

Keywords: breast cancer; young patients; prognosis; adjuvant! online; Nottingham Prognostic Index; adjuvant treatment; subtypes

The prognostic performance of Adjuvant! Online and Nottingham Prognostic Index in young breast cancer patients

Matteo Lambertini^{1,2,12}, Ana Catarina Pinto^{1,12}, Lieveke Ameye³, Lynn Jongen⁴, Lucia Del Mastro⁵, Fabio Puglisi^{6,7}, Francesca Poggio², Marta Bonotto^{6,7}, Giuseppe Floris^{8,9}, Kathleen Van Asten⁴, Hans Wildiers^{4,10}, Patrick Neven^{4,11}, Evandro de Azambuja¹, Marianne Paesmans³ and Hatem A Azim Jr^{*1}

¹Department of Medicine, BrEAST Data Centre, Institut Jules Bordet, and l'Université Libre de Bruxelles (U.L.B.), Boulevard de Waterloo 121, 1000 Brussels, Belgium; ²Department of Medical Oncology, U.O. Oncologia Medica 2, IRCCS AOU San Martino–IST, Largo Rosanna Benzi 10, 16132 Genova, Italy; ³Data Centre, Institut Jules Bordet, and l'Université Libre de Bruxelles (U.L.B.), Boulevard de Waterloo 121, 1000 Brussels, Belgium; ⁴Department of Oncology, KU Leuven–University of Leuven, University Hospitals Leuven, B-3000 Leuven, Belgium; ⁵Department of Medical Oncology, U.O. Sviluppo Terapie Innovative, IRCCS AOU San Martino–IST, Largo Rosanna Benzi 10, 16132 Genova, Italy; ⁶Department of Medical and Biological Sciences, University of Udine, Piazzale Santa Maria della Misericordia 15, 33010 Udine, Italy; ⁷Department of Oncology, University Hospital of Udine, Piazzale Santa Maria della Misericordia 15, 33010 Udine, Italy; ⁸Department of Imaging and Pathology, Laboratory of Translational Cell & Tissue Research, KU Leuven–University of Leuven, B-3000 Leuven, Belgium; ⁹Department of Pathology, KU Leuven–University of Leuven, University Hospitals Leuven, B-3000 Leuven, Belgium; ¹⁰Department of General Medical Oncology, KU Leuven–University of Leuven, University Hospitals Leuven, B-3000 Leuven, Belgium and ¹¹Department of Gynecology and Obstetrics, KU Leuven–University of Leuven, University Hospitals Leuven, B-3000 Leuven, Belgium

Background: Limited data are available on the prognostic performance of Adjuvant! Online (AOL) and Nottingham Prognostic Index (NPI) in young breast cancer patients.

Methods: This multicentre hospital-based retrospective cohort study included young (≤ 40 years) and older (55–60 years) breast cancer patients treated from January 2000 to December 2004 at four large Belgian and Italian institutions. Predicted 10-year overall survival (OS) and disease-free survival (DFS) using AOL and 10-year OS using NPI were calculated for every patient. Tools ability to predict outcomes (i.e., calibration) and their discriminatory accuracy was assessed.

Results: The study included 1283 patients, 376 young and 907 older women. Adjuvant! Online accurately predicted 10-year OS (absolute difference: 0.7%; $P=0.37$) in young cohort, but overestimated 10-year DFS by 7.7% ($P=0.003$). In older cohort, AOL significantly underestimated both 10-year OS and DFS by 7.2% ($P<0.001$) and 3.2% ($P=0.04$), respectively. Nottingham Prognostic Index significantly underestimated 10-year OS in both young (8.5%; $P<0.001$) and older (4.0%; $P<0.001$) cohorts. Adjuvant! Online and NPI had comparable discriminatory accuracy.

Conclusions: In young breast cancer patients, AOL is a reliable tool in predicting OS at 10 years but not DFS, whereas the performance of NPI is sub-optimal.

*Correspondence: Dr HA Azim Jr; E-mail: hatemazim@icloud.com

¹²These authors contributed equally to this work.

Received 20 July 2016; revised 27 September 2016; accepted 4 October 2016; published online 1 November 2016

© 2016 Cancer Research UK. All rights reserved 0007–0920/16

Breast cancer represents the most frequently diagnosed malignant tumour in women younger than 40 years (DeSantis *et al*, 2016). Treatment recommendations in early stage breast cancer are based on both patient's prognosis and expected benefit of adjuvant therapies (Coates *et al*, 2015). Young patients have an increased risk of disease recurrence and death compared with the old counterpart (Azim Jr and Partridge, 2014); hence, a larger fraction of patients with early stage disease are offered chemotherapy (Azim Jr *et al*, 2016). However, these treatments cause significant long-term side effects and negatively influence quality of life (Azim Jr *et al*, 2011; Lambertini *et al*, 2016). Thus, adequate estimation of their prognosis is very important to help optimising adjuvant management strategies.

Adjuvant! Online (AOL) (Ravdin *et al*, 2001) and Nottingham Prognostic Index (NPI) (Haybittle *et al*, 1982) are two widely used tools based on traditional clinicopathological factors developed to predict an individual's prognosis and to aid clinical decision-making (Engelhardt *et al*, 2014). Although performing globally well, some concerns have been raised regarding their applicability in populations other than those used in their validation studies (de Glas *et al*, 2014) and in specific subgroups such as women younger than 40 years (Engelhardt *et al*, 2014). In addition, in light of advances in adjuvant systemic therapies (Coates *et al*, 2015), differences in outcome according to breast cancer subtypes (Engström *et al*, 2013), and improvement in survival (DeSantis *et al*, 2016), it is legitimate to understand whether these tools are still offering reliable predictions for long-term outcomes.

The present study evaluated the prognostic performance of AOL and NPI in young breast cancer patients (≤ 40 years at diagnosis). We examined the ability of both tools to predict outcomes (i.e., calibration) and their discriminatory accuracy in a cohort of young breast cancer patients. We further assessed the prognostic performance of both tools in an older cohort of patients (aged 55–60 years at diagnosis) that acted as a control group.

MATERIALS AND METHODS

Study design and patient selection. The present study was a multicentre hospital-based retrospective cohort study, which was conducted in four Belgian and Italian referral institutions specialised in managing breast cancer patients.

All consecutive women treated for breast cancer from 1 January 2000 to 31 December 2004 were identified in the hospitals' cancer registries. This restriction was chosen to allow at least a 10-year follow-up for included patients. Eligible patients were women diagnosed with invasive unilateral primary breast cancer treated with upfront surgery, and with complete information on pathology and treatment. Two cohorts were included, patients aged ≤ 40 years at diagnosis (young cohort) and those aged 55 until 60 years (older cohort). Patients were excluded if they received neoadjuvant systemic therapy, presented with newly diagnosed stage IV disease or non-invasive breast cancer, had unknown data on tumour size, nodal status and/or adjuvant treatments. Adjuvant trastuzumab was not approved at that time; women who received trastuzumab in the context of a clinical trial were excluded.

The study is reported according to the STROBE statement (von Elm *et al*, 2007). The institutional review boards of participating centres approved the study protocol and the retrospective review of patients' medical records. Signed informed consent was not required by the respective institutional review boards.

Study procedure. The medical records of all eligible patients were retrieved and coded data were entered into an excel database. Adjuvant! Online and NPI estimates for each patient were calculated separately by two investigators (ML and ACP) blinded to patient outcomes.

Patient and tumour characteristics were entered in the AOL programme version 8.0 to calculate the predicted 10-year overall survival (OS) and 10-year disease-free survival (DFS) for each patient; the 'average for age' category of comorbidity was used for

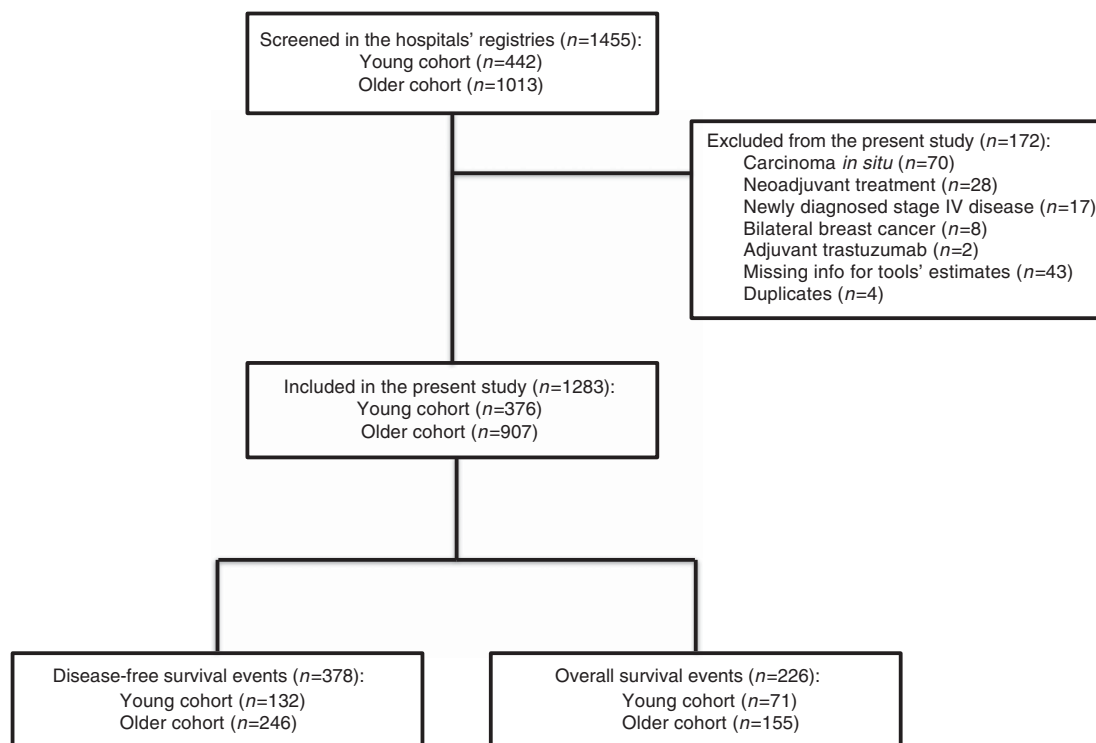


Figure 1. Study flow chart.

all patients. For every patient, NPI score was calculated with the following algorithm: maximum tumour size in centimetres \times 0.2 + nodal status + tumour grade. The numeric score obtained was used to stratify patients in risk categories. The following six risk categories with the respective likely predicted 10-year OS were considered (Quintyne *et al*, 2013): excellent, good, moderate group 1, moderate group 2, poor and very poor (Supplementary Table A1). No DFS estimates for the different categories of NPI were available in the literature.

Study objectives and end points. The primary objective of the study was to evaluate the prognostic performance of AOL and NPI in young breast cancer patients (\leq 40 years), and to compare it to the one obtained in the older cohort (55–60 years). We opted to avoid the inclusion of an older age group to circumvent any impact of associated morbidity on long-term outcome.

The study end points were calibration and discriminatory accuracy. Calibration refers to the agreement between the predicted and observed survival rates. Discriminatory accuracy refers to the ability of distinguishing individuals who will survive 10 years compared with those who will not (i.e., the ability to discern patients with good outcomes from those with poor outcomes at the individual patient level).

The secondary objective of the study was to describe the differences in prognostic accuracy according to breast cancer subtypes defined on the basis of oestrogen receptor (ER) and HER2 status by immunohistochemistry as defined locally (ER+/HER2-, HER2+ and ER-/HER2-). Other secondary objectives included describing the characteristics of tumours arising in young women, the type of adjuvant systemic treatments offered and to compare these findings with those of an older cohort of patients.

Statistical analyses. The observation time for each patient was defined as the time between the date of diagnosis and an event. Disease-free survival event was defined as the occurrence of any of the following: local recurrence, distant metastases, contralateral or ipsilateral breast tumour, second primary malignancy or death from any cause. Overall survival event was defined as death from any cause. Vital status was cross-checked with the national registries in Belgium and Italy. The date of last follow-up vital status was 01 March 2015. For DFS, observation times of patients without the event or who did not reach the 10-year follow-up were censored on the date of their last contact in the participating institutions. Disease-free survival and OS curves for the two cohorts of patients were calculated using the Kaplan–Meier method, and log-rank test was used for comparing the curves between the six NPI classes.

For assessing calibration, the trimmed mean of predicted 10-year survival probabilities (by AOL and NPI) was compared with the observed 10-year survival rates (Kaplan–Meier curves). A one-sample *t*-test for proportions was used, assuming AOL and NPI predicted values to be the population values (under the assumption that the models are true) and thus fixed. Calibration plots for AOL were constructed by visualising mean predicted *vs* observed survival outcomes (OS or DFS) by deciles of predicted outcomes (OS or DFS). Calibration plots for NPI were constructed by visualising the 10-year observed OS by the six categories of NPI predicted OS. Error bars in the calibration plots are s.e. of the observed 10-year outcomes.

For assessing discriminatory accuracy, the area under the receiver-operator characteristic curve (AUC under the ROC) and corresponding 95% confidence intervals (CI) for 10-year predicted OS and 10-year predicted DFS were calculated. The AUC translates into the probability that the predicted 10-year outcome of a randomly selected patient who indeed had a good outcome at 10 years is higher than that of a patient who did not; the higher the AUC the better the tools are at identifying patients with a better

outcome (being alive without disease or longer survival). The area under the time-dependent ROC curves was compared by using the nonparametric method of Chiang and Hung (2010).

To test for heterogeneity among the breast cancer subtypes with regard to calibration and discriminatory accuracy, the Cochran's χ^2 -test was used. Heterogeneity between breast cancer subtypes was assessed in the young and older cohorts, for AOL and NPI, considering as outcome OS and DFS, respectively.

Patient cohorts were analysed by applying descriptive statistics: χ^2 -test was used for correlations of nominal variables and Fisher exact test was used for small sample sizes. All analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA), GraphPad Prism software and R version 3.2.2, package time ROC (GraphPad Software Inc., San Diego, CA, USA).

RESULTS

Out of 1455 patients screened, 1283 were included in the present study (376 in the young and 907 in the older cohorts; Figure 1). All patients had 10-year follow-up for OS calculations. For DFS,

Table 1. Tumour baseline characteristics and treatments

	Young cohort (n = 376) n (%)	Older cohort (n = 907) n (%)	P-value
Type of surgery			
Breast-conserving surgery	241 (64)	633 (70)	0.06
Mastectomy	133 (36)	274 (30)	
Unknown	2 (0.5)	0 (0.0)	
Tumour size (mm)			
1–10	55 (15)	243 (27)	<0.001
11–20	150 (40)	377 (42)	
24–50	151 (40)	251 (28)	
>50	20 (5)	36 (4)	
Nodal status			
0	206 (55)	585 (65)	0.002
1–3	116 (31)	226 (25)	
>3	54 (14)	96 (11)	
Tumour grade			
G1	27 (7)	210 (23)	<0.001
G2	152 (40)	447 (49)	
G3	197 (52)	250 (28)	
ER status			
ER-	120 (32)	130 (14)	<0.001
ER+	255 (68)	772 (86)	
Unknown	1 (0.3)	5 (0.6)	
HER2 status			
HER2-	267 (78)	668 (87)	<0.001
HER2+	74 (22)	102 (13)	
Unknown	35 (9)	137 (15)	
Tumour subtype			
ER+/HER2-	181 (53)	591 (77)	<0.001
ER-/HER-	86 (25)	76 (10)	
HER2+	74 (22)	102 (13)	
Unknown	35 (9)	138 (15)	
Systemic treatment			
None	14 (4)	42 (5)	<0.001
Chemotherapy only	104 (28)	119 (13)	
Endocrine therapy only	44 (12)	437 (48)	
Chemotherapy and endocrine therapy	214 (57)	309 (34)	
Abbreviations: ER = oestrogen receptor; G = grade; HER2 = human epidermal growth factor receptor 2.			

10-year follow-up was reached in 80% and 77% of patients in the young and older cohorts, respectively.

Patient outcome was homogeneous among the four centres with no difference in OS (Supplementary Figures A1 and A2).

As compared with patients in the older cohort, those in the young cohort had more aggressive tumour characteristics at diagnosis with larger tumours ($P < 0.001$), more nodes involved ($P = 0.002$), higher tumour grade ($P < 0.001$), higher incidence of ER- ($P < 0.001$) and HER2+ ($P < 0.001$) disease (Table 1). Moreover, young patients were more often treated with chemotherapy than older women ($P < 0.001$).

In the young and older cohorts, respectively, 10-year OS was 84.6 and 88.4% ($P = 0.07$), and 10-year DFS was 68.3 and 76.6% ($P < 0.001$). Women with ER+/HER2- tumours showed the longest survival, with a 10-year OS of 87.3% and 89.7% and a 10-year DFS of 70.8% and 79.0% in the young and older cohorts, respectively (Table 2).

Figure 2 reports OS stratified by NPI class in the young ($P < 0.001$; Figure 2A) and older cohorts ($P < 0.001$; Figure 2B).

Calibration

Overall survival. In the young cohort, there was no significant difference between predicted survival by AOL (83.9%) and observed OS (84.6%; $P = 0.37$; Table 2). Conversely, NPI underestimated 10-year OS by 8.5% (76.1% vs 84.6%; $P < 0.001$). The discrepancy in NPI estimation was mainly observed in women with ER+/HER2- (absolute difference: -8.5%, $P = 0.001$) and HER2+ tumours (absolute difference: -12.4%, $P = 0.01$).

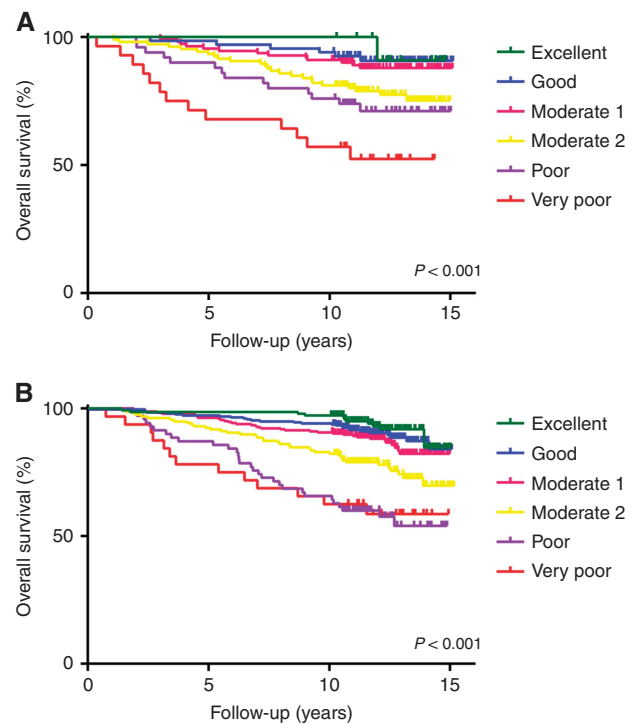


Figure 2. Overall survival stratified by Nottingham Prognostic Index class in patients in the young cohort (A) and in the older cohort (B).

Table 2. Observed and predicted 10-year overall survival for Adjuvant! Online and Nottingham Prognostic Index, and observed and predicted 10-year disease-free survival for Adjuvant! Online in the two cohorts and by tumour subtype

	% 10-year overall survival					% 10-year disease-free survival				
	Predicted	Observed	s.e.	Difference (95% CI)	P-value	Predicted	Observed	s.e.	Difference (95% CI)	P-value
Young cohort										
All (n = 376)										
AOL	83.9	84.6	1.86	-0.7 (-4.3 to 3.0)	0.37	76.0	68.3	2.46	7.7 (2.8 to 12.5)	0.003
6-class NPI	76.1	84.6	1.86	-8.5 (-12.1 to -4.8)	<0.001					
ER+/HER2-(n = 181)										
AOL	86.9	87.3	2.48	-0.4 (-5.2 to 4.5)	0.39	79.2	70.8	3.46	8.4 (1.6 to 15.2)	0.02
6-class NPI	78.8	87.3	2.48	-8.5 (-13.4 to -3.7)	0.001					
ER-/HER2-(n = 86)										
AOL	79.2	79.1	4.39	0.1 (-8.5 to 8.7)	0.40	72.2	69.0	5.07	3.2 (-6.7 to 13.2)	0.33
6-class NPI	73.8	79.1	4.39	-5.3 (-13.9 to 3.3)	0.19					
HER2+(n = 74)										
AOL	79.8	81.1	4.55	-1.3 (-10.2 to 7.6)	0.38	72.2	58.2	5.88	14.0 (2.5 to 25.6)	0.02
6-class NPI	68.7	81.1	4.55	-12.4 (-21.3 to -3.5)	0.01					
Older cohort										
All (n = 907)										
AOL	81.2	88.4	1.06	-7.2 (-9.3 to -5.1)	<0.001	73.4	76.6	1.44	-3.2 (-5.9 to -0.3)	0.04
6-class NPI	84.4	88.4	1.06	-4.0 (-6.1 to -2.00)	<0.001					
ER+/HER2-(n = 591)										
AOL	82.7	89.7	1.25	-7.0 (-9.5 to -4.6)	<0.001	74.7	79.0	1.72	-4.3 (-7.6 to -0.9)	0.02
6-class NPI	86.2	89.7	1.25	-3.5 (-5.9 to -1.0)	0.009					
ER-/HER2-(n = 76)										
AOL	71.1	81.6	4.45	-10.5 (-19.2 to -1.8)	0.03	65.5	71.9	5.22	-6.4 (-16.6 to 3.9)	0.19
6-class NPI	73.6	81.6	4.45	-8.0 (-16.7 to 0.7)	0.08					
HER2+(n = 102)										
AOL	78.4	83.3	3.69	-4.9 (-12.2 to 2.3)	0.16	71.4	68.0	4.78	3.4 (-6.0 to 12.8)	0.31
6-class NPI	78.9	83.3	3.69	-4.4 (-11.6 to 2.8)	0.20					

Abbreviations: AOL = Adjuvant! Online; CI = confidence interval; ER = oestrogen receptor; HER2 = human epidermal growth factor receptor 2; NPI = Nottingham Prognostic Index.

In the older cohort, both tools underestimated 10-year OS by 7.2% for AOL ($P < 0.001$) and 4.0% for NPI ($P < 0.001$; Table 2). A greater discrepancy was shown for both tools in patients with ER+/HER2- (absolute difference for AOL: -7.0%; $P < 0.001$; absolute difference for NPI: -3.5%; $P = 0.009$) and ER-/HER2- (absolute difference for AOL: -10.5%; $P = 0.03$; absolute difference for NPI: -8.0%; $P = 0.08$) tumours.

In both the young (AOL: $P = 0.98$; NPI: $P = 0.53$) and older (AOL: $P = 0.63$; NPI: $P = 0.61$) cohorts, there was no significant heterogeneity in the observed vs predicted OS difference across the three subtypes.

The calibration plots showing the observed vs predicted 10-year OS in both the young and older cohorts are reported in Supplementary Figures A3 and A4 for AOL and in Supplementary Figures A5 and A6 for NPI, respectively.

Disease-free survival. In the young cohort, there was a discrepancy between predicted (76.0%) and observed (68.3%) 10-year DFS for AOL, with a significant overestimation of 7.7% ($P = 0.003$). The discrepancy in AOL estimation was mainly observed in patients with ER+/HER2- (absolute difference: 8.4%; $P = 0.02$) and HER2+ disease (absolute difference: 14.0%; $P = 0.02$; Table 2).

Conversely, AOL underestimated 10-year DFS by 3.2% in older patients ($P = 0.04$; Table 2). A greater discrepancy was shown in women with ER+/HER2- disease (absolute difference: -4.3%; $P = 0.02$).

In both the young ($P = 0.38$) and the older ($P = 0.28$) cohorts, there was no significant heterogeneity in the observed vs predicted DFS difference across the three subtypes.

The calibration plots showing the observed vs predicted 10-year DFS in both the young and older cohorts are reported in Supplementary Figures A7 and A8.

Discriminatory accuracy

Overall survival. Tool discriminatory accuracy was 73.1 (95% CI, 66.1-80.2) for AOL and 71.9 (95% CI, 65.0-78.7) for NPI in the young cohort ($P = 0.45$; Table 3 and Figure 3A). There was tendency of higher accuracy for AOL over NPI in patients with HER2+ tumours ($P = 0.06$).

No difference in tool discriminatory accuracy was shown between AOL (74.0; 95% CI, 69.0-79.1) and NPI (74.1; 95% CI, 69.1-79.1) in the older cohort ($P = 0.94$; Table 3 and Figure 3B). This was consistent across the different breast cancer subtypes.

Disease-free survival. No difference in the discriminatory accuracy was shown between AOL (66.0; 95% CI, 59.6-72.4) and NPI (64.6; 95% CI, 58.2-71.1) in the young cohort ($P = 0.51$; Table 3 and Figure 3C). However, for patients with HER2+ tumours, discriminatory accuracy of AOL was significantly better than NPI ($P = 0.03$).

No difference in tool discriminatory accuracy was shown between AOL (63.5; 95% CI, 58.8-68.2) and NPI (64.9; 95% CI, 60.3-69.5) in the older cohort ($P = 0.35$; Table 3 and Figure 3D). This was consistent across the different breast cancer subtypes.

Despite the observed trend in HER2+ tumours in the young cohort, no significant interaction was observed between cohorts and the prognostic performance of AOL and NPI across the different subtypes.

Table 3. Area under the receiver-operator characteristic curve for 10 years regarding overall survival and disease-free survival for Adjuvant! Online and Nottingham Prognostic Index by age group and tumour subtype

	Overall survival		Disease-free survival	
	AUC for time-point 10 years (95% CI)	P-value AOL vs NPI	AUC for time-point 10 years (95% CI)	P-value AOL vs NPI
Young cohort				
All (n = 376)				
AOL	73.1 (66.1-80.2)	0.45	66.0 (59.6-72.4)	0.51
6-class NPI	71.9 (65.0-78.7)		64.6 (58.2-71.1)	
ER+/HER2- (n = 181)				
AOL	67.2 (56.2-78.2)	0.44	64.2 (54.9-73.5)	0.45
6-class NPI	69.0 (58.6-79.4)		66.5 (57.5-75.5)	
ER-/HER2- (n = 86)				
AOL	76.4 (63.9-88.8)	0.91	60.9 (46.9-74.9)	0.90
6-class NPI	76.7 (63.8-89.6)		60.4 (46.1-74.7)	
HER2+ (n = 74)				
AOL	75.4 (60.7-90.1)	0.06	73.8 (60.5-87.1)	0.03
6-class NPI	68.2 (51.1-85.2)		62.9 (48.0-77.8)	
Older cohort				
All (n = 907)				
AOL	74.0 (69.0-79.1)	0.94	63.5 (58.8-68.2)	0.35
6-class NPI	74.1 (69.1-79.1)		64.9 (60.3-69.5)	
ER+/HER2- (n = 591)				
AOL	71.9 (64.9-78.9)	0.80	63.7 (57.5-69.8)	0.59
6-class NPI	72.2 (65.3-79.1)		64.8 (58.8-70.8)	
ER-/HER2- (n = 76)				
AOL	61.5 (45.8-77.3)	0.98	59.8 (44.1-75.5)	0.31
6-class NPI	61.6 (45.6-77.6)		63.6 (48.2-79.0)	
HER2+ (n = 102)				
AOL	76.2 (62.8-89.5)	0.24	63.0 (49.7-76.2)	0.22
6-class NPI	80.2 (68.9-91.4)		68.4 (56.0-80.9)	

Abbreviations: AOL = Adjuvant! Online; AUC = area under the receiver-operator characteristic curve; CI = confidence interval; ER = oestrogen receptor; HER2 = human epidermal growth factor receptor 2; NPI = Nottingham Prognostic Index.

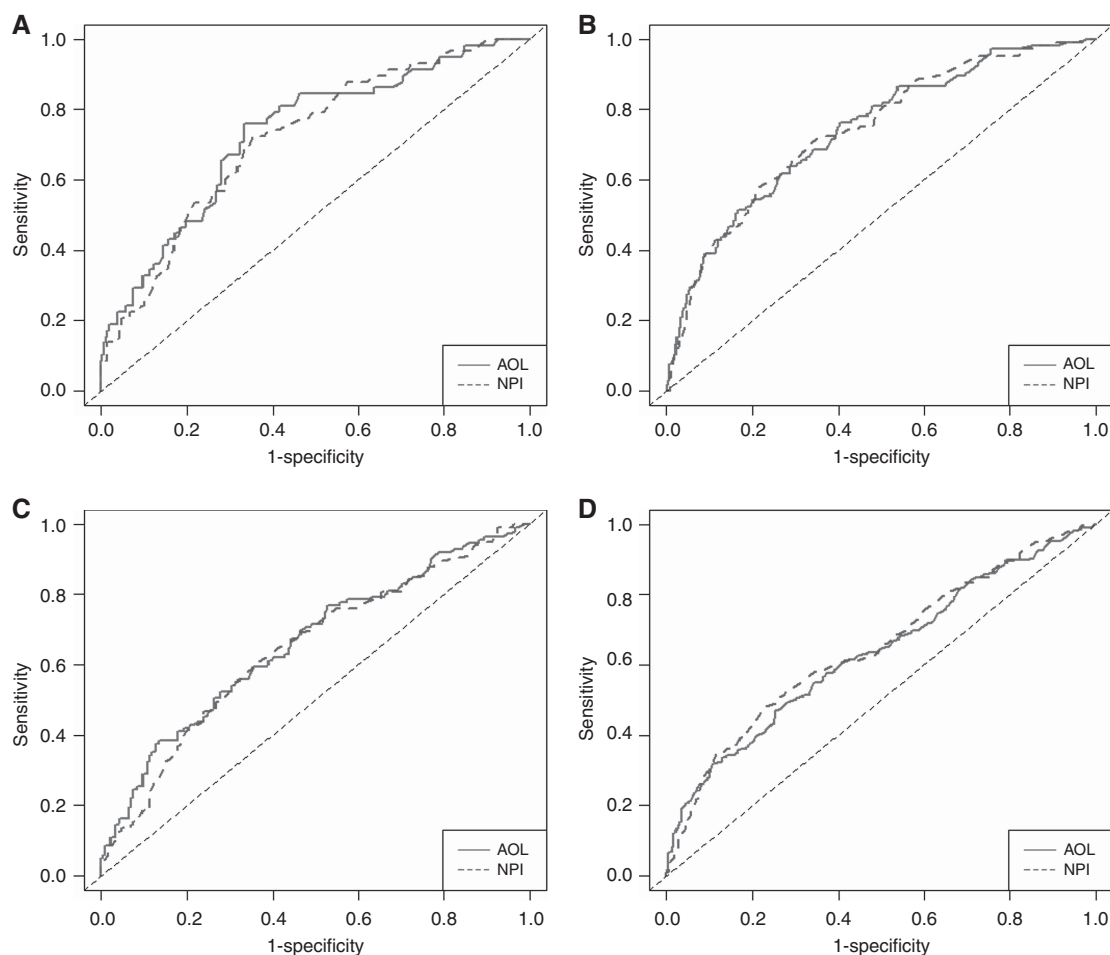


Figure 3. Receiver-operator characteristic (ROC) plot of Adjuvant! Online and Nottingham Prognostic Index to predict 10-year overall survival in patients in the young cohort (A) and in the older cohort (B); ROC plot of Adjuvant! Online and Nottingham Prognostic Index to predict 10-year disease-free survival in patients in the young cohort (C) and in the older cohort (D). AOL = Adjuvant! Online; NPI = Nottingham Prognostic Index.

DISCUSSION

Our study showed that AOL is a reliable tool in predicting OS at 10 years in young breast cancer patients but not DFS, whereas the performance of NPI is sub-optimal. Both tools overestimated the risk of death in the older cohort. We also showed that AOL and NPI had comparable discriminatory accuracy.

As expected, patients in the young cohort were diagnosed with more advanced and biologically aggressive tumours as compared with those in the older cohort, and were more often treated with adjuvant chemotherapy (85% vs 47%). However, their survival outcomes were quite impressive with a 10-year OS of 84.6% reaching up 87.3% in the subgroup of women with luminal disease. Ten-year OS of HER2+ patients was 81.1%, which is expected to be even better in the current era with trastuzumab (Vici *et al*, 2014). The cross-check of OS information with national registries further strengthens the robustness of this observation.

Adjuvant! Online and NPI are widely accepted as reliable prognostic tools that could aid treatment decision-making in the adjuvant setting (Engelhardt *et al*, 2014). However, only a small proportion of patients in the validation populations were younger than 40 years (Engelhardt *et al*, 2014). In addition, these studies suggested that AOL estimates might be less accurate in young women (Olivotto *et al*, 2005; Campbell *et al*, 2009; Mook *et al*, 2009; Hajage *et al*, 2011; Bhoo-Pathy *et al*, 2012), and the software alerts to

inaccuracy of results in this subgroup (Mook *et al*, 2009). Specifically, in a large study reporting the prognostic performance of AOL in young patients ($n = 592$), AOL was found to overestimate 10-year OS by 4.2% (Mook *et al*, 2009). Of note, patients enrolled in this study were treated in a relatively older era, approximately half of them received no adjuvant systemic therapy, and most of the treated patients had received CMF-based chemotherapy and tamoxifen as adjuvant treatments. Hence, the clinical validity of AOL predictions in young breast cancer patients treated in the modern era could not be ascertained. Moreover, in the study by Mook *et al* (2009), no information on DFS nor on the discriminatory accuracy of the tool in the young population was provided. On the other hand, very limited evidence is available on the prognostic performance of NPI in young breast cancer patients (Sundquist *et al*, 1999). Thus, we believe that our results provide important information on the reliability of these tools in young women that could help physicians in managing their patients in daily practice.

Recently, a single-centre observational study evaluated the prognostic performance of AOL and NPI in the specific subgroup of young breast cancer patients ($n = 92$) (Hearne *et al*, 2015). No significant difference between the actual 10-year OS (77.2%) and that predicted by AOL (82.1%) or NPI (77.3%) was observed (Hearne *et al*, 2015). Both tools showed a strong correlation and predicted cumulative survival curves accurately; however, this study did not include a control group and no information on 10-year DFS estimates or discriminatory accuracy of the tools was provided. Our study, which included nearly

four times the number of young patients, confirmed that AOL accurately predicts 10-year OS ($P = 0.37$). However, we showed that AOL overestimated 10-year DFS by 7.7% ($P = 0.003$). As observed in previous studies evaluating the performance of AOL in unselected breast cancer patients, DFS estimates are less accurate than OS prediction (Engelhardt *et al*, 2014). Specifically, as shown in the study by Olivotto *et al* (2005), a 14% overestimation of DFS was observed in patients ($n = 127$) under the age of 35 years. Finally, our study showed the ability of NPI to identify six classes of patients with distinct outcome discriminating between those with good and poor prognosis ($P < 0.001$; Figure 2). However, NPI significantly underestimated 10-year OS ($P < 0.001$): yet, it should be noted that we observed higher survival rates in young breast cancer patients as compared with what was previously presented in older women for the six classes of NPI (Quintyne *et al*, 2013), particularly for the two moderate groups and, above all, for the two poor prognostic groups (Supplementary Table A1).

Adjuvant! Online and NPI had comparable discriminatory accuracy for both OS and DFS in the cohort of young patients, although there was a trend for AOL being slightly better than NPI in patients with HER2+ tumours. Of note, both tools were developed prior to the trastuzumab era and HER2 status is not included in the prognostic calculation. Hence, their performance in patients with HER2+ disease should be considered with caution and remains an open question. The upcoming version of AOL incorporating information on HER2 status is largely awaited.

Our results raise some concerns regarding the prognostic performance of AOL and NPI in the 55–60 years patient cohort, with both tools underestimating survival outcomes, mainly in patients with ER+/HER2– breast cancer. Recently, several molecular assays became available and appear to refine prognostication beyond what can be provided by clinical prognostic models such as AOL in patients with ER+/HER2– tumours (Azim Jr *et al*, 2013). With the increasing use of these assays, it is likely that the reliability on tools such as AOL and NPI would be reduced. Nevertheless, it is crucial to understand the prognostic accuracy of these tools being ‘free’ and easy to use and their integration with the molecular assays (Cardoso *et al*, 2016). In addition, they remain informative in ER– tumours, in which molecular assays are of very limited value, if any.

Some limitations of the present study should be acknowledged. The retrospective design led the study prone to bias and confounding. Approximately 20% of patients did not have 10-year follow-up data for DFS, and for 13.5% of the tumours the specific subtype could not be defined. However, the study has important strength points. It is the first study investigating calibration and discriminatory accuracy of AOL and NPI in young breast cancer patients. It was a multicentre study, including a large cohort of young women and a comparator group of older patients. Overall survival data have been cross-checked with the national registries to confirm the reliability of the findings: hence, the study results can give a point estimate for expected OS and benefit from modern adjuvant treatments in young women with breast cancer.

In conclusion, our results suggest that AOL can reliably be used for OS prognostication in young women with breast cancer, but is less performant for DFS estimate. In this population, NPI is sub-optimally performing. Due to the relative underestimation of long-term outcomes in patients aged 55–60 years, the role of AOL and NPI deserves further investigations.

ACKNOWLEDGEMENTS

Matteo Lambertini is supported by a Translational Research Fellowship from the European Society for Medical Oncology (ESMO). No funding was received for this study.

CONFLICT OF INTEREST

HAA reports employment at Innate Pharma at the end of this study; this employment is not related in any sort to the subject of the current study. All the remaining authors declare no conflict of interest.

REFERENCES

- Azim HA Jr, Davidson NE, Ruddy KJ (2016) Challenges in Treating Premenopausal Women with Endocrine-Sensitive Breast Cancer. *Am Soc Clin Oncol Educ Book* 35: 23–32.
- Azim HA Jr, de Azambuja E, Colozza M, Bines J, Piccart MJ (2011) Long-term toxic effects of adjuvant chemotherapy in breast cancer. *Ann Oncol* 22: 1939–1947.
- Azim HA Jr, Michiels S, Zagouri F, Delaloge S, Filipits M, Namer M, Neven P, Symmans WF, Thompson A, André F, Loi S, Swanton C (2013) Utility of prognostic genomic tests in breast cancer practice: the IMPAKT 2012 Working Group Consensus Statement. *Ann Oncol* 24: 647–654.
- Azim HA Jr, Partridge AH (2014) Biology of breast cancer in young women. *Breast Cancer Res* 16: 427.
- Bhoo-Pathy N, Yip C-H, Hartman M, Saxena N, Taib NA, Ho G-F, Looi L-M, Bulgiba AM, van der Graaf Y, Verkooijen HM (2012) Adjuvant! Online is overoptimistic in predicting survival of Asian breast cancer patients. *Eur J Cancer* 48: 982–989.
- Campbell HE, Taylor MA, Harris AL, Gray AM (2009) An investigation into the performance of the Adjuvant! Online prognostic programme in early breast cancer for a cohort of patients in the United Kingdom. *Br J Cancer* 101: 1074–1084.
- Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, Pierga J-Y, Brain E, Causeret S, DeLorenzi M, Glas AM, Golinopoulos V, Goulioti T, Knox S, Matos E, Meulemans B, Neijenhuis PA, Nitz U, Passalacqua R, Ravdin P, Rubio IT, Saghachian M, Smilde TJ, Sotiriou C, Stork L, Straehle C, Thomas G, Thompson AM, van der Hoeven JM, Vuylsteke P, Bernardis R, Tryfonidis K, Rutgers E, Piccart M. MINDACT Investigators (2016) 70-Genes signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med* 375: 717–729.
- Chiang CT, Hung H (2010) Non-parametric estimation for time-dependent AUC. *J Stat Plan Inference* 140: 1162–1174.
- Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, Thürlimann B, Senn H-J. Panel Members (2015) Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Ann Oncol* 26: 1533–1546.
- DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A (2016) Breast cancer statistics, 2015: convergence of incidence rates between black and white women. *CA Cancer J Clin* 66: 31–42.
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. STROBE Initiative (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 370: 1453–1457.
- Engelhardt EG, Garvelink MM, de Haes JHCJM, van der Hoeven JJM, Smets EMA, Pieterse AH, Stiggelbout AM (2014) Predicting and communicating the risk of recurrence and death in women with early-stage breast cancer: a systematic review of risk prediction models. *J Clin Oncol* 32: 238–250.
- Engström MJ, Opdahl S, Hagen AI, Romundstad PR, Akslen LA, Haugen OA, Vatten LJ, Bofin AM (2013) Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients. *Breast Cancer Res Treat* 140: 463–473.
- de Glas NA, van de Water W, Engelhardt EG, Bastiaannet E, de Craen AJM, Kroep JR, Putter H, Stiggelbout AM, Weijl NI, van de Velde CJH, JEA Portielje, Liefers G-J (2014) Validity of adjuvant! Online program in older patients with breast cancer: a population-based study. *Lancet Oncol* 15: 722–729.
- Hajage D, de Rycke Y, Bollet M, Savignoni A, Caly M, Pierga J-Y, Horlings HM, Van de Vijver MJ, Vincent-Salomon A, Sigal-Zafrani B, Senechal C, Asselain B, Sastre X, Reyat F (2011) External validation of adjuvant! Online breast cancer prognosis tool. Prioritising recommendations for improvement. *PLoS One* 6: e27446.

- Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, Nicholson RI, Griffiths K (1982) A prognostic index in primary breast cancer. *Br J Cancer* **45**: 361–366.
- Hearne BJ, Teare MD, Butt M, Donaldson L (2015) Comparison of Nottingham Prognostic Index and adjuvant online prognostic tools in young women with breast cancer: review of a single-institution experience. *BMJ Open* **5**: e005576.
- Lambertini M, Del Mastro L, Pescio MC, Andersen CY, Azim HA, Peccatori FA, Costa M, Revelli A, Salvagno F, Gennari A, Ubaldi FM, La Sala GB, De Stefano C, Wallace WH, Partridge AH, Anserini P (2016) Cancer and fertility preservation: international recommendations from an expert meeting. *BMC Med* **14**: 1.
- Mook S, Schmidt MK, Rutgers EJ, van de Velde AO, Visser O, Rutgers SM, Armstrong N, van't Veer LJ, Ravdin PM (2009) Calibration and discriminatory accuracy of prognosis calculation for breast cancer with the online adjuvant! program: a hospital-based retrospective cohort study. *Lancet Oncol* **10**: 1070–1076.
- Olivetto IA, Bajdik CD, Ravdin PM, Speers CH, Coldman AJ, Norris BD, Davis GJ, Chia SK, Gelmon KA (2005) Population-based validation of the prognostic model ADJUVANT! for early breast cancer. *J Clin Oncol* **23**: 2716–2725.
- Quintyne KI, Woulfe B, Coffey JC, Gupta RK (2013) Correlation between Nottingham Prognostic Index and adjuvant! Online prognostic tools in patients with early-stage breast cancer in mid-Western Ireland. *Clin Breast Cancer* **13**: 233–238.
- Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, Parker HL (2001) Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol* **19**: 980–991.
- Sundquist M, Thorstenson S, Brudin L, Nordenskjöld B (1999) Applying the Nottingham Prognostic Index to a Swedish breast cancer population. South East Swedish Breast Cancer Study Group. *Breast Cancer Res Treat* **53**: 1–8.
- Vici P, Pizzuti L, Natoli C, Moscetti L, Mentuccia L, Vaccaro A, Sergi D, Di Lauro L, Trenta P, Seminara P, Santini D, Iezzi L, Tinari N, Bertolini I, Sini V, Mottolese M, Giannarelli D, Giotta F, Maugeri-Saccà M, Barba M, Marchetti P, Michelotti A, Sperduti I, Gamucci T (2014) Outcomes of HER2-positive early breast cancer patients in the pre-trastuzumab and trastuzumab eras: a real-world multicenter observational analysis. The RETROHER study. *Breast Cancer Res Treat* **147**: 599–607.
- This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

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