ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN CANCER HORIZONS

ORIGINAL RESEARCH

A clinical validation study of MammaPrint in hormone receptor-positive breast cancer from the Austrian Breast and Colorectal Cancer Study Group 8 (ABCSG-8) biomarker cohort

P. Dubsky[1,2*], L. Van't Veer[3], M. Gnant[4], M. Rudas[5], Z. Bago-Horvath[5], R. Greil[6], E. Lujinovic[7], J. Buresch[8], G. Rinnerthaler[6], W. Hulla[9], F. Moinfar[10], D. Egle[11], W. Herz[12], C. Dreezen[13], S. Frantal[14] & M. Filipits[4,8]

[1]Department of Surgery and Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria; [2]Breast Center St. Anna, Lucerne, Switzerland; [3]Helen Diller Family Comprehensive Cancer Center, University California San Francisco, San Francisco, USA; [4]Comprehensive Cancer Center, Medical University of Vienna, Vienna; [5]Department of Pathology, Breast Health Center and Comprehensive Cancer Center, Medical University of Vienna, Vienna; [6]Department of Internal Medicine III with Hematology, Medical Oncology, Hemostaseology, Infectious Disease, Rheumatology, Oncologic Center, Laboratory for Immunological and Molecular Cancer Research, Paracelsus Medical University, Salzburg Cancer Research Institute-CCCIT, Cancer Cluster Salzburg, Salzburg, Austria; [7]Medical Affairs Department, Agendia NV, Amsterdam, the Netherlands; [8]Institute of Cancer Research, Department of Medicine I, Medical University of Vienna, Vienna; [9]Department of Pathology, Federal Hospital Wiener Neustadt, Wiener Neustadt; [10]Department of Clinical Pathology, Ordensklinikum Linz, Barmherzige Schwestern, Linz; [11]Department of Obstetrics and Gynecology, Medical University Innsbruck, Innsbruck; [12]Department of Surgery, LKH Hochsteiermark-Leoben, Leoben, Austria; [13]Statistics Department, Agendia NV, Amsterdam, the Netherlands; [14]Department of Statistics, Austrian Breast and Colorectal Cancer Study Group, Vienna, Austria

Available online xxx

Background: MammaPrint is a prognostic assay based on gene expression in tumors from patients with early breast cancer. MammaPrint has been extensively validated and Food and Drug Administration cleared in fresh and formalin-fixed and paraffin-embedded (FFPE) tissue. We aimed to assess its prognostic performance in the biomarker cohort of the Austrian Breast and Colorectal Cancer Study Group 8 (ABCSG-8) patient population, and to obtain a higher level of evidence with regard to its clinical validity after RNA extraction from FFPE biobank tissue.

Patients and methods: A prespecified retrospective analysis to test the prognostic performance of the MammaPrint test to predict distant recurrence-free survival at 5 and 10 years as primary end point was carried out. MammaPrint risk, clinicopathological factors (after central pathological review), and clinical risk (using a modified version of Adjuvant! Online) were evaluated by Cox regression analyses.

Results: From 1347 available samples, 607 (45%) failed quality control after RNA extraction. In total, 658 (49%) patients were included in survival analyses: MammaPrint low risk versus high risk is a significant prognostic factor for distant recurrence-free survival at 5 years (94.0% versus 91.6%) with a significant risk reduction of 6.5% at 10 years (log-rank P value = 0.017, low risk 91.3% versus high risk 84.8%). The multivariable models suggest that hazard ratio (HR) is primarily driven by tumor stage (5-year HR 3.89; confidence interval 1.97-7.71) and nodal status (5-year HR 1.73; confidence interval 0.91-3.21). After adjustment for clinical risk groups, MammaPrint HRs remain stable with values just below 2.0 after the first 3 years.

Conclusions: The MammaPrint test showed significant prognostic performance at 5 and 10 years of follow-up. In the particular cohort of ABCSG-8, the statistical independence from clinically assessed covariates remains unclear, and no conclusions concerning the clinical validity of the test can be drawn.

Key words: ER-positive/HER2-negative, early breast cancer, ABCSG-8, prognostic biomarkers, MammaPrint, clinical validation

INTRODUCTION

MammaPrint (i.e. 70-gene signature) is a diagnostic assay that uses expression levels of the 70 MammaPrint genes to assess distant recurrence risk in early stage breast cancer, providing a binary low- or high-risk prediction of breast cancer recurrence.[1] The MicroarRAy PrognoSTics in Breast CancER (RASTER) trial provided the first prospective results showing that low-risk patients had a 97.0% distant

*Correspondence to: Peter C. Dubsky, Department of Surgery, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria; Breast Center St. Anna, Lucerne, Switzerland. Tel: +41-41208-3754

E-mail: leitung-brustzentrum.stanna@hirslanden.ch (P. Dubsky).

recurrence-free interval at 5 years[2] and a 93.7% distant recurrence-free interval at 10 years.[3]

Most notably, MammaPrint was validated in the 6693 patients of the prospective randomized phase III 'Microarray In Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid ChemoTherapy' (MINDACT) trial. In this study, patients who had high clinical risk and low MammaPrint genomic risk and who were treated without chemotherapy showed a 94.7% rate of distant metastasis-free survival at 5 years, thus providing the clinical evidence for clinical utility of the test by omitting chemotherapy in a relevant subgroup of women.[4] The assay has also been validated in a number of node-positive and node-negative cohorts treated outside of prospective clinical trials.[5-7] Importantly, most of this work was done on fresh tissue. One exception is the randomized prospective Stockholm Tamoxifen (STO-3) trial biomarker cohort which showed that low-risk patients who received 2-5 years of tamoxifen had an excellent 93% distant metastasis-free survival at 10 years and a 90% breast cancer-specific survival at 20 years without chemotherapy—these analyses were done on formalin-fixed and paraffin-embedded (FFPE) archived tissue.[8]

The biomarker cohort of the Austrian Breast and Colorectal Cancer Study Group 8 (ABCSG-8) has previously served as a validation cohort for prognostic gene expression tests at both early and late timepoints.[9-11] This cohort is highly representative of the entire ABCSG-8 phase III trial[12] that included postmenopausal estrogen receptor-positive (ER+) patients, with low to intermediate grading tumors (G1-G2 only) and low to intermediate clinical risk, treated with endocrine therapy in the absence of chemotherapy. The ABCSG-8 biobank offers the opportunity to test the prognostic performance of the MammaPrint test on (i) FFPE tumor samples, (ii) in a highly homogenous cohort, and (iii) with prospectively recorded and monitored clinical outcome data.

The aim of this study was to assess the prognostic performance of the MammaPrint in the biomarker cohort of the ABCSG-8 patient population and thereby obtain a higher level of evidence with regard to the clinical validity of the MammaPrint performed on FFPE tissue.

PATIENTS AND METHODS

ABCSG-8

The study population consisted of patients with retrospectively collected FFPE breast tumor samples from the ABCSG-8 trial archived in the ABCSG tumor bank. ABCSG-8 included postmenopausal patients, with lymph node-negative and lymph node-positive disease. Patients with ER-positive tumors and G1 and G2 disease were randomized to 5 years of adjuvant tamoxifen or 2 years of tamoxifen followed by 3 years of anastrozole. Adjuvant chemotherapy was an exclusion criterion.[12]

All procedures concerning the collection of samples and consent of patients to translational research have previously been described.[11]

MammaPrint assay and MammaPrint risk score

The MammaPrint gene expression profile has been previously described.[1,6] In brief, MammaPrint is a microarray-based assay that uses RNA extracted from tumor tissue (FFPE). This RNA is labeled, hybridized, and run on custom designed microarray platform, manufactured by Agilent (Santa Clara, CA), containing specific probes for 70 prognostic genes. The oligonucleotide microarrays assess the gene expression of the 70 MammaPrint and 80 BluePrint subtype genes, 465 normalization genes, and more than 250 probes for hybridization and printing quality control. Tumor classification is subsequently read out using the MammaPrint algorithm, which has been described previously.[6] The sample processing for this study was carried out in the Institute of Cancer Research, Medical University of Vienna.

Study methodology

This study was designed as a 'retrospective-prospective' subanalysis, using archived specimens from both arms of the ABCSG-8 study cohort that have been obtained at the time of surgery before adjuvant therapy.

The MammaPrint assay is Food and Drug Administration (FDA) cleared in all specimen types,[13] and the samples are processed centrally in Clinical Laboratory Improvements Amendments (CLIA)/College of American Pathologists (CAP)-certified laboratories. Diagnostic validation was performed according to FDA and National Committee for Clinical Laboratory Standards (NCCLS) guidelines.[14,15]

Sample processing for MammaPrint and BluePrint were largely performed according to standard protocols as described.[1,6] In total, 1347 FFPE tissue samples from the ABCSG-8 research tumor bank were available. The laboratory was trained and subsequently validated in accordance with the standard protocol from Agendia. Isolated RNA obtained centrally at the ABCSG-8 biorepository laboratory (Biobank) from the available tissue samples were hybridized onto the Agendia proprietary full-genome microarray slides provided by Agilent. After hybridization the slides were washed and stored under vacuum at −70°C. Slides were sent to Agendia's lab in Amsterdam, the Netherlands, under vacuum at room temperature. MammaPrint was read out (masked with respect to patient clinical data including outcome) when all RNA quality controls were passed. Drying of the slides as well as the shipment process has been thoroughly validated as having no impact on gene expression.

The patient tumors were subsequently classified into risk categories as either low risk or high risk using the thresholds previously developed.[1]

Central pathology review of ER, PR, HER2, Ki-67 and presence of ductal carcinoma *in situ* component were performed by a breast cancer-dedicated pathologist. For Ki-67, a cut-off of 14% was used to dichotomize into low and high expression.[16]

To assess the clinical risk, the modified version of Adjuvant! Online, version 8.0 with HER2 status, was used (www.adjuvantonline.com, Table S13 in Cardoso et al.[4]). The assay

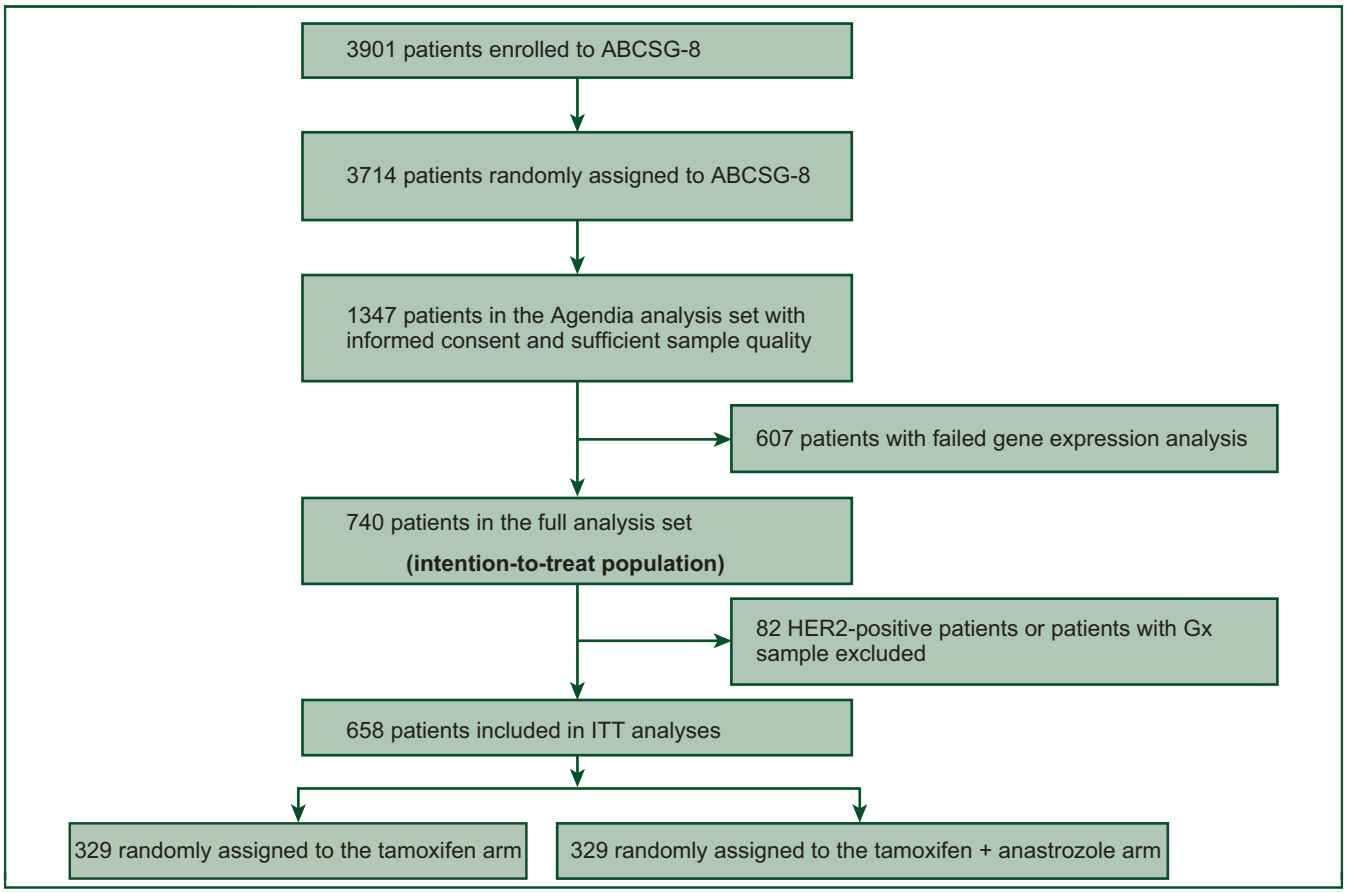

Figure 1. Consort flow chart.
Gx, grade could not be assessed; ITT, intention to treat.

results, pathology information, and clinical data were merged and analyzed at the ABCSG Statistical Center.

Statistical analysis

The primary analysis was designed to determine the prognostic performance of MammaPrint risk groups at 5 years. For this analysis, the 5-year distant recurrence was the event of interest and is defined as the time in years from random assignment to first occurrence of distant metastases, where all secondary carcinoma (including contralateral breast cancer) and death due to any cause were censored. In case of no distant metastases, secondary carcinoma, or death, the patient was censored at the last contact date. Results are reported according to the treatment assignment at the time of enrollment (intention-to-treat population, Figure 1). No values are imputed for missing data, including events. Only partially incomplete dates (day missing) are imputed for dates related to events and follow up information.

A descriptive statistics table of available clinical, pathological, therapeutic, and event variables divided in MammaPrint risk groups was set up. Categorical variables were expressed in numbers and percentages and were tested for difference with a chi-square test, or a Fisher's exact test in case cell counts are not sufficiently high. Continuous variables were expressed with median, minimum, and maximum, and a nonparametric test was used to test for differences in distribution between risk groups.

Distant recurrence rates were estimated using the Kaplan—Meier method. Reported *P* values in Kaplan—Meier figures are results of two-sided log-rank tests.

For each univariable and multivariable Cox regression analysis the proportional hazard (PH) assumption was tested by analyzing a premodel including interaction of the variable(s) of interest with time as time-dependent covariate(s). Where the time-dependent effect showed a *P* value <0.05, the time-dependent effect was additionally included in the actual model (PH-corrected Cox model). For multivariable Cox regression analyses the correlation between covariates was examined. In case of highly correlated variables, the most significant in univariable analysis was kept for the multivariable model. The multivariable models included all available clinical, pathological, and therapeutic variables, with the exception that variables may be dropped if problems with the models (e.g. quasi-complete separation) occur. Additional prognostic value of MammaPrint was assessed by partial likelihood-ratio test statistics (LRTS) and according to *P* values.

All *P* values <0.05 were considered statistically significant. Analyses were done by members of the biostatistics group at ABCSG using Statistical Analysis System (SAS) software (version 9.3 or higher).

RESULTS

Study population

From 3901 patients enrolled into ABCSG-8, 3714 were randomly assigned to endocrine treatment arms. A total of 1147 patients consented to translational biomarker studies and had adequate tumor available; the earliest FFPE blocks are dated from January 1996.

Of these, 607 (45.1%) samples failed the quality control standards for MammaPrint gene expression analysis due to degeneration of the RNA and deviation from the standardized Normscore based on control genes. A further 82 patients were excluded; 54 patients with HER2 overexpressing tumors and/or 30 patients with unknown tumor grade, mostly lobular differentiation. As a result, 658 patients were included from the intention-to-treat analyses, with an—by chance—equal distribution in the tamoxifen-alone arm versus the tamoxifen followed by anastrozole arm ($n = 329$ versus $n = 329$; Figure 1).

Demography

Mean age of the study population was 63 years. Hormone receptor expression was ER++ and PgR++ or higher in 82.4%, and 72.6% had a low Ki-67 value (<14%). Four of five patients (79.6%) had a moderately differentiated tumor grading, 65.8% had tumors smaller than 2 cm, and 68.5% showed node-negative disease. In general, the clinical risk parameter, assessed from Adjuvant! Online, classified 54% of the study population as low risk and 45% as high risk; for 0.5% clinical risk assessment was missing. Breast conservation was carried out in 79.6% and 71% received radiation therapy.

Within the 658 samples, 77.8% showed a MammaPrint low-risk profile. This profile was typically and significantly associated with less aggressive tumor features such as high ER expression, low Ki-67, good differentiation, and tumor size below 2 cm (Table 1). It is noteworthy that node-positive disease was only slightly elevated in the genomic high-risk group. However, the clinical risk score assessment showed that 41.6% of genomic low-risk women had high-risk clinical scores.

Univariable analyses of distant recurrence

Analysis of survival data showed that for patients in the MammaPrint low-risk group, the 5- and 10-year distant recurrence-free survival was 94.0% and 91.3%, respectively. For patients in the high-risk group, the 5- and 10-year survival was 91.6% and 84.8%, respectively. The Kaplan—Meier curve shows a statistically significant distinction between both groups' survival data ($P = 0.0171$, absolute risk reduction at 10 years 6.5%, at 5 years 2.4%; Figure 2).

The unadjusted hazard ratios (HRs) in high- versus low-risk patient groups as defined by clinicopathologic factors and MammaPrint are shown in Table 2. The results are based on a univariable Cox regression analysis for distant recurrence at 5 and 10 years. The PH assumption was met for each variable at both timepoints but was violated for the MammaPrint variable at the 5-year timepoint. Therefore, an additional time-dependent effect was included in the 5-year Cox model with MammaPrint (PH-corrected Cox model).

Results from the PH-corrected 5-year Cox model show that the prognostic value of MammaPrint was stronger than most of the individual traditional risk factors and variables in this analysis [HR 22.0, 95% confidence interval (CI) 3.26-148; $P = 0.0015$]. By contrast, the effect at 5 years without time-dependent modeling was not statistically significant (HR 1.52, 95% CI 0.78-2.98; $P = 0.2224$), whereas the genomic high-risk group showed almost a twofold increase in the distant recurrence risk when censored at 10 years (HR 1.91, 95% CI 1.11-3.27; $P = 0.0191$).

Other clinical variables that were significant were T-stage, N-stage, and Ki-67 (for 10 years only). The clinical risk classification parameter, mainly based on T-stage and N-stage, showed significant prognostic value.

Multivariable analyses of distant recurrence

In analogy to the univariable analysis a time-dependent effect for MammaPrint (PH-corrected model censored at 5 years) needed to be included into the 5-year multivariable model, but not into the 10-year multivariable model (Table 3): Results censored at 5 years show that the prognostic value of MammaPrint was stronger than most of the individual traditional risk factors and variables in the multivariable analysis (HR 13.3, 95% CI 1.92-92.7; $P = 0.0088$). Interestingly, with the exception of T-stage (which showed significant HR effects in all models at 5 and 10 years) and nodal status (HR 1.7 at both timepoints; significant at 10 years), none of the other variables (age, hormone receptor, and Ki-67) had significant independent prognostic value.

Because the predefined model selection process according to PH violation resulted in artificially high HR and high CI, we provide an overview of all results without PH correction in Supplementary Figure 1A and B, available at https://doi.org/10.1016/j.esmoop.2020.100006. These results without time-dependent modeling do not show a major prognostic impact of MammaPrint; the model is mostly driven by the anatomic features tumor size and lymph node status. The additional prognostic value of MammaPrint to a model including T-stage and N-stage is small (LRTS = 0.26, $P = 0.61$ at 5 years and LRTS = 2.85, $P = 0.09$ at 10 years). This underlines that the highly statistically significant results (with high HRs and CIs) in the corrected univariable Cox models for MammaPrint seem to be artificial and based on calculational effects.

Having established T-stage and N-stage as stable prognostic factors for distant recurrence, HRs of MammaPrint adjusted for clinical risk classification for these clinical factors were calculated (Figure 3). From the timepoint there appear enough events (>3 years) MammaPrint shows a fairly stable HR (<2.0) over time.

DISCUSSION

The main aims of this study were to validate the molecular biomarker MammaPrint assessed from FFPE archived

Table 1. Demography

Variable	MammaPrint Low risk *N* = 512 (77.8%)	MammaPrint High risk *N* = 146 (22.2%)	Total *N* = 658	*P* value
Randomized treatment arm				
Tam	257 (50.2)	72 (49.3)	329 (50.0)	0.8512*
Tam + Ana	255 (49.8)	74 (50.7)	329 (50.0)	
Age, years	63.0 (46.0-79.0)	64.0 (41.0-79.0)	63.0 (41.0-79.0)	0.2181**
Estrogen receptor				
Negative	2 (0.4)	9 (6.2)	11 (1.7)	<0.0001***
Positive	508 (99.2)	136 (93.2)	644 (97.9)	
Missing	2 (0.4)	1 (0.7)	3 (0.5)	
Progesterone receptor				
Negative	26 (5.1)	17 (11.6)	43 (6.5)	0.0026*
Positive	482 (94.1)	121 (82.9)	603 (91.6)	
Missing	4 (0.8)	8 (5.5)	12 (1.8)	
Hormone receptor				
Low HR[a]	73 (14.3)	33 (22.6)	106 (16.1)	0.0106*
High HR[b]	434 (84.8)	108 (74.0)	542 (82.4)	
Missing	5 (1.0)	5 (3.4)	10 (1.5)	
Ki-67				
Low[c]	416 (81.3)	62 (42.5)	478 (72.6)	<0.0001*
High[d]	96 (18.8)	84 (57.5)	180 (27.4)	
Tumor grade				
G1	115 (22.5)	19 (13.0)	134 (20.4)	0.0124*
G2	397 (77.5)	127 (87.0)	524 (79.6)	
T-stage				
pT1	360 (70.3)	73 (50.0)	433 (65.8)	<0.0001***
pT2	146 (28.5)	71 (48.6)	217 (33.0)	
pT3	6 (1.2)	2 (1.4)	8 (1.2)	
N-stage				
N0	355 (69.3)	96 (65.8)	451 (68.5)	0.0179***
N1	147 (28.7)	40 (27.4)	187 (28.4)	
N2	10 (2.0)	10 (6.8)	20 (3.0)	
Clinical risk				
C-low	297 (58.0)	61 (41.8)	358 (54.4)	0.0006*
C-high	213 (41.6)	84 (57.5)	297 (45.1)	
Missing	2 (0.4)	1 (0.7)	3 (0.5)	
Radiotherapy				
No	152 (29.7)	39 (26.7)	191 (29.0)	0.4848*
Yes	360 (70.3)	107 (73.3)	467 (71.0)	
Type of surgery				
Breast conserving	409 (79.9)	115 (78.8)	524 (79.6)	0.7678*
Mastectomy	103 (20.1)	31 (21.2)	134 (20.4)	

Categorical variables are expressed as *n* (%) and continuous variables as median (range). Missing categories are not used for testing.
ER, estrogen receptor; HR, hormone receptor; PR, progesterone.
[a] High HR: ER++ and PgR++ or higher.
[b] Low HR: lower than ER++ or PR++.
[c] Low Ki-67: <14%.
[d] High Ki-67: ≥14%.
* Chi-square test.
** Wilcoxon test.
*** Fisher's test.

samples in a homogenous cohort of hormone receptor-positive and HER2-negative patients. The biomarker cohort of ABCSG-8 provided the opportunity to study the genomic biomarker in comparison to clinical and pathologic variables in a prospectively randomized, ER-positive, G1/G2 cohort of patients treated with endocrine therapy.

The earliest FFPE blocks used date from January 1996: 45.1% of samples (601/1347) failed the quality control standards for MammaPrint gene expression analysis due to degeneration of the RNA and deviation from the standardized Normscore based on control genes. The long storage may have impacted the RNA quality and analysis of the 70-gene test. Analysis of more recent samples[17] have not shown such lack of robustness but certainly the relatively high sample failure rate in this cohort impacted on the results of this preplanned statistical analysis (Figure 1).

Almost 78% of all valid samples showed a low-risk genomic score; this was confirmed by a 5-year point estimate of distant recurrence of 94% (10-year distant recurrence rate 91.3%) and a statistically significant difference to the high-risk group with an absolute difference in risk at 10 years of 6.5%. The primary end point of the study, described as the prognostic performance of the test at 5 years, has thus been met (Figure 2).

The clinical validity of the test is however additionally judged by its independence from routinely assessed

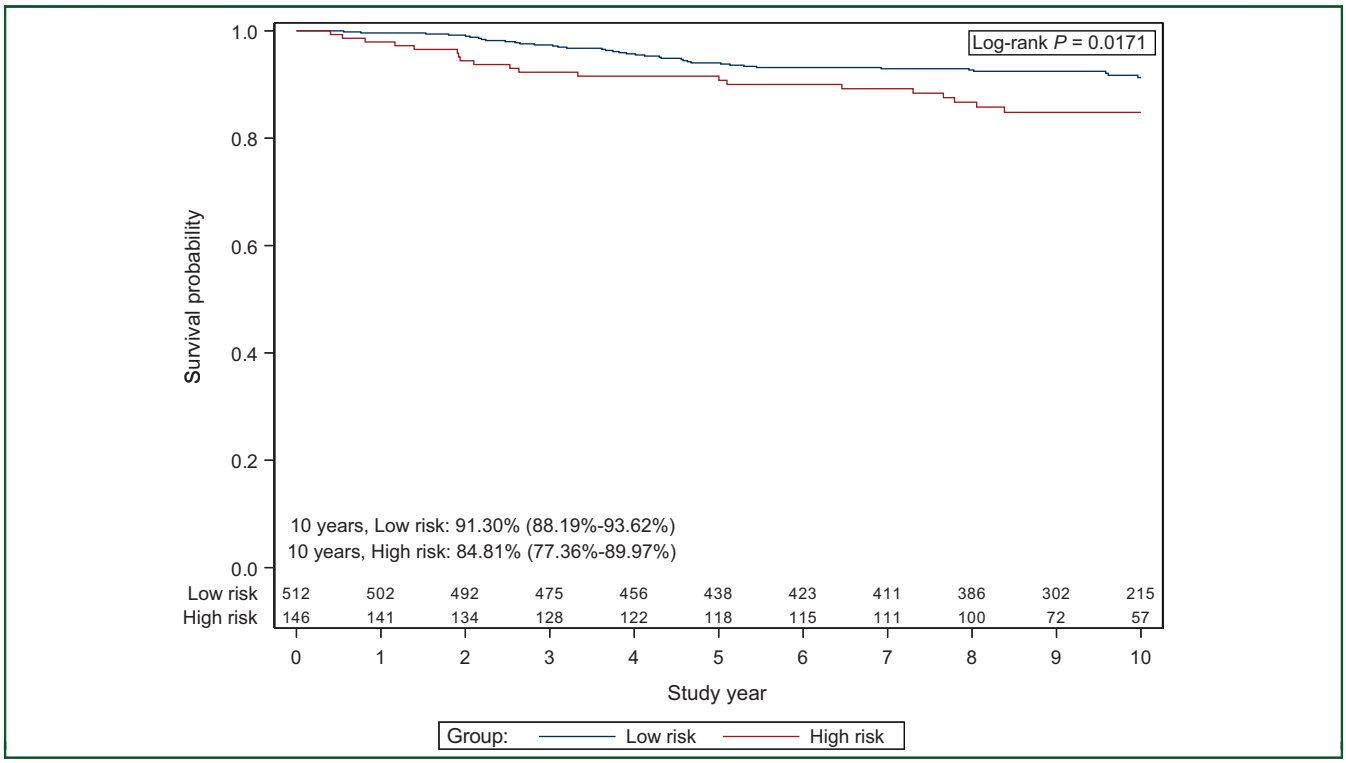

Figure 2. **Time to first distant recurrence (censored at 10 years) by the MammaPrint risk group.**

Table 2. **Univariable Cox regression analysis, 5-year, and 10-year distant recurrence censoring**

Variable	*N*	Censored at 5 years		Censored at 5 years PH corrected		Censored at 10 years	
		Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
MammaPrint[a]							
Low risk	512	1		1		1	
High risk	146	1.52 (0.78-2.98)	0.2224	22.0 (3.26-148)	0.0015	1.91 (1.11-3.27)	0.0191
Age, years	658	1.02 (0.98-1.06)	0.3741			1.02 (0.99-1.05)	0.1956
HR							
Lower ER++ or PR++	106	1				1	
Higher ER++ and PR++	542	0.56 (0.27-1.14)	0.1098			0.61 (0.33-1.13)	0.1179
Ki-67							
Low	478	1				1	
High	180	1.79 (0.95-3.35)	0.0698			2.11 (1.26-3.54)	0.0045
Tumor grade							
G1	134	1				1	
G2	524	0.91 (0.43-1.90)	0.8003			1.26 (0.64-2.48)	0.5117
T-stage							
pT1	433	1				1	
pT2/pT3	225	4.08 (2.14-7.78)	<0.0001			2.94 (1.75-4.93)	<0.0001
N-stage							
Negative	451	1				1	
Positive	207	2.21 (1.20-4.08)	0.0110			2.01 (1.20-3.35)	0.0076
Clinical risk							
C-low	358	1				1	
C-high	297	4.10 (2.01-8.36)	0.0001			3.10 (1.78-5.41)	<0.0001

CI, confidence interval; ER, estrogen receptor; HR, hormone receptor; *N*, number of patients in analysis set with available covariate information; distant recurrence including distant metastases; censored for secondary carcinoma and death due to any cause; PR, progesterone.

[a] Proportional hazard (PH)-corrected model for MammaPrint at 5 years.

variables: taking into account the violation of the proportional risk assumption over time, there is a statistically independent contribution of the test at 5 years. These data are not confirmed when censored at 10 years (Table 2) and the multivariable model without correction would suggest that tumor size and nodal status in this cohort of patients, rather than proliferation index and hormone receptor status, drive the occurrence of events. This finding is in contrast to prior validation studies carried out on cohorts with a more heterogeneous patient

Table 3. Multivariable Cox regression analysis, 5- and 10-year distant recurrence censoring

Variable	*N*	Censored at 5 years		Censored at 5 years PH correction		Censored at 10 years	
		Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
MammaPrint[a]							
Low risk	507	1		1		1	
High risk	141	1.06 (0.50-2.22)	0.8798	13.3 (1.92-92.7)	0.0088	1.28 (0.69-2.35)	0.4368
Age, years	648	1.01 (0.97-1.05)	0.5621	1.01 (0.97-1.05)	0.5178	1.01 (0.98-1.05)	0.4164
HR							
Lower ER++ or PR++	106	1		1		1	
Higher ER++ and PR++	542	0.65 (0.32-1.34)	0.2453	0.65 (0.32-1.35)	0.2505	0.71 (0.38-1.33)	0.2840
Ki-67							
Low	473	1		1		1	
High	175	1.62 (0.80-3.24)	0.1775	1.60 (0.80-3.22)	0.1859	1.74 (0.97-3.12)	0.0633
Tumor grade							
G1	132	1		1		1	
G2	516	0.67 (0.32-1.43)	0.3026	0.69 (0.33-1.47)	0.3375	0.91 (0.46-1.83)	0.8016
T-stage							
pT1	427	1		1		1	
pT2/pT3	221	3.89 (1.97-7.71)	<0.0001	3.84 (1.94-7.61)	0.0001	2.53 (1.47-4.36)	0.0008
N-stage							
Negative	444	1		1		1	
Positive	204	1.73 (0.92-3.24)	0.0894	1.71 (0.91-3.21)	0.0959	1.72 (1.01-2.92)	0.0453

CI, confidence interval; ER, estrogen receptor; HR, hormone receptor; PH, proportional hazard; PR, progesterone. distant recurrence including distant metastases; censored for secondary carcinoma and death due to any cause; *N*, number of patients in the analysis set with available covariate information.
[a] Variables included: MammaPrint (low risk and high risk), age, HR (low and high), Ki-67 (low and high), tumor grade (G1 and G2), T-stage (pT1 and pT2/pT3), and N-stage (negative and positive).

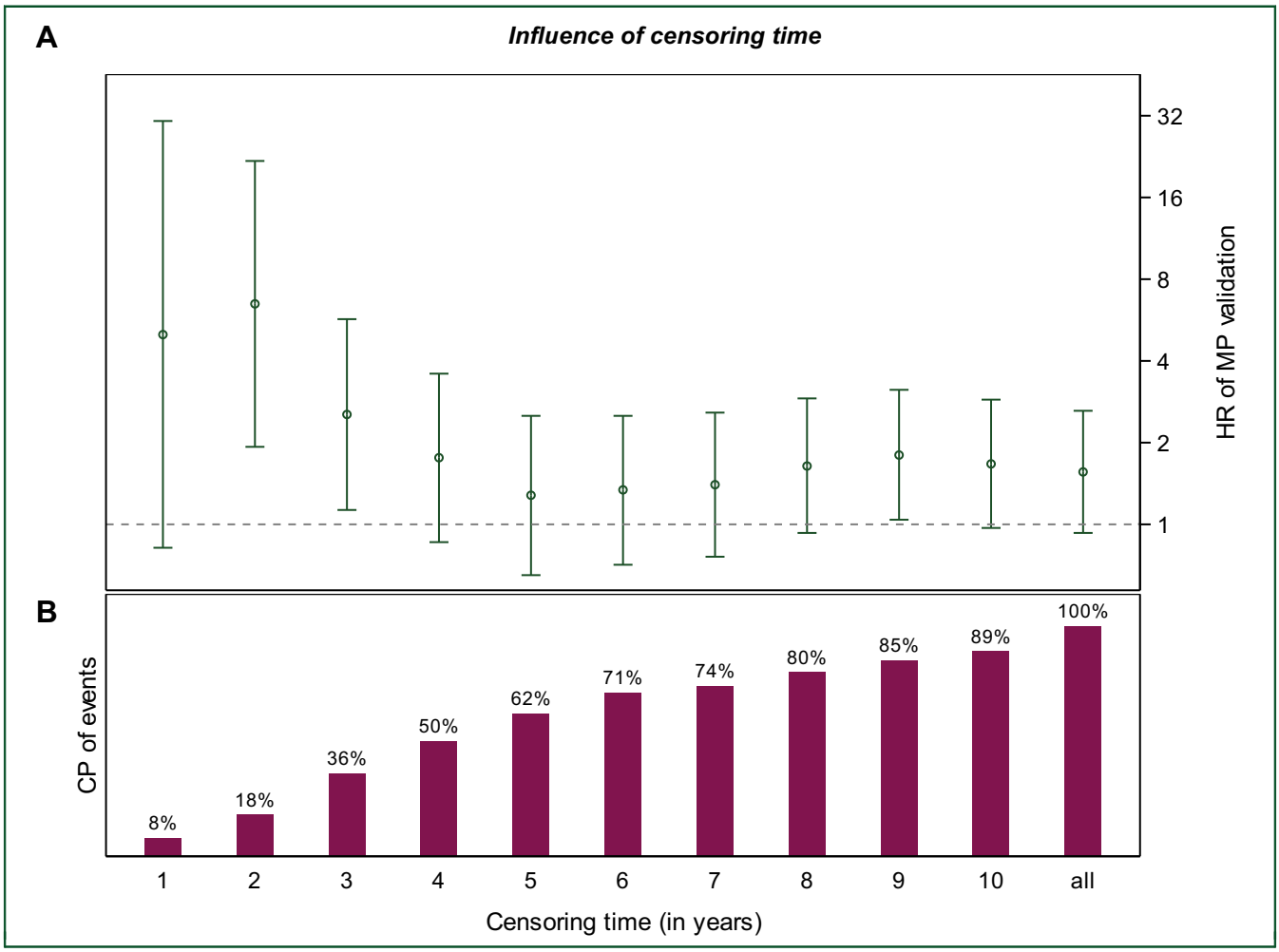

Figure 3. Censoring time, events, and HR for distant recurrence over time.
(A) HR and 95% confidence intervals for time to DR comparing MP high risk versus low risk, adjusted for clinical risk according to tumor size and nodal status and for increasing arbitrary censoring times. (B) Cumulative proportions of events over all time points. CP, cumulative proportion; DR, distant recurrence; HR, hazard ratio; MP, MammaPrint.

population, including such with node-positive patients[18] and including a single study that was carried out on FFPE archived samples.[19]

In contrast to prior validation studies in the ABCSG-8 biomarker cohort, this analysis comprises a small subset. Figure 3 illustrates large CIs due to the small number of events, especially during the first 3 years. In addition, the timing of events plays a major role in outcome analyses and this subset may have led to a different distribution over time compared with other validation analyses using the entire cohort. Indeed, in other validation studies we had not experienced a PH violation over time.[10,11]

When using the approach of testing the PH assumption by analyzing a premodel including interaction of the variable(s) of interest with time as time-dependent covariate(s), the resulting HR shows the baseline effect of the examined variable(s) of interest. By contrast, results without time-dependent modeling show the time-averaged effect. In this study the baseline effect of MammaPrint is quite high and decreases over time. This can also be seen in Figure 3, where HRs for MammaPrint (adjusted for clinical risk groups) for increasing arbitrary censoring times are shown.

The ABCSG-8 biomarker cohort has previously been used to investigate the clinical validity of two other commercially available gene expression tests.[10,11] The most striking differences in comparison to the current study are (i) the higher rate of successful RNA extraction and analysis (leading to a larger cohort and higher statistical power), (ii) the clear independence of the prognostic value from clinical variables, as shown in multivariable models or similar statistical tests. In the case of EndoPredict, we were also able to show that the test improves prognostic classification in comparison to clinical guidelines.[20]

The most striking difference from prior validation cohorts of the MammaPrint test itself, however, is the homogeneity of this small subset: ABCSG-8 patients were randomized to two different endocrine treatments but did not receive adjuvant chemotherapy. Undifferentiated tumors (G3) were an exclusion criterion. As a result, ABCSG-8 comprises patients that had a rather low rate of nodal involvement; indeed, only 3% showed involvement of more than three nodes. Keeping in mind that patients with HER2 overexpression were also removed from the cohort, it is not surprising that within this biologically homogenous, lower-risk cohort the prognosis is driven by anatomy rather than biological factors such as the proliferation index, the expression of hormone receptors, and MammaPrint profile (Supplementary Figure 1B, available at https://doi.org/10.1016/j.esmoop.2020.100006).

The strengths of this analysis include the predefined statistical analysis plan which retrospectively analyzed a prospectively randomized cohort from a large phase III endocrine trial. This cohort also provided FFPE tumor samples with monitored clinical data including long-term survival events. At the same time, the test could only be carried out in a relatively small sample of patients and given the high sample failure rate this provides a smaller chance of showing statistically significant findings.

In future validation projects of the test it will be important to obtain FFPE samples from more recent cohorts and thus obtain a clearer understanding of the prognostic test in otherwise similar ER-positive cohorts of patients. Another weakness of this study concerns the biomarker cohort itself: none of these patients were treated with chemotherapy; therefore, we were not able to investigate predictive effects of the test.

The MammaPrint test showed significant prognostic performance at 5 and 10 years of follow-up. In the particular cohort of ABCSG-8, the statistical independence from clinically assessed covariates remains unclear, and thus no conclusions concerning the clinical validity of the test can be drawn.

FUNDING

This work was supported by funding from Agendia (no grant number) to ABCSG.research in the framework of 'Gene Expression Analysis to investigate survival outcomes in ER+ breast cancers: GiAntEss'.

DISCLOSURE

PD reports that funding for this project was provided by Agendia to ABCSG.research in the framework of 'Gene Expression Analysis to investigate survival outcomes in ER+ breast cancers: GiAntEss', grants from Cepheid/Danaher and Myriad during the conduct of the study. LVV reports personal fees, a part-time employment, and is stakeholder at Agendia NV during the conduct of the study. MG reports personal fees/travel support from Amgen, Daiichi Sankyo, AstraZeneca, Eli Lilly, Lifebrain, NanoString, Novartis, TLC Biopharmaceuticals; an immediate family member is employed by Sandoz. ZB-H reports personal fees and other from Roche, and personal fees from Novartis, Biomedica outside the submitted work. RG reports personal fees from Celgene, Roche, Merck, Takeda, Astra Zeneca, Novartis, Amgen, BMS, MSD, Sandoz, AbbVie, Gilead, Daiichi Sankyo, and Janssen C outside the submitted work. EL reports his employment in Medical Affairs by Agendia. GR reports personal fees from Pierre Fabre, Eli Lilly, AstraZeneca, personal fees and nonfinancial support from Roche, Novartis, Pfizer, Amgen, and nonfinancial support from Bristol-Myers Squibb outside the submitted work. WH reports personal fees from Roche Austria GmbH, MSD, Bioproducts GmbH, Boehringer Ingelheim and Biomedica Medizinprodukte GmbH outside the submitted work. DE reports personal fees from AstraZeneca, Novartis, Myriad, MSD, Roche, Pfizer outside the submitted work. CD reports to work as a consultant for Agendia. MF reports personal fees from AstraZeneca, Bayer, Biomedica, Bio-Rad, Boehringer Ingelheim, Myriad Genetics Inc., Pfizer and Roche outside the submitted work. The remaining authors have declared no conflicts of interest.

REFERENCES

1. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415(6871):530-536.
2. Drukker CA, Bueno-de-Mesquita JM, Retel VP, et al. A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. *Int J Cancer*. 2013;133(4):929-936.
3. Vliek SB, Retel V, Drukker C, et al. 10 years follow up of the RASTER study; implementing a genomic signature in daily practice. *Ann Oncol*. 2017;28(5):V46 [Abstract].
4. Cardoso F, van 't Veer LJ, Bogaerts J, et al. 70-Gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med*. 2016;375(8):717-729.
5. Buyse M, Loi S, van 't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst*. 2006;98(17):1183-1192.
6. van de Vijver MJ, He YD, van 't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999-2009.
7. Brandao M, Ponde N, Piccart-Gebhart M. MammaPrint: a comprehensive review. *Future Oncol*. 2019;15(2):207-224.
8. van 't Veer LJ, Yau C, Yu NY, et al. Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. *Breast Cancer Res Treat*. 2017;166:593-601.
9. Filipits M, Nielsen TO, Rudas M, et al. The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. *Clin Cancer Res*. 2014;20(5):1298-1305.
10. Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res*. 2011;17(18):6012-6020.
11. Gnant M, Filipits M, Greil R, et al. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 risk of recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Ann Oncol*. 2014;25:339-345.
12. Dubsky PC, Jakesz R, Mlineritsch B, et al. Tamoxifen and anastrozole as a sequencing strategy: a randomized controlled trial in postmenopausal patients with endocrine-responsive early breast cancer from the Austrian Breast and Colorectal Cancer Study Group. *J Clin Oncol*. 2012;30:722-728.
13. US Food and Drug Administration (FDA). Available at https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?id=K141142; 2016. Accessed October 16, 2020.
14. The Clinical & Laboratory Standards Institute (CLSI). Available at https://clsi.org; 2020. Accessed October 16, 2020.
15. US Food and Drug Administration (FDA). Statistical guidance on reporting results from studies evaluating diagnostic tests—guidance for industry and FDA staff. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda; 2020. Accessed October 16, 2020.
16. Dowsett M, Nielsen TO, A'Hern R, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst*. 2011;103(22):1656-1664.
17. Beumer I, Witteveen A, Delahaye L, et al. Equivalence of MammaPrint array types in clinical trials and diagnostics. *Breast Cancer Res Treat*. 2016;156(2):279-287.
18. Mook S, Schmidt MK, Viale G, et al. The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. *Breast Cancer Res Treat*. 2009;116(2):295-302.
19. Esserman LJ, Yau C, Thompson CK, et al. Use of molecular tools to identify patients with indolent breast cancers with ultralow risk over 2 decades. *JAMA Oncol*. 2017;3(11):1503-1510.
20. Dubsky P, Filipits M, Jakesz R, et al. EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer. *Ann Oncol*. 2013;24(3):640-647.