Methodological quality assessment of studies investigating the OncotypeDX test

Study feature	Qualities sought	Ademuyiwa *et al.* (2011)[82]	Albain *et al.* (2010)[83]	Cuzick *et al.* (2011)[84]	Dowsett *et al.* (2010)[79]	Geffen *et al.* (2009)[77]	Holt *et al* (2011)[78] (abstract only)	Kelly *et al.* (2010)[85]	Lo *et al.* (2010)[76]	Tang *et al.* (2011);[81] Mamounas *et al.* (2010)[80]	Tang *et al.* (2010)[86] (abstract only)	Toi *et al.* (2010)[87]	Yorozuya *et al.* (2009)[88]
Sample of patients	Inclusion criteria defined	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Sample selection explained	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Adequate description of diagnostic criteria	N	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y
	Clinical and demographic characteristics fully described	Y	Y	N	Y	Y	N	Y	Y	N	N	Y	Y
	Representative (selected by random selection or as consecutive cases)	Y	U	U	U	Y	U	Y	Y	U	U	U	N
	Assembled at a common (usually early) point in the course of their disease	Y (ER+, HER2–, LN–)	U (LN+, ER+, postmenopausal)	U (all ER+ or PR+)	U (postmenopausal HR+ women only)	Y (T1N0M0)	Y	U (HR+ cancers only)	Y (LN–, ER+ only)	Y (LN–, ER+)	Y (LN–, ER+)	Y (early stage, ER+, LN–)	Y (LN–, ER+, stage I or IIA)
	Complete (all eligible patients were included)	Y	N	Y	N	N	N	Y	N	N	U	N	Y
Follow-up of patients	Sufficiently long	Y	Y	Y	Y	N	U	Y	N	Y	U	U	N

Study feature	Qualities sought	Ademuyiwa *et al.* (2011)[82]	Albain *et al.* (2010)[83]	Cuzick *et al.* (2011)[84]	Dowsett *et al.* (2010)[79]	Geffen *et al.* (2009)[77]	Holt *et al* (2011)[78] (abstract only)	Kelly *et al.* (2010)[85]	Lo *et al.* (2010)[76]	Tang *et al.* (2011);[81] Mamounas *et al.* (2010)[80]	Tang *et al.* (2010)[86] (abstract only)	Toi *et al.* (2010)[87]	Yorozuya *et al.* (2009)[88]
Outcome	Objective	Y	U	Y	Y	U	Y	Y	Y	Y	Y	U	Y
	Unbiased (e.g. assessment blinded to prognostic information)	Y	U	U	U	U	Y	N	Y	U	U	U	Y
	Fully defined	Y	Y	Y	N	U	Y	Y	Y	Y	U	Y	N
	Appropriate	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
	Known for all or a high proportion of patients	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Prognostic variable	Fully defined, including details of method of measurement if relevant	Y	Y	Y	Y	N	Y	N	N	Y	U	N	Y
	Precisely measured	Y	Y	Y	Y	Y	Y	U	Y	Y	U	Y	Y
	Available for all or a high proportion of patients	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
	If relevant, cut-point(s) defined and justified	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	U	Y (reference provided)	Y (reference provided)

Study feature	Qualities sought	Ademuyiwa *et al.* (2011)[82]	Albain *et al.* (2010)[83]	Cuzick *et al.* (2011)[84]	Dowsett *et al.* (2010)[79]	Geffen *et al.* (2009)[77]	Holt *et al* (2011)[78] (abstract only)	Kelly *et al.* (2010)[85]	Lo *et al.* (2010)[76]	Tang *et al.* (2011);[81] Mamounas *et al.* (2010)[80]	Tang *et al.* (2010)[86] (abstract only)	Toi *et al.* (2010)[87]	Yorozuya *et al.* (2009)[88]
Analysis	Continuous predictor variable analysed appropriately	U	Y	Y	Y	U	Y	Y	U	Y	U	U	Y
	Statistical adjustment for all important prognostic factors	Y	U	Y	Y	N	U	U	U	Y	U	Y	Y
Intervention subsequent to inclusion in cohort	Fully described	Y	Y	Y	Y	U	U	N	U	Y	Y	U	N
	Intervention standardised or randomised	N	Y	Y	Y	U	U	N	N	Y	Y	U	N

HR, hormone receptor; N, no; U, unclear/not reported; Y, yes.

Summary of results: OncotypeDX test (new data)

Study	Outcomes/end points	Results	Author conclusions	Comments
Ademuyiwa *et al.* (2011)[82]	1. Impact on clinical decision-making in terms of recommending chemotherapy (CT)	**1a. OncotypeDX (ODX)-blinded recommendation vs. ODX risk group and actual treatment received** Low (0–17): $n=142$; recommended CT: 52 (37%); actually received CT: 13 (9%) Intermediate (18–30): $n=110$; recommended CT: 52 (47%); actually received CT: 52 (47%) High (>30): $n=24$; recommended CT: 21 (87%); actually received CT: 23 (96%) **1b. ODX blinded recommendation vs. ODX score-based actual treatment** ODX-blinded 'no' and ODX-based 'no': 117/276 (42.3%) ODX-blinded 'yes' and ODX-based 'no': 71/276 (25.7%) ODX-blinded 'no' and ODX-based 'yes': 34/276 (12.3%) ODX-blinded 'yes' and ODX-based 'yes': 54/276 (19.7%) 37 fewer patients (71 − 34) received CTX using ODX score to help decide CTX use 38% of patients (25.7% + 12.3%) had a change in management as a result of ODX score **1c. ODX-blinded recommendation vs. NPI category** Low (0–17): $n=142$; excellent/good NPI: 123; moderate NPI: 19 Intermediate (18–30): $n=110$; excellent/good NPI: 86; moderate NPI: 24 High (>30): $n=24$; excellent/good NPI: 11; moderate NPI: 13 $p<0.001$	The ODX score had a significant impact on the receipt of adjuvant CT and altered management for 38% of women	

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Albain et al. (2010)[83]</td><td>1. The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes</td><td>
1a. RS for DFS
In tamoxifen (TAM)-alone group stratified by number of positive nodes, log-rank, $p=0.017$
DFS estimate at 10 years: low RS: 60%; intermediate RS: 49%; high RS: 43%
Cox regression model, continuous RS highly significant, $p=0.006$, HR = 2.64 (95% CI 1.33 to 5.27) for 50-point difference
Proportional hazards showed test not consistent over time ($p=0.0016$)
HR for those surviving beyond 5 years = 0.86 (95% CI 0.27 to 2.74, $p=0.8$)
DFS HRs adjusted for number of positive nodes, for chemotherapy benefit, by RS over time
All years interaction p-value = 0.053
5 years interaction p-value = 0.029
10 years interaction p-value = 0.58 (i.e. RS not good predictor for chemotherapy benefit over 5 years)
Treatment effect overall: DFS HRs (95% CIs) adjusted for number of positive nodes, for chemotherapy benefit, by RS over time:
<table>
<tr><th></th><th>All years HR</th><th>5 years HR</th><th>After 5 years HR</th></tr>
<tr><td>Entire RS sample</td><td>0.72 (0.51 to 1.00)</td><td>0.79 (0.51 to 1.23)</td><td>0.63 (0.39 to 1.04)</td></tr>
<tr><td>At selected RS values</td><td></td><td></td><td></td></tr>
<tr><td>10</td><td>0.95 (0.59 to 1.52)</td><td>1.24 (0.62 to 2.48)</td><td>0.72 (0.38 to 1.36)</td></tr>
<tr><td>18</td><td>0.83 (0.56 to 1.22)</td><td>1.03 (0.58 to 1.81)</td><td>0.67 (0.40 to 1.14)</td></tr>
<tr><td>25</td><td>0.74 (0.53 to 1.04)</td><td>0.87 (0.53 to 1.42)</td><td>0.64 (0.39 to 1.05)</td></tr>
<tr><td>31</td><td>0.67 (0.48 to 0.93)</td><td>0.75 (0.48 to 1.18)</td><td>0.61 (0.35 to 1.04)</td></tr>
<tr><td>40</td><td>0.57 (0.39 to 0.83)</td><td>0.61 (0.38 to 0.96)</td><td>0.56 (0.28 to 1.11)</td></tr>
</table>
There are data looking at the validity of RS in the cyclophosphamide, doxorubicin and fluorouracil followed by tamoxifen (CAF-T) group alongside the data for the TAM group, but these data seem to show that chemotherapy has a benefit over TAM alone; do not give HR for CAF-T group alone
RS was a strong predictor of benefit from CAF-T for DFS; only those in high-risk groups gain benefit (CAF-T vs. TAM DFS at 10 years, stratified by number of nodes, log-rank test):
Low RS: not significantly different ($p=0.97$): 64% survival in CAF-T group, 60% in TAM group
Intermediate RS: not significantly different ($p=0.48$)
High RS: significantly different ($p=0.033$): 55% survival in CAF-T group, 43% in TAM group
</td><td></td><td></td></tr>
</table>

Study	Outcomes/end points	Results	Author conclusions	Comments
		1b. RS for OS		
		In TAM-alone group, stratified by number of positive nodes, log-rank, $p=0.003$		
		DFS estimate at 10 years: low RS: 77%; intermediate RS: 68%; high RS: 51%		
		HR after adjustment for number of positive nodes = 4.42 (95% CI 1.96 to 9.97, $p=0.0006$) for 50-point difference		
		Proportional hazards showed not consistent over time ($p=0.0005$)		
		RS was a strong predictor of benefit from CAF-T for OS; only those in high-risk groups gain benefit (CAF-T vs. TAM OS at 10 years, stratified by number of nodes, log-rank test):		
		Low RS: not significantly different ($p=0.63$)		
		Intermediate RS: not significantly different ($p=0.85$)		
		High RS: significantly different ($p=0.027$)		
		1c. RS for BCSS		
		RS was a predictor of benefit from CAF-T for BCSS; only those in high-risk groups gain benefit (CAF-T vs. TAM BCSS at 10 years, stratified by number of nodes, log-rank test):		
		Low RS: not significantly different ($p=0.56$)		
		Intermediate RS: not significantly different ($p=0.89$)		
		High RS: significantly different ($p=0.033$)		

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Cuzick et al.[84] (2011)</td><td>Distant recurrence (within 10 years)
TTDR</td><td>G1 cohort: 195 recurrences of which 145 distant recurrences; in LN– women 101 recurrences of which 67 distant recurrences
The mean change in likelihood ratio chi-squared (95% CI) for addition of GHI-RS to the classical score in the validation halves of 100 random splits of the data (higher values indicate more added prognostic information):
<table>
<tr><th colspan="2">TTDR (months)</th><th colspan="2">Time to recurrence (all recurrences)</th></tr>
<tr><th>All patients</th><th>LN–</th><th>All patients</th><th>LN–</th></tr>
<tr><td>25.3 (25.2–25.9)</td><td>20.9 (20.7–21.6)</td><td>25.6 (25.2–25.9)</td><td>25.7 (25.4–26.4)</td></tr>
</table>
9-year distant recurrence probabilities for 25th and 75th percentiles of GHI-RS scores for different grades and nodal status for women aged >65 years with a 1–2 cm tumour treated with anastrozole:
<table>
<tr><th></th><th></th><th colspan="3">Grade (%)</th></tr>
<tr><th>Nodal status</th><th>Percentile</th><th>Poor or undifferentiated</th><th>Moderate</th><th>Well differentiated</th></tr>
<tr><td>Negative</td><td>25</td><td>8.3</td><td>5.8</td><td>2.5</td></tr>
<tr><td></td><td>75</td><td>12.1</td><td>8.4</td><td>3.6</td></tr>
<tr><td>Positive</td><td>25</td><td>12.1</td><td>8.4</td><td>3.6</td></tr>
<tr><td></td><td>75</td><td>17.3</td><td>12.2</td><td>5.3</td></tr>
</table></td><td>NR for GHI-RS alone</td><td></td></tr>
</table>

GHI-RS, Genomic Health Recurrence Score.

Study	Outcomes/end points	Results	Author conclusions	Comments
Dowsett *et al* 2010[79]	1. Degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	**1a. RS and risk of distant recurrence (DR)** Risk score for a 50-point change (e.g. RS = 55 vs. RS = 5) was significantly associated with risk of DR (HR 3.92; 95% CI 2.08 to 7.39; $\Delta\chi^2 = 15.5$; $p < 0.001$) when adjusted for the effects of tumour size, local grade, age and treatment When local grade replaced with central grade in multivariate analysis, adjusted RS also significantly associated with risk of DR (HR 5.25; 95% CI 2.84 to 9.73; $\Delta\chi^2 = 22.7$; $p < 0.001$) **1b. RS and TTDR** In N0 patients: HR = 5.25 (95% CI 2.84 to 9.73); $\Delta\chi^2 = 22.7$; $p < 0.001$ In N+ patients: HR = 3.47 (95% CI 1.64 to 7.38); $\Delta\chi^2 = 9.4$; $p < 0.002$ **1c. Differences in absolute DR rates for N0 and N+ patients** *DR at 9 years N0 patients* RS < 18: 4% (95% CI 3% to 7%) RS 18–30: 12% (95% CI 8% to 18%) RS ≥ 31: 25% (95% CI 17% to 34%) HR adjusted for clinical variables (tumour size, grade, age, treatment and number of positive nodes): between high and low RS groups = 5.2 (95% CI 2.7 to 10.1); between intermediate and low RS groups = 2.5 (95% CI 1.3 to 4.5) *DR at 9 years N+ patients* RS < 18: 17% (95% CI 12% to 24%) RS 18–30: 28% (95% CI 20% to 39%) RS ≥ 31: 49% (95% CI 35% to 64%) HR adjusted for clinical variables (tumour size, grade, age, treatment and number of positive nodes): between high and low RS groups = 2.7 (95% CI 1.5 to 5.1); between intermediate and low RS groups = 1.8 (95% CI 1.0 to 3.2)	This study confirmed the performance of RS in postmenopausal hormone receptor-positive patients treated with tamoxifen in a large contemporary population and demonstrated that RS is an independent predictor of DR in N0 and LN+ hormone receptor-positive patients treated with anastrozole, adding value to estimates with standard clinicopathological features	

Study	Outcomes/end points	Results	Author conclusions	Comments
		1d. OS at 9 years for N0 and N+ patients		
		N0 patients		
		RS < 18: 88% (95% CI NR)		
		RS 18–30: 84% (95% CI NR)		
		RS ≥ 31: 73% (95% CI NR)		
		HR adjusted for clinical variables (tumour size, grade, age, treatment and number of positive nodes): between high and low RS groups = 2.5 (95% CI 1.5 to 4.0); between intermediate and low RS groups = 1.2 (95% CI 0.8 to 1.9)		
		N+ patients		
		RS < 18: 74% (95% CI NR)		
		RS 18–30: 69% (95% CI NR)		
		RS ≥ 31: 54% (95% CI NR)		
		HR adjusted for clinical variables (tumour size, grade, age, treatment and number of positive nodes): between high and low RS groups = 2.1 (95% CI 1.2 to 3.8); between intermediate and low RS groups = 1.4 (95% CI 0.9 to 2.4)		
		Data to show that treatment group (tamoxifen vs. anastrozole) did not interact with RS prediction of DR		
		1e. RS, Adjuvant! Online and DR		
		Correlation between RS-predicted DR and Adjuvant! Online-predicted recurrence was low but statistically significant by central grade (Spearman's rank correlation = 0.23, $p < 0.001$) or local grade (Spearman's rank correlation = 0.22, $p < 0.001$). Only approx. 5% of variability explained by each other, therefore have independent prognostic value		
Geffen *et al* 2009[77]	1. Impact on clinical decision-making	**1. Impact on clinical decision-making**	NR for this outcome	
		25 patients had RS assay; nine patients' (36%) treatment recommendations were changed based on the scores, six from chemotherapy to no chemotherapy		

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Holt et al 2011[78]</td><td>1. Impact on clinical decision-making</td><td>1a. Change in initial recommendations pre RS assay to post RS assay
All patients have hormone therapy as standard
No change no chemotherapy (CT): 49 (46.23%)
Change CT to no CT: 25 (23.6%)
Change no CT to CT: 10 (9.43%)
No change CT: 22 (20.75%)
1b. Change in patient choices pre RS assay to post RS assay by NPI score
<table>
<tr><th>NPI</th><th>No CT (unchanged) (n)</th><th>CT to no CT (changed) (n)</th><th>No CT to CT (changed) (n)</th><th>CT (unchanged) (n)</th></tr>
<tr><td><2.4</td><td>9</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>2.4–3.4</td><td>31</td><td>8</td><td>4</td><td>5</td></tr>
<tr><td>3.4–4.4</td><td>8</td><td>15</td><td>5</td><td>10</td></tr>
<tr><td>4.4–5.4</td><td>1</td><td>2</td><td>0</td><td>6</td></tr>
<tr><td>>5.4</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>Total</td><td>49</td><td>25</td><td>10</td><td>22</td></tr>
</table>
1c. Spearman's rank correlation comparing RS with individual components of NPI
Of size, LN status and grade, only grade was significantly correlated</td><td>Early results of study suggest that OncotypeDX is applicable and feasible to perform in the UK setting with a reduction in the use of adjuvant CT consistent with the findings of other reported studies. RS added prognostic information beyond that from NPI alone</td><td></td></tr>
<tr><td>Kelly et al 2010[85]</td><td>1. Correlation with Adjuvant! Online
2. Risk prediction</td><td>1. Correlation between predicted risk of recurrence and death after 5 years of tamoxifen therapy vs. RS = 0.13 and 0.18 respectively
2. Assumes cohort of patients sent for OncotypeDX testing are clinically intermediate patients. Of these, OncotypeDX was able to dichotomise 52% (n = 160) to low-risk group and 9% (n = 27) to high-risk group; 39% (n = 122) were judged at intermediate risk</td><td>Authors concluded that OncotypeDX yielded potentially informative risk assignments in patients who may be considered at indeterminate risk by routine clinical variables. However, 40% of the time they remain intermediate risk using RS thresholds; this increases to 66% when using revised TAILORx thresholds</td><td></td></tr>
</table>

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Lo et al 2010[76]</td><td>Impact of the 21-gene RS assay on clinical decision-making and patient preferences. End points include (1) changes in physician treatment recommendations, (2) physician self-assessed changes in long-term adjuvant treatment, (3) patient anxiety, (4) quality of life, (5) relapse data</td><td>1a. Whole cohort – changes in physician treatment recommendations
From hormone therapy (HT) to chemotherapy and hormone therapy (CHT): 3/89 (3.4%)
From CHT to HT: 20/89 (22.5%)
From HT to equipoise:[a] 3 (3.4%)
From CHT to equipoise:[a] 2 (2.2%)
No change HT: 40 (44.9%)
No change CHT: 20 (22.5%)
No change equipoise:[a] 1 (1.1%)
1b. By RS category – changes in physician treatment recommendations
<table><tr><th>Physician pre- to post-RS assay treatment recommendation</th><th>Low RS</th><th></th><th>Intermediate RS</th><th></th><th>High RS</th><th></th><th>Total</th><th></th></tr><tr><th></th><th>n</th><th>%</th><th>n</th><th>%</th><th>n</th><th>%</th><th>n</th><th>%</th></tr><tr><td>HT to HT</td><td>21</td><td>52.5</td><td>19</td><td>47.5</td><td>0</td><td>0</td><td>40</td><td>100</td></tr><tr><td>HT to CHT</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>100</td><td>3</td><td>100</td></tr><tr><td>CHT to HT</td><td>12</td><td>60</td><td>8</td><td>40</td><td>0</td><td>0</td><td>20</td><td>100</td></tr><tr><td>CHT to CHT</td><td>3</td><td>15</td><td>11</td><td>55</td><td>6</td><td>30</td><td>20</td><td>100</td></tr><tr><td>HT to equipoise</td><td>1</td><td>33.3</td><td>2</td><td>66.7</td><td>0</td><td>0</td><td>3</td><td>100</td></tr><tr><td>HT to equipoise</td><td>1</td><td>50</td><td>1</td><td>50</td><td>0</td><td>0</td><td>2</td><td>100</td></tr><tr><td>Equipoise to equipoise</td><td>0</td><td>0</td><td>1</td><td>100</td><td>0</td><td>0</td><td>1</td><td>100</td></tr><tr><td>Total</td><td>38</td><td>42.7</td><td>42</td><td>47.2</td><td>9</td><td>10.1</td><td>89</td><td>100</td></tr></table>Difference between mean RS for recommendation of CHT vs. HT alone: 29 vs. 16 ($p=0.0001$)
Difference between mean RS for recommendation of CHT vs. equipoise: 29 vs. 19 ($p=0.001$)
Difference between mean RS for HT alone vs. equipoise: 16 vs. 19 ($p=0.288$)
1c. Correlation between treatment and RS category
High-risk RS: 9/9 (100%) CHT
Intermediate RS: 11 (26.2%) CHT
Low-risk RS: 3 (7.9%) CHT</td><td>The RS assay impacts significantly on physician and patient adjuvant treatment decision-making. Most of the treatment changes were from a pretreatment recommendation of CHT to HT alone for both physicians and patients. In addition, RS results have an enduring impact on physician confidence in their treatment recommendations, patient satisfaction and patient anxiety</td><td></td></tr>
</table>

Study	Outcomes/end points	Results	Author conclusions	Comments
		2. Physician self-assessed changes in long-term adjuvant treatment 16 (94%) physicians completed a follow-up questionnaire; 15/16 (94%) of these stated that the assay provided additional information for adjuvant decision-making; 14/16 believed that it had influenced their recommendations; 16/16 (100%) would use it again **3. DCS and anxiety** Mean DCS pre RS: 1.99 (SD 0.62); mean DCS post RS: 1.69 (SD 0.5) ($p<0.001$) STAI pre RS, post RS and at 12-month follow-up: state: 39.6 (SD 14.5), 36 (SD 12.6), 34 (SD 11.5) ($p=0.007$); trait: 32.2 (SD 14.5), 31.7 (SD 13.3), 33.2 (SD 11.0) ($p=0.27$) **4. Quality of life** FACT-B pre RS: mean 112.2 (SD 17.4), FACT-B 12 months post RS: mean 114.3 (SD 18.6) ($p=0.55$) FACT-G pre RS: mean 88.7 (SD 12.3), FACT-G 12 months post RS: mean 87.6 (SD 14.9) ($p=0.49$) **5. Relapse data** Of the 67 patients who completed the 12-month questionnaire, none had experienced a relapse. The status of the remaining 22 is unknown		

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Mamounas et al 2010[80]</td><td>1. The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes</td><td>1a. Association between RS and locoregional recurrence by treatment group
All groups showed significant associations
Kaplan–Meier estimates and 95% CIs of the proportion of patients with locoregional recurrence at 10 years for 355 placebo-treated patients (NSABP B14), 895 tamoxifen-treated patients (NSABP B14 and B20) and 424 tamoxifen plus chemotherapy-treated patients (NSABP B20)
<table>
<tr><th>Treatment group and RS group</th><th>10-year Kaplan–Meier estimate (%)</th><th>95% CI</th><th>Log-rank p-value</th><th>No. of events/no. at risk</th></tr>
<tr><td>Placebo</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Low (<18)</td><td>10.8</td><td>5.8 to 15.8</td><td>0.022</td><td>19/171</td></tr>
<tr><td>Intermediate (18–30)</td><td>20.0</td><td>9.9 to 30.0</td><td></td><td>15/85</td></tr>
<tr><td>High (≥31)</td><td>18.4</td><td>9.5 to 27.4</td><td></td><td>19/99</td></tr>
<tr><td>Tamoxifen</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Low (<18)</td><td>4.3</td><td>2.3 to 6.3</td><td>0.001</td><td>24/473</td></tr>
<tr><td>Intermediate (18–30)</td><td>7.2</td><td>3.4 to 11.0</td><td></td><td>6/194</td></tr>
<tr><td>High (≥31)</td><td>15.8</td><td>10.4 to 21.2</td><td></td><td>33/228</td></tr>
<tr><td>Chemotherapy + tamoxifen</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Low (<18)</td><td>1.6</td><td>0.0 to 3.5</td><td>0.028</td><td>4/218</td></tr>
<tr><td>Intermediate (18–30)</td><td>2.7</td><td>0.0 to 6.4</td><td></td><td>2/89</td></tr>
<tr><td>High (≥31)</td><td>7.8</td><td>2.6 to 13.0</td><td></td><td>8/117</td></tr>
</table>
Note: Results are given for all patients and for the prespecified RS risk categories.</td><td>Similar to the association between RS and risk for distant recurrence, a significant association exists between RS and risk for locoregional recurrence. This information has biologic consequences and potential clinical implications relative to locoregional therapy decisions for patients with LN– and ER+ breast cancer</td><td>Locoregional relapse, not distant relapse</td></tr>
</table>

Study	Outcomes/end points	Results	Author conclusions	Comments

1b. Multivariate Cox regression analysis of predictors of locoregional recurrence

Cohort of 895 tamoxifen-treated patients from NSABP trials B14 and B20

Variable	Hazard	95% CI	Wald test *p*-value
Age (≥50 vs. <50)	0.40	0.25 to 0.65	0.0002
Mastectomy vs. L + XRT	0.62	0.39 to 0.99	0.047
Clinical tumour size (>2 vs. ≤2 cm)	0.98	0.61 to 1.59	0.933
Tumour grade (moderate vs. well)	1.10	0.54 to 1.92	0.113
Tumour grade (poor vs. well)	1.76	0.89 to 3.48	
Recurrence score[a]	2.16	1.26 to 3.68	0.005

L, lumpectomy; LRR, locoregional recurrence; XRT, radiation therapy.

a RS was a continuous variable, with the HR for LRR calculated relative to an increment of 50 units (chosen to dichotomise the RS and thus improve comparability of the HR with the HRs based on the clinical covariates). The *p*-value for the likelihood ratio test on RS is 0.007.

Study	Outcomes/end points	Results	Author conclusions	Comments
Tang *et al* 2011[81]	1. The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	**1a. Comparison between point estimates for RS risk group and recurrence interval (RI) (Adjuvant! Online) risk group for DRFI in NSABP B14 tamoxifen-treated patients ($n=668$)** RS low overall: $n=338$; RS low, RI low: $n=216$; RS low, RI intermediate: $n=57$; RS low, RI high: $n=65$ RS intermediate overall: $n=149$; RS intermediate, RI low: $n=84$; RS intermediate, RI intermediate: $n=24$; RS intermediate, RI high: $n=41$ RS high overall: $n=181$; RS high, RI low: $n=52$; RS high, RI intermediate: $n=43$; RS high, RI high: $n=86$ Concordance between RS and RI was 0.49, correlation was modest (Spearman's correlation coefficient of 0.38) RI low overall: $n=332$; RI low, RS low ($n=216$) point estimate distant recurrence (DR) 10 years: 5.6%; RI low, RS intermediate ($n=84$) point estimate DR 10 years: 10%; RI low, RS high ($n=52$) point estimate DR 10 years: 18.2% RI intermediate overall: $n=146$; RI intermediate, RS low ($n=57$) point estimate DR 10 years: 13.4%; RI intermediate, RS intermediate ($n=24$) point estimate DR 10 years: 13.9%; RI intermediate, RS high ($n=43$) point estimate DR 10 years: 43.2% RI high overall: $n=190$; RI high, RS low ($n=65$) point estimate DR 10 years: 5%; RI high, RS intermediate ($n=41$) point estimate DR 10 years: 23.4%; RI high, RS high ($n=86$) point estimate DR 10 years: 31.5% **1b. Cox models of HRs in B14 tamoxifen-treated patients ($n=668$)** RI percentile as sole predictor, using 50-point increment in score, HR = 2.87 (95% CI 1.95 to 4.23) RS percentile as sole predictor, using 50-point increment in score, HR = 3.61 (95% CI 2.49 to 5.24)		

Study	Outcomes/end points	Results	Author conclusions	Comments

1c. Multivariate Cox models assessing relative associations of RI and RS using 50-point increment in score in B14 tamoxifen-treated patients ($n=668$)

Model 1 – not relevant

Model 2 – not relevant

Model 3 – RS percentile using 50-point increment in score, HR = 3.51 (95% CI 2.49 to 5.24), $p<0.001$

Model 4 – RI and RS percentiles using 50-point increment in score, HR for RS = 2.83 (95% CI 1.91 to 4.18), $p<0.001$

Model 5 – RI and RS percentiles using 50-point increment in score, age, tumour size, grade (moderate vs. well), grade (poor vs. well), HR for RS = 2.37 (95% CI 1.58 to 3.55)

Model 6 – as model 5 but without RI percentile using 50-point increment in score, HR for RS = 2.34 (95% CI 1.56 to 3.5)

Model 7 – not relevant

1d. Multivariate Cox models assessing relative associations of RI and RS using risk groups in B14 tamoxifen-treated patients ($n=668$)

Model	Variables	HR (95% CI)	*p*-value
1	RI intermediate vs. low	2.51 (1.55 to 4.21)	0.001
	RI high vs. low	2.01 (1.25 to 3.23)	
	RS intermediate vs. low	2.21 (1.28 to 3.81)	<0.001
	RS high vs. low	3.8 (2.36 to 6.1)	
2	Age (>50 vs. ≤50 years)	0.76 (0.52 to 1.13)	0.173
	Tumour size	1.2 (1.07 to 1.36)	0.003
	Grade moderate vs. well	1.51 (0.75 to 3.05)	0.003
	Grade poor vs. well	3.18 (1.42 to 7.15)	
	RI intermediate vs. low	1.51 (0.82 to 2.78)	0.176
	RI high vs. low	0.95 (0.52 to 1.76)	
	RS intermediate vs. low	2.07 (1.18 to 3.61)	<0.001
	RS high vs. low	2.88 (1.74 to 4.76)	

Study	Outcomes/end points	Results	Author conclusions	Comments
		1e. Multivariate Cox models assessing relative associations of RS using 50-point increment in score (RS/50) in B14 tamoxifen-treated patients with breast cancer-specific mortality as the end point		
		RS/50 alone, HR = 3.32 (95% CI 2.29 to 4.81), $p<0.001$		
		RS/50 (RI/50 in model), HR = 2.45 (95% CI 1.66 to 3.61), $p<0.001$		
		RS/50 (age, tumour size, grade and RI/50 in model), HR = 2.02 (95% CI 1.35 to 3.0), $p<0.001$		
		RS/50 (age, tumour size, grade in model), HR = 2.01 (95% CI 1.35 to 2.98), $p<0.001$		
		1f. Multivariate Cox models assessing relative associations of RS/50 in B14 tamoxifen-treated patients with OS as the end point		
		RS/50 alone, HR = 1.95 (95% CI 1.51 to 2.52), $p<0.001$		
		RS/50 (RI/50 in model), HR = 1.77 (95% CI 1.35 to 2.33), $p<0.001$		
		RS/50 (age, tumour size, grade and RI/50 in model), HR = 1.67 (95% CI 1.26 to 2.22), $p<0.001$		
		RS/50 (age, tumour size, grade in model), HR = 1.65 (95% CI 1.24 to 2.19), $p<0.001$		
		1g. Multivariate Cox models assessing relative associations of RS/50 in B14 tamoxifen-treated patients with DFS as the end point		
		RS/50 alone, HR = 1.77 (95% CI 1.44 to 2.18), $p<0.001$		
		RS/50 (RI/50 in model), HR = 1.75 (95% CI 1.4 to 2.18); $p<0.001$		
		RS/50 (age, tumour size, grade and RI/50 in model), HR = 1.69 (95% CI 1.34 to 2.14); $p<0.001$		
		RS/50 (age, tumour size, grade in model), HR = 1.67 (95% CI 1.32 to 2.11); $p<0.001$		

Study	Outcomes/end points	Results	Author conclusions	Comments

1h. Cox models assessing relative associations of RI and RS using risk groups in B20 chemotherapy patients ($n=651$) and outcomes of DRFI, OS and DFS

End point and cohort		B20 patients with RS assessment ($n=651$)		All B20 patients with tumour grade ($n=1952$)	
		HR for benefit from MF/CMF (95% CI)	Pa (interaction)	HR for benefit from MF/CMF (95% CI)	Pa (interaction)
DRFI	Overall	0.56 (0.34 to 0.91)		0.62 (0.47 to 0.81)	
	RS low	1.31 (0.46 to 3.78)	0.031	NA	
	RS intermediate	0.61 (0.24 to 1.59)			
	RS high	0.26 (0.13 to 0.53)			
	Adjuvant! Online low	0.58 (0.23 to 1.42)	0.99	0.92 (0.53 to 1.62)	0.219
	Adjuvant! Online intermediate	0.54 (0.2 to 1.46)		0.52 (0.29 to 0.93)	
	Adjuvant! Online high	0.53 (0.25 to 1.1)		0.53 (0.36 to 0.77)	
OS	Overall	0.76 (0.49 to 1.17)		0.74 (0.58 to 0.95)	
	RS low	1.37 (0.63 to 3.01)	0.011	NA	
	RS intermediate	0.94 (0.4 to 2.25)			
	RS high	0.31 (0.16 to 0.6)			
	Adjuvant! Online low	1.16 (0.55 to 2.45)	0.311	1.26 (0.81 to 1.95)	0.009
	Adjuvant! Online intermediate	0.7 (0.3 to 1.61)		0.53 (0.31 to 0.9)	
	Adjuvant! Online high	0.53 (0.26 to 1.07)		0.57 (0.4 to 0.82)	

Study	Outcomes/end points	Results						Author conclusions	Comments
				B20 patients with RS assessment ($n=651$)		All B20 patients with tumour grade ($n=1952$)			
		End point and cohort		HR for benefit from MF/CMF (95% CI)	p-value[a] (interaction)	HR for benefit from MF/CMF (95% CI)	p-value[a] (interaction)		
		DFS	Overall	0.73 (0.54 to 0.99)		0.75 (0.63 to 0.89)			
			RS low	0.91 (0.57 to 1.45)	0.082	NA			
			RS intermediate	0.79 (0.43 to 1.47)					
			RS high	0.41 (0.23 to 0.71)					
			Adjuvant! Online low	0.97 (0.59 to 1.61)	0.357	0.91 (0.69 to 1.21)	0.099		
			Adjuvant! Online intermediate	0.6 (0.33 to 1.09)		0.75 (0.5 to 1.13)			
			Adjuvant! Online high	0.62 (0.36 to 1.05)		0.59 (0.45 to 0.78)			
		NA, not applicable. a From likelihood ratio test.							

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td></td><td></td><td>All RS tests are significant; Adjuvant! Online tests are significant when larger cohort is used
1i. Cox models assessing relative associations of RI and RS using risk groups in B20 chemotherapy patients (n=651) and outcome of breast cancer-specific mortality
<table>
<tr><th>Cohort</th><th>HR for benefit from MF/CMF (95% CI)</th><th>p-value[a]</th><th>p-value[b] (interaction)</th></tr>
<tr><td>Overall</td><td>0.62 (0.36 to 1.06)</td><td>0.081</td><td></td></tr>
<tr><td>RS low</td><td>1.86 (0.38 to 9.19)</td><td>0.449</td><td>0.025</td></tr>
<tr><td>RS intermediate</td><td>0.94 (0.32 to 2.82)</td><td>0.918</td><td></td></tr>
<tr><td>RS high</td><td>0.27 (0.13 to 0.55)</td><td><0.001</td><td></td></tr>
<tr><td>Adjuvant! Online low</td><td>1.03 (0.35 to 3.01)</td><td>0.96</td><td>0.463</td></tr>
<tr><td>Adjuvant! Online intermediate</td><td>0.62 (0.23 to 1.71)</td><td>0.358</td><td></td></tr>
<tr><td>Adjuvant! Online high</td><td>0.44 (0.19 to 1.02)</td><td>0.054</td><td></td></tr>
</table>
a From Wald tests.
b From likelihood ratio tests.</td><td></td><td></td></tr>
<tr><td>Tang 2010[86] (abstract only)</td><td>Distant recurrence
Value of RSPC in the prediction of chemotherapy benefit in reducing risk of recurrence</td><td>60/625 distant recurrences occurred
RS showed a significant interaction with chemotherapy treatment ($p=0.037$) with a standardised HR of 0.836. Interaction of RSPC with treatment not significant ($p=0.10$) although trend was in the same direction as RS (HR 0.833)</td><td>RS used alone remains the best predictor of chemotherapy benefit in ER+, LN- breast cancer</td><td></td></tr>
</table>

Study	Outcomes/end points	Results	Author conclusions	Comments
Toi *et al* 2010[87]	1. The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	**1a. Kaplan–Meier plot of DRFI by RS**		

RS category	No. in category at year 0	No. of distant recurrences 0–5 years	No. in category at year 5	No. of distant recurrences 5–10 years	No. in category at year 10
Low	95	2	90	1	70
Intermediate	40	0	40	0	31
High	65	9	52	6	36

Low-risk category patients had a significantly lower risk of distant recurrence than patients in the high-risk category ($p<0.001$, log-rank test)

No recurrences in the intermediate RS group

1b. Univariate Cox proportional hazards model of DRFI – continuous risk score

50-point increase in RS, HR = 6.20 (95% CI 2.27 to 17.0)

1c. Multivariate cox model adjusting for age (< 50 vs. ≥ 50 years) and clinical tumour size (≤ 2 cm vs. > 2 cm)

50-point increase in RS, HR = 6.03 (95% CI 2.17 to 16.7)

1d. Kaplan–Meier estimates of other event rates by RS group

End point	Event	Low (RS < 18) ($n=95$), % (95% CI)	Intermediate (RS 18–30) ($n=40$), % (95% CI)	High (RS ≥ 31) ($n=65$), % (95% CI)
DRFI	Distant recurrence	3.3 (1.1 to 10.0)	0 (NA)	24.8 (15.7 to 37.8)
RFI	Recurrence	5.5 (2.3 to 12.8)	2.5 (0.4 to 16.5)	24.6 (15.6 to 37.6)
RFS	Recurrence or death	9.6 (5.1 to 17.6)	5.1 (1.3 to 18.8)	23.4 (14.8 to 35.9)
OS	Death	6.4 (2.9 to 13.6)	2.6 (0.4 to 16.8)	19.1 (11.3 to 31.3)

NA, not available.

1e. Cox proportional hazards models, adjusting for age (< 50 vs. ≥ 50 years) and clinical tumour size (≤ 2 cm vs. > 2 cm)

Risk of recurrence: HR = 3.38 (95% CI 1.32 to 8.69)

Risk of recurrence or death: HR = 2.09 (95% CI 0.84 to 5.20)

Risk of death: HR = 2.67 (95% CI 0.93 to 7.62)

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Yorozuya et al. 2009[88]</td><td>1. Degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes</td><td>1a. Difference in mean RS value between cases and control subjects
Cases: mean RS = 40.0 (95% CI 21.1 to 58.9); control subjects: mean RS = 17.8 (95% CI 13.8 to 21.9) ($p<0.001$)
1b. Proportion of patients in risk category groups
Cases: low risk (0 to <18): 3 (30%), intermediate risk (18–30): 1 (10%), high risk (RS ≥ 31): 6 (60%); control subjects: low risk (0 to <18): 19 (63%), intermediate risk (18–30): 8 (27%), high risk (RS ≥ 31): 3 (10%) ($p = 0.005$)
1c. Multivariate logistic regression analysis of age, ER score, PR score, RS, histological grade, Ly vs. distant metastases
<table>
<tr><th>Variable</th><th>p-value</th><th>Odds ratio (95% CI)</th></tr>
<tr><td>Age at diagnosis</td><td>0.195</td><td>0.90 (0.764 to 1.057)</td></tr>
<tr><td>ER score</td><td>0.651</td><td>1.33 (0.389 to 4.53)</td></tr>
<tr><td>PR score</td><td>0.378</td><td>0.65 (0.246 to 1.702)</td></tr>
<tr><td>RS ≥ 50 vs. RS < 50</td><td>0.579</td><td>2.85 (0.07 to 115.552)</td></tr>
<tr><td>Histological grade II vs. histological grade I</td><td>0.369</td><td>7.48 (0.093 to 602.504)</td></tr>
<tr><td>Histological grade II vs. histological grade I</td><td>0.041</td><td>222.0 (1.243 to 39,647.336)</td></tr>
<tr><td>Ly(+) vs. Ly(–)</td><td>0.557</td><td>0.37 (0.013 to 10.312)</td></tr>
</table>
Ly, lymphatic invasion.

RS is not significant but study concludes that odds ratio indicates that it has value</td><td>Both histological grade and risk category classification were effective in identifying women at risk of developing distant metastases after initial therapy for ER+, LN– stage I or IIA breast cancer. These patients may benefit from the addition of adjuvant therapy at diagnosis</td><td></td></tr>
</table>

NR, not reported; SD, standard deviation.

a Equipoise defined as equal options of CHT, HT or enrolment onto the TAILORx clinical trial (random assignment to HT alone or CHT then HT).