0–230) and 47.83 (range: 0–310) in those subgroups, respectively. Among the 87 patients who underwent conservative surgery, 32% (28 out of 87), 53% (46 out of 87), and 15% (13 out of 87) underwent unilateral or bilateral cystectomy, unilateral salpingo-oophorectomy, unilateral salpingo-oophorectomy plus contralateral cystectomy, respectively. Among the 114 patients who presented a mucinous BOT, respectively, 42% (48 out of 114)

Table 1. Patients' characteristics of the Obermair *et al* ($N=801$) and the validation ($N=314$) cohorts

	Obermair cohort *N* %	Validation cohort *N* %	
Parameters	***N*** **= 801**	***N*** **= 314**	***P*-value**
Age at diagnosis: mean years	49.1 (s.d. 16.1)	37.23 range (14–84)	–
FIGO stage			
I II–IV Stage not known	629 (78.5) 89 (11.1) 83 (10.4)	142 (45) 161 (51) 11 (4)	<0.001
Conservative surgery			
Yes No	198 (24.7) 603 (75.3)	87 (28) 227 (72)	0.30
Histologic subtype			
Serous Mucinous Others	334 (41.7) 443 (55.3) 24 (3.0)	193 (61,5) 114 (36,3) 7 (2,2)	<0.001
CA125 median (IQR)	36 (17.9–92)	77.62 (94.00)	–
Outcomes			
Total relapse Invasive relapse BOT relapse Deaths (from any cause)	44 (5.5%) – – 42 (5.0)	94 (29.9) 15/314(4.7) 79/314 (25.1) 13 (4.1)	–

Abbreviations: BOT = borderline ovarian tumour; FIGO = International Federation of Gynecology and Obstetrics; IQR = interquartile range; s.d. = standard deviation.

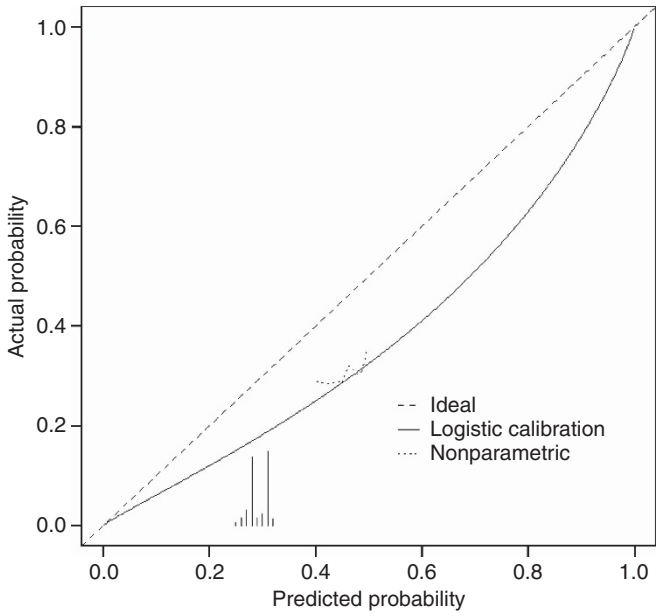

Figure 1. Calibration plot of the Obermair *et al* nomogram for the entire cohort of 314 patients.

and 58% (66 out of 114) underwent conservative and non-conservative surgery. Overall survival at 5 years was 95.3% (95% confidence interval, 93.2–97.4%).

Validation. The nomogram concordance probability was 0.54 (95% CI, 0.52–0.56) in the validation cohort compared with 0.668 in the Obermair *et al* (2013) cohort. The calibration plot (Figure 1) illustrates how the prediction of the nomogram compares with the actual outcomes. The mean error rate in the whole population was 16%. The accuracy of the nomogram within each FIGO stage shows poor discriminative and calibration abilities.

DISCUSSION

This external multicentre validation study of the Obermair nomogram failed to confirm good generalisability to predict risk of recurrence in an independent population of women with BOTs. Identification of patients with a high risk of relapse is a major goal for a physician faced with a patient with BOTs particularly if the patient wishes to preserve her child-bearing potential. During the last decade, changes in the histologic diagnostic criteria for BOTs have been observed (Morice *et al*, 2012). Previously, only prognostic factors based on histology and less frequently on tumour markers were taken into account. However, a recent review of the literature on BOTs underlined that among classic histologic criteria for serous BOTs – including micropapillary patterns, microinvasion, and the presence of peritoneal implants – the sole impacting factor on the relapse rate was the presence of peritoneal implants. For mucinous BOTs, despite a relatively higher risk of relapse with the invasive form compared with serous BOTs, no prognostic histologic factors of relapse – including intraepithelial carcinoma and microinvasion – have been identified. This highlighted the need to find another model to predict relapse in patients with BOTs (Obermair *et al*, 2013). The use of predictive tools such as nomograms meant that optimising management strategies became a realistic objective for BOT patients. The Obermair tool, based on evidence-based risk factors of relapse, thus constitutes a valuable contribution for improving healthcare for women with BOTs (Obermair *et al*, 2013). By combining commonly available parameters, it offers the advantage of condensing the high heterogeneity of the disease into a simple and easily interpretable format. However, we were unable to confirm the validity of the nomogram in our study. These findings may be explained by differences in the epidemiological and surgical characteristics and histologic pattern of the two populations. First, the relatively low incidence of patients with stage II–IV in the Obermair *et al* (2013) cohort is a potential cause of under-estimating the relapse rate. Indeed, Morice *et al* (2012) underlined that the main prognostic factor of relapse was the presence of peritoneal implants (i.e. advanced stages of BOTs) and that their presence was correlated with the presence of micropapillary patterns in patients with serous BOTs. 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 which participated in the study are reference centres.

Second, the fusion of mucinous with other BOT histologic types is a confounding factor to predict relapse rates. Indeed, it has been demonstrated that mucinous BOTs have a higher rate of relapse into the invasive form compared with endometrioid, Brenner or clear-cell BOTs for which only three cases have been reported (Uzan *et al*, 2011; Morice *et al*, 2012). Third, the nomogram did not distinguish between patients with stage IA and IB BOTs. This point is crucial as it has been demonstrated that serous BOTs are more frequently bilateral (Camatte *et al*, 2004). In this specific setting, a prospective randomized trial on the surgical management of serous BOTs in women wishing to became pregnant (Palomba *et al*, 2010) showed that ultra-conservative treatment based on bilateral cystectomy compared with conservative treatment of unilateral salpingo-oophorectomy on the larger BOTs associated with contralateral cystectomy was associated with a significantly higher pregnancy rate but conveyed a higher incidence of relapse. Finally, no difference between relapse of BOTs or into the invasive form was made in the Obermair nomogram although it has been demonstrated that a second conservative operation following BOT relapse is associated with a high overall survival. It could be reasonably argued that a potentially heterogeneous predictive model designed to predict relapse events based on a data set from 23 years ago derived from a heterogeneous (Asian, Australian, and European) population might not be able to accurately predict the end point, especially if the quality of pathology, the type of conservative surgery (cystectomy versus adnexectomy), the comprehensive initial or secondary staging, and the quality of surgery (residual tumour or not) is not taken into account. Such differences may in turn affect the applicability of the nomogram to our patients and in fine its generalisability. To provide a better evaluation of the true accuracy of the model, we tested the calibration ability upon the whole population and for each FIGO stage and our results suggest a low predictive ability of this tool (Table 2) despite subset analysis. According to the results of du Bois *et al* (2013), we hypothesised that risk grouping and more accurate identification of individual prediction may be enriched by tumour-related (i.e. implants/micropapillary patterns or with stromal microinvasion), surgery-related (i.e. residual tumour, staging quality, and nodal involvement), and treatment-related prognostic factors.

Some limitations of the present study have to be underlined. First, as for the Obermair *et al* (2013), the retrospective nature of our study cannot exclude bias. Second, during the data collection period, modifications in staging modalities and surgical techniques (surgical sparing surgery or staging) occurred. However, the rate of conservative treatment of BOTs was similar in the two populations. Third, a long follow-up period is required to evaluate the recurrence rate for BOTs. Silva *et al* (2006) have reported the impact of the long follow-up in the true recurrence rate and have underlined both the necessity to follow for a minimum of

Table 2. Nomogram performance according to FIGO stages

	Patient number	Predicted probability mean (%)	Observed probability mean (%)	Error between predicted and observed probability mean (%)	*P*-value of calibration curve	AUC	Error maximal (%)	Error average (%)
Whole population	314	13	29	16	10^{-5}	0.54	19	16
FIGO stage I	142	4	30	26	10^{-5}	0.61	100	20.8
FIGO stage II	40	20	35	15	10^{-5}	0.57	42	20.8
FIGO stage III	121	22	24	4	10^{-5}	0.52	100	30

Abbreviation: AUC = area under the curve.

10 years to evaluate recurrences and for 20 years to evaluate for survival. Both studies with, respectively, 57.5 (range: 0–60) and 46.43 (range: 0.1–360) months of median follow-up seems to be limited. The true recurrence rate may have been underestimated. Finally, as previously mentioned, the inclusion of patients over a long period implies that changes in the histologic criteria potentially overestimated the risk of diagnosing BOTs compared with the strict criteria proposed by the NCI (National Cancer Institute) and NIH (National Institutes of Health) workshop in 2004 (Bell *et al*, 2004; Seidman *et al*, 2004; Silverberg *et al*, 2004).

In conclusion, although nomogram may improve medical management and guide the decision-making process towards the most adapted treatment options or follow-up strategies, concerns to use it routinely have been raised especially with regard to the clinical relevance of that tool. The Obermair nomogram (Obermair *et al*, 2013) needs to be improved before being used to identify eligible women for clinical trials or guiding the physician in decisions about post-treatment follow-up. Other external validation based on populations with better differentiated stage I disease as well as histologic and surgical procedures are essential to complete this work. Further studies designed to evaluate such prediction tools are needed in the BOT area.

REFERENCES

Bell DA, Longacre TA, Prat J, Kohn EC, Soslow RA, Ellenson LH, Malpica A, Stoler MH, Kurman RJ (2004) Serous borderline (low malignant potential, atypical proliferative) ovarian tumors: workshop perspectives. *Hum Pathol* **35**: 934–948.

Bendifallah S, Genin AS, Naoura I, Chabbert Buffet N, Clavel Chapelon F, Haddad B, Luton D, Darai E, Rouzier R, Koskas M (2012) A nomogram for predicting lymph node metastasis of presumed stage I and II endometrial cancer. *Am J Obstet Gynecol* **207**: 197.e1–8.

Camatte S, Morice P, Thoury A, Fourchotte V, Pautier P, Lhomme C, Duvillard P, Castaigne D (2004) Impact of surgical staging in patients with macroscopic 'stage I' ovarian borderline tumours: analysis of a continuous series of 101 cases. *Eur J Cancer* **40**: 1842–1849.

du Bois A, Ewald-Riegler N, de Gregorio N, Reuss A, Mahner S, Fotopoulou C, Kommoss F, Schmalfeldt B, Hilpert F, Fehm T, Burges A, Meier W, Hillemanns P, Hanker L, Hasenburg A, Strauss HG, Hellriegel M, Wimberger P, Keyver-Paik MD, Baumann K, Canzler U, Wollschlaeger K, Forner D, Pfisterer J, Schröder W, Münstedt K, Richter B, Kommoss S, Hauptmann S. Arbeitsgmeinschaft Gynäkologische Onkologie (AGO) Study Group (2013) Borderline tumours of the ovary: a cohort study of the Arbeitsgmeinschaft Gynäkologische Onkologie (AGO) Study Group. *Eur J Cancer* **49**: 1905–1914.

Fauvet R, Poncelet C, Boccara J, Descamps P, Fondrinier E, Daraï E (2005) Fertility after conservative treatment for borderline ovarian tumors: a French multicenter study. *Fertil Steril* **83**: 284–290.

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**: 29–36.

Hart WR (2005) Borderline epithelial tumors of the ovary. *Mod Pathol* **18**: S33–S50.

Isariyawongse BK, Kattan MW (2012) Prediction tools in surgical oncology. *Surg Oncol Clin N Am* **21**: 439–447.

Morice P, Uzan C, Fauvet R, Gouy S, Duvillard P, Darai E (2012) Borderline ovarian tumour: pathological diagnostic dilemma and risk factors for invasive or lethal relapse. *Lancet Oncol* **13**: 103–115.

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.

Palomba S, Falbo A, Del Negro S, Rocca M, Russo T, Cariati F, Annunziata G, Tolino A, Tagliaferri P, Zullo F (2010) Ultra-conservative fertility-sparing strategy for bilateral borderline ovarian tumours: an 11-year follow-up. *Hum Reprod* **25**: 1966–1972.

Seidman JD, Soslow RA, Vang R, Berman JJ, Stoler MH, Sherman ME, Oliva E, Kajdacsy-Balla A, Berman DM, Copeland LJ (2004) Borderline ovarian tumors: diverse contemporary viewpoints on terminology and diagnostic criteria with illustrative images. *Hum Pathol* **35**: 918–933.

Silva EG, Gershenson DM, Malpica A, Deavers M (2006) The recurrence and the overall survival rates of ovarian serous borderline neoplasms with non invasive implants is time dependent. *Am J Surg Pathol* **30**: 1367–1371.

Silverberg SG, Bell DA, Kurman RJ, Seidman JD, Prat J, Ronnett BM, Copeland L, Silva E, Gorstein F, Young RH (2004) Borderline ovarian tumors: key points and workshop summary. *Hum Pathol* **35**: 910–917.

Tavassoli FA, Devilee P (2003) *World Health Organization Classification of Tumours: Pathology and Genetics-Tumours of the Breast and Female Genital Organs*. IARC Press: Lyon.

Trimble CL, Kosary C, Trimble EL (2002) Long-term survival and patterns of care in women with ovarian carcinoma: a population-based analysis. *Gynecol Oncol* **86**: 34–37.

Uzan C, Kane A, Rey A, Gouy S, Camatte S, Pautier P, Lhommé C, Haie-Meder C, Duvillard P, Morice P (2011) Prognosis and prognostic factors of the micropapillary pattern in patients treated for stage II and III serous borderline tumors of the ovary. *Oncologist* **16**: 189–196.