Supplementary File 3: Narrative synthesis and additional tables for Chapter 2, Chemotherapy benefit: Oncotype DX and RSPC

Study designs: Oncotype DX chemotherapy benefit

Two data sets[1-4] were re-analyses of RCTs, which provide evidence relating to the extent of any interaction between the effect of chemotherapy and Oncotype DX on outcome (i.e. whether the result of the test is able to predict a differential treatment effect).

Albain *et al*. 2010[1] conducted a re-analysis of the Southwest Oncology Group (SWOG)-8814 study, a Phase 3, open-label, parallel-group RCT. Two arms of the trial were reanalysed: the tamoxifen only arm and the tamoxifen plus cyclophosphamide, doxorubicin and fluorouracil (CAF-T) arm.

Paik *et al*. 2006,[2] Tang *et al*. 2011a[4] and Tang *et al*. 2011b[3] re-analysed the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-20 trial in which patients were randomised to tamoxifen alone, or to tamoxifen plus clyclophosphamide, methotrexate and fluorouracil (CMF-T), or to tamoxifen plus methotrexate and fluorouracil (MF-T). It should be noted that some of the patients of the B-20 trial were used to derive the Oncotype DX score.[2] Tang *et al*. 2011b[3] derived the prognostic Oncotype DX RSPC algorithm using the TransATAC and NSABP B-14 data sets, and then tested the ability of the RSPC to predict benefit from chemotherapy in the NSABP B-20 data set.

The remaining three data sets (MD Anderson Center,[5, 6] Clalit Health Services[7, 8] and SEER registry)[9, 10] were retrospective observational studies where patients were treated according to routine practice and their Oncotype DX score.

Patients: Oncotype DX chemotherapy benefit

The RCTs comprised one data set in LN+[1] and one in LN0[2-4] patients; however, neither data set matched the decision problem exactly in other respects.

The SWOG-8814[1] data set comprised all HR+, LN+ patients with 38.1% having four or more positive lymph nodes. All patients were post-menopausal and 12% were HER2+. A total of 367 (40%) out of the 927 patients recruited to the original trial were included in the analysis, with attrition due to missing samples, insufficient tissue and test failures. Analyses in this study were adjusted for LN1-3 and ≥4.

The NSABP B-20[2-4] data set comprised ER+, LN0 patients, with an unreported percentage being HER2+. A total of 651 (28%) out of the 2363 patients recruited to the original trial were included in

the analysis, with attrition due to missing clinical variables, missing samples, insufficient tissue, and test failures.

The three observational studies are described in detail in Supplementary File 4 and Chapter 2, Clinical utility: Oncotype DX.

Quality assessment: Oncotype DX chemotherapy benefit

Table 4 presents the quality assessment of the included studies. The two reanalyses of RCTs[1-4] were at some risk of bias, largely because of patient spectrum bias, where those individuals excluded because of insufficient tissue may be systematically different to the included patients and no attempt was made to account for missing data. Other sources of bias arising from the analysis of the data include not accounting for stratification factors used in the randomisation of patients to treatment, excluding potentially relevant prognostic variables and treatment effect modifiers, and not considering higher order and non-linear terms in the Cox regression. Blinding of test assessors to clinical outcomes was only conducted in Albain *et al*. 2010.[1]

The three observational studies[5-11] are limited by their non-randomised design, whereby patients who received chemotherapy are likely to be systematically different in terms of known (and potentially unknown) prognostic variables (e.g. age) and treatment effect modifiers to those who did not, leading to a high risk of confounding. They also only recruited patients for whom an Oncotype DX test had been ordered and it is unclear how this may have affected the patient spectrum and generalisability to the decision problem. Three studies did, however, due to their prospective use of the test in clinical practice, blind the test assessors to the long-term outcomes.[5, 6, 9-12]

Results: Oncotype DX chemotherapy benefit

Table 7 of the main report presents data from RCTs relating to the ability of Oncotype DX to predict benefit from chemotherapy.

DRFI: This was the primary outcome in Tang *et al* 2011a,[4] but was not reported by Albain *et al*. 2010,[1] where an exploratory analysis of BCSS was presented instead. For DRFI in Tang *et al*. 2011a,[4] HRs for no chemotherapy compared with chemotherapy showed a similar trend as DFS, with the Oncotype DX high-risk category showing a statistically significant effect of chemotherapy (HR 0.26 (95% CI: 0.13, 0.53); unadjusted); an unadjusted test of the interaction between treatment and recurrence score was also statistically significant (p=0.031).[4] Paik *et al*. 2006[2] performed a Cox regression adjusted for age, tumour size, ER, PR, tumour grade, recurrence score as a continuous variable, treatment and the interaction between treatment and recurrence score (interaction p-values 0.035 to 0.068).In a personal communication with the study statistician (Prof Tang, University of

Pitsburgh, via NICE, February 2018), additional data were provided for a HER2- subgroup; HRs were similar to the whole cohort data (Table 7 of the main report). However, the interaction tests in the HER2- subgroup had p values of p=0.007, 0.018, 0.022 (depending on how tumour grade was assessed), hence providing stronger evidence of an interaction between treatment and recurrence score in this subgroup.

Tang *et al.* 2011a[4] also reported the effect of chemotherapy by AOL risk groups (data not tabulated) in patients with RS scores and reported a test for the interaction between treatment and AOL risk group (*p*=0.99), indicating that it was unable to predict the benefit of chemotherapy; HRs were low-risk 0.58 (95% CI: 0.23, 1.42); intermediate-risk 0.54 (95% CI: 0.20, 1.46); high-risk 0.53 (95% CI: 0.25, 1.1)). In an additional analysis of 1952 patients from B-20 with tumour grade, the test for the interaction between treatment and AOL risk group was statistically non-significant (*p*=0.219). However, although the effects of treatment were similar in patients at intermediate- and high-risk by AOL, there was evidence of no effect of treatment in patients at low-risk; HRs low-risk 0.92 (95% CI: 0.53, 1.62); intermediate-risk 0.52 (95% CI: 0.29, 0.93); high-risk 0.53 (95% CI: 0.36, 0.77).

DFS: Albain *et al.* 2010 reported 5 and 10 year DFS and Tang *et al*. 2011a report 10 year DFS.[1-4] 10-year HRs for the effect of chemotherapy compared with no chemotherapy showed a progressively greater effect on DFS when moving from low-risk to high-risk Oncotype DX categories in both studies (Table 7 of the main report) but only the high-risk group in Tang *et al*. 2011a (HR 0.41 (95% CI: 0.23, 0.71); unadjusted)[4] and Albain *et al*. 2010 (HR 0·59 (95% CI: 0·35, 1·01); log-rank p-value=0.033; adjusted for the number of positive nodes) were statistically significant.[1] Formally, the test for the interaction between treatment and RS risk group test was not statistically significant in Tang *et al*. 2011a (p=0.082).[4] Albain *et al.* 2010 assessed the effect of RS on the continuous scale and its interaction with treatment adjusted for the number of positive nodes and found the interaction to be borderline statistically non-significant (p=0.053).[1] However, Albain *et al*. 2010[1] also found that the effect of recurrence score on treatment varied over time and that recurrence score is a treatment effect modifier in the first 5 years (interaction p-value=0.029) but not after 5 years (interaction p-value=0.580). Within the first 5 years, they performed a series of Cox regression analyses adjusting individually for age, ethnic origin, tumour size, progesterone status, grade, P53 and HER2, treatment, continuous recurrence score and the interaction between continuous recurrence score and treatment, and found that the interaction remained statistically significant (p-value not presented). However, afteradjustment for ER status only (by Allred-scoring), the interaction was not statistically significant (p=0.15).

Breast Cancer Specific Survival: BCSS also showed a statistically significant effect in the high-risk group in Albain *et al.* 2010 (p=0.033; adjusted for the number of positive nodes), although no interaction test was reported and data was not reported for intermediate and low risk patients.[1]

Overall survival: HRs were reported for both data sets for chemotherapy compared with no chemotherapy in low-, intermediate- and high-risk groups (Table 7 of the main report). HRs showed the greatest effect of chemotherapy in the high-risk groups; the HR was statistically significant in Tang *et al*. (HR 0.31 (95% CI: 0.16, 0.60); unadjusted) [4] and borderline statistically significant in Albain *et al.* 2010 (HR 0·56 (95% CI: 0·31, 1·02), p=0·057; adjusted for the number of positive nodes).[1] In Tang *et al*,[4] the test for the interaction between treatment and recurrence score (i.e. low-, intermediate- and high-risk) was statistically significant (p=0.011). Albain *et al*. 2010 assessed the effect of RS on the continuous scale and its interaction with treatment adjusted for the number of positive nodes only and found the interaction with treatment statistically significant over 10 years (p=0.026) and within the first 5 years (p=0.016).

Tang 2011a[4] also reported the effect of chemotherapy by AOL risk groups (data not tabulated) in patients with RS scores and reported a test for the interaction between treatment and AOL risk group (p=0.311). In an additional analysis of 1952 patients from B-20 with tumour grade, the test for the interaction between treatment and AOL risk group was significant (p=0.009); HRs low-risk 1.26 (95% CI: 0.81, 1.95); intermediate-risk 0.53 (95% CI: 0.31, 0.9); high-risk 0.57 (95% CI: 0.40< 0.82).

Whilst the results from Tang *et al*. 2011a suggest that Oncotype DX is better at identifying individuals who would benefit from chemotherapy than AOL, the authors did not provide a formal comparison of the performance of the models and the relative benefit of Oncotype DX over AOL remains unclear.

Cut-off below which chemotherapy has no benefit: Albain *et al.* 2010 suggested that within the first 5 years, the effect of chemotherapy on DFS was clinically equivalent to the effect of no chemotherapy for recurrence scores up to about 20 but that chemotherapy performed better at higher scores. Paik *et al*. 2006[2] explored the effect of treatment, Oncotype DX score as a continuous variable and their interaction on distant recurrence but were unable to estimate the cut-off below which there was no benefit from chemotherapy as chemotherapy provided a benefit at all risk scores.

Observational studies

Data relating to the ability of Oncotype DX to predict benefit from chemotherapy from observational studies is presented in Table 8 of the main report. These studies are at high risk from confounding.

DRFS, IDFS, RFS and BCSS: The MD Anderson study reported DRFS[5, 6] (using Cox regression by risk group adjusted for treatment, age at diagnosis, tumour size, grade, histologic subtype, Ki-67 expression, LVI, type of surgery and endocrine therapy at both the 18-30 RS cut-off and the 11-25 RS cut off).

The Clalit Health Services study reported a subgroup of patients with one micro metastasis up to 3 lymph node metastases (LN1micro to LN3);[8] and a subgroup of patients with no lymph node metastases or one micrometastasis.[11] Other analyses were provided by the company as Academic in Confidence data and could not be reported here (Stemmer 2017[7]). In the LN1micro to LN3 group, rates of DR and BC death for chemotherapy-treated and untreated patients were reported as exploratory analyses in patients with Oncotype DX RS scores 18-30 and scores 11-25 only (i.e. no data for low risk or high risk patients). Statistical tests were not conducted, but for both endpoints, those LN1micro to LN3 pateints treated with chemotherapy had more favourable results compared with those not treated with chemotherapy, and this was more evident in the subgroup of patients with Oncotype DX RS scores 18-30 (DRFS 97.8% compared with 90.4%; BCSS 98.9% compared with 96.3%, respectively) than in the group with score 11-25 (DR 97.3% compared with 95.9%; BC death 100% compared with 98.8%). LN0/1micro patients[8, 11] did not appear to receive benefit from chemotherapy in the intermediate group (DRFI for chemotherapy versus no chemotherapy: 94.4% versus 94.7% respectively) whilst those in the high risk group did (DRFI for chemotherapy versus no chemotherapy: 86.7% versus 78.9% respectively).

The MD Anderson study[5, 6] presented Kaplan-Meier survival functions by risk group and 5-year DRFS, IDFS, RFS and OS rates for LN0 patients only. At both RS cut offs, event rates were too few in the low-risk categories to allow an analysis. Kaplan Meier survival functions indicated no difference between chemotherapy and no chemotherapy for any outcome and unadjusted log-rank tests were not statistically significant. The observed event rates were similar or worse in chemotherapy treated patients in the intermediate RS category (11-25). Analyses using the 11-25 RS cut off reported HRs>1 for the effect of chemotherapy in the intermediate-risk group, and HRs<1 for the effect of chemotherapy in the high-risk group, across all outcomes, although p-values were not statistically significant. Analyses using the 18-30 RS cut-off reported HRs <1 in all risk categories (except the RS <18 risk group, where the HR was 1.09 (95% CI: 0.14, 8.62, p=0.938), though HRs were closer to 1 in the intermediate-risk groups than in the high-risk groups. P-values were non-significant and no tests for the interactions between treatment and RS were reported. Results are presented in Table 8 of the main report.

A further analysis, unadjusted for potential prognostic variables and treatment effect modifiers, was conducted which split the Stage 1 disease patients in the intermediate-risk group (RS 18-30) by

tumour size, and found the effect of chemotherapy versus no chemotherapy (HR not reported) was statistically significant in the pT1c (tumour size >10mm, log rank test p=0.02) patients, but not in pT1b (tumour size >5mm, ≤10mm, log-rank test p=0.752) patients. However, the direction of effect was not clear because of conflicting statements within the published report.[6]

The SEER registry study[9, 10] used Cox regression adjusted for treatment, age, tumour size, and recurrence score risk group with and without terms for the interaction between treatment and recurrence score risk group. They found that the association between RS and BCSS remained prognostic, but was attenuated for those with chemotherapy compared to those reported as having no chemotherapy or unknown treatment (interaction p=0.03). They also fitted recurrence score as a continuous variable, although no details were provided of the extent of the interaction with treatment.

One further study (Sparano 2012; ECOG trial E2197)[13] noted that their data were consistent with previous reports indicating greater chemotherapy treatment effect for high RS (RS>20), based on the levelling off of a plot (see source paper),[13] but offered no formal analysis.

Results: RSPC

RSPC was derived in the TransATAC and NSABP B-14 data sets[3] and is based on the Oncotype DX score with the addition of clinicopathological variables (namely RS using a natural cubic spline with 2 degrees-of-freedom with knots at 5, 18 and 50; age; tumour size and grade; nodal status; and hormonal treatment) formally incorporated. Data are available only in LN0 patients. The prognostic ability of RSPC is reported in Chapter 2, Prognostic performance: Oncotype DX. In the same publication,[3] the NSABP B-20 data set (which was used to derived Oncotype DX) was used to assess the score's abilities to predict chemotherapy benefit based on 625 (26%) of 2,362 randomised individuals who had available tumour blocks, Oncotype DX ER expression ≥6.5 and complete information on tumour grade and size, and age. Whilst there was a weak statistically significant interaction between treatment effect and Oncotype DX RS risk score (p=0.037) with a standardised HR of 0.66 (95% CI: 0.44, 0.97), there was insufficient evidence of an interaction between treatment and RSPC risk score (p=0.10) with a standardised HR of 0.65 (95% CI: 0.39, 1.09) (data not tabulated).

Discussion

Key limitations of studies assessing chemotherapy benefit

a) *Lack of data on chemotherapy benefit for the clinically intermediate-risk group:* NICE currently recommends Oncotype DX only for patients who are clinically intermediate-risk, for whom the chemotherapy decision is uncertain. This is a key subgroup for the economic modelling (defined as NPI>3.4). There are no data on the chemotherapy effect in patients who are Oncotype DX low-risk

but clinically intermediate-risk. It is plausible that even if there is no chemotherapy benefit for clinically-low Oncotype DX-low patients, there could be benefit for clinically-intermediate (NPI>3.4) Oncotype DX-low patients.

b) *Statistical significance of interaction tests:* Most unadjusted interaction tests were statistically significant (Table 7 of the main report). In terms of adjusted interaction tests, these were significant or borderline significant in B20 (LN0); and more clearly significant for the new HER2- subgroup (personal communication via NICE with Prof Tang, University of Pitsburgh, February 2018). One of the key concerns in the EAG report was that it was unclear whether all factors were adjusted for simultaneously in B20; however, personal communication via NICE with the biostatistician confirms that this was the case. This, along with the new HER2- subgroup analysis, provides stronger evidence for an interaction than presented in the EAG report.

However, in SWOG-8814 (LN+), it is now apparent after clarification from the lead biostatistician (Professor Barlow, University of Washington School of Public health, personal communication, March 2018) that interaction tests were adjusted for each clinicopathological factor individually (not all together, as initially thought by the EAG). All were individually significant except for the interaction test adjusted for Allred-scored ER status (p=0.15). As such, it remains unclear whether the interaction test would remain significant after adjustment for all relevant clinicopathological variables.

This also raises an interesting point as to whether results should be adjusted for ER status. On the one hand, test results should be adjusted to account for the effect of clinicopathological factors for which data are available in routine practice. On the other hand, it is not clear to what extent quantitative ER results are routinely available in UK practice, or their level of analytic validity.

c) *Possible overestimation of chemotherapy benefit due to B20 being derivation study:* Patients from the no-chemotherapy arm of B20 were used to derive the Oncotype DX score. Therefore, Oncotype DX may be overfitted in this study arm (i.e. recurrence rates may be artificially low in Oncotype low-risk patients and artificially high in Oncotype DX high-risk patients). This could lead to an overestimate of chemotherapy benefit since the chemotherapy arm was not used in derivation, therefore recurrence rates in this arm may show less separation between the low and high risk groups.

B14 (Paik 2004)[14] is a validation study of Oncotype DX (tamoxifen only; no chemotherapy arm). It can be seen that the prognostic effect of Oncotype DX in the no-chemotherapy arm of B20 is greater than that in B14 (Table 1); in other words, low-risk patients have a better 10-year recurrence-free rate

in B20 (96.8%) than B14 (93.2%), while high-risk patients have a worse recurrence-free rate in B20 (60.5%) than B14 (69.5%).

In terms of prediction of chemotherapy benefit, B20 has a worse recurrence-free rate in the chemotherapy arm in low-risk patients (95.6% with chemotherapy vs. 96.8% without). This is counter-intuitive, and gives a corresponding HR greater than 1 (HR=1.31). However, comparing the chemotherapy arm of B20 (95.6% recurrence-free) with the no-chemotherapy arm of B14 (93.2% recurrence-free) indicates a small benefit in low-risk patients, though this breaks randomisation and may be affected by population differences between trials.

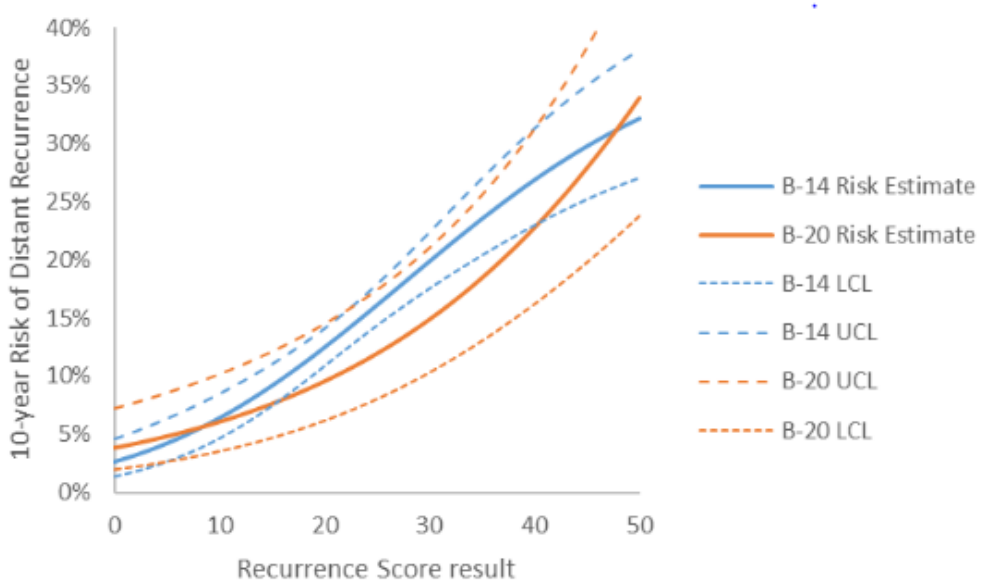
Additional data (personal communication with Prof Tang, University of Pitsburgh, via NICE, February 2018) compares the recurrence rates for a range of Oncotype DX scores in B14 and B20 (Figure 1). This analysis (which uses continuous Oncotype DX scores) is interpreted by Prof Tang as suggesting that the range of distant recurrence risk estimates, and slopes, are very similar between B20 and B14. However, the EAG still note that recurrence rates per risk group do appear to show greater separation in B20 than B14 (Table 1).

Table 1: Comparison of Oncotype prognostic ability in B14 and B20

Oncotype risk group	NSABP-B14 (Paik 2004)[14]		NSABP-B20 (Paik 2006[2])			
	Tamoxifen		Tamoxifen		Tamoxifen + chemotherapy	
	% patients per risk group (n)	% recurrence-free 10yr	% patients per risk group (n)	% recurrence-free 10yr	% patients per risk group (n)	% recurrence-free 10yr
Low	51% (388)	93.2%	60% (135)	96.8%	51% (218)	95.6%
Intermediate	22% (149)	85.7%	20% (45)	90.9%	21% (89)	89.1%
High	27% (181)	69.5%	21% (47)	60.5%	28% (117)	88.1%

Data from Table 12 in EAG report (also comment 161a in Comments on Diagnostics Consultation Document)

Figure 1: **10yr risk of distant recurrence in tamoxifen-alone groups: B20 and B14. Reproduction of Figure provided via personal communication with Prof Tang, University of Pitsburgh, via NICE, February 2018**



d) *Clinical relevance of chemotherapy benefit is unclear for the Oncotype DX intermediate-risk group:* Hazard ratios for chemotherapy benefit are available for this group, but it is unclear how they should be interpreted in clinical practice, i.e., would patients be treated, not treated, or would other clinicopathological variables be taken into consideration when making a decision?

e) *The number of events per subgroup is relatively low*, particularly for the B20 study (

Table **2**). Confidence intervals for the hazard ratios in low-risk and intermediate-risk groups are very wide in both B20 and SWOG-8814 (

Table **2**).

f) Use of RS in clinical practice alongside clinicopathological factors: The RSPC algorithm (Oncotype DX plus age, tumour size and grade) showed a non-significant interaction test between chemotherapy benefit and RSPC risk group,[3] indicating that the incorporation of clinicopathological factors may reduce prediction of chemotherapy benefit, and therefore if chemotherapy decisions are based on an informal consideration of clinicopathological factors alongside the Oncotype DX score, this may reduce the predictive ability of Oncotype DX in clinical practice.

Table 2: Event rates for B14, B20 and SWOG-8814

Oncotype risk group	Treatment	N events / N patients		
		B14 (Paik 2004)[14] LN0	B20 (Paik 2006)[2] LN0	SWOG-8814 (Albain 2010),[1] LN+
Low	Chemo	-	10 / 218	26 / 91
Low	No chemo	28 / 338	5 / 135	15 / 55
Intermediate	Chemo	-	9 / 89	20 / 57
Intermediate	No chemo	25 / 149	7 / 45	22 / 46
High	Chemo	-	13 / 117	28 / 71
High	No chemo	56 / 181	18 / 47	26 / 47

Table 3: Study and patient characteristics: Oncotype DX and RSPC for chemotherapy benefit

Reference; N	Cohorts	Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Reanalysis of RCT – Oncotype DX								
Albain 2010[1] N=367	SWOG-8814	USA	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health	18-30	100% HR+ 12% HER2+ Postmenopausal 100% Female	LN+, 100% LN>3, 38%	1) tamoxifen monotherapy 2) Tamoxifen plus cyclophosphamide
Paik 2006[2] Tang 2011a[4] N= 651	NSABP B-20	USA	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health	18-30	100% ER+ % NR HER2+/- Meno NR Female 100%	LN0	1)tamoxifen monotherapy (N=227) 2) Tamoxifen plus cyclophosphamide (N=424)
Observational studies – Oncotype DX								
Barcenas 2017[5] Le Du 2015[5, 6] N=1424	MD Anderson Centre	USA	Retrospective cohort study	NR	11-25	100% HR 100% HER2- 67% postmeno 99% female Had O-DX test	LN0	91% ET 22% CT Treated according to usual practice with O-DX test
Stemmer 2016[8] Stemmer 2016[11] 1)LN0-1mic, N=1594[11] 2)LN1mic–LN3, N=627[8]	Clalit Health Services	Israel	Retrospective cohort study	NR	18-30	100% ER+ 100% HER2- Meno NR Had O-DX test	1)LN0-LNmic 2)LNmic-LN3	Treated according to usual practice with O-DX test 1)% ET NR 20% CT 2)% ET NR 27% CT
Petkov 2016[15] Roberts 2016[10] Roberts2017[12]N=40,134	SEER registry	USA	Retrospective cohort study	NR Genomic health	18-30	100% HR+ 100% HER2- 40-85 years old Unclear if only those with O-DX test	LN0	% ET NR CT 23% Treated according to usual practice with O-DX test
Reanalysis of RCT – RSPC								

Tang 2011b[3] **B-20:** n=625	NSABP B-20	USA	Reanalysis of prospective trials (RCT); archive tissue	FFPE Genomic Health	**RSPC:** 12% - 20%	100% ER+ HER2+/-, % NR	**B-20:** LN0	**B-20:** 36% ET; 64% CT&ET

CT, chemotherapy; ET, endocrine therapy; pts, patients; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; RS or O-DX, Oncotype DX recurrence score; FFPE, formalin fixed paraffin embedded; postmeno, postmenopausal; Meno, menopausal status; RSPC, recurrence score- clinical-pathological score

Table 4: Quality assessment of studies reporting the ability of Oncotype DX and RSPC to predict chemotherapy responsiveness

Author, Year	Cohort name	Derivation or validation?	Study design appro priate?	All eligible patients included?	Blinding (of test assessors to outcomes)	Definition of outcome standardised or a priori?	Applicability: Patient Spectrum	Applicability : Test as per decision problem?
O-DX								
Albain 2010[1]	SWOG-8814	V	Y, R-RCT	N InT, TF	Y	Y	N: >20% >LN3+[a]	Y
Paik 2006[2] Tang 2011a[4]	NSABP B-20	V **	Y, R-RCT	N InT, <5% cancer cells, MS	UC	Y	UC, % HER2+ NR	Y
Barcenas 2017[5] Le Du 2015[5, 6]	MD Anderson Cancer Centre	V	N, not RCT	N, SFT	Y	Y	No, SFT	Y
Stemmer 2016[8] Stemmer 2016[11]	Clalit Health Services[7, 8]	V	N, not RCT	N, SFT	Y	Y	No, SFT	Y
Petkov 2016[15] Roberts 2016[10] Roberts2017[12]	SEER registry[9, 10]	V	N, not RCT	N, SFT	Y	Y	No, SFT	Y
O-DX RSPC								
Tang 2011b[3]	NSABP B-20 cohort	D & V of RSPC[b]	Y, R-RCT	N Pts ER+ by RS only; MS	UC	Y	Unclear - % HER2+ NR	Y

Y, Yes; N, No; UC, unclear; R-RCT, Reanalysis of RCT; InT, insufficient tissue; TF, test failure; MS, missing samples; D, Development; V, validation; SFT, only those sent for test included;
[a] Most/all analyses adjusted for number of positive nodes (1 to 3 and 4 or more); [b]used some of O-DX derivation sample

REFERENCES

1. Albain K, Barlow W, Shak S, Hortobagyi G, Livingston R, Yeh IT, *et al.* Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* 2010;11:55-65.
2. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, *et al.* Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *Journal of Clinical Oncology* 2006;24:3726-34.
3. Tang G, Cuzick J, Costantino JP, Dowsett M, Forbes JF, Crager M, *et al.* Risk of recurrence and chemotherapy benefit for patients with node-negative, estrogen receptor-positive breast cancer: recurrence score alone and integrated with pathologic and clinical factors. *J Clin Oncol* 2011b;29:4365-72.
4. Tang G, Shak S, Paik S, Anderson SJ, Costantino JP, Geyer CE, Jr., *et al.* Comparison of the prognostic and predictive utilities of the 21-gene Recurrence Score assay and Adjuvant! for women with node-negative, ER-positive breast cancer: results from NSABP B-14 and NSABP B-20. *Breast Cancer Res Treat* 2011a;127:133-42.
5. Barcenas CH, Raghavendra A, Sinha AK, Syed MP, Hsu L, Patangan MG, Jr., *et al.* Outcomes in patients with early-stage breast cancer who underwent a 21-gene expression assay. *Cancer* 2017;15:15.
6. Le Du F, Gonzalez-Angulo AM, Park M, Liu DD, Hortobagyi GN, Ueno NT. Effect of 21-gene RT-PCR assay on adjuvant therapy and outcomes in patients with stage I breast cancer. *Clin Breast Cancer* 2015;15:458-66.
7. Stemmer S, Steiner M, Rizel S, Soussan-Gutman L, Ben-Baruch N, Bareket-Samish A, *et al.* Clinical outcomes in early-stage breast cancer treated according to the recurrence score results: Evidence from a large prospectively-designed cohort study. In. Academic-In-Confidence; 2017.
8. Stemmer SM, Steiner M, Rizel S, Geffen D, Nisenbaum B, Peretz T, *et al.* First prospectively-designed outcome study in estrogen receptor (ER)+ breast cancer (BC) patients (pts) with N1mi or 1-3 positive nodes in whom treatment decisions in clinical practice incorporated the 21-gene recurrence score (RS) result. *Ann Oncol Conference: 41st European Society for Medical Oncology Congress, **ESMO*** 2016;27.
9. Petkov VI, Miller DP, Howlader N, Gliner N, Howe W, Schussler NC, *et al.* Outcome disparities by age and 21-gene recurrence score (RS) in hormone receptor positive (HR+) breast cancer (BC). *J Clin Oncol Conf* 2016;34.
10. Roberts M, Petkov VI, Miller DP, Shak S, Howlader N, Cronin K, *et al.* Breast cancer specific survival in patients with node-positive hormone receptor positive invasive breast cancer and Oncotype DX recurrence score results in the SEER database. *J Clin Oncol Conf* 2016;34.
11. Stemmer SM, Steiner M, Rizel S, Soussan-Gutman L, Geffen DB, Nisenbaum B, *et al.* Real-life analysis evaluating 1594 N0/Nmic breast cancer patients for whom treatment decisions incorporated the 21-gene recurrence score result: 5-year KM estimate for breast cancer specific survival with recurrence score results <30 is >98%. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United Stat* 2016;76.
12. Roberts MC, Miller DP, Shak S, Petkov VI. Breast cancer-specific survival in patients with lymph node-positive hormone receptor-positive invasive breast cancer and Oncotype DX Recurrence Score results in the SEER database. *Breast Cancer Res Treat* 2017;163:303-10.
13. Sparano JA, O'Neill A, Gray RJ, Perez EA, Shulman LN, Martino S, *et al.* 10-year update of E2197: Phase III doxorubicin/docetaxel (AT) versus doxorubicin/cyclophosphamide (AC) adjuvant treatment of LN+ and high-risk LN-breast cancer and the comparison of the prognostic utility of the 21-gene recurrence score (RS) with clinicopathologic features. *J Clin Oncol Conf* 2012;30.

14. 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. *New England Journal of Medicine* 2004;351:2817-26.
15. Petkov VI, Miller DP, Howlader N, Gliner N, Howe W, Schussler N, *et al.* Breast-cancer-specific mortality in patients treated based on the 21-gene assay: a SEER population-based study. *NPJ Breast Cancer* 2016;2:16017.