External validation of three prognostic models for overall survival in patients with advanced-stage epithelial ovarian cancer

R van de Laar*[,1,2,3], J IntHout[4], T Van Gorp[2,3], S Verdonschot[1], A M van Altena[1], C G Gerestein[5], L F A G Massuger[1], P L M Zusterzeel[1] and R F P M Kruitwagen[2,3]

[1]Department of Obstetrics and Gynaecology, Radboud University Medical Centre, Geert Grooteplein 10, 6500 HB Nijmegen, The Netherlands; [2]Department of Obstetrics and Gynecology, Maastricht University Medical Centre, Maastricht, The Netherlands; [3]GROW, School for Oncology and Developmental Biology, Maastricht University, P Debyelaan 25, 6229 HA Maastricht, The Netherlands; [4]Department for Health Evidence, Radboud University Medical Centre, Geert Grooteplein 10, 6500 HB Nijmegen, The Netherlands and [5]Department of Obstetrics and Gynecology, Meander Medical Centre, Utrechtseweg 160, 3818 ES Amersfoort, The Netherlands

Background: For various malignancies, prognostic models have shown to be superior to traditional staging systems in predicting overall survival. The purpose of this study was to validate and compare the performance of three prognostic models for overall survival in patients with advanced-stage epithelial ovarian cancer.

Methods: A multi-institutional epithelial ovarian cancer database was used to identify patients and to evaluate the predictive performance of two nomograms, a prognostic index and FIGO (International Federation of Obstetrics and Gynecology) stage. All patients were treated for advanced-stage epithelial ovarian cancer between January 1996 and January 2009 in 11 hospitals in the eastern part of The Netherlands.

Results: In total, 542 patients were found to be eligible. Overall performance did not differ between the three prognostic models and FIGO stage. The discriminative performance for Chi's model was moderately good (c indices 0.65 and 0.68) and for the models of Gerestein and Teramukai reasonable (c indices between 0.60 and 0.62). The c indices of FIGO stage ranged between 0.54 and 0.62. After recalibration, the three models showed almost perfect calibration, whereas calibration of FIGO stage was reasonable.

Conclusion: The three prediction models showed general applicability and a reasonably well-predictive performance, especially in comparison to FIGO stage. To date, there are no studies available that analyse the impact of these prognostic models on decision-making and patient outcome. Therefore, the usefulness of these models in daily clinical practice remains to be investigated.

Prediction is ubiquitous in oncology: innumerable decisions by patients, oncologists and other care providers depend on assessing the likelihood of future events. For patients with advanced-stage epithelial ovarian cancer (International Federation of Obstetrics and Gynecology (FIGO) stage IIb–IV), decisions about life-extending *vs* palliative care inherently involve a survival prediction (Vickers, 2011).

Survival after primary treatment for advanced-stage epithelial ovarian cancer is associated with multiple prognostic factors (Voest *et al*, 1989; Omura *et al*, 1991; Chi *et al*, 2001).

Although FIGO stage is an implicit predictor for survival (Benedet et al, 2000; Heintz et al, 2006), researchers have developed five different statistical prognostic tools that provide a quantitative estimate of the survival probability for an individual patient with epithelial ovarian cancer (Clark et al, 2001; Teramukai et al, 2007; Chi et al, 2008; Gerestein et al, 2009; Barlin et al, 2012). Three of them, two nomograms (Chi et al, 2008; Gerestein et al, 2009) and one prognostic index (Teramukai et al, 2007), were developed to predict survival exclusively in patients with advanced-stage disease. To our knowledge, only one of these three models was externally validated (Clark et al, 2007). However, none of these models has gained widespread acceptance or utilisation. In the current study, we evaluated the predictive performance of these three models for advanced-stage ovarian cancer in an independent external validation cohort and compared the predictive performance of the three models with each other and with FIGO stage.

MATERIALS AND METHODS

Study population. All patients diagnosed with advanced-stage (FIGO IIb–IV) epithelial ovarian cancer between 1 January 1996 and 1 January 2009 in 11 hospitals (1 university hospital and 10 community hospitals) in the eastern part of The Netherlands were identified at the Netherlands Cancer Registry (NCR) (van Altena et al, 2013).

Data on patient characteristics, tumour characteristics, therapy and recurrence were retrospectively collected. Follow-up data were available until 1 January 2011. None of these patients were used in the development of any of the three prognostic models, and as such, this population represents an independent external validation cohort. To validate the predictive performance of the three prognostic models, we developed two validation cohorts for each model: a 'specific' validation cohort, limited to a subset of patients with advanced-stage disease according to the in- and exclusion criteria of the original population in which the prognostic model was created, and a 'general' validation cohort, consisting of a subset of patients with advanced-stage epithelial ovarian cancer treated with primary surgical debulking followed by platinum-based chemotherapy. Because each model contained different prognostic factors with possibly missing values, this resulted in three different 'general' cohorts, as patients with missing predictor values were excluded from external validation.

Prognostic models. The nomogram of Gerestein (Gerestein et al, 2009) was developed in a patient population consisting of women with advanced-stage epithelial ovarian cancer and uses three prognostic variables: preoperative haemoglobin serum concentrations ($mmol\,l^{-1}$), preoperative blood platelet counts ($10^9\,l^{-1}$) and postoperative residual tumour ($<1\,cm$), to predict 5-year overall survival (OS) probability. The reported c index is 0.67.

Chi (Chi et al, 2008) developed a nomogram in a patient population consisting of women with FIGO stage IIIC epithelial ovarian cancer. To predict 5-year OS probability, six prognostic variables are used: age at surgery, preoperative platelet count ($10^9\,l^{-1}$), tumour grade (1–2 vs 3), histology (serous vs non-serous), presence of ascites and residual disease status (0 vs <0.5 vs 0.5–1.0 vs 1–2 vs >2 cm). The reported c index is 0.67.

The prognostic index of Teramukai (Teramukai et al, 2007) was developed in a patient population consisting of women with FIGO stage III or IV epithelial ovarian cancer. The prognostic index predicts 5-year OS probability in three risk groups (low, intermediate and high). It uses four characteristics: age ($\leqslant 69$ vs $\geqslant 70$ years), World Health Organisation (WHO) performance status (0 vs 1 or 2 vs 3 or 4), cell type (mucinous or clear cell vs others) and residual tumour size (0 vs $\geqslant 0.1$ cm). The reported c index is 0.65.

Further details including in- and exclusion criteria of the three development cohorts are shown in Supplementary File 1.

External validation/statistical analysis. All statistical analyses and plotting were performed using SPSS version 19.0 and R version 2.15.2 (Team, 2012). Overall survival was calculated from the date of surgery to either the date of death or the last follow-up date. Five-year OS probabilities were calculated for every patient using the nomograms, the prognostic index and FIGO stage (Netherlands CCCt, 2013, NCR). Observed 5-year OS rates were obtained using the method of Kaplan–Meier (KM). The discriminative ability was measured with the c index. The c index is similar to the area under the receiver operating characteristic curve for logistic models and is measured on a scale ranging from 0.5 (no better than chance) to 1 (perfect prediction) (Pencina and D'Agostino, 2004).

Calibration, which refers to how closely the predicted 5-year survival probability by the models and FIGO stage agree with the observed 5-year survival probability, was assessed graphically (Steyerberg et al, 2010) by means of the R package rms (Harrell, 2013). The predicted 5-year survival probabilities (x axis) are plotted against the observed probabilities (y axis). Ideally, if predicted and observed 5-year survival probabilities agree over the whole range of probabilities, the plots shows a 45° line. We recalibrated the models using a practical approach. We determined in our validation cohort the baseline 5-year survival probability, that is, the probability that a patient with zero points would survive at least 5 years. We also determined a shrinkage factor, to be used for multiplication with the linear part of the prognostic model, to correct for possible overfitting (Steyerberg et al, 2010). The shrinkage factor was set equal to the slope of the linear predictor of the prognostic model after Cox's regression in the validation cohort. Both the baseline 5-year survival probability of the validation cohort and the shrinked linear part were used in the recalibrated models for the calculation of the predicted 5-year survival probabilities

RESULTS

Validation cohorts. During the study period, 1554 patients were diagnosed with epithelial ovarian cancer in the 11 hospitals, of whom 1095 (70%) were with advanced-stage epithelial ovarian cancer (FIGO IIB–IV). Five hundred and forty-two (49%) of these 1095 patients were treated with primary surgical debulking and adjuvant platinum-based chemotherapy.

The 'general' advanced-stage validation cohorts consisted of 323 (Gerestein), 279 (Chi) and 459 (Teramukai) patients, as 40%, 49% and 15% of the patients were excluded due to missing predictor values, respectively (Supplementary File 2). In the validation cohorts for the models of Gerestein and Chi, absence of preoperative platelet count caused most missing values (92% and 73%, respectively). In the validation cohort for the Teramukai model, 67% of missing values was due to the absence of performance status. The 'specific' validation cohorts, with in- and exclusion criteria conform the prognostic model, consisted of 227, 154 and 393 patients, respectively (Supplementary File 3). The amount of missing values was comparable with the general cohorts.

Differences between the three development cohorts and their validation cohorts are provided in Tables 1, 2 and 3. The development cohort of Gerestein contains a significantly higher rate of patients with FIGO stage IV, poorly differentiated tumours and a significantly lower rate of optimal ($<1\,cm$) primary debulking with a corresponding lower 5-year OS probability compared with our two validation cohorts. Both validation cohorts were larger than the development cohort. The development cohort

Table 1. Patient and study characteristics of the Gerestein *et al* (2009) original development cohort and the 'general' and 'specific' advanced-stage validation cohorts

Parameter	Gerestein original cohort (*N* = 118)	General validation cohort (*N* = 323)	*P*-value[a]	Specific validation cohort (*N* = 227)	*P*-value[b]
Age at diagnosis, median (interquartile range) (years)	62 (NR)	61 (52–69)	—	60 (52–69)	—
FIGO stage, *n* (%)					
II	10 (9)	50 (15)	0.01	36 (16)	0.009
III	89 (75)	248 (77)		175 (77)	
IV	19 (16)	25 (8)		16 (7)	
Tumour grade, *n* (%)					
1	6 (5)	27 (9)	0.3	16 (8)	0.6
2	34 (29)	82 (29)		54 (27)	
3	78 (66)	176 (62)		127 (65)	
Histology, *n* (%)					
Serous	90 (76)	229 (71)	0.3	161 (71)	0.3
Non-serous	28 (24)	94 (29)		66 (29)	
Residual tumour after primary debulking, *n* (%)					
<1 cm	57 (48)	235 (73)	<0.001	165 (73)	<0.001
⩾1 cm	61 (52)	88 (27)		62 (27)	
Preoperative haemoglobin serum concentrations, mean (s.d.) (mmol l^{-1})	7.9 (0.8)	7.8 (0.9)	0.3	7.9 (0.9)	1.0
Preoperative blood platelet count, mean (s.d.) (10^9 l^{-1})	303 (121)	362 (131)	<0.001	364 (132)	<0.001
Follow-up time for survivors, median (interquartile range) (months)	40 (NR)	36 (24–54)	—	36 (24–52)	—
Number of death, *n*	63	172	—	113	—
5-Year OS probability	0.32	0.40	—	0.42	—

Abbreviations: FIGO = International Federation of Gynaecology and Obstetrics; NR = not reported; OS = overall survival; s.d. = standard deviation.

P-values were obtained using the χ^2 test (categorical variables) and the Student's *t*-test (continuous variables).

[a]*P*-value for the comparison between original cohort and general validation cohort.

[b]*P*-value for the comparison between original cohort and specific validation cohort.

of Chi also shows a significantly higher rate of poorly differentiated tumours and a significantly lower rate of complete and optimal primary debulking (<1 cm) with a non-corresponding higher 5-year OS probability in comparison with the two validation cohorts. Both validation cohorts contained fewer patients than the development cohort. The most important differences between the development cohort of Teramukai and the two validation cohorts are a significantly higher rate of patients with FIGO stage IV and a significantly lower rate of poorly differentiated tumours and complete and optimal (<1 cm) primary debulking with a non-corresponding higher 5-year OS probability in the development cohort. Both validation cohorts contained fewer patients than the development cohort. In the validation cohorts, median follow-up times for surviving patients ranged between 36 and 42 months, percentage of observed deaths between 50 and 64 and the observed 5-year OS probabilities between 0.31 and 0.42.

External validation. Table 4 represents the discriminative performance (*c* index) for the three prognostic models and FIGO stage. The *c* indices for the general validation cohorts were 0.62, 0.68 and 0.62 for the Gerestein, Chi and Teramukai model, respectively. Discrimination of all three models was better than that of FIGO stage, with *c* indices for FIGO stage 0.60, 0.59 and 0.58, respectively. The *c* indices for the specific validation cohorts were 0.62, 0.65 and 0.60 for the Gerestein, Chi and Teramukai model, respectively. Compared with FIGO stage, discrimination of the model of Teramukai was better (*c* index for FIGO stage 0.54), whereas discrimination of the model of Gerestein showed the same

c index as FIGO stage. The prognostic index of Teramukai was previously validated by Clark (Clark *et al*, 2007) and showed a *c* index of 0.625.

Figure 1 shows the calibration plots of FIGO stage and of the three prognostic models before and after recalibration. In contrast to the acceptable calibration of the nomogram of Gerestein and FIGO stage, the models of Chi and Teramukai systematically overestimated the 5-year OS probability. This probably is a result of the lower 5-year OS probabilities in our validation cohorts compared with the development cohorts. We therefore recalibrated the models without changing the underlying parameters. After recalibration, the three models showed almost perfect calibration. Compared with the recalibrated models, calibration of FIGO stage was reasonable.

DISCUSSION

This study aimed external validation as well as comparison of three formerly developed prognostic models for 5-year OS and FIGO stage, in patients with advanced-stage epithelial ovarian cancer treated with primary surgical debulking followed by platinum-based chemotherapy.

We did not review two other published models that predict 5-year OS probability in patients with both early- and advanced-stage disease (Clark *et al*, 2001; Barlin *et al*, 2012). To our opinion, early- and advanced-stage epithelial ovarian cancer are such

Table 2. Patient and study characteristics of the Chi *et al* (2008) original development cohort and the 'general' and 'specific' advanced-stage validation cohorts					
Parameter	Chi original cohort (*N* = 424)	General validation cohort (*N* = 279)	*P*-value[a]	Specific validation cohort (*N* = 154)	*P*-value[b]
Age at diagnosis, median (interquartile range) (years)	60 (NR)	61 (52–69)	—	61 (52–70)	—
FIGO stage, *n* (%)					
II IIIa IIIb IIIc IV	424 (100)	42 (15) 3 (1) 40 (14) 177 (62) 22 (8)	<0.001	154 (100)	—
Tumour grade, *n* (%)					
1 2 3	11 (3) 84 (20) 329 (78)	26 (9) 82 (29) 171 (62)	<0.001	10 (7) 34 (22) 110 (71)	0.06
Histology, *n* (%)					
Serous Non-serous	308 (73) 116 (27)	204 (73) 75 (27)	0.9	99 (64) 55 (36)	0.05
Residual tumour after primary debulking, *n* (%)					
Complete >0<1 cm ≥1 cm	57 (13) 158 (37) 209 (50)	107 (38) 98 (35) 74 (27)	<0.001	39 (25) 59 (38) 56 (37)	0.001
Preoperative blood platelet count, median (interquartile range), $10^9 l^{-1}$	375 (NR)	341 (261–438)	—	356 (266–470)	—
Presence of ascites, *n* (%)					
Yes No	355 (84) 69 (16)	221 (79) 58 (21)	0.13	131 (85) 23 (15)	0.7
Follow-up time for survivors, median (interquartile range) (months)	NR	36 (25–58)	—	36 (39–42)	—
Number of death, *n*	NR	150	—	97	—
5-Year OS probability	0.51	0.39	—	0.31	—

Abbreviations: FIGO = International Federation of Gynaecology and Obstetrics; NR = not reported; OS = overall survival; s.d. = standard deviation.
P-values were obtained using χ^2 test (categorical variables) and Student's *t*- test (continuous variables).
[a]*P*-value for the comparison between original cohort and general validation cohort.
[b]*P*-value for the comparison between original cohort and specific validation cohort.

different entities regarding their tumour biology, treatment, prognostic factors and survival that both require stage-specific prognostic models for survival.

Study design (selection method of participants), participants (in- and exclusion criteria), predictors and outcome were adequately described in all three prognostic model development studies. All studies assessed model performance by internal validation and they all reported optimism corrected c-statistics (Bouwmeester *et al*, 2012). Calibration was only reported by Chi and Teramukai. Both the original development cohorts and the current validation cohorts were retrospective cohorts that allows for a longer follow-up period but at the expense of selection bias and missing data. Missing data also may create bias or the need for multiple imputation techniques (Moons *et al*, 2009b). According to a frequently mentioned rule of thumb that states that 10 events are needed per candidate predictor, sample sizes of the development cohorts were sufficient (Concato *et al*, 1995) but the development cohort of Gerestein *et al* (2009) still can be defined as small. Application of prognostic models requires unambiguous definitions of prognostic factors and standardised and reproducible measurements using methods available in clinical practice (Moons *et al*, 2009b). Unfortunately, the definition of residual tumour

after primary debulking, one of the most powerful prognostic factors for survival and used in all three prognostic models, differs substantially between the models. Similarly, presence of ascites, prognostic factor in the nomogram of Chi was not clearly defined. Moreover, the model should be developed in a population that resembles the population for whom the model estimates will be applied (Iasonos *et al*, 2008). Nowadays, more and more patients with advanced-stage ovarian cancer are treated with neoadjuvant chemotherapy followed by interval debulking surgery. None of the models can be applied to this category of patients because prognostic factors like preoperative blood platelet count, preoperative haemoglobin, presence of ascites, performance status and postoperative residual tumour can be substantially different after treatment with neoadjuvant chemotherapy.

Although our general advanced-stage epithelial ovarian cancer cohort originally consisted of 542 patients, all six validation cohorts were significantly smaller owing to a high percentage (15–49%) of patients with missing predictor values. We had to decide whether we were going to use modern statistical techniques like multiple imputation to replace missing values in order to create less biased estimates of accuracy and obtain more powerful sample sizes for external validation (Vergouwe *et al*, 2010). Until now there is little

Table 3. Patient and study characteristics of the Teramukai et al (2007) original development cohort and the 'general' and 'specific' advanced-stage validation-n cohorts

Parameter	Teramukai original cohort (N = 768)	General validation cohort (N = 459)	P-value[a]	Specific validation cohort (N = 393)	P-value[b]
Age at diagnosis, median (interquartile range) (years)	NR	60 (52–69)	—	61 (52–70)	—
FIGO stage, n (%)					
II	—	66 (14)	<0.001	—	<0.001
III	614 (80)	362 (79)		362 (92)	
IV	154 (20)	31 (7)		31 (8)	
Tumour grade, n (%)					
1	121 (16)	37 (9)	<0.001	26 (8)	<0.001
2	146 (19)	111 (28)		92 (27)	
3	235 (31)	249 (63)		221 (65)	
Missing	266 (34)	0		0	
Histology, n (%)					
Serous	505 (66)	329 (72)	0.03	296 (75)	0.001
Non-serous	263 (34)	130 (28)		97 (25)	
Residual tumour after primary debulking, n (%)					
Complete	119 (16)	158 (34)	<0.001	102 (26)	<0.001
>0<1 cm	129 (17)	174 (38)		166 (42)	
⩾1 cm	520 (67)	127 (28)		125 (32)	
Performance status, n (%)					
0	308 (40)	194 (42)	<0.001	157 (40)	<0.001
1–2	395 (51)	263 (57)		233 (59)	
3–4	65 (9)	2 (<1)		3 (0.8)	
Follow-up time for survivors, median (interquartile range) (months)	49 (NR)	42 (28–62)	—	39 (26–61)	—
Number of death, n	408	269	—	250	—
5-Year OS probability	0.45	0.41	—	0.36	—

Abbreviations: FIGO = International Federation of Gynaecology and Obstetrics; NR = not reported; OS = overall survival; s.d. = standard deviation.
P-values were obtained using χ^2 test (categorical variables) and Student's t-test (continuous variables).
[a]P-value for the comparison between original cohort and general validation cohort.
[b]P-value for the comparison between original cohort and specific validation cohort.

experience with the use of multiple imputation in an external validation setting. We therefore restricted ourselves to a comparison of cases with and without missing predictor values. Patients with and without missing predictor values did not notably differ from each other with respect to FIGO stage, histology, tumour grade, performance status, residual tumour after primary debulking and 5-year OS probability. However, finding proof that predictors are randomly missing is impossible (Sterne et al, 2009). To obtain a reasonable power of 80% to detect substantial changes in model performance, it has been suggested that a validation sample should contain at least 100 events and 100 non-events (Vergouwe et al, 2005). In this study, all validation cohorts except the specific validation cohort of Chi consisted of at least 100 events (deaths) and 100 non-events (survivors).

Although our validation cohorts showed more favourable tumour characteristics and more frequent complete cytoreductive surgery than the original cohorts of Chi and Teramukai, the KM estimates for survival were lower. This perhaps may be related to differences in chemotherapeutic treatment, age or a shorter follow-up period, resulting in more censored observations and thus higher impact of a few deaths on the 5-year KM estimates.

Noteworthy is the similar or even better performance in the general cohorts as compared with the specific cohorts, which is a positive finding that enables us to use these models in general practice. Compared with FIGO stage, the models of Chi and Teramukai showed clearly better discriminative performance. The three models systematically showed miscalibration to a greater or lesser extent. Nevertheless, all models showed good calibration after correction for differences in 5-year survival between the development and validation cohorts, and shrinkage of the effect of the predictors on survival. The near-perfect recalibration indicates that the miscalibration is a result of the lower 5-year OS probabilities in our validation cohorts, and an overfit of the prognostic models. However, these recalibrated models also need validation (Royston and Altman, 2013). Nevertheless, our results provide evidence that the prognostic models overall can perform reasonably well irrespective of tumour and treatment characteristics. Indicating one of the three models as the best model is difficult as all three models have different positive and negative properties.

To be considered useful, a prognostic model must not only show generalisability (external validation) but also clinical effectiveness. For the three validated models, an impact study can determine whether use of the model has added value, including the acceptability of the prognostic model to clinicians and ease of use (Moons et al, 2009a).

Table 4. Discrimination statistics for the three prediction models and FIGO stage in estimating 5-year OS probability

|	Nomogram Gerestein (Gerestein et al, 2009)	Nomogram Chi (Chi et al, 2008)	Prognostic index Teramukai (Teramukai et al, 2007)
Development cohort	**N = 118**	**N = 424**	**N = 768**
Discrimination			
c index (95% CI)			
Prediction model	0.67 (NR)	0.67 (NR)	0.65 (NR)
FIGO stage	NR	NR	NR
General validation cohort	**N = 323**	**N = 279**	**N = 459**
Discrimination			
c index (95% CI)			
Prediction model	0.62 (0.58–0.67)	0.68 (0.63–0.73)	0.62 (0.59–0.65)
FIGO stage	0.60 (0.56–0.64)	0.59 (0.55–0.63)	0.58 (0.55–0.62)
Specific validation cohort	**N = 227**	**N = 154**	**N = 393**
Discrimination			
c index (95% CI)			
Prediction model	0.62 (0.56–0.68)	0.65 (0.59–0.71)	0.60 (0.57–0.64)
FIGO stage	0.62 (0.57–0.67)	—	0.54 (0.51–0.57)
Other external validation cohort	—	—	**Clark (Clark et al, 2007) (N = 894)**
Discrimination			
c index (95% CI)			
Prediction model	NA	NA	0.625 (NR)
FIGO stage	NA	NA	NR
Abbreviations: CI = confidence interval; FIGO = International Federation of Gynaecology and Obstetrics; NA = not applicable; NR = not reported; OS = overall survival.			
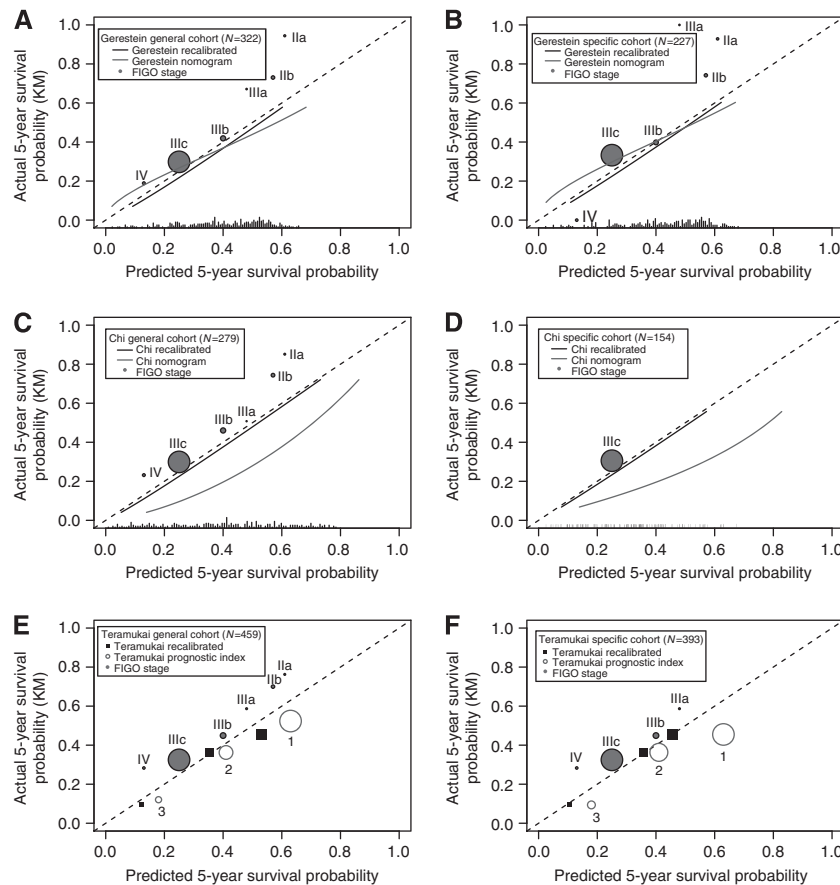
Figure 1. Calibration of the two nomograms, prognostic index and FIGO stage in their 'general' and 'specific' validation cohorts. Five-year survival probabilities based on the three prognostic models and FIGO stage are plotted against the KM survival rates. Dots and blocks show the relation between the predicted and observed survival probabilities for FIGO stage (A–F) and for the prognostic index before and after recalibration (E and F). A spline shows the continuous relation between the predicted and observed survival probabilities for the two nomograms before and after recalibration (A–D) (F Harrell, rms package). The dashed line represents the perfect calibration slope. Distribution of the recalibrated predicted probabilities is indicated with vertical lines at the bottom of the plot (A–D).

CONCLUSION

In this study, we externally validated three prognostic models for survival in patients with advanced-stage ovarian cancer. They all showed reasonably well accuracy, especially in comparison to FIGO stage. Impact studies, quantifying the effect of using the prognostic models on decision-making and patient outcome, are warranted.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Barlin JN, Yu C, Hill EK, Zivanovic O, Kolev V, Levine DA, Sonoda Y, Abu-Rustum NR, Huh J, Barakat RR, Kattan MW, Chi DS (2012) Nomogram for predicting 5-year disease-specific mortality after primary surgery for epithelial ovarian cancer. *Gynecol Oncol* **125**: 25–30.

Benedet JL, Bender H, Jones 3rd H, Ngan HY, Pecorelli S (2000) FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. FIGO Committee on Gynecologic Oncology. *Int J Gynaecol Obstet* **70**: 209–262.

Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, Altman DG, Moons KG (2012) Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* **9**: 1–12.

Chi DS, Liao JB, Leon LF, Venkatraman ES, Hensley ML, Bhaskaran D, Hoskins WJ (2001) Identification of prognostic factors in advanced epithelial ovarian carcinoma. *Gynecol Oncol* **82**: 532–537.

Chi DS, Palayekar MJ, Sonoda Y, Abu-Rustum NR, Awtrey CS, Huh J, Eisenhauer EL, Barakat RR, Kattan MW (2008) Nomogram for survival after primary surgery for bulky stage IIIC ovarian carcinoma. *Gynecol Oncol* **108**: 191–194.

Clark TG, Stewart M, Rye T, Smyth JF, Gourley C (2007) Validation of a new prognostic index for advanced epithelial ovarian cancer: results from its application to a UK-based cohort. *J Clin Oncol* **25**: 5669–5670 (author reply 5670–5671).

Clark TG, Stewart ME, Altman DG, Gabra H, Smyth JF (2001) A prognostic model for ovarian cancer. *Br J Cancer* **85**: 944–952.

Concato J, Peduzzi P, Holford TR, Feinstein AR (1995) Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. *J Clin Epidemiol* **48**: 1495–1501.

Gerestein CG, Eijkemans MJ, de Jong D, van der Burg ME, Dykgraaf RH, Kooi GS, Baalbergen A, Burger CW, Ansink AC (2009) The prediction of progression-free and overall survival in women with an advanced stage of epithelial ovarian carcinoma. *Br J of Gynecol* **116**: 372–380.

Heintz AP, Odicino F, Maisonneuve P, Quinn MA, Benedet JL, Creasman WT, Ngan HY, Pecorelli S, Beller U (2006) Carcinoma of the ovary. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. *Int J Gynaecol Obstet* **95**(Suppl 1): S161–S192.

Iasonos A, Schrag D, Raj GV, Panageas KS (2008) How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* **26**: 1364–1370.

Harrell Jr FEH (2013) *rms: Regression Modeling Strategies, R package version 3.6-3.*

Moons KG, Altman DG, Vergouwe Y, Royston P (2009a) Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* **338**: b606.

Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG (2009b) Prognosis and prognostic research: what, why, and how? *BMJ* **338**: b375.

Omura GA, Brady MF, Homesley HD, Yordan E, Major FJ, Buchsbaum HJ, Park RC (1991) Long-term follow-up and prognostic factor analysis in advanced ovarian carcinoma: the Gynecologic Oncology Group experience. *J Clin Oncol* **9**: 1138–1150.

Pencina MJ, D'Agostino RB (2004) Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* **23**: 2109–2123.

Royston P, Altman DG (2013) External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* **13**: 33.

Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* **338**: b2393.

Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* **21**: 128–138.

Team RDC (2012) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, Austria.

Teramukai S, Ochiai K, Tada H, Fukushima M (2007) PIEPOC: a new prognostic index for advanced epithelial ovarian cancer – Japan Multinational Trial Organization OC01-01. *J Clin Oncol* **25**: 3302–3306.

van Altena AM, van den Akker PA, Hullu de JA, Ottevanger PB, Aalders AL, Gerritse R, Happel M, Hoekstra MP, Janssen MJ, Samlal RA, Smeets KM, Snijders MP, Vasmel MJ, Vollebergh JH, Kiemeney LA, Massuger LF (2013) The efficacy of a regional network for ovarian cancer care. *Obstet Gynecol* **122**(3): 668–675.

Vergouwe Y, Royston P, Moons KG, Altman DG (2010) Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol* **63**: 205–214.

Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD (2005) Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* **58**: 475–483.

Vickers AJ (2011) Prediction models in cancer care. *CA Cancer J Clin* **61**: 315–326.

Voest EE, van Houwelingen JC, Neijt JP (1989) A meta-analysis of prognostic factors in advanced ovarian cancer with median survival and overall survival (measured with the log (relative risk)) as main objectives. *Eur J Cancer Clin Oncol* **25**: 711–720.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)