The Breast 24 (2015) 705–712

Contents lists available at ScienceDirect

The Breast

journal homepage: www.elsevier.com/brst

Original article

St. Gallen endocrine response classes predict recurrence rates over time

R.H.T. Koornstra [a, b, *], K.J. Beelen [a, b], A.D. Vincent [c], V. van der Noort [c], P.J. van Diest [d], S.C. Linn [a, b, d]

[a] Department of Molecular Biology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
[b] Department of Medical Oncology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
[c] Department of Biometrics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
[d] Department of Pathology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 22 December 2014
Received in revised form
6 July 2015
Accepted 21 August 2015
Available online 1 October 2015

Keywords:
Breast cancer
Endocrine responsiveness
Tamoxifen
Adjuvant therapy
Estrogen receptor
Progesterone receptor

ABSTRACT

Background: In 2007 the St. Gallen consensus panel defined three endocrine response classes: highly endocrine responsive (ER-H), incomplete endocrine responsive (ER-I) and non-endocrine responsive tumours (ER-N). However, it is uncertain whether ER-I tumours are less responsive than ER-H tumours. We investigated whether recurrence rates vary over time between response classes. Additionally, we investigated the most predictive response class definition for tamoxifen benefit.
Patients and methods: We recollected tumours from 646 patients who participated in a randomized trial of adjuvant tamoxifen vs. observation. Estrogen receptor (ER), progesterone receptor (PgR), HER2 status and tumour grade were revised centrally. St. Gallen classes were evaluated for recurrence free interval (RFI). Change in hazards over time was assessed. Subsequently, 6 alternative response class definitions were compared to optimize the cut-off for PgR and ER.
Results: Schoenfeld residuals indicate a failure of proportional hazards between the endocrine response groups ($p = 0.0001$). The HR for recurrence risk shifted over time with the ER-H group initially being at lower risk (HR ER-H vs. ER-I 0.5), but after six years the recurrence risk increased (HR 1.9). The cut-off values for ER and PgR that statistically best discriminated RFI in the first 4 years for lymph node positive patients were ER $\geq$ 50% and PgR $\geq$ 75%.
Conclusion: We demonstrated a marked variability in endocrine therapy benefit. Patients with ER-H tumours have a larger benefit during adjuvant tamoxifen and in the first years after accomplishing of the therapy, but suffer from late recurrences. This might have implications for optimal treatment duration.

Background

Estrogen receptor (ER) positive breast cancer patients are generally recommended adjuvant endocrine therapy. Although recurrence rates and mortality have decreased substantially, still more than 30% of these patients relapse [1].

While predictive biomarkers in breast cancer are studied widely, ER is still the most powerful individual predictor of benefit from endocrine therapy [2]. However many questions concerning technique and scoring of ER are still unanswered. Already in 1999 Harvey et al. described and validated a composite score (a combination of intensity and percentage of positive staining invasive cells) of immunohistochemistry (IHC) staining for ER in breast cancer [3]. However currently, with automated staining, oversaturated conditions to allow rapid reaction (about 30 min or less) for antibody–antigen binding lead to oversaturation on intensity. Therefore, in the 2010 ASCO pathology guidelines on ER testing the need and utility to report composite scores (e.g. H-score, Allred, or quick score) is questioned [4].

Furthermore, the cut-off of positive staining cells varies greatly. In the clinical guidelines the hormone receptor (HR) status is assessed as a dichotomous variable with a tumour considered hormone receptor positive when $\geq$1% of the invasive cells stain

* Corresponding author. Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. Tel.: +31 20 5122951; fax: +31 20 5122572.
E-mail address: r.koornstra@nki.nl (R.H.T. Koornstra).

http://dx.doi.org/10.1016/j.breast.2015.08.009
0960-9776/

positive for ER and/or progesterone receptor (PgR) [4–7]. In current clinical practice, a range of thresholds is used, varying from 1% to 10%. However prospective data addressing the optimal cutoff level correlated with the efficacy of endocrine treatment are lacking. In retrospective studies survival rates of patients with tumours expressing ER in 1–9% did not significantly differ from patients with ER <1% tumours [8]. In molecular subtyping studies most tumours expressing ER 1–9% show ER–, basal-like molecular characteristics [9]. Additionally, while the incidence of low ER score (1–9%) is only approximately 1% [10], this represents a very little clinical problem. In the Netherlands the cut-off to consider a tumour HR+ is defined at 10% of invasive cells staining positive [11]. A positive hormone receptor status is considered sufficient to justify 5 years of adjuvant endocrine therapy [12]. However, this completely disregards the potential variance in treatment sensitivity of individual HR+ tumours.

In current clinical practice, several decision aids are being developed. For example the 21-gene recurrence score has been shown to predict distant recurrence in tamoxifen-treated patients [13]. Since ER, PgR and HER2 are major components of gene-expressing based assays, it might be speculated that standardized, quality-controlled pathological tests may have a similar predictive power. Actually, Allison et al. showed that features like PgR, grade and Ki67 can predict recurrence score in a subset of ER+ patients [14]. Whether or not the combined analysis of precise percentages of receptor expression levels can also predict benefit from endocrine therapy is unknown.

Already in 2007 the St. Gallen consensus panel defined three endocrine response classes for decision making on adjuvant endocrine and chemotherapeutic treatment [15–17]. It is accepted to define the St. Gallen classes as follows: tumours lacking both ER and PgR (using a cut-off of 1% of positive nuclei) are “non endocrine-responsive” (ER-N); tumours expressing both ER and PgR in 50% or more of the tumour cells are “highly endocrine responsive” (ER-H); all others are “incomplete endocrine responsive” (ER-I) [18,19]. However, validation of these definitions is lacking [18]. And though its nomenclature suggests this, it is not clear whether ER-I tumours are less endocrine responsive than ER-H tumours. In a recent publication, Dowsett et al. demonstrated, that ER mRNA expression levels are positively correlated with tamoxifen benefit [20]. We investigated whether there is a difference in treatment benefit and whether there are differences in treatment benefit over time based on St. Gallen classes, defined with IHC. Furthermore we investigated the optimal cut-off values for defining the response classes.

Methods

Patients and material

We recollected primary tumours from stage I–III postmenopausal breast cancer patients from the IKA tamoxifen trial (IKA trial, 1982–1994). Patients were randomized (2:1) between one year tamoxifen (30 mg per day) versus no adjuvant therapy, followed by a second randomization, after one year, for the tamoxifen treated patients to receive another two years of tamoxifen or to stop further treatment. Patients were eligible if they were postmenopausal, less than 76 years of age and had a $T_{1–4}$, $N_{0–3}$, M_0 breast tumour [21] but no mastitis or palpable supra- or infraclavicular lymph nodes. From 1989, based on two interim analyses showing a significant improvement in recurrence free survival in node positive patients, these patients skipped the first randomization and all received 1 year of tamoxifen. Eventually 1662 patients were enrolled with a 3:1 distribution (tamoxifen vs. no treatment), and none received adjuvant chemotherapy. The study data and patient characteristics of the original study group [22] and different subgroups have been presented elsewhere [23–26]. The data were also part of the Oxford systematic review [1].

The central ethics committee of the Netherlands Cancer Institute approved the original trial. Informed consent was obtained from all study participants. For this retrospective study, no additional consent was required according to the Dutch Act on medical research involving humans, and in compliance with Good Clinical Practice guidelines [27] since archival pathology left-over material handling does not interfere with patient care. Tumour tissue was used according to the code of conduct of Human Tissue and Medical Research: Code of conduct for responsible use (2011)” by the Federa (http://www.federa.org/codes conduct).

Pathology

Tissue blocks from 739 participating patients could be traced. After construction of a tissue microarray (TMA), 646 tumours were left with sufficient material for both ER, PgR and HER2 analysis. Median follow-up of this series for RFI is 9.56 years (95%CI: 9.14–9.90). Median follow-up for overall survival is 13.2 years (95%

Table 1
Patient characteristics vs. St. Gallen response groups.

	ER-N	ER-I	ER-H	All	p-value
	N = 135	N = 329	N = 182	N = 646	
Age					
Median	62	66	65	65	
(Range)	(47–78)	(45–80)	(49–77)	(45–80)	0.04
Treatment					
No Tamoxifen	36 (27%)	74 (22%)	43 (24%)	153 (24%)	
Tamoxifen	99 (73%)	255 (78%)	139 (76%)	493 (76%)	0.58
Tamoxifen duration (mths)					
Median	14	13	14	13	
(Range)	(0.8–76)	(0.1–59)	(0.2–76)	(0.1–76)	
Nodal status					
N+	74 (55%)	156 (47%)	78 (43%)	308 (48%)	
N–	61 (45%)	173 (53%)	104 (57%)	338 (52%)	0.04
ER (cut-off 1%)					
0%	135 (100%)	9 (3%)	0 (0%)	144 (22%)	
1–100%	0 (0%)	320 (97%)	182 (100%)	502 (78%)	
ER (St. Gallen crit.)					
0%	135 (100%)	9 (3%)	0 (0%)	144 (22%)	
1–50%	0 (0%)	27 (8%)	0 (0%)	27 (4%)	
>50%	0 (0%)	293 (89%)	182 (100%)	475 (74%)	
PgR (cut-off 1%)					
0%	135 (100%)	200 (61%)	0 (0%)	335 (52%)	
1–100%	0 (0%)	129 (39%)	182 (100%)	311 (48%)	
PgR (St. Gallen crit.)					
0%	135 (100%)	200 (61%)	0 (0%)	335 (52%)	
1–50%	0 (0%)	122 (37%)	0 (0%)	122 (19%)	
>50%	0	7 (2%)	182 (100%)	189 (29%)	
HER2					
Negative	95 (70%)	277 (84%)	169 (93%)	541 (84%)	
Positive	36 (27%)	38 (12%)	6 (3%)	80 (12%)	
Unknown	4 (3%)	14 (4%)	7 (4%)	25 (4%)	<0.0001
T-stage (clinical)					
T1	32 (24%)	104 (32%)	47 (26%)	183 (28%)	
T2	84 (62%)	186 (57%)	114 (63%)	384 (59%)	
T3	12 (9%)	25 (8%)	16 (9%)	53 (8%)	
T4	5 (4%)	10 (3%)	5 (3%)	20 (3%)	
Unknown	2 (1%)	4 (1%)	0 (0%)	6 (<1%)	0.74
Grade					
1	4 (3%)	72 (22%)	57 (31%)	133 (21%)	
2	33 (24%)	119 (36%)	81 (45%)	233 (36%)	
3	98 (73%)	138 (42%)	44 (24%)	280 (43%)	<0.0001

ER-H = highly endocrine responsive patients; ER-I = incomplete endocrine responsive patients; ER-N = non-endocrine responsive patients.

CI: 12.2–14.7). For the remaining patient characteristics of the current analysis see Table 1.

One observer (PJvD) centrally revised all tumours. TMAs were stained for ER, PgR and HER2. The percentage of positive nuclei was scored for ER and PgR. HER2 was considered positive when membranous staining was DAKO score 3. In case of a DAKO score 2, Silver In Situ Hybridization was performed using UltraView SISH Detection Kit (Ventana®) according to the manufacturer's instructions.

In order to identify the optimal cut-off values for defining the response classes, we predefined six possible definitions of cut-off values for ER and PgR to compare. Definition 1: ER $\geq$ 50% & PgR $\geq$ 75%; definition 2: ER $\geq$ 50% & PgR $\geq$ 50% (current definition); definition 3: ER $\geq$ 50% & PgR $\geq$ 25%; definition 4: ER $\geq$ 75%; definition 5: ER $\geq$ 50%; definition 6: ER $\geq$ 25%.

Statistical methods

The associations between endocrine response groups and grade, tumour stage, nodal stage and HER2 status were tested using linear-by-linear tests. The association between ER $\geq$ 50% and PgR $\geq$ 50% was tested using a Fisher exact test. Recurrence free interval (RFI) was defined as the time from the first randomization until the occurrence of a local, regional or distant recurrence or breast cancer specific death [28]. Effect of tamoxifen treatment was assessed using Cox proportional hazard models. Due to the change in randomization of the original trial analyses are stratified by nodal status, where appropriate.

Multivariable Cox regressions include tumour grade, T-stage, HER2 status and age as covariates. The proportional hazards assumption was tested using the Schoenfeld's partial residuals technique [29]. The assessment of hazard ratios during particular periods (e.g., 0–2, 2–4, 4–6 and >6 years) were assessed using landmark analyses including patients who were event free at the start of each period, and being censored at the end of each period. The model log-likelihood chi-squares and concordance indices were used to determine the best-predictive endocrine response definition using differing ER and PgR thresholds. RFI curves were constructed using the Kaplan–Meier method. All analyses were performed using R (v3.0.1).

Results

Endocrine response groups correlate with grade and HER2

ER-N patients (N = 135, 21%) did not exhibit significant benefit from tamoxifen (multivariable HR = 0.8 (0.4–1.5)) and were excluded from further analyses. 329 tumours were classified as ER-I (51%) and 182 tumours were classified as ER-H (28%). ER-I tumours were more often HER2 positive ($p < 0.0001$) and were associated with higher tumour grade ($p < 0.0001$).

Overall tamoxifen benefit in ER positive patients

In this representative subset (n = 511) of the total study population, the multivariable recurrence free interval hazard ratio (HR) for tamoxifen in ER+ patients was 0.3 (95%CI 0.2–0.6; p = 0.003) for lymph node positive patients and 0.7 (95%CI 0.4–1.4; p = 0.36) for lymph node negative patients. When stratified by nodal status, the tamoxifen effect in ER+ patients was 0.5 (95% CI 0.3–0.8, p = 0.002). The median duration of tamoxifen treatment was 13 months (range 0.1–76 months).

Hazard ratios for recurrence vary over time

Examination of Schoenfelds partial residuals indicated a failure of proportional hazards for recurrence between the ER-H and ER-I groups (p = 0.0001). Fig. 1a presents the univariate Kaplan–Meier curves for the three endocrine response groups, both treated versus untreated and nodal positive versus negative. The RFI-curves for the ER-H group exhibit a decline after approximately six years. In Fig. 1b the time periods are separated in two periods (0–5 years and 5 year-later) in order to visualize the effect of the non-proportional hazards more clearly. However, since these are univariate curves one should be careful for over-interpretation of the curves only and focus more on adjusted hazard ratios. The failure of proportional hazards was still present after the inclusion of tumour grade, T-stage, HER2 status, age and the pairwise treatment-St. Gallen class interaction (p = 0.0003). Landmark analyses for the hazard ratios for recurrence risk during different time periods illustrate this shift in recurrence risk. The ER-N group was initially at higher risk for recurrence when compared with the ER-I group, however this increased risk faded over time. Even more striking was the difference between the ER-H and ER-I group over time, with the ER-H group initially at lower risk (HR 0.5), but over time the recurrence risk increased (HR 1.9) (Table 2a). Restricting the analysis to N+ patients did not affect this result (Table 2b). This increased risk over time is also reflected in approaching Kaplan–Meier curves for OS in tamoxifen treated patients (Fig. S1).

Refining the St. Gallen response class definitions: best predicting tamoxifen benefit

For testing 6 alternative receptor level cut-off definitions, best discriminating tamoxifen benefit, we restricted the analysis to the tamoxifen treated ER-H and ER-I N+ patients (according to the protocol amendment in 1989, identifying those patients at highest risk) noting that ER and PgR levels were not independent ($p < 0.0001$, Fig. 2). We censored at four years in order to prevent dilution of the treatment effect in time. With this restriction we found the most pronounced separation between the ER-I and the ER-H patients concerning RFI to be with ER $\geq$ 50 & PgR $\geq$ 75 (definition 1) (Chi sq = 3.27, p = 0.07, C-index 0.682; Table 3). This definition also provided the most pronounced separation of the survival curves for the following years (also see Table 4 and Fig. S2).

Discussion

Our results indicate that, although high levels of hormone receptors are associated with tamoxifen benefit during treatment and in the first years after accomplishing therapy, after this period, the recurrence risk for patients with ER-H tumours exceeded that of patients with ER-I tumours.

Generally, it is known that HR+ breast cancer has a propensity for late recurrences. More than half of all disease recurrences in ER+ breast cancer occur 6 years or more after diagnosis, particularly following 5 years of adjuvant anti-estrogen therapy [1,30]. In several studies it has been shown that HR+ tumours have better prognosis during the first years of follow up and subsequently showing late recurrences, while in HR-tumours the mortality rate and relapse rate are initially high and then progressively level off over time [31–33].

The increased rate of late recurrences in ER+ breast cancer is a well-known, significant clinical challenge. However, little attention has been paid to non-proportional hazards when analysing the differences in recurrence risk for HR+ breast cancer patients according to their receptor levels.

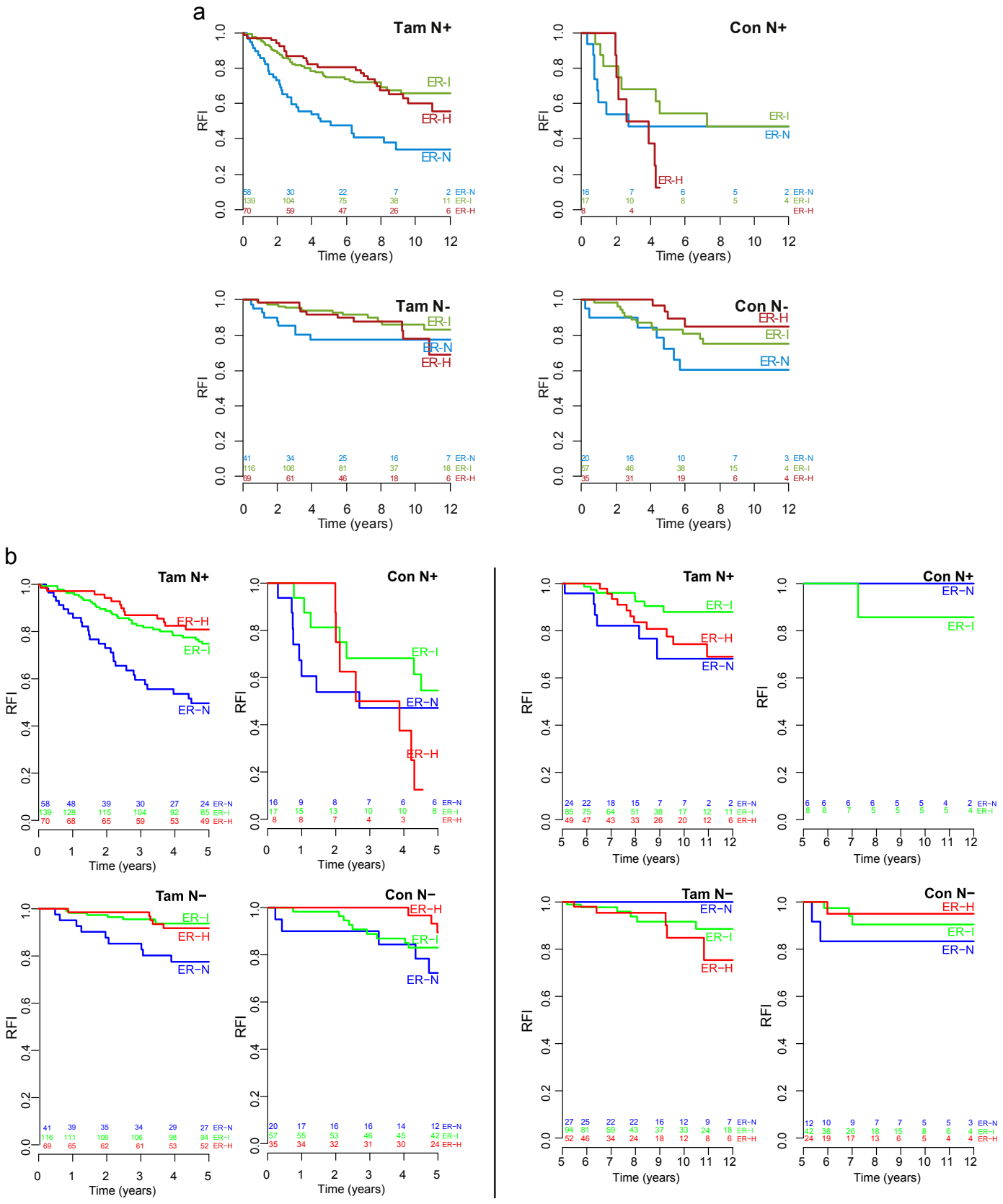

Fig. 1. a: Kaplan–Meier curves (RFI) for the three endocrine response groups, both treated vs. untreated and nodal positive vs. negative (RFI = relapse free interval; Tam = tamoxifen treated; Con = control group, untreated patients; N+ = lymph node positive disease; N− = lymph node negative disease; ER-H = highly endocrine responsive patients; ER-I = incomplete endocrine responsive patients; ER-N = non-endocrine responsive patients). b: Kaplan–Meier curves (RFI), same as Fig. a, now separated in two time periods. Left panel: 0–5 years. Right panel: 5–12 years (100% is defined as all patients without recurrence at 5 years followup).

Table 2a
Landmark analyses hazard ratios (90% confidence intervals) for the Cox proportional hazard models including St. Gallen groups and treatment, tumour grade, HER2 status and age and stratifying for nodal status. Time is divided into four period, patients are only included in a section if they are recurrence free at the start of that period, and patients without events are censored at the end of each period.

	0–2 years	2–4 years	4–6 years	6+ years
ER-N vs. ER-I	3.3 (2–5.2)	1.9 (1.1–3.2)	1.6 (0.7–3.6)	1.1 (0.5–2.7)
ER-H vs. ER-I	0.5 (0.2–1.1)	1.1 (0.7–1.9)	1.3 (0.6–2.8)	1.9 (1–3.6)
Tam vs. Con	0.7 (0.4–1.1)	0.7 (0.4–1.1)	0.2 (0.1–0.3)	1.8 (0.6–5.1)

Table 2b
Sensitivity analysis, same as in Table 2a, however restricted to only N+ patients.

	0–2 years	2–4 years	4–6 years	6+ years
ER-N vs. ER-I	2.8 (1.6–4.7)	2 (1–3.9)	1 (0.3–3.3)	2.4 (0.9–6.3)
ER-H vs. ER-I	0.5 (0.2–1.2)	1.5 (0.8–2.8)	1 (0.3–3.2)	2.4 (1.1–5.3)
Tam vs. Con	0.5 (0.3–0.9)	0.5 (0.3–1.1)	0.2 (0.1–0.6)	2.7 (0.5–15.4)

ER-H = highly endocrine responsive patients; ER-I = incomplete endocrine responsive patients; ER-N = non-endocrine responsive patients.

Results of previous studies on the distinctive nature of receptor levels are contradictory, leading to a comprehensive discussion concerning prognostic versus predictive value of ER and PgR expression. Tovey et al. found favourable outcome for tamoxifen treated PgR+ patients, compared to PgR− patients, during the first years after diagnosis with an increased risk after 3 years of tamoxifen [34], which is in accordance with our results. However the data from the Oxford systematic review showed that the overall benefit from 5 years tamoxifen was not different between ER+ patients with PgR+ versus PgR− tumours [30]. These contradictory results are possibly confounded by adjuvant chemotherapy benefit, since Tovey et al. do not report on the percentage of patients receiving adjuvant chemotherapy. In our current series, patients received no adjuvant chemotherapy, whereby our results might be a more pure reflection of prognosis versus tamoxifen benefit. Additionally, the conflicting results may also partly be due to the assay used for ER and PgR receptor assessment. In the Oxford systematic review most tumours had been assessed with the ligand-binding assay [1], while in our study all cases have been analysed using standardized immunohistochemistry.

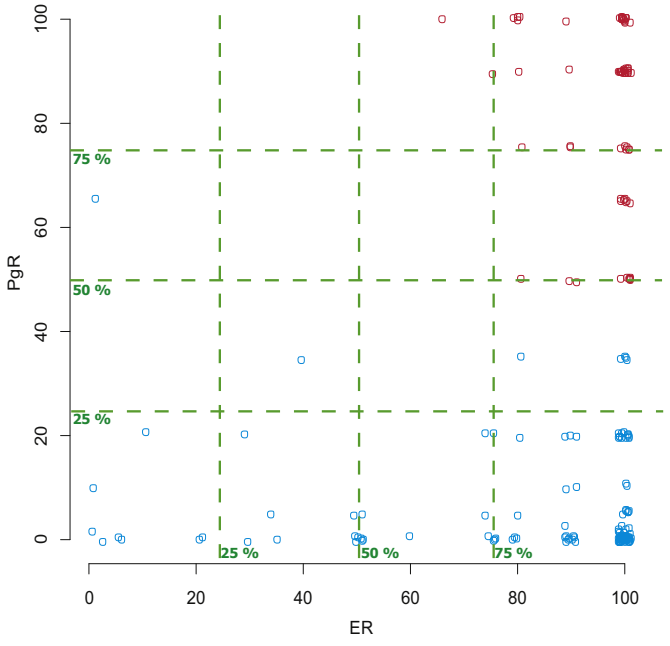

Fig. 2. ER vs. PgR status. Blue dots are ER-I and red dots are ER-H patients using the current definition. Green dotted lines are indicating the different cut-off values as discussed in the section on alternative definitions of the St. Gallen criteria.

Also the scoring method might influence results. The different composite scores are all somewhat differently calculated and the most recent ASCO guideline warns of confusion across institutions [4]. Moreover, with the current automated staining techniques, oversaturated conditions to allow rapid reaction (about 30 min or less) for antibody–antigen binding lead to oversaturation on intensity. In contrast to IHC4 [35,36], based on composite IHC score (ER, PgR, HER2 and Ki67) we could identify patients at higher risk for late recurrences. A possible explanation is that IHC4 used H-score with three predefined quartile cut-off points, but inaccuracy due to oversaturation might also be an explanation.

Recently several authors reported on the capability of genomic prognostic classifiers to provide information for late distant recurrences after adjuvant therapy [36–39]. Although only the Breast Cancer Index-linear [39] could add independent prognostic information regarding the 10-year recurrence-free interval [36], all described gene classifiers (EndoPredict, PAM50-ROR score, I/H-ratio and BCI) offer additional significant prognostic information concerning the late recurrence-risk, beyond the clinical–pathological predictors [36–39]. Additionally, none of the gene classifiers did evaluate the *levels* of ER and PgR expression as predictors for late recurrences. All these reports focus on the low-risk patients and the intention to prevent those patients from overtreatment, since they seem not to benefit from additional, extended treatment. In contrast, we report on the increased risk over time in the specific ER-H patients. Moreover, this additional information is already available for every patient without extra costs. And although molecular tests might be more accurate than expression levels of ER and PgR, an assumption that still needs to be tested, we do see an application of the cheap, additional information provided by hormone receptor expression levels to guide therapy decisions, especially in areas of the world where availability of expensive molecular tests is non-existent.

Several trials have been conducted to study the effect of extended adjuvant endocrine treatment. Lately the ATLAS-trial and the aTTom-trial reported modest benefit of extending endocrine therapy beyond five years in unselected populations of mainly post-menopausal patients [40,41]. Two recent meta-analyses also reported contradictory results [42,43]. Petrelli and co-workers conclude that extended endocrine therapy an opportunity is for ER+ breast cancer patients to reduce recurrence risk and mortality. Al Mubarak and colleagues found no significant reduction in recurrence and all-cause death in unselected patients [43]. Some possible explanations for the different conclusions are differences in statistical methods (fixed effect modelling was not feasible for Al Mubarak; individual patient data was not available; the choice of endpoint differs in the different studies, with locoregional recurrence versus distant recurrence). The large proportion ER− or ER unknown breast cancer patients in the aTTom trial, who were excluded from the analyses by Petrelli could have biased the conclusion of Al Mubarak. Furthermore, according to Petrelli the subgroup benefitting most are the postmenopausal, N+ patients. It might also be (partly) due to a different beneficial effect in different ER+ breast cancer subgroups. Unfortunately, still no means are available to precisely identify the actual subgroup benefitting. An exploratory, retrospective analysis of the MA.17 data showed an improvement in disease free survival due to extended adjuvant endocrine treatment for patients with ER+/PgR+ tumours and not for patients with ER+/PgR− tumours [44]. Furthermore, Dowsett

Table 3
Hazard ratios (95% confidence intervals) and concordance indices for Cox proportional hazards models for the 6 *alternative* definitions of incomplete and high responsive groups. The models contain the response definition (ER-I vs. ER-H), treatment group (control vs. tamoxifen), tumour grade, HER2-status, age, and their interaction. The column Chi sq gives the log likelihood chi-square of the models. The cohort is restricted to node positive patients with RFI censored at 4 years.

		HR	95% CI	Chi sq	df	p-value	C-index
Definition 1	Tam-H	Ref		3.27	1	0.07	0.682
ER ≥ 50% & PgR ≥ 75%	Tam-I	1.7	(0.7–4.3)				
	Con-H	7.1	(2.1–23.8)				
	Con-I	2.9	(0.8–9.8)				
Definition 2	Tam-H	Ref		2.29	1	0.13	0.678
ER ≥ 50% & PgR ≥ 50% (*current*)	Tam-I	1.3	(0.6–2.6)				
	Con-H	5.5	(1.9–16.1)				
	Con-I	2.2	(0.7–6.5)				
Definition 3	Tam-H	Ref		1.67	1	0.2	0.676
ER ≥ 50% & PgR ≥ 25%	Tam-I	1.4	(0.7–2.9)				
	Con-H	4.9	(0.9–14.1)				
	Con-I	2.6	(0.9–7.7)				
Definition 4	Tam-H	Ref		2.16	1	0.14	0.678
ER ≥ 75%	Tam-I	0.4	(0.1–1.7)				
	Con-H	2.1	(0.9–5.2)				
	Con-I	3.5	(1.2–10.3)				
Definition 5	Tam-H	Ref		0.82	1	0.37	0.667
ER ≥ 50%	Tam-I	0.7	(0.2–2.9)				
	Con-H	2.3	(0.9–5.5)				
	Con-I	3.7	(1.2–10.8)				
Definition 6	Tam-H	Ref		1.63	1	0.2	0.677
ER ≥ 25%	Tam-I	0.5	(0.1–3.4)				
	Con-H	2.2	(1.0–5.1)				
	Con-I	4.6	(1.3–15.9)				

Tam-H = highly endocrine responsive patients, treated with Tamoxifen; Tam-I = incomplete endocrine responsive patients, treated with Tamoxifen; Con-H = highly endocrine responsive patients, untreated; Con-I = incomplete endocrine responsive patients, untreated.

et al. have shown that women with ER+/HER− breast cancer had more than double the risk of their cancer recurring between five and ten years after surgery and five years of adjuvant hormone therapy [20]. This is completely in line with our results. This shift in the relative risk of recurrence in ER-H patients may have biologic and therapeutic relevance. Our data support the hypothesis that patients with ER-H tumours might be candidates for prolonged treatment. Therefore, we would like to encourage the authors of the NSABP B-14 trial [45] and of the extended adjuvant endocrine therapy trials [40,41] to perform a subset analysis based on the St. Gallen classes to confirm our results.

Furthermore, the best cut-off of ER and PgR levels for endocrine response classes appears to be definition 1 (ER ≥ 50%; PgR ≥ 75%). Although none of the definitions differed statistical significantly from definition 2, definition 1 showed a trend towards providing the most pronounced separation of the survival curves with a distinct decline for the ER-H curve after six years (Fig. S2). However, the maximum C-index difference is very limited (0.667 versus 0.682) and it is unclear whether this is clinically relevant. Within HR+ patients these cut-off values best divided patients into two groups with intermediate or excellent tamoxifen-benefit during treatment. Stendahl et al. found the same cut-off level for PgR to be most predictive for tamoxifen benefit in a series of 500 premenopausal patients [46]. Though, our results show a marked variability in therapy sensitivity in HR+ breast cancer patients, the reviewed cut-off values should only be hypothesis generating and an

Table 4
Landmark analyses hazard ratios (90% confidence intervals) for the Cox proportional hazard models for the 6 alternative St. Gallen definitions including treatment, tumour grade, HER2 status and age and stratifying for nodal status. Time is divided into four period, patients are only included in a section if they are recurrence free at the start of that period, and patients without events are censored at the end of each period. (Note: definition 2 is the current definition.)

		0–2 years	2–4 years	4–6 years	6-Inf years
Definition 1	ER-I vs ER-H (all patients)	1.7 (0.6–5.2)	0.9 (0.4–2)	0.4 (0.2–1.1)	0.4 (0.2–0.9)
ER ≥ 50% & PgR ≥ 75%	Con vs Tam (all patients)	1.3 (0.5–3.1)	2.5 (1.3–4.6)	6.5 (2.8–15)	0.5 (0.1–1.8)
	ER-I vs ER-H (Tam. treated)	2.0 (0.6–7.1)	1.1 (0.4–2.8)	0.7 (0.1–3.5)	0.4 (0.2–0.9)
Definition 2	ER-I vs ER-H (all patients)	2.2 (0.8–6)	0.7 (0.4–1.4)	0.5 (0.2–1.3)	0.6 (0.3–1.3)
ER ≥ 50% & PgR ≥ 50% (*current*)	Con vs Tam (all patients)	1.2 (0.5–2.9)	2.5 (1.3–4.7)	6.4 (2.8–14.9)	0.5 (0.1–1.7)
	ER-I vs ER-H (Tam. treated)	2.4 (0.8–7.4)	0.6 (0.3–1.4)	1.1 (0.2–5.7)	0.5 (0.2–1.2)
Definition 3	ER-I vs ER-H (all patients)	1.9 (0.8–5)	0.8 (0.4–1.6)	0.7 (0.3–1.7)	0.7 (0.3–1.4)
ER ≥ 50% & PgR ≥ 25%	Con vs Tam (all patients)	1.3 (0.5–3)	2.5 (1.3–4.6)	6.1 (2.7–14)	0.5 (0.1–1.6)
	ER-I vs ER-H (Tam. treated)	2.7 (0.9–8.2)	0.7 (0.3–1.6)	1.4 (0.3–6.8)	0.6 (0.3–1.3)
Definition 4	ER-I vs ER-H (all patients)	1.5 (0.5–4.1)	0.7 (0.2–2)	0.6 (0.1–2.6)	1.0 (0.3–3.6)
ER ≥ 75%	Con vs Tam (all patients)	1.2 (0.5–2.9)	2.5 (1.4–4.7)	6.1 (2.7–13.8)	0.4 (0.1–1.6)
	ER-I vs ER-H (Tam. treated)	0.9 (0.2–3.8)	0.6 (0.1–2.4)	1.0 (0.1–8.6)	1.5 (0.4–5.2)
Definition 5	ER-I vs ER-H (all patients)	2.2 (0.8–6.1)	0.5 (0.1–2)	0.8 (0.2–3.5)	1.1 (0.2–4.8)
ER ≥ 50%	Con vs Tam (all patients)	1.1 (0.4–2.6)	2.6 (1.4–4.9)	6.0 (2.6–13.8)	0.4 (0.1–1.6)
	ER-I vs ER-H (Tam. treated)	1.5 (0.3–6.4)	0.5 (0.1–3.4)	1.7 (0.2–15.3)	1.6 (0.4–7.2)
Definition 6	ER-I vs ER-H (all patients)	2.5 (0.8–8)	0.0 (0–Inf)	0.7 (0.1–5.6)	1.5 (0.3–6.9)
ER ≥ 25%	Con vs Tam (all patients)	1.1 (0.4–2.6)	2.6 (1.4–4.8)	5.9 (2.6–13.3)	0.4 (0.1–1.5)
	ER-I vs ER-H (Tam. treated)	1.1 (0.1–8.4)	0.0 (0–Inf)	2.2 (0.2–22.2)	2.3 (0.5–10.9)

Tam = tamoxifen treated; Con = control group/untreated patients; ER-H = highly endocrine responsive patients; ER-I = incomplete endocrine responsive patients; ER-N = non-endocrine responsive patients.

incentive for other authors to investigate the cut-off levels of the St. Gallen response classes.

Currently, IHC staining and scoring for hormone receptor status is standardized, in order to optimize the prognostic and predictive value [4,5,47]. However, a dichotomous technique completely neglects the variable treatment sensitivity of ER+ tumours. Subsequently, a second step in treatment decision-taking can be considered after the simple dichotomous determination of hormone receptor status, namely the classification conform the St. Gallen response classes in order to decide on duration. Extended endocrine treatment in all ER+ breast cancer patients with tumours expressing higher PgR levels can than be considered.

Our study has several limitations. First of all, sufficient tumour material was available from only 646 patients, and only 511 HR+ patients. However, this subgroup was comparable to the total original study population (data not shown). Furthermore, the time on tamoxifen therapy is, with a median duration of 13 months, less then the current standard of 5 years. However, we hypothesize that our results can be extrapolated over time, based on the results of Dowsett et al. and the observation that the increased recurrence risk occurs after treatment cessation [20]. Next, it is not clear whether our results also apply for aromatase inhibitor based therapy. And although our results are considered to mainly reflect intrinsic tumour biology rather than a specific treatment effect, it has to be determined whether our data can also be reproduced with an aromatase inhibitor treatment.

In conclusion, we have shown a marked variability in therapy sensitivity within the group of patients with HR+ breast tumours. Patients with ER-H tumours have an increased tamoxifen benefit during therapy and in the first years after completion of adjuvant tamoxifen, but suffer from late recurrences. We propose that these patients may benefit from prolonged endocrine treatment.

Conflict of interest statement

None of the authors have conflicts of interest to declare.

Acknowledgements

We would like to thank J.B. Vermorken for designing the original trial and thereby offering us the opportunity to perform these analyses. We thank all pathology departments throughout the Netherlands for submission of FFPE tumour blocks. This work was supported by a grant from A Sisters' Hope.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.breast.2015.08.009.

References

[1] Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005;365:1687–717.
[2] Payne SJL, Bowen L, Jones L, Wells A. Predictive markers in breast cancer – the present. Histopathology 2007;52:82–90.
[3] Harvey JM, Clark GM, Osborne CK, Allred DC. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol 1999;17:1474–81.
[4] Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol 2010;28:2784–95.
[5] Aebi S, Davidson T, Gruber G, Cardoso F, ESMO Guidelines Working Group. Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2011;22(Suppl. 6):i12–24.
[6] Burstein HJ, Prestrud AA, Seidenfeld J, Anderson H, Buchholz TA, Davidson NE, et al. American Society of Clinical Oncology clinical practice guideline: update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. J Clin Oncol 2010;28:3784–96.
[7] Carlson RW, Hudis CA, Pritchard KI, St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer. Adjuvant endocrine therapy in hormone receptor-positive postmenopausal breast cancer: evolution of NCCN, ASCO, and St Gallen recommendations. J Natl Compr Cancer Netw 2006;4:971–9.
[8] Yi M, Huo L, Koenig KB, Mittendorf EA, Meric-Bernstam F, Kuerer HM, et al. Which threshold for ER positivity? A retrospective study based on 9639 patients. Ann Oncol 2014;25:1004–11.
[9] Iwamoto T, Booser D, Valero V, Murray JL, Koenig K, Esteva FJ, et al. Estrogen receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry. J Clin Oncol 2012;30:729–34.
[10] Collins LC, Botero ML, Schnitt SJ. Bimodal frequency distribution of estrogen receptor immunohistochemical staining results in breast cancer: an analysis of 825 cases. Am J Clin Pathol 2005;123:16–20.
[11] http://www.oncoline.nl/mammacarcinoom. Mammacarcinoom, landelijke richtlijn (versie 2.0), Integraal Kankercentrum Nederland, Richtlijnen Oncologische Zorg [date retrieved 06.07.15.].
[12] Goldhirsch A, Ingle JN, Gelber RD, Coates AS, Thürlimann B, Senn HJ, et al. Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. Ann Oncol 2009;20:1319–29.
[13] Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26.
[14] Allison KH, Kandalaft PL, Sitlani CM, Dintzis SM, Gown AM. Routine pathologic parameters can predict Oncotype DX recurrence scores in subsets of ER positive patients: who does not always need testing? Breast Cancer Res Treat 2012;131:413–24.
[15] Bauer K, Parise C, Caggiano V. Use of ER/PR/HER2 subtypes in conjunction with the 2007 St Gallen Consensus Statement for early breast cancer. BMC Cancer 2010;10:228.
[16] Goldhirsch A, Wood WC, Gelber RD, Coates AS, Thürlimann B, Senn HJ, et al. Progress and promise: highlights of the international expert consensus on the primary therapy of early breast cancer 2007. Ann Oncol 2007;18:1133–44.
[17] Parise CA, Bauer KR, Brown MM, Caggiano V. Breast cancer subtypes as defined by the estrogen receptor (ER), progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2) among women with invasive breast cancer in California, 1999–2004. Breast J 2009;15:593–602.
[18] Colleoni M, Bagnardi V, Rotmensz N, Gelber RD, Viale G, Pruneri G, et al. Increasing steroid hormone receptors expression defines breast cancer subtypes non responsive to preoperative chemotherapy. Breast Cancer Res Treat 2009;116:359–69.
[19] Gnant M, Harbeck N, Thomssen C, St. Gallen. Summary of the consensus discussion. Breast Care (Basel) 2011;6:136–41.
[20] Dowsett M, Sestak I, Buus R, Lopez-Knowles E, Mallon E, Howell A, et al. Estrogen receptor expression in 21-gene recurrence score predicts increased late recurrence for estrogen-positive/HER2-negative breast cancer. Clin Cancer Res 2015;21:2763–70.
[21] American Joint Committee for Cancer Staging and End-Results Reporting, Manual for Staging of Cancer. first ed. Chicago, Illinois: American Joint Committee; 1977.
[22] Vermorken JB, Burgers JMV, Taat CW, Van de Slee P, Hennipman A, Nortier JWR, et al. Adjuvant tamoxifen in breast cancer: interim results of a comprehensive cancer center Amsterdam trial. Breast Cancer Res Treat 1998;50 (abstr #329).
[23] Beelen K, Opdam M, Severson TM, Koornstra RH, Vincent AD, Hauptmann M, et al. CYP2C19*2 predicts substantial tamoxifen benefit in postmenopausal breast cancer patients randomized between adjuvant tamoxifen and no systemic treatment. Breast Cancer Res Treat 2013;139:649–55.
[24] Beelen K, Opdam M, Severson TM, Koornstra RH, Vincent AD, Wesseling J, et al. Phosphorylated p-70S6K predicts tamoxifen resistance in postmenopausal breast cancer patients randomized between adjuvant tamoxifen versus no systemic treatment. Breast Cancer Res 2014;16. R6.
[25] Beelen K, Opdam M, Severson TM, Koornstra RH, Vincent AD, Wesseling J, et al. PIK3CA mutations, phosphatase and tensin homolog, human epidermal growth factor receptor 2 and insulin-like growth factor 1 receptor and adjuvant tamoxifen resistance in postmenopausal breast cancer patients. Breast Cancer Res 2014;16. R13.
[26] Michalides R, van Tinteren H, Balkenende A, Vermorken JB, Benraadt J, Huldij J, et al. Cyclin A is a prognostic indicator in early stage breast cancer with and without tamoxifen treatment. Br J Cancer 2002;86:402–8.
[27] http://wetten.overheid.nl/BWBR0009408/geldigheidsdatum_06-07-2015. Wet Medisch-wetenschappelijk Onderzoek Met Mensen – Wetten.nl – BWBR0009408 [date retrieved 06.07.15.].
[28] Hudis CA, Barlow WE, Costantino JP, Gray RJ, Pritchard KI, Chapman JA, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. J Clin Oncol 2007;25:2127–32.
[29] Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994;81:515–26.

[30] Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet 2011;378:771–84.
[31] Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med 2010;7:e1000279.
[32] Dignam JJ, Dukic V, Anderson SJ, Mamounas EP, Wickerham DL, Wolmark N. Hazard of recurrence and adjuvant treatment effects over time in lymph node-negative breast cancer. Breast Cancer Res Treat 2009;116:595–602.
[33] Hilsenbeck SG, Ravdin PM, de Moor CA, Chamness GC, Osborne CK, Clark GM. Time-dependence of hazard ratios for prognostic factors in primary breast cancer. Breast Cancer Res Treat 1998;52:227–37.
[34] Tovey S, Dunne B, Witton CJ, Forsyth A, Cooke TG, Bartlett JM. Can molecular markers predict when to implement treatment with aromatase inhibitors in invasive breast cancer? Clin Cancer Res 2005;11:4835–42.
[35] Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the genomic health recurrence score in early breast cancer. J Clin Oncol 2011;29:4273–8.
[36] Sgroi DC, Sestak I, Cuzick J, Zhang Y, Schnabel CA, Schroeder B, et al. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. Lancet Oncol 2013;14:1067–76.
[37] Filipits M, Nielsen TO, Rudas M, Greil R, Stöger H, Jakesz R, 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;20:1298–305.
[38] Sestak I, Cuzick J, Dowsett M, Lopez-Knowles E, Filipits M, Dubsky P, et al. Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and Arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. J Clin Oncol 2015;33:916–22.
[39] Zhang Y, Schnabel CA, Schroeder BE, Jerevall PL, Jankowitz RC, Fornander T, et al. Breast cancer index identifies early-stage estrogen receptor-positive breast cancer patients at risk for early- and late-distant recurrence. Clin Cancer Res 2013;19:4196–205.
[40] Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. Lancet 2012;381:805–16.
[41] Gray RG, Rea D, Handley K, Bowden SJ, Perry P, Earl HM, et al. aTTom: long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years in 6,953 women with early breast cancer. J Clin Oncol 2013;31. abstr #5.
[42] Petrelli F, Coinu A, Cabiddu M, Ghilardi M, Lonati V, Barni S. Five or more years of adjuvant endocrine therapy in breast cancer: a meta-analysis of published randomised trials. Breast Cancer Res Treat 2013;140:233–40.
[43] Al-Mubarak M, Tibau A, Templeton AJ, Cescon DW, Ocana A, Seruga B, et al. Extended adjuvant tamoxifen for early breast cancer: a meta-analysis. PLoS One 2014;9:e88238.
[44] Goss PE. Extending the benefits of adjuvant therapy in early HR+ breast cancer. Breast Cancer Res Treat 2008;112(Suppl. 1):45–52.
[45] Fisher B, Dignam J, Bryant J, Wolmark N. Five versus more than five years of tamoxifen for lymph node-negative breast cancer: updated findings from the National Surgical Adjuvant Breast and Bowel Project B-14 randomized trial. J Natl Cancer Inst 2001;93:684–90.
[46] Stendahl M, Rydén L, Nordenskjöld B, Jönsson PE, Landberg G, Jirström K. High progesterone receptor expression correlates to the effect of adjuvant tamoxifen in premenopausal breast cancer patients. Clin Cancer Res 2006;12:4614–8.
[47] http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf. W.J. Gradishar, NCCN clinical practice guidelines in oncology – breast cancer, version 2.2015 (date retrieved 06.07.15.).