



UvA-DARE (Digital Academic Repository)

Prognostic factors in breast cancer: one fits all?

Mook, S.

Publication date
2011

[Link to publication](#)

Citation for published version (APA):

Mook, S. (2011). *Prognostic factors in breast cancer: one fits all?* [Thesis, fully internal, Universiteit van Amsterdam].

General rights

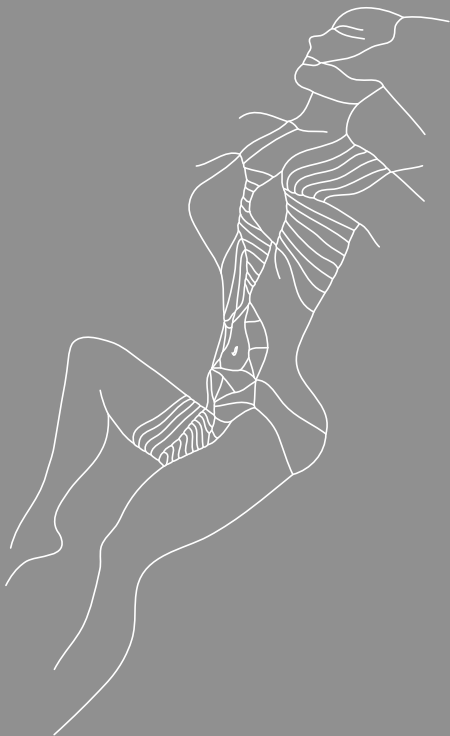
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Chapter 8

Calibration and discriminatory accuracy of prognosis calculation for breast cancer with the online Adjuvant! program: a hospital-based retrospective cohort study



Stella Mook*
Marjanka K. Schmidt*
Emiel J.Th. Rutgers
Anthonie O. van de Velde
Otto Visser
Sterre M. Rutgers
Nicola Armstrong
Laura J. Van 't Veer
Peter M. Ravdin
* Contributed equally

Summary

Background

Adjuvant! is a web-based program that calculates individualized 10-year survival probabilities and predicted benefit of adjuvant systemic therapy. The Adjuvant! model has not been validated in any large European series. The aim of our study was to validate Adjuvant! in Dutch patients, investigating both its calibration and discriminatory accuracy.

Methods

Patients who were at least partly treated at the Netherlands Cancer Institute for breast cancer between 1987 and 1998 were included if they met the following criteria: tumor size T1 (≤ 2 cm), T2 (2–5 cm), or T3 (> 5 cm), invasive breast carcinoma, with information about involvement of axillary lymph nodes available, no distant metastases, primary surgery, axillary staging, and radiotherapy according to national guidelines. Clinicopathological characteristics and adjuvant treatment data were retrieved from hospital records and medical registries and were entered into the Adjuvant! (version 8.0) batch processor with blinding to outcome. Endpoints were overall survival and the proportion of patients that did not die from breast cancer (breast cancer-specific survival [BCSS]).

Findings

5380 patients were included with median follow-up of 11.7 years (range 0.03–21.8). The 10-year observed overall survival (69.0%) and BCSS (78.6%) and Adjuvant! predicted overall survival (69.1%) and BCSS (77.8%) were not statistically different ($p = 0.87$ and $p = 0.18$, respectively). Moreover, differences between predicted and observed outcomes were within 2% for most relevant clinicopathological subgroups. In patients younger than 40 years, Adjuvant! overestimated overall survival by 4.2% ($p = 0.04$) and BCSS by 4.7% ($p = 0.01$). The concordance index, which indicates discriminatory accuracy at the individual level, was 0.71 for BCSS in the entire cohort.

Interpretation

Adjuvant! accurately predicted 10-year outcomes in this large-scale Dutch validation study and is of use for adjuvant treatment decision making, although the results may be less reliable in some subgroups.

Introduction

Treatment recommendations for early-stage breast cancer are based on prognosis (*i.e.*, the estimated risk of relapse and death after primary surgery and radiotherapy) and expected benefit of adjuvant therapy. Treatment guidelines qualitatively incorporate prognosis and treatment efficacy without quantitative estimates.¹⁻³ Nevertheless, methods that give quantitative estimates of prognosis exist, such as the Nottingham Prognostic Index and Adjuvant!.⁴⁻⁶ These quantitative methods include several assumptions, the most crucial being that the populations for which the models were developed are representative of others.

Adjuvant! is a computer program that is freely accessible on the internet (www.adjuvantonline.com). The program provides estimated 10-year survival probabilities and risk of relapse on the basis of a model incorporating patient's age, co-morbidity, tumor size, tumor grade, estrogen-receptor status, and number of involved lymphnodes.⁶ The program calculates the expected efficacy of adjuvant therapy (chemotherapy, hormonal therapy, or both) for different classes of regimens.⁷⁻⁹ The program gives the estimated prognosis and expected treatment benefit in a comprehensive format and can help to inform patients and to involve them in decision making about therapeutic options.¹⁰⁻¹²

Adjuvant! was largely developed with information from the Surveillance, Epidemiology and End Results (SEER) registry. The SEER registry has data for about 10% of patients with breast cancer in the USA.⁶ Olivotto and colleagues¹³ validated Adjuvant! (version 5.0) in a population-based series of 4083 early-stage patients with breast cancer registered in the British Columbia Breast Cancer Outcomes Unit database. They showed that the Adjuvant! model was well calibrated - *i.e.*, it accurately predicted the number of breast cancer-related deaths observed in the whole study cohort and subsets of their population (predicted and observed outcomes were within 2%). However, because European populations might differ from those in the USA and Canada, whether outcome predictions of the Adjuvant! model are applicable in Europe is unknown: differences in incidence of obesity, duration and type of adjuvant and salvage treatment, ethnic background, and intrinsic tumor characteristics might affect prognosis.¹⁴⁻¹⁹ Furthermore, although Olivotto and colleagues¹³ showed the goodness of fit of the Adjuvant! model, no information about its discriminatory accuracy was given. Because Adjuvant! is used in Europe and the USA to support treatment decisions in clinical practice and randomized trials,²⁰⁻²² we aimed to test the validity of Adjuvant! in a large cohort of Dutch patients with breast cancer, determining its ability to predict outcomes in groups of patients (calibration) and to distinguish individuals who will experience different outcomes (discriminatory accuracy).^{23,24}

Methods

Patients

All women who were at least partly treated for breast cancer at the Netherlands Cancer Institute at the Antoni van Leeuwenhoek Hospital (NKI-AVL) from 1987 to 1998 were identified in the hospital's Medical Registry. Patients were included if they had tumor size T1 (≤ 2 cm), T2 (2–5 cm), or T3 (>5 cm), unilateral tumors, invasive breast carcinoma, information about involvement of axillary lymph nodes available, no distant metastases, primary surgery, axillary staging, and radiotherapy according to national guidelines. Patients with previous malignant disease and those who received neoadjuvant therapy were excluded, as were those with unknown tumor size, unknown nodal status, unknown adjuvant systemic therapy, no definitive axillary surgery (axillary-lymph-node dissection with fewer than six nodes examined; *Figure 1*). Information about adjuvant systemic treatment was derived from the medical registry. Adjuvant treatment was given according to national guidelines, taking into account patients' wishes and preferences.¹⁷

The study is reported according to the STROBE statement.²⁵ No ethical review was required according to Dutch legislation.

Procedures

Histology, tumor size, tumor grade, and number of positive lymph nodes were retrieved from three sources and entered in the database according to the following hierarchy of preference for data source: first, personal logbook from pathologists at NKI-AVL containing pathology revisions of breast cancer diagnosed between 1994–96; second, the PALGA system (Dutch network and National Database for Pathology); third, medical registry of the NKI-AVL. Information about estrogen-receptor status was retrieved from three sources and entered into a database according to the following hierarchy: first, estrogen-receptor ligand-binding assays (breast cancers diagnosed 1987–95); second, Pathologist logbook (breast cancers diagnosed 1995–96), and third the PALGA system (Dutch network and National Database for Pathology; *Supplements*).

Outcome data (date of first local, regional and distant recurrence, second malignancies, contralateral breast cancer, and date of last follow-up or death) were obtained from the medical registry. These data were completed by linking patient records to the Dutch municipal registry, which contains the date of death or emigration if applicable, for *all* Dutch citizens. For patients not in this national registry as having died or emigrated, the date of last follow-up was recorded as Feb 1, 2007 (*i.e.*, 2 months before the date of linkage). Cause of death was retrieved from the medical registry if available and from individual patients' files when no cause of death was entered in the registry ($n=1090$ breast cancer-specific death; $n=188$ other causes). If neither the medical registry nor patients' files contained

cause of death, we used the presence of distant metastases as a surrogate: patients without known cause of death (n=875) were assigned to breast cancer-specific mortality if they were diagnosed with distant metastases during follow-up (n=191 breast cancer-specific death; n=684 other causes). Patients with registered breast cancer-specific death and patients who were assigned to breast cancer-specific death based on the presence of distant metastases were pooled for analyses.

10-year predicted overall survival and breast cancer-specific survival (BCSS) were calculated for each patient individually. Data on age, co-morbidity, tumor size, tumor grade, number of positive axillary lymph-nodes, estrogen-receptor status, and adjuvant systemic treatment were entered in the Adjuvant! (version 8.0) batch processor, with blinding to patient outcomes. The model's estimation of prognosis is based on 10-year observed overall survival of women diagnosed with breast cancer between 1988 and 1992 in the USA and recorded in the SEER database.⁶ The estimations of treatment efficacy are mainly based on the proportional risk reductions derived from the Early Breast Cancer Trialists' Collaborative Group 1998 meta-analysis and recently updated with the meta-analysis data from 2005.⁷⁻⁹ Because we could not retrieve reliable data for co-morbidity, we used the default assumption of minor health problems. For patients with no data on estrogen-receptor status, the status was entered in the model as unknown.

Statistical analyses

Overall survival and BCSS were derived from Kaplan-Meier survival analyses of the entire group and various subsets.²⁶ For the same datasets, the average predicted overall survival and BCSS were calculated from individual predicted outcomes by Adjuvant!. To assess the calibration of the model (goodness of fit), observed and average predicted outcomes were compared by use of a one-sample t-test for proportions, assuming the Adjuvant! predicted value to be the population value (under the assumption that the model is true) and thus fixed. In addition, we plotted averages of observed *versus* predicted outcomes, grouped by deciles of predicted outcomes.²³ The slope of the fitted line was compared with the slope of the line indicating a perfect relationship ($y=x$).

To assess discriminatory accuracy of Adjuvant! (its ability to discern patients having good outcomes from those having poor outcomes), we calculated an index of predictive discrimination, the concordance index (c-index).²³ The c-index was corrected for overfitting by bootstrapping with 200 resamples each. A c-index of 1 means that the model perfectly ranks patients according to survival (*i.e.*, patients having a better outcome also having a better predicted outcome), 0.5 means the model does no better than chance. The predictive accuracy and proportion of explained variation, as defined by Schemper and colleagues,²⁴ was also calculated. SEs were estimated by bootstrapping with 200 resamples each. Known prognostic factors (*i.e.*, age, tumor size, tumor grade, number of positive lymph-nodes, histology, estrogen-receptor status, and adjuvant systemic therapy) were used in the Cox

multivariate model. Furthermore, on the basis of backward multivariate Cox regression analyses, year of diagnosis was added to construct the best predictive model for BCSS and overall survival in our dataset.²⁷ Analyses were done with SPSS version 15.0 and R statistical software (www.r-project.org).

Role of the funding source

The funding sources had no role in study design; collection, analysis, or interpretation of data; writing of the paper; or in decisions relating to publication. SM, MKS, and LVV had full access to all data. SM, MKS, LVV, and PMR took final responsibility for the decision to submit the paper for publication.

Results

Our database included 5380 patients, 2604 of whom (48%) received no adjuvant systemic therapy. The algorithm in Adjuvant! attributes different efficacy estimates depending on type of chemotherapy and hormonal treatment. Among 1961 patients treated with endocrine therapy, 1908 (97%) received tamoxifen (2–5 years); therapy was not specified for 13 (0.7%). 892 (82%) of 1084 patients treated with adjuvant chemotherapy received cyclophosphamide, methotrexate, and fluorouracil, 122 (11%) received fluorouracil, epirubicin, and cyclophosphamide, 42 (4%) received high-dose chemotherapy, and 11 (1%) received cyclophosphamide and doxorubicin. For the remaining 16 patients (2%), type of chemotherapy was unspecified. 2276 (42%) of 5380 patients had complete data for all factors used in the Adjuvant! model to predict outcome. Grade was unknown for 1379 patients (26%), and estrogen-receptor status unknown for 2253 (42%).

During a median follow-up of 11.7 years (range 0.03–21.8), 2153 (40%) of 5380 patients died; 3032 (94%) of 3227 patients alive at last follow-up had 10 years or more follow-up (*Figure 1*). *Table 1* shows the distribution of demographic, pathological, and primary treatment data for our study cohort. For all patients, the 10-year observed overall survival (69.0%) and BCSS (78.6%) rates as compared with the 10-year overall survival (69.1%) and BCSS (77.8%) rates predicted by Adjuvant! were within 1% and not significantly different ($p > 0.05$; *Table 1*). In general, Adjuvant! predicted overall survival accurately in the various subsets of patients (*i.e.*, differences between predicted and observed outcomes were within 2%), whereas Adjuvant! underestimated BCSS in some subsets (*Table 1*). Subsets of patients for whom there was a discrepancy between predicted outcomes by Adjuvant! and actual observed outcomes included patients under 40 years, for whom both predicted overall survival and BCSS were overly optimistic (4.2% and 4.7%, respectively; $p = 0.04$ and $p = 0.01$). For patients older than 69 years the program also overestimated overall survival by 3.4% ($p = 0.05$), but BCSS was accurately predicted in this group (predicted–observed –1.7%).

In subgroups of nodal status, overall survival was accurately predicted by Adjuvant!; however, the program underestimated BCSS by 3.1% ($p = 0.002$) in patients with one to three positive lymph-nodes.

Although Adjuvant! predicted overall survival accurately for subsets of tumor size, a discrepancy between predicted and observed BCSS was noted (-5.8% to 2.4% ; *Table 1*). In particular, predicted BCSS was optimistic for patients with tumors with diameter 11–20 mm, although it was pessimistic in patients with tumors 21–50 mm in diameter. In patients with estrogen-receptor-negative tumors Adjuvant! underestimated BCSS by 4.1% ($p = 0.02$). This underestimation of outcome by Adjuvant!, although non-significant, was also seen for overall survival (-3.2% , $p = 0.07$).

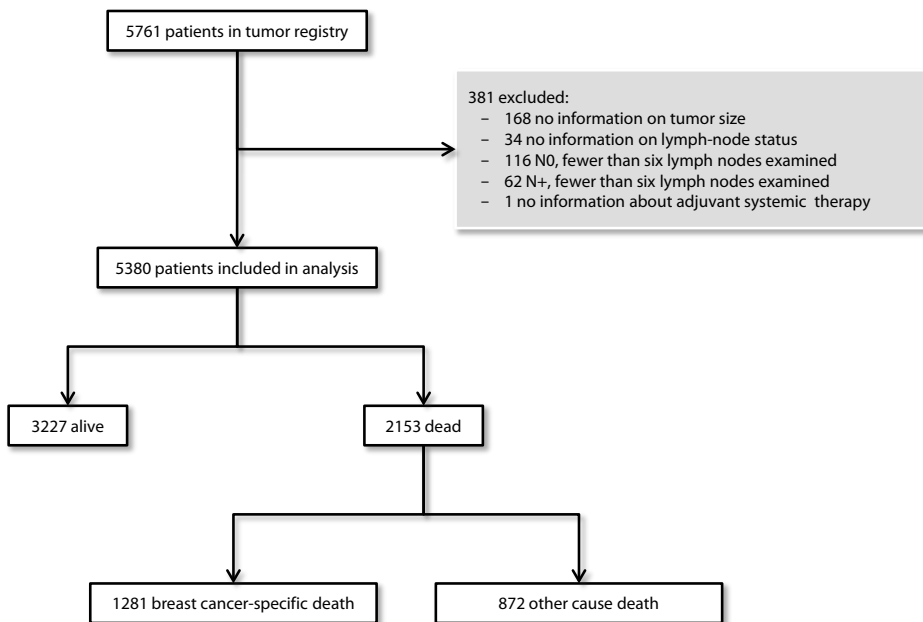


Figure 1. Study profile

Table 1. Baseline characteristics and Adjuvant! predicted versus observed overall survival and breast cancer-specific survival.

	Patients		Overall survival				Breast cancer-specific survival			
	No.	%	Adjuvant! Predicted	Observed (SE)	Predicted - Observed (95% CI)	P value	Adjuvant! Predicted	Observed (SE)	Predicted - Observed (95% CI)	P value
All patients	5380	100	69.1	69 (0.6)	0.1 (-1.08 to 1.28)	0.87	77.8	78.6 (0.6)	-0.8 (-1.98 to 0.38)	0.18
Year of diagnosis										
1987-1989	1169	21.7	67.6	67.7 (1.4)	-0.1 (-2.85 to 2.65)	0.94	76.4	77.0 (1.3)	-0.6 (-3.15 to 1.95)	0.64
1990-1993	2017	37.5	69.2	67.8 (1.0)	1.4 (-0.56 to 3.36)	0.16	78.1	77.7 (1.0)	0.4 (-1.56 to 2.36)	0.69
1994-1997	2194	40.8	69.8	70.8 (1.0)	-1.0 (-2.96 to 0.96)	0.32	78.4	80.3 (0.9)	-1.9 (-3.67 to -0.14)	0.04
Type of surgery										
Breast-conserving surgery	2885	53.6	76.5	76 (0.8)	0.5 (-1.07 to 2.07)	0.53	83.8	82.8 (0.7)	1.0 (-0.37 to 2.73)	0.15
Mastectomy	1978	36.8	59.1	60 (1.1)	-0.9 (-3.06 to 1.26)	0.41	69.7	72.6 (1.0)	-2.9 (-4.86 to -0.94)	0.004
Unknown	517	9.6	66.1	64.3 (2.1)	1.8 (-2.33 to 5.73)	0.39	75.7	77.7 (1.9)	-2.0 (-5.73 to 1.73)	0.29
Age (years)										
<40	572	10.6	72.4	68.2 (2.0)	4.2 (0.27 to 8.13)	0.04	73.5	68.8 (1.9)	4.7 (0.97 to 8.43)	0.01
40-49	1448	26.9	76.6	78.6 (1.1)	-2.0 (-4.16 to 0.16)	0.07	78.8	81.2 (1.0)	-2.4 (-4.36 to -0.44)	0.02
50-59	1369	25.4	72.0	73.6 (1.2)	-1.6 (-3.95 to 0.75)	0.18	76.8	77.7 (1.1)	-0.9 (-3.06 to 1.26)	0.41
60-69	1174	21.8	68.9	68.3 (1.4)	0.6 (-2.15 to 3.35)	0.67	80.1	80.9 (1.2)	-0.8 (-3.15 to 1.55)	0.51
≥70	817	15.2	49.1	45.7 (1.7)	3.4 (0.06 to 6.74)	0.05	77.8	79.5 (1.6)	-1.7 (-4.84 to 1.44)	0.29
Histology										
Mainly DCIS	55	1.0	83.8	87.2 (4.5)	-3.4 (-12.42 to 5.62)	0.45	90.2	96.2 (2.6)	-6.0 (-11.21 to -0.79)	0.03
IDC	4001	74.4	68.7	68.3 (0.7)	0.4 (-0.97 to 1.77)	0.57	77.2	77.6 (0.7)	-0.4 (-1.77 to 0.97)	0.57
ILC	614	11.4	65.3	66.2 (1.9)	-0.9 (-4.63 to 2.83)	0.64	75.6	77.8 (1.7)	-2.2 (-5.54 to 1.14)	0.20
ID/LC	318	5.9	70.4	70.1 (2.6)	0.3 (-4.82 to 5.42)	0.91	78.5	78.5 (2.4)	0.0 (-4.72 to 4.72)	1.00
Tubular	114	2.1	88.8	95.6 (1.9)	-6.8 (-10.56 to -3.04)	0.0005	95.2	100 (0.0)	-4.8 (-4.80 to -4.79)	< 0.0001
Mucinous	76	1.4	74.0	65.6 (5.5)	8.4 (-2.56 to 19.36)	0.13	87.6	88.1 (4.0)	-0.5 (-8.47 to 7.47)	0.90
Medular	68	1.3	74.5	80.9 (4.8)	-6.4 (-15.98 to 3.18)	0.19	79.4	86.4 (4.2)	-7.0 (-15.38 to 1.38)	0.10
Others	134	2.5	67.5	64.1 (4.2)	3.4 (-4.91 to 11.71)	0.42	78.8	76.7 (3.8)	2.1 (-5.42 to 9.62)	0.58

	Patients		Overall survival				Breast cancer-specific survival			
	No.	%	Adjuvant! Predicted	Observed (SE)	Predicted - Observed (95% CI)	P value	Adjuvant! Predicted	Observed (SE)	Predicted - Observed (95% CI)	P value
Tumor size (mm)										
1-10	724	13.5	85.5	85.2 (1.3)	0.3 (-2.25 to 2.85)	0.82	93.2	93.4 (0.9)	-0.2 (-1.97 to 1.57)	0.82
11-20	2177	40.5	77.7	76.1 (0.9)	1.6 (-0.17 to 3.37)	0.07	86.3	83.9 (0.8)	2.4 (0.83 to 3.97)	0.003
21-50	2244	41.7	58.6	59.9 (1.0)	-1.3 (-3.26 to 0.66)	0.19	67.8	71.0 (1.0)	-3.2 (-5.16 to -1.24)	0.001
>50	235	4.4	39.4	39.9 (3.2)	-0.5 (-6.80 to 5.80)	0.88	47.7	53.5 (3.4)	-5.8 (-12.50 to 0.90)	0.09
Grade										
1	777	14.4	83.8	83.6 (1.3)	0.2 (-2.35 to 2.75)	0.88	94.1	94.3 (0.9)	-0.2 (-1.97 to 1.57)	0.82
2	1935	36.0	71.8	71.3 (1.0)	0.5 (-1.46 to 2.46)	0.62	81.0	81.4 (0.9)	-0.4 (-2.17 to 1.37)	0.66
3	1289	24.0	57.4	56.6 (1.4)	0.8 (-1.95 to 3.55)	0.57	64.4	64.4 (1.4)	0 (-2.75 to 2.75)	1.00
Unknown	1379	25.6	67.9	69.2 (1.2)	-1.3 (-3.65 to 1.05)	0.28	76.8	78.9 (1.1)	-2.1 (-4.26 to 0.06)	0.06
Number of positive lymph-nodes										
0	2704	50.3	79.7	78.5 (0.8)	1.2 (-0.37 to 2.77)	0.13	88.5	87.0 (0.7)	1.5 (0.13 to 2.87)	0.03
1-3	1720	32.0	66.0	67.7 (1.1)	-1.7 (-3.86 to 0.46)	0.12	75.1	78.2 (1.0)	-3.1 (-5.06 to -1.14)	0.002
4-9	707	13.1	49.8	49.7 (1.9)	0.1 (-3.63 to 3.83)	0.96	58.0	60.9 (1.9)	-2.9 (-6.63 to 0.83)	0.13
>9	249	4.6	30.7	30.0 (2.9)	0.7 (-5.01 to 6.41)	0.81	37.5	37.7 (3.2)	-0.2 (-6.50 to 6.10)	0.95
Estrogen-receptor status										
Positive	2417	44.9	71.6	70.9 (0.9)	0.7 (-1.07 to 2.47)	0.44	81.1	80.3 (0.8)	0.8 (-0.77 to 2.37)	0.32
Negative	710	13.2	60.4	63.6 (1.8)	-3.2 (-6.73 to 0.33)	0.07	66.4	70.5 (1.7)	-4.1 (-7.44 to -0.76)	0.02
Unknown	2253	41.9	69.2	68.6 (1.0)	0.6 (-1.36 to 2.56)	0.55	78.0	79.4 (0.9)	-1.4 (-3.17 to 0.37)	0.12

Continued ▲

Table 1. Continued.

	Patients		Overall survival				Breast cancer-specific survival			
	No.	%	Adjuvant! Predicted	Observed (SE)	Predicted - Observed (95% CI)	P value	Adjuvant! Predicted	Observed (SE)	Predicted - Observed (95% CI)	P value
Systemic treatment										
None	2604	48.4	76.8	75.8 (0.8)	1.0 (-0.57 to 2.57)	0.21	85.3	84.6 (0.7)	0.7 (-0.67 to 2.07)	0.32
Chemotherapy only	815	15.1	67.4	69.3 (1.6)	-1.9 (-5.04 to 1.24)	0.24	69.4	71.7 (1.6)	-2.3 (-5.44 to 0.84)	0.15
Hormonal therapy only	1692	31.4	58.8	58.4 (1.2)	0.4 (-1.95 to 2.75)	0.74	72.1	73.7 (1.1)	-1.6 (-3.76 to 0.56)	0.15
Chemo and hormonal therapy	269	5.0	64.5	69.1 (2.8)	-4.6 (-10.11 to 0.91)	0.10	67.0	71.4 (2.8)	-4.4 (-9.91 to 1.11)	0.12
Completeness of data										
Missing estrogen-receptor status and/or grade	3104	57.7	69.0	68.7 (0.8)	0.3 (-1.27 to 1.87)	0.71	77.8	78.9 (0.8)	-1.1 (-2.67 to 0.47)	0.17
Complete	2276	42.3	69.2	69.4 (1.0)	-0.2 (-2.16 to 1.76)	0.84	77.8	78.2 (0.9)	-0.4 (-2.17 to 1.37)	0.66

P-values calculated with one sample t-test.

Figure 2 shows the observed outcome *versus* the average predicted outcome for the cohort, grouped by deciles of predicted overall survival or BCSS probabilities. The slope of the line representing a perfect fit of predicted with observed outcomes ($y=x$) and the slope of the actual line fitted to our data for overall survival was not significantly different. This indicates that the calibration of Adjuvant! is similarly good in patients with poor overall survival and patients with excellent overall survival. However, for BCSS, the model tended to underestimate and to overestimate BCSS in the extremes of the distribution of poor and good survival, respectively (slope was significantly different $p < 0.0001$).

To assess discriminatory accuracy of the model (*i.e.*, its ability to separate patients who will die from breast cancer from those who will not), we calculated Harrell's c-index (0.71 for BCSS), as well as the predictive accuracy and explained variation (0.73 and 13% for BCSS, respectively). Hence, the predictive accuracy for BCSS increased from 0.69 for a model without predictors to 0.73 for the Adjuvant! model (Table 2). In various clinical subgroups (Table 1), the c-index varied from 0.65 to 0.75 (data not shown). The c-index for a multivariate Cox regression model best fitted to the outcome of the 5380 patients with a backward approach (model included age, tumor size, tumor grade, number of positive lymph nodes, estrogen-receptor status, histology, type of adjuvant systemic therapy and year of diagnosis) was similar to the Adjuvant! model (*i.e.*, 0.72 and 0.71, respectively for BCSS).

Table 2. Discriminatory accuracy of Adjuvant! and a multivariate Cox model fitted to the outcome.

	Overall survival		Breast cancer-specific survival	
	Adjuvant!	Cox model*	Adjuvant!	Cox model*
C index	0.70	0.69	0.71	0.72
Predictive accuracy model without predictors [1-D0]	0.64	0.64	0.69	0.69
Predictive accuracy [1-Dx] (SE)	0.69 (0.008)	0.69 (0.007)	0.73 (0.02)	0.74 (0.02)
Explained variation [(Dx-D0)/D0] (SE)	15% (1%)	15% (1%)	13% (1%)	16% (2%)

For overall survival the model included age, tumor size, tumor grade, number of positive lymph-nodes, histology, adjuvant systemic therapy and year of diagnosis. For breast cancer-specific survival the model included age, tumor size, tumor grade, number of positive lymph-nodes, histology, estrogen-receptor status, adjuvant systemic therapy, and year of diagnosis.

C-index = Harrell's concordance index. Dx = predictive accuracy of model with predictors. D0 = predictive accuracy of model without predictors.

*Best-fitted multivariate Cox model.

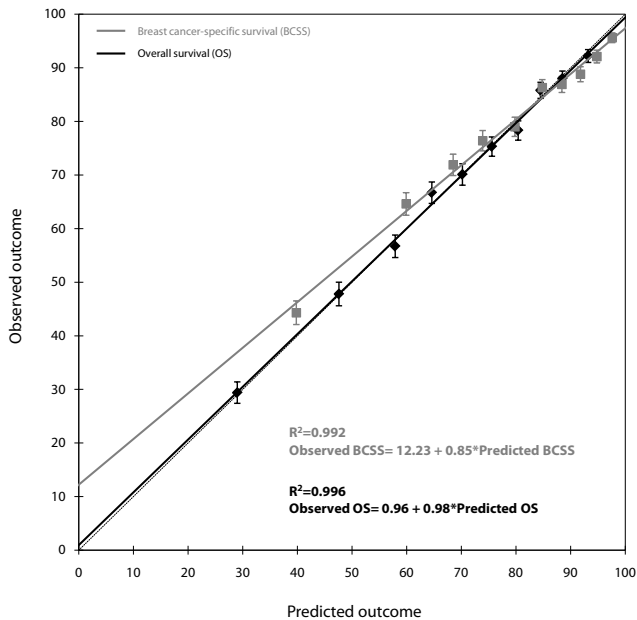


Figure 2. Mean predicted *versus* observed outcomes by deciles of predicted outcome. Error bars are SE.

Discussion

Overall projections of overall survival and BCSS with Adjuvant! were within 1% of observed results and estimates within most subgroups seemed reasonably accurate (within 2% or not significantly different from observed estimates). The conservative Dutch guidelines for adjuvant systemic therapy used in the era of the study cohort results in a large proportion of patients (2604 [48%] of 5380) who received no adjuvant systemic therapy. For this group of patients, we could assess the prognostic value of Adjuvant! (*i.e.*, the prediction of disease outcome in the absence of adjuvant systemic therapy). Although most patients with early breast cancer now receive some form of adjuvant systemic treatment, the confirmation of the prognostic value is important for the decision whether or not to treat. Moreover, when the program predicts prognosis accurately, the potential benefit of different types of adjuvant treatment is also predicted more accurately, because the latter depends on the a priori risk of recurrence. Adjuvant! is commonly used to decide whether patients who will be treated with endocrine therapy are candidates for additional chemotherapy. Because a small proportion of patients received chemotherapy with or without endocrine therapy (815 [15%] and 269 [5%] of 5380, respectively) the prediction of chemotherapy benefit in

addition to endocrine therapy is less robust. However, the decision to add chemotherapy to the endocrine treatment regimen depends also on the predicted outcome of a patient when treated with hormonal therapy only. Adding chemotherapy to endocrine therapy will be more beneficial for patients who will have an a priori poorer predicted disease outcome. Our results show that Adjuvant! predicted both overall survival and BCSS accurately in patients treated with endocrine therapy (1692 [31%] of 5380).

One of the limitations of our study is that in the era of this study standard adjuvant chemotherapy consisted of cyclophosphamide, methotrexate, and fluorouracil for six cycles and standard adjuvant endocrine therapy consisted of tamoxifen for which we evaluated Adjuvant!. Therefore, future studies are required to validate Adjuvant! predictions of currently used therapies, such as taxane-based chemotherapy and aromatase inhibitors. The largest discrepancy between subgroups between our study and the Canadian validation study¹³ was in patients younger than 35 years (10% for BCSS). As a consequence of this disagreement and after further review of the SEER registry data, Adjuvant! was modified to give more pessimistic estimates for estrogen-receptor-positive patients under 35 years of age. Even after this adjustment (the major difference between Adjuvant! 5.0 and Adjuvant! 8.0), the predicted outcomes still seem too optimistic, albeit less so than in the original validation study (by 5% for BCSS). When results were stratified for estrogen-receptor status in patients younger than 35 years and age 35–40 years, the overestimation was exclusively seen in patients with estrogen-receptor-positive tumors in both age groups (Supplements). This suggests that the correction factor of 1.5 for patients under 35 years is insufficient and that an additional correction for patients between 35–40 years with estrogen-receptor-positive tumors might be justified.

Both studies showed that the outcomes of ductal and lobular cancers were accurately predicted, but for other histological subtypes, the predicted outcomes by Adjuvant! are too pessimistic. At present, histology is not incorporated in Adjuvant!; however, the program warns the user that some histological subtypes might warrant an adjustment (*e.g.*, medullary cancers where high grade does not confer high risk).²⁸ Other discrepancies between observed outcomes and outcomes predicted by Adjuvant! seem modest or inconsistent between our study and the original validation study. For example, underestimation of BCSS was seen in one subgroup of tumor size. Olivotto and colleagues¹³ noted no such underestimation, although the distributions of tumor size were similar. This suggests that the discrepancy in predicted and observed BCSS is not caused by a suboptimum incorporation of size in the Adjuvant! model.

A second subgroup in which the underestimation of BCSS was significant is patients with one to three positive lymph nodes. Although we do not have information about the extent of lymph-node involvement (*i.e.*, isolated tumor cells, micrometastases, or macrometastases), until the late 1990s lymph nodes that contained only isolated tumor cells were assessed as positive lymph-nodes in the Netherlands. Consequently, the group of patients with one to three positive lymph nodes in our database probably includes some patients with only

isolated cells who have a better disease outcome than patients with macrometastases in one to three lymph-nodes (Supplements).^{29,30} As for younger patients, the predicted overall survival by Adjuvant! was too optimistic for patients older than 70 years, which is possibly caused by the lack of data on co-morbidity status. Co-morbidity was entered as the default assumption of minor health problems, which is likely to be an underestimation in older patients and therefore to result in overestimation of overall survival in these patients. The accurate prediction of BCSS in patients older than 70 years supports this hypothesis.

The proportion of missing data is one of the limitations of our retrospective cohort for the validation of Adjuvant!; we lacked data on estrogen-receptor status, tumor grade, or both for 3104 (58%) of 5380 patients. Missing information about tumor grade or estrogen-receptor status will now be less common. However, patients with incomplete data had similar disease outcome as patients with complete data, indicating that including patients with missing data did not induce a selection bias (*Table 1*). Information on HER2 status was not available in this cohort and will be incorporated in an upcoming version of Adjuvant!. The program predicted disease outcome accurately for patients with unknown estrogen-receptor status. By contrast, the model underestimated BCSS in patients with estrogen-receptor-negative tumors. Detailed analysis of this subgroup revealed that the underestimation was exclusively seen in patients treated with hormonal therapy (n=250; *Supplements*). This particular group had better outcome than predicted, suggesting that these tumors could have been erroneously scored or coded in the registry as estrogen-receptor-negative.

Patients who were partly treated at NKI-AVL were mainly referred from regional hospitals to our institute for radiotherapy. All diagnostic information was made available and reviewed by the NKI-AVL. Adjuvant! predicted overall survival and BCSS accurately in these patients and those treated at NKI-AVL (*Supplements*). As a consequence, the population includes a much wider representation of patients, and selection bias of our cohort is likely to be less pronounced than in a single-institute cohort.

This large-scale validation study of Adjuvant! in a hospital-based population of Dutch patients with breast cancer showed that the calculated predictions by Adjuvant! agreed with the observed outcomes and that the predictions are applicable to a Dutch population, and presumably to a European population, corroborating that populations of patients with European ancestry in different continents have similar disease. Potential differences between US and European patients with breast cancer could have resulted in deviations of outcome in both directions and therefore would level out in an overall comparison of predicted and observed outcomes. The good performance of the model in American and European settings implies that the prognostic features and disease course are broadly similar in both settings.

The model's success in these settings does not ensure success in other uses, because, for example, time, changes in exogenous exposures (hormone replacement therapy), diagnostic techniques (types and intensity of screening), and surgical staging could affect

the prognosis of patients with otherwise similar tumors. Furthermore, although we used a large validation set, somewhat larger than the validation in British Columbia Breast Cancer patients published by Olivotto and colleagues (n=4083), some subgroups were small and findings in those groups should be considered with at least some caution.

To validate the Adjuvant! model in our study population we investigated both the calibration (goodness of fit) and the discriminatory accuracy of the model. Although the latter is rarely tested, it is of paramount importance to justify the use of prognostic models for clinical outcome prediction.^{23,24} Results of the discriminatory accuracy of the Adjuvant! model showed that in addition to good calibration, the model was capable of separating individuals with a poor outcome from those with a good outcome with moderate power (c-index 0.71). Remarkably, the discriminatory accuracy of a multivariate Cox model fitted to our dataset was similar to that of Adjuvant!, indicating that the prognostic information of the variables used in Adjuvant! was incorporated in the model in the best way possible. Furthermore, the maximum explained variation by clinicopathological variables is about 15%, irrespective of whether they are incorporated in Adjuvant! or a model fitted to our dataset. that the unexplained variation remains relatively large is supported by the observation that patients with identical clinicopathological variables can have strikingly different outcomes and proves that the information captured by these criteria can only explain part of the differences in outcome. incorporation of biological markers, such as molecular profiles and germline variants, in the model will likely increase the explained variation and therefore result in a more rigorous prediction of outcome at the individual patient level in the near future.

Contributors

SM, MKS, ER, LVV, and PMR designed the study. SM, MKS, AOV, OV, SMR, and PMR collected data. SM, MKS, and NA analyzed data. SM, MKS, ER, NA, LVV, and PMR interpreted data. SM, MKS, ER, LVV, and PMR wrote the paper.

Funding

Dutch National Genomics Initiative-Cancer Genomics Center, Dutch Cancer Society-KWF grant NKI 2007-3839.

Conflicts of interest

PMR owns and is paid in part by Adjuvant Inc, which owns the rights to Adjuvant! online. The other authors have no conflicts of interest to declare.

Acknowledgments

We thank Hans Peterse (deceased), Otilia Dalesio, and the Medical Registry staff for providing baseline data, Hans Bonfrer and Tiny Korse for providing the estrogen-receptor ligand binding assay data, Matti Rookus, Flora van Leeuwen, and Marieke Vollebergh for helpful discussions. This study was financially supported by the Dutch National Genomics Initiative-Cancer Genomics Center (SM and SMR) and the Dutch Cancer Society-KWF grant NKI 2007-3839 (MKS).

References

1. NCCN Clinical Practice Guidelines in Oncology. Breast Cancer V.1.2009. www.nccn.org. (Last accessed June 2, 2009).
2. Eifel P, Axelson JA, Costa J, et al. National Institutes of Health consensus development conference statement: Adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst* 2001; 93: 979-989.
3. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Senn HJ. Meeting Highlights: International consensus panel on the treatment of primary breast cancer. *J Clin Oncol* 2001; 19: 3817-3827.
4. Haybittle JL, Blamey RW, Elston CW, et al. A prognostic index in primary breast cancer. *Br J Cancer* 1982; 45: 361-366.
5. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat* 1992; 22: 207-219.
6. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol* 2001; 19: 980-991.
7. Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 1998; 352: 930-942.
8. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998; 351: 1451-1467.
9. Early Breast Cancer Trialists' Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005; 365: 1687-1717.
10. Siminoff LA, Gordon NH, Silverman P, Budd T, Ravdin PM. A decision aid to assist in adjuvant therapy choices for breast cancer. *Psychooncology* 2006; 15: 1001-1013.
11. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Improving understanding of adjuvant therapy options by using simpler risk graphics. *Cancer* 2008; 113: 3382-3390.
12. Belkora J, Rugo HS, Moore DH, Hutton D, Esserman L. Risk communication with patients with breast cancer: cautionary notes about printing Adjuvant! estimates. *Lancet Oncol* 2008; 9: 602-603.
13. Olivotto IA, Bajdik CD, Ravdin PM, et al. Population-based validation of the prognostic model ADJUVANT! for early breast cancer. *J Clin Oncol* 2005; 23: 2716-2725.
14. Coleman MP, Quaresma M, Berrino F, et al. Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol* 2008; 9: 730-756.
15. Sant M, Allemani C, Berrino F, et al. Breast carcinoma survival in Europe and the United States. *Cancer* 2004; 100: 715-722.
16. Chlebowski RT, Aiello E, McTiernan A. Weight loss in breast cancer patient management. *J Clin Oncol* 2002; 20: 1128-1143.
17. CBO Richtlijn Mammacarcinoom 2008. http://www.cbo.nl/product/richtlijnen/folder20021023121843/rl_mamma_08.pdf (Last accessed, June 2, 2009).
18. Harlan LC, Clegg LX, Abrams J, Stevens JL, Ballard-Barbash R. Community-based use of chemotherapy and hormonal therapy for early-stage breast cancer: 1987-2000. *J Clin Oncol* 2006; 24: 872-877.

19. Chlebowski RT, Chen Z, Anderson GL, et al. Ethnicity and breast cancer: factors influencing differences in incidence and outcome. *J Natl Cancer Inst* 2005; 97: 439-448.
20. Love N. Management of breast cancer in the adjuvant and metastatic settings: Patterns of care in medical oncology. <http://www.paternsofcare.com/2005/3/adjuvant.htm> (Last accessed June 2, 2009).
21. Bogaerts J, Cardoso F, Buyse M, et al. Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol* 2006; 3: 540-551.
22. Cardoso F, Van 't Veer L, Rutgers E, et al. Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol* 2008; 26: 729-735.
23. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15: 361-387.
24. Schemper M and Henderson R. Predictive accuracy and explained variation in Cox regression. *Biometrics* 2000; 56: 249-255.
25. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; 370: 1453-1457.
26. Kaplan EL and Meier P. Nonparametric estimation from incomplete observation. *J Am Stat Assoc* 1958; 53: 457-481.
27. Cox DR. Regression models and lifetables. *J R Statist Soc B* 1972; 34: 187-220.
28. Pedersen L, Holck S, Schiodt T, Zedeler K, Mouridsen HT. Medullary carcinoma of the breast, prognostic importance of characteristic histopathological features evaluated in a multivariate Cox analysis. *Eur J Cancer* 1994; 30A: 1792-1797.
29. Cote RJ, Fpeterston H, Chaiwun B, et al. Role of immunohistochemical detection of lymph-node metastases in management of breast cancer. *Lancet* 1999; 354: 896-900.
30. Reed J, Rosman M, Verbanac K, et al. Prognostic implications of isolated tumor cells and micrometastases in sentinel nodes of patients with invasive breast cancer: 10-year analysis of patients enrolled in the prospective East Carolina University/Anne Arundel Medical Center Sentinel Node Multicenter Study. *J Am Coll Surg* 2009; 208: 333-340.

Supplements Chapter 8

Supplementary Table 1. Different sources of pathology data.

	Dept. Clin. Chem.	Pathologist logbook*	PALGA	Med. Reg.	Unknown	Total
	N (%)	N (%)	N (%)	N (%)	N (%)	N
Histology		482 (9.0)	4692 (87.2)	206 (3.8)	0	5380
Tumor size		429 (8.0)	4339 (80.6)	612 (11.4)	0	5380
Grade		397 (7.5)	3252 (60.4)	352 (6.5)	1379 (25.6)	5380
Number of positive lymph-nodes		486 (9.0)	4478 (83.2)	416 (7.8)	0	5380
Estrogen-receptor status	643 (11.9)	270 (5.0)	2214 (41.2)	0	2253 (41.9)	5380

Dept. Clin. Chemistry, Department of Clinical Chemistry; PALGA, Dutch network and National Database for Pathology; Med. Reg., Medical Registry of Netherlands Cancer Institute-Antoni van Leeuwenhoek hospital (NKI-AVL).

* Personal logbook of NKI-AVL pathologist J.L. Peterse.

Supplementary Table 2. Adjuvant! predicted *versus* observed OS and BCSS stratified by age and estrogen-receptor status.

	No.	Overall survival				Breast cancer-specific survival			
		Adjuvant! predicted	Observed (SE)	Predicted - Observed	P-value	Adjuvant! predicted	Observed (SE)	Predicted - Observed	P-value
Age ≤ 35 yrs	272	69.5	67.4 (2.9)	2.1	0.47	70.4	67.8	2.6	0.35
Estrogen-receptor status									
Positive	96	73.8	64.5 (4.9)	9.3	0.06	74.8	64.5 (4.9)	10.3	0.04
Negative	58	66.0	69.0 (6.1)	-3.0	0.62	66.8	69.0 (6.1)	-2.2	0.72
Unknown	118	67.8	69.1 (4.3)	-1.3	0.76	68.7	69.9 (4.3)	-1.2	0.78
Age 35-40 yrs	300	75.0	68.9 (2.7)	6.1	0.02	76.3	69.8 (2.7)	6.5	0.02
Estrogen-receptor status									
Positive	119	78.9	68.9 (4.2)	10.0	0.02	80.2	69.6 (4.2)	10.6	0.01
Negative	58	68.4	70.7 (6.0)	-2.3	0.70	69.5	70.7 (6.0)	-1.2	0.84
Unknown	123	74.4	68.1 (4.2)	6.3	0.14	75.7	69.5 (4.2)	6.2	0.14

OS, Overall survival; BCSS, Breast cancer-specific survival; SE, Standard error.

P-values based on one sample t-test.

Supplementary Table 3. Adjuvant! predicted *versus* observed BCSS stratified by number of positive lymph-nodes.

	Patients		Breast cancer-specific survival			
	No.	%	Adjuvant! predicted	Observed (SE)	Predicted - Observed	P-value
Number of positive lymph-nodes						
0	2704	50.3	88.5	87.0 (0.7)	1.5	0.03
1	876	16.3	75.8	81.8 (1.3)	-6.0	<0.0001
2	538	10.0	74.7	76.0 (1.9)	-1.3	0.49
3	306	5.7	73.6	71.9 (2.7)	1.7	0.53
4	213	4.0	60.6	65.8 (3.4)	-5.2	0.13
5	168	3.1	57.9	66.0 (3.8)	-8.1	0.03
6	111	2.1	57.4	60.7 (4.9)	-3.3	0.50
7	86	1.6	56.9	55.0 (5.5)	1.9	0.73
8	70	1.3	54.5	53.9 (6.2)	0.6	0.92
9	59	1.1	55.3	45.7 (6.8)	9.6	0.16
>9	249	4.6	37.5	37.7 (3.2)	-0.2	0.95

BCSS, Breast cancer-specific survival; SE, Standard error.

P-values based on one sample t-test.

Supplementary Table 4. Adjuvant! predicted *versus* observed BCSS stratified by treatment and estrogen-receptor status.

	Patients		Breast cancer-specific survival			
	No.	%	Adjuvant! predicted	Observed (SE)	Predicted - Observed	P-value
Untreated patients	2604					
Estrogen-receptor status						
Positive	1143	43.9	87.0	85.0 (1.1)	2.0	0.07
Negative	331	12.7	76.8	77.7 (2.3)	-0.9	0.70
Unknown	1130	43.4	86.2	86.2 (1.1)	0.0	1.00
Chemotherapy only	815					
Estrogen-receptor status						
Positive	349	42.8	72.0	74.0 (2.4)	-2.0	0.41
Negative	129	15.8	62.9	64.0 (4.3)	-1.1	0.80
Unknown	337	41.3	69.2	72.3 (2.5)	-3.1	0.22
Hormonal therapy only	1692					
Estrogen-receptor status						
Positive	787	46.5	77.8	77.4 (1.6)	0.4	0.80
Negative	192	11.3	55.5	64.9 (3.6)	-9.4	0.01
Unknown	713	42.1	70.3	72.1 (1.8)	-1.8	0.32
Chemo- & hormonal therapy	269					
Estrogen-receptor status						
Positive	138	8.2	74.1	73.6 (3.8)	0.5	0.90
Negative	58	3.4	50.7	62.0 (6.4)	-11.3	0.08
Unknown	73	4.3	66.3	75.0 (5.1)	-8.7	0.09

BCSS, Breast cancer-specific survival; SE, Standard error.

P-values based on one sample t-test.

Supplementary table 5. Adjuvant! predicted *versus* observed outcomes stratified by location (within NKI-AVL or elsewhere) of primary surgery.

	Patients		Adjuvant! predicted	Observed (SE)	Predicted - Observed	P-value
	No.	%				
Overall survival						
Location of primary treatment	5380					
Primary surgery NKI-AVL	1659	30.8	71.1	72.6 (1.1)	-1.5	0.17
Primary surgery elsewhere	3721	69.2	68.2	67.4 (0.8)	0.8	0.32
Breast cancer-specific survival						
Location of primary treatment	5380					
Primary surgery NKI-AVL	1659	30.8	80.4	80.4 (1.0)	0.0	1.00
Primary surgery elsewhere	3721	69.2	77.1	77.8 (0.7)	-0.7	0.32

NKI-AVL, Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital; SE, Standard error.

P-values based on one sample t-test.