Supplementary File 2: Narrative synthesis and additional Tables from Chapter 2, Results: Oncotype DX

Development: Oncotype DX

Oncotype DX was developed through the selection of 250 candidate genes from the published literature, genomic databases, pathway analyses and microarray-based gene expression profiling studies.[1] Three independent breast cancer cohorts (N=447 patients, NSABP-20; Rush University Medical Centre, Chicago, USA; Providence St Joseph's Hospital, Burbank, USA)[2-4] were then used to identify genes that were highly associated with recurrence in all three cohorts, and for which the assay methods performed consistently, and an algorithm derived to fit the data from the three cohorts, using correlational analysis, concordance measurese of accuracy and bootstrap resampling. Data from NSABP-20 were more highly weighted in the derivation set, as the validation set was to be a trial with similar patient characteristics, NSABP-14. Oncotype DX was derived to predict prognosis in patients with HR+ disease who have been treated with endocrine therapy for 5 years.

Study designs and patients: Oncotype DX for prognosis

Oncotype DX was validated in eleven distinct data sets reported across twelve publications[1, 5-16] and a personal communication with the TransATAC team via NICE (Ivana Sestak, Queen Mary University of London, July 2017).[17] Study and patient characteristics are presented in Table 1.

Study design: Seven studies were reanalyses of prospectively collected RCT data[1, 5-7, 9-12, 15, 17] using archived tissue samples; one of these adopted a nested case-control design.[7, 12] The remaining four data sets were retrospective studies using routinely collected data and archived samples.[8, 13, 14, 16] Data sets ranged from 93[13] to 1065 patients.[9, 15]

Five RCTs were from the USA:

- NSABP-B-14[1] – the National Surgical Adjuvant Breast and Bowel Project (NSABP) which recruited patients between 1982 and 1988 and randomised them to placebo or tamoxifen. Only the tamoxifen arm is included in this analysis. Patients were LN0.
- NSABP-B20[10] – another NSABP trial which recruited patients between 1988 and 1993 and randomised them to either tamoxifen alone, or tamoxifen plus chemotherapy (cyclophosphamide, methotrexate, and fluorouracil (CMF) or methotrexate and fluorouracil (MF). Only analyses of the tamoxifen plus chemotherapy arm are presented here, as the tamoxifen monotherapy arm was the training set for Oncotype DX (and therefore does not count as a validation cohort). Patients were LN0.

- SWOG-8814[5] – the Southwest Oncology Group trial 8814, which recruited patients between 1989 and 1995 and randomised them to one of three arms: (1) tamoxifen only; (2) cyclophosphamide, doxorubicin and fluorouracil (CAF) followed by tamoxifen, or (3) CAF with concurrent tamoxifen. Only the tamoxifen arm was included in Albain 2010.[5] Patients were LN+.
- NSABP-28[9, 15] – a third NSABP trial which recruited patients between 1995 and 1998 and randomised them to one of two chemotherapy arms (doxorubicin plus cyclophosphamide (AC; 4 cycles) or four cycles of AC followed by four cycles of paclitaxel). Patients analysed received both endocrine therapy and chemotherapy. Patients were LN+.
- E2197[7, 12] – an Eastern Cooperative Oncology Group (ECOG) trial, which recruited patients between 1997 and 1999 and randomised them to one of two chemotherapy (doxorubicin or docetaxel) plus tamoxifen arms. Patients analysed received both endocrine therapy and chemotherapy. The analysis reported here is a nested case-control study using the trial data. Patients were a mix of LN0/LN+.

The two remaining RCTs were from the UK[6, 17] and France,[11] respectively:

- TransATAC[6, 17] was an international trial, but only UK samples were included in this analysis. The trial evaluated anastrozole, tamoxifen, or the combination of both treatments. Recruitment ended in 2006. Only the tamoxifen arm is included in this analysis.
- PACS01[11] was a French trial which recruited patients between 1997 and 2000 and randomised them to one of two chemotherapy treatment arms. All patients analysed (ER+, HER2-) received chemotherapy and 74% received endocrine therapy (after a protocol amendment, ER+ patients received endocrine therapy). Patients were LN+.

There were four retrospective studies.[8, 13, 14, 16] Importantly, archival tissue samples were analysed and as such patients were not treated according to Oncotype DX scores. Studies in which patients were treated according to test results may be confounded, and are therefore excluded from analysis of prognostic performance, but included in the analysis of clinical utility in Chapter 2, Clinical utility of Oncotpye DX. One retrospective study[16] was from the USA, while three[8, 13, 14] were from China or Japan:

- Russell *et al*. 2016[16] recruited patients from two hospitals in the USA (University of South Florida and Morton Plan Hospital. The lymph node status, HER2 status and treatments received were not reported.
- Gong *et al*. 2016[8] (China) recruited patients from Sun Yat-sen Memorial Hospital and the Third Hospital of Nanchang City. Three separate cohorts were recruited, but Oncotype DX data were only reported for one cohort, which recruited post-menopausal LN0 patients. A

second cohort reported IHC4 data (see Chapter 2, Results: IHC4 and IHC4+C). Patients received varying levels of endocrine and chemotherapy according to local practice.

- Sun *et al.* 2011[13] (China) recruited patients from the Hospital Affiliated Academy of Military Medical Science, Beijing. Patients were a mix of LN0 and LN+, with over 18% having more than three positive nodes. Patients received varying levels of endocrine and chemotherapy according to local practice.
- Toi *et al.* 2010[14] (Japan) recruited patients diagnosed between 1992 and 1998 who were treated with tamoxifen, but it is unclear whether any were also treated with chemotherapy. Patients were LN0.

Clinical advice received by the EAG suggests that the three studies from China or Japan[8, 13, 14] may be less generalisable to the English context because (a) patients were treated according to usual clinical practice and this may differ in these countries compared to the UK enough to affect prognostic outcomes, and (b) it is possible that people of different ethnicities have different underlying risk profiles and disease natural history. For this reason, data from these studies should be interpreted with caution and with reference to data from studies where the ethnic profile and clinical practice is similar to the UK.

Lymph node status: Amongst the RCT reanalysis studies, TransATAC[6, 17] and E2197[7, 12] recruited patients regardless of lymph node status (E2197 specifically recruited patients with LN1-3 or LN0 with tumour ≥1.1cm); NSABP B-14[1, 15] and NSABP B-20[10] recruited LN0 patients; and SWOG-8814,[5] NSABP B-28[9, 15] and PACS01[11] recruited LN+ patients. Amongst the retrospective studies, two studies recruited LN0 patients[8, 14] and one[13] recruited patients with any LN status, with patients with LN>3 making up 18% of the cohort. Lymph node status was not reported by Russell *et al.* 2016.[16]

Hormone receptor status: All studies recruited either ER+ or HR+ patients.

Menopausal status: Across the eleven data sets, TransATAC and SWOG-8814 recruited only postmenopausal patients.[5, 6, 17] The remainder either did not report the proportion of patients who were post-menopausal,[1, 7, 9-12, 14, 15] or recruited regardless of menopausal status.[8, 13]

HER2 status: Only TransATAC[6, 17] and Gong *et al.* 2016[8] recruited or reported a subgroup of exclusively HER2- patients. Six studies[1, 9-11, 14-16] did not report HER2 status, probably because patients were recruited before this information was routinely collected.

Treatments: Oncotype DX was derived to predict prognosis in patients with HR+ disease who have been treated with endocrine therapy for 5 years. Treatment with chemotherapy, especially if the effect of chemotherapy is differential across risk categories, could potentially reduce the apparent prognostic performance of the test as it could affect event rates. As such, validation cohorts should treat patients with endocrine monotherapy, but not chemotherapy. TransATAC,[6, 17] SWOG-8814 (subgroup 1)[5] and NSABP B-14[1, 15] all treated patients with endocrine monotherapy, whilst E2197[7, 12], SWOG-8814 (subgroup 2, not included here),[5] NSABP B-20[10] and NSABP B-28[9, 15] treated all patients with endocrine and chemotherapy. PACS01[11] treated all patients with chemotherapy and 74% with endocrine therapy. Gong *et al*. 2016[8] treated all patients with endocrine therapy and 79% with chemotherapy, whilst Sun *et al* 2011[13] treated 75% with endocrine therapy and 81% with chemotherapy and Toi *et al*. 2010[14] treated all with endocrine therapy but did not report whether patients were also given chemotherapy. Russell *et al*. 2016[16] did not report the proportion of patients receiving chemotherapy or endocrine therapy.

Tests and comparators: Oncotype DX prognostic performance

Two studies did not report how the test was conducted (PACS01 study; Russell *et al*. 2016).[11, 16] In all but three other cases the test was performed on fixed, paraffin embedded tissue by Genomic Health using the commercial Oncotype DX assay. The three exceptions were the two studies from China where the test was not performed by Genomic Health,[8, 13] and Paik *et al.* 2004,[1] as Paik *et al.* 2006[10] described the assay used in Paik *et al.* 2004[1] as being "*a preliminary version of the RT-PCR assay (lacking standardized reagents, calibrators, and controls)*". In these three studies, the equivalence of the tests to the commercially offered Oncotype DX assay is unknown.

An analysis of NSABP B-14 included comparison to a "clinical integrator" based on AOL, where the integrator was adjusted to 5-year outcomes rather than the 10 year outcomes used in AOL. The bespoke TransATAC data request provided to the EAG included a comparison of Oncotype DX to three of the tests (IHC4, Prosigna and EndoPredict) and this is presented in Appendix 5.

Quality assessment: Oncotype DX prognostic performance

Quality assessment is summarised in Table 2. All studies were validation studies. Only three studies[1, 5, 6, 17] used an appropriate study design, as eight[7-13, 15, 16] included patients who had been treated with chemotherapy or did not report the proportion treated. No studies included all eligible patients and only three[5-7, 17] stated that they blinded test assessors to patient outcomes. There are concerns about patient spectrum bias in all studies, mainly due to the retrospective nature of the studies and the exclusion of tumour samples with insufficient tissue probably leading to the loss of patients with smaller tumours.

Results: Oncotype DX prognostic performance

Distribution of patients across risk categories

In LN0 cohorts, the proportion of patients ranged from 48%[14] to 64%[17] in the low-risk category, from 20%[10, 14] to 27%[17] in the intermediate-risk category and 9%[17] to 33%[14] in the high-risk category. It is interesting to note that the distribution in the Japanese cohort (Toi *et al*. 2010)[14] indicates more high-risk and fewer low-risk patients that the other LN0 cohorts.

In LN+ cohorts, the proportion of patients ranged from 36%[9, 15] to 57%[17] in low-risk patients, from 30%[11] to 34%[9, 15] in intermediate-risk patients and from 11%[17] to 32%[5] in high-risk patients. The proportion of low-risk patients was generally lower in LN+ than LN0 cohorts, and the proportion of intermediate- and high-risk patients was generally higher.

Prognostic performance: unadjusted analyses

This section reports unadjusted analyses. Adjusted analyses, which show whether the test has prognostic value over clinicopathological variables, are reported in the section "Additional prognostic value".

DRFS: Table 4 of the main report presents DRFS data. One study from China[8] reported 5-year DRFS, with HR for high vs. low-risk of 2.2 (95% CI: 1.11, 4.30, p=0.004) and a C-index (AUC) of 0.685 (95% CI: 0.540, 0.830) indicating the model is better than chance at placing patients into appropriate risk categories.

DRFI: Data relating to DRFI are presented in Table 4 of the main report. Three studies[1, 6, 10, 14, 17] in LN0 patients receiving 100% endocrine monotherapy reported DRFI; all showed a statistically significant prognostic effect. For 5-year DFRI, the HR for a 50-point difference in RS was 6.04 (3.88, 9.41, p<0.001) in one study,[1, 15] while in another the HR for high versus low-risk was 12.39 (95% CI: 4.05, 37.89).[6, 17] For 10-year DFRI, the HR for high versus low-risk was 3.8 (95% CI: 2.36, 6.1; p<0.001) in one study[1, 15] and 5.43 (95% CI: 2.84, 10.35) at 10 years in another,[6, 17] while in a third study the HR for a 50-point difference in RS was 6.20 (95% CI: 2.27, 17.0, p<0.001).[14] Intermediate versus low HRs were lower at 6.37 (95% CI: 2.27,17.87)[17] and 2.21 (1.28, 3.81)[1, 15] at 5 years and 2.67 (95% CI: 1.53, 4.68) at 10 years.[17] Across all three studies, estimates of DRFI at 10 years ranged from 93.2%[1] to 96.7%[14] in low-risk patients, from 85.7% [1, 15] to 100%[14] in intermediate-risk patients, and from 69.5%[1] to 77.2%[6, 17] in high-risk patients.

Two studies of LN0 patients[10, 13] who were treated with endocrine therapy and chemotherapy in varying proportions (Table 4 of the main report) reported 10 year DRFI, with one reporting 5 year DRFI also.[13] Sun *et al*. 2011[13] (China) reported particularly poor DRFI at both time points in

comparison with other studies. DRFI was progressively worse with increasing risk category in both studies (see Table 4 of the main report) and the difference was statistically significant (p=0.02) in the one study that reported this.[13] In the other study (NSABP B-20),[10] survival in the high-risk group was higher (88.1 (95% CI: 82.0, 94.2)) than in other studies where patients were not treated with chemotherapy.

In LN+ patients (Table 4 of the main report), only the TransATAC analysis included 100% patients with endocrine monotherapy.[17] In this study, 5-year DRFI was statistically significantly different for high versus low-risk (4.45 (95% CI: 1.19, 16.58)) and intermediate- versus low-risk (3.84 (95 % CI: 1.31, 11.23)) whereas at 10 years these differences were borderline statistically significant (2.35 (95% CI: 0.99, 5.60) and 1.66 (95% CI: 0.86, 3.23) respectively).

Three LN+ studies[9, 11, 13, 15] treated patients with variable endocrine therapy and chemotherapy and each reported statistically significant differences in DRFI between risk groups. For 5-year DRFI, the HR for a 50-point difference in RS was 4.1 (CI: NR, p<0.001) in one study[11] and 4.22 (2.93, 6.07, p<0.001) in another.[9, 15] DRFI rates were generally lower than LN0 groups, again with Sun *et al*. 2011[13] (China) reporting very poor survival rates compared with other studies.

DFS: Table 4 of the main report presents DFS data. One study[5] in LN+ patients reported a statistically significant 10-year HR for a 50-point difference in RS (2.64 (95% CI: 1.33, 5.27, p=0.006)) but the assumption of proportional hazards was not met with a 5-10 year HR of 0.86 (95% CI: 0.27, 2.74, p=0.80). One study[16] (in patients of unknown LN status and treatment status) reported statistically significant differences between high- and low-risk patients (p=0.760) but not between high- and intermediate-risk, or low- and intermediate-risk groups (p=0.072 and p=0.760 respectively). Two studies[9, 11, 15] in LN+ patients receiving variable levels of endocrine therapy and chemotherapy reported that RS was statistically significantly prognostic for DRFI (p<0.001 in each case);[9, 11, 15] one reported an HR for a 50-point difference in RS of 3.3 (CI: NR, p<0.001)[11] while the other did not report an HR.[9, 15]

OS and BCSS: Table 3 presents OS and BCSS data. Two studies of LN0 patients treated with endocrine monotherapy reported progressively worse survival with increasing risk category.[14, 17] The TransATAC analysis reported statistically significant HRs for intermediate- versus low-, and high- versus low- risk, comparisons for OS at both 5 and 10 years (Table 3),[17] and the other study reported a statistically significant difference between high and low-risk groups (p=0.008).[14]

The TransATAC study[17] of LN+ patients treated with endocrine monotherapy reported statistically significant HRs for intermediate versus low, and high versus low, risk group comparisons for OS, at

both 5 and 10 years (Table 3), whilst Albain *et al*. 2010[5] (LN+) reported an HR for 10-year OS for a 50-point difference in RS of 4.42 (95% CI: 1.96, 9.97, p=0.0006).

In LN+ patients variably treated with endocrine and chemotherapy, one study[11] reported a statistically significant difference in OS (7.7 year median) with an HR for a 50-point difference in RS of 5.0 (CI: NR, p<0.001). Another study reported a statistically significant effect on 10-year OS (p<0.001).[9, 15]

RFI and RFS: Table 4 and Table 5 present RFI and RFS data, respectively. Two studies reported data for these outcomes. Toi *et al*. (Japan)[14] reported a statistically significant difference between high- and low-risk patients for 10-year RFI and RS (both p<0.05). The E2197 analysis[7, 12] reported very similar rates of 5- and 10-year RFI across subgroups of LN0, LN+ and LN+/- patients who were all treated with endocrine and chemotherapy; survival was progressively worse with increasing risk category but no significance tests were reported (the C-index (AUC) for 5-year RFI was 0.69).

Additional prognostic value

This section reports adjusted analyses, which indicate the additional prognostic value of IHC4 over clinicopathological factors. The clinicopathological factors adjusted for vary from study to study, and are detailed in the footnotes to the tables.

Table 5 of the main report presents data relating to the additional prognostic value of Onctoype-Dx RS over clinicopathological variables. One study (E2197)[7, 12] reported RFI for a mixed cohort of LN+/- patients. For RFI, HRs for a 50-point differerence in RS (adjusted for number of positive nodes, tumour size, age, HER2 status and grade) were borderline statistically significant at 5 years (2.12; 95% CI: 0.97, 4.65, p=0.06) and 10 years (2.27; 95% CI: 1.04, 4.97). However, in a subgroup of HER2- patients, the adjusted HR for a 50-point differerence in RS was not statistically significant (data NR).

Two studies (NSABP B-14 and the Japanese study)[1, 14] reported analyses of LN0 patients who received endocrine monotherapy. Both reported analyses adjusted for clinicopathological variables. HRs for a 50-point differerence in RS were statistically significant in all DRFI and RFI analyses,[1, 14] with a statistically significant increase in likelihood ratio χ^2 (p<0.001) over age and tumour size alone, and over age, tumour size, tumour grade, HER2 amplification, ER and PR.[1] HRs for a 50-point differerence in RS adjusted for age and tumour size were not statistically significant for RFS and OS in one study.[14]

In a study of LN0 patients[13] some of whom had endocrine and/or chemotherapy the HR for DRFS, for a 1-point difference in RS, was 1.03 (95% CI: 1.01, 1.06, p=0.017), but it was unclear if all

clinicopathological variables listed were included in the model (age, tumour size, nodal status, ER, PR, HER2, endocrine therapy, chemotherapy, St Gallen), or just endocrine therapy and chemotherapy.

Three studies assessed LN+ patients, some or all of whom were treated with endocrine and chemotherapy. HRs for Oncotype DX RS adjusted for clinicopathological variables (see footnote to Table 5 of the main report) were statistically significant in all three studies[9, 11, 13, 15] for outcomes including DRFI, DRFS, DFS and OS; only one reported an HR, which was for a 1-point difference in RS (1.03 (95% CI: 1.00, 1.07), p=0.039).[13] Notably, of these three studies, only Sun *et al*. (2011) adjusted for ER, PR and HER2.[13]

The use of the 50-point difference in the adjusted analyses of prognostic performance indicate that RS is prognostic after adjusting for clinicopathological factors, but does not provide information about the clinical significance of the 18 -30 RS cut points.

Oncotype DX versus Adjuvant! Online

Two studies (E2197 and NSABP-B-14)[1, 7, 12] compared Oncotype Dx RS with AOL (Table 5 of the main report). The E2197[7, 12] study (LN0/+, 100% endocrine and chemotherapy) compared Oncotype DX against a model (the "clinical integrator") based on AOL, where the integrator was adjusted to 5-year outcomes rather than AOL's 10 year outcomes. For RFI, based on the C-indexes (AUC) reported (Oncotype DX 0.69; Integrator 0.61; p-values NR) and on HRs (Oncotype DX HR for 50-point difference: 2.51 (95% CI: 1.71, 3.70; p<0.001); integrator HR: 1.51 (95% CI: 1.07, 2.13; p=0.02)), the integrator performed less well than Oncotype DX (statistical significance NR). Analyses (not in Table) where patients were sub-grouped by the integrator or RS risk groups, and the other test applied to the patients in that risk group, showed that both tests provided additional prognostic information over the other.

The NSABP B-14 analysis[1] of LN0 patients treated with endocrine monotherapy showed that Oncotype DX was statistically significantly prognostic for DRFI when adjusted for AOL (HR for 50-point difference 2.83 (95% CI: 1.91, 4.18, p<0.001). In addition, AOL was statistically significantly prognostic for DRFI when adjusted for Oncotype DX (HR 1.93; 95% CI: 1.27, 2.91, p=0.002) (Table 5 of the main report). When clinical variables were added into the model, the HR for AOL was no longer statistically significant (HR 0.86 (95% CI: 0.45, 1.62, p=0.636)) whereas that for Oncotype DX was (HR 2.37 (95% CI: 1.58, 3.55, p<0.001)).

Oncotype DX versus CTS and NPI

The TransATAC analysis[17] reports a reduced dataset of patients where data for all four in-scope tests are available. Additional prognostic value over NPI or the Clinical Treatment Score (CTS, a

combination of nodal status, tumour size, grade, age and treatment) was assessed via increases in likelihood ratio χ^2 for 10-year DRFI, for Oncotype DX plus NPI or CTS, over NPI or CTS alone (Table 5 of the main report). Increases in likelihood ratio χ^2 were statistically significant for the mixed cohort of LN0/+ patients: 15.22 (p=0.0001) over CTS and 11.89 (p=0.0006) over NPI, as well as for LN0 patients: 10.64 (p=0.001) over CTS and 8.82 (p=0.003) over NPI. However, increases in likelihood ratio χ^2 were not statistically significant for LN+ patients: 3.56 (p=0.06) over CTS and 2.14 (p=0.1) over NPI.[17]

Prognostic performance: Oncotype RSPC

The Oncotype RSPC algorithm includes Oncotype RS plus age, tumour size and grade.[18] Table 5 of the main report presents data relating to Oncotype RSPC. One study (Tang *et al*. 2011b)[18] derived the RSPC score in a meta-analysis of NSABP B-14 and TransATAC (LN+/- patients, 100% endocrine monotherapy), and performed a limited validation in NSABP B-20 (LN0 patients, 100% treated with endocrine therapy; 64% also with chemotherapy).

Derivation: In the derivation cohort, both RSPC and Oncotype DX RS had statistically significantly (p<0.001) worsening 10-year DRFI rates as test scores increased (HR/CI NR). However, DRFI rates were not significantly different between RSPC and RS within each risk group (respectively, 93.5% vs. 94.1%, p=0.68 for low-risk; 82.4% vs. 86.2%, p=0.27 for intermediate-risk; and 73.8% vs. 70.5%, p=0.42 for high-risk. RSPC was able to reclassify RS intermediate patients as 16.9% (n=46) high-risk RSPC and 55.1% (n=150) low-risk RSPC; RS low-risk patients as 1.9% (n=15) high-risk RSPC and 8.9% (n=70) intermediate-risk RSPC; and RS high-risk patients as 28.6% (n=NR) intermediate-risk RSPC. The increase in likelihood ratio χ^2 for 10-year DRFI was 76.9 (p<0.001) for RSPC over RS, and 45.4 (p<0.001) for RSPC over grade, tumour size and age.

Vlaidation: NSABP B-20 was the derivation set for Oncotype DX and is therefore not an entirely independent dataset for validation purposes. Only HRs were reported and these were 2.43 (p<0.001) for RSPC and 2.22 (p<0.001) for RS.

Further data relating to the RSPC were reported in the TransATAC data request. However, as the original derivation cohort for RSPC includes TransATAC patients, these data are not included in this analysis. They are included in the section on multiple tests (Appendix 5).

Table 1: Study and patient characteristics: Oncotype DX prognostic performance

Reference; N	Cohorts	Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Reanalyses of RCTs: LN status mixed								
100% ET monotherapy								
Sestak 2017 (data request)[17] Dowsett 2010[6 a] N=1048	TransATAC	UK	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health	18-30	100% HR+ 100% HER2- Postmenopausal 100% Female	LN+/- LN0, 79.1% LN1-3, 20.9%	100% ET monotherapy
Variable ET&CT								
Goldstein 2008 (5 year) ; [7]Sparano 2012[12] (10-year) N=465	E2197 (ECOG trial)	USA	Nested Case-Control from prospective RCT; archive tissue	FFPE Genomic Health	18-30	100% HR+ 44% HER2- (34.1% unknown), Meno NR 100% Female If LN0, tumour ≥1.1cm	LN0, 56.5% LN1-3, 43.5%	100% ET & CT 40% aromatase inhibitor
Reanalyses of RCTs: LN0 studies								
100% ET monotherapy								
Paik 2004;[1] Wolmark 2016[15] N= 668	NSABP B-14	USA	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health[b]	18-30	100% ER+ HER2+/-, % NR Meno NR Female NR	LN0	100% ET monotherapy
100% ET & CT								
Paik 2006[10] N= 424	NSABP B-20	USA	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health	18-30	100% ER+ HER2+/-, % NR Meno NR Female 100%	LN0	100% ET + 100% CT (N=424)
Reanalyses of RCTs: LN+ studies								
100% ET monotherapy								
Albain 2010[5] N=148 (tamoxifen monotherapy subgroup)	SWOG-8814	USA	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health	18-30	100% HR+ 91% HER2- Postmenopausal 100% Female	LN+, 100% LN>3, 37%	100% ET monotherapy

Reference; N	Cohorts	Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
100% CT&ET								
Wolmark 2016[15 c] Mamounas 2012[9 d] N=1065	NSABP B-28 (Also reports NSABP B-14, listed here under Paik 2004)	USA	Reanalysis of prospective trial (RCT); RS available	FFPE Genomic Health (Assumed for B-28)	18-30	100% ER+ HER2 NR Meno NR Female NR	LN+	100% CT & ET
Variable ET&CT								
Penault-Llorca 2014[11] N=530	PACS01	France	Reanalysis of prospective trial (RCT); unclear if archive tissue	NR	NR	100% HR+ HER2 NR Meno NR Female NR	LN+	100% CT 74.2% ET
Retrospective studies								
Gong 2016[8] O-DX subgroup N=153	SYSMH; CCSYU; 3rdHNC	China	Retrospective reanalysis of routinely collected data; archive tissue	FFPE Multiplex branched-DNA liquid chip technology Surexam, Guangzhou, China	NR, assume 18-30	100% HR+ 100% HER2- 61% post meno % female NR non-metastatic	LN0	100% ET; 79% CT
Russell 2016[16] N=135	University of South Florida; Morton Plan Hospital	USA	Observational study (not treated according to O-DX)	NR	NR	100% ER+ HER2- NR Meno NR Female NR	NR	NR – usual practice guided by MMP
Sun 2011[13] N=93	Hospital Affiliated Academy of Military Medical Science, Beijing	China	Retrospective reanalysis of routinely collected data; consecutive	FFPE qRT-PCR (not Genomic Health)	18-30	100% HR+ 86% HER2- (7.5% unclear) 82.6% Premeno 100% Female	LN+/- LN0, 61.3% LN1-3, 19.4% LN>3, 18.3%	75.3% ET 80.6% CT
Toi 2010[14] N=200	8 Japanese hospitals (unnamed)	Japan	Retrospective reanalysis of routinely collected data; archive tissue	FFPE Genomic Health	18-30	100% ER+ HER2 NR Meno NR % Female NR T1-T2	LN0	100% ET % CT NR

Reference; N	Cohorts	Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Oncotype DX RSPC data								
Tang 2011b[18] **B-14:** n=647 **TransATAC:** n=1088 **B-20:** n=625	NSABP B-14 & TransATAC meta-analysis NSABP B-20	NSABP: USA TransATAC: UK	Reanalysis of prospective trials (RCT); archive tissue	FFPE Genomic Health	**RSPC:** 12% - 20% **RS:** 18-30	100% ER+ **B-14:** HER2+/-, % NR **TransATAC:** HER2+/- **B-20:** HER2+/-, % NR	**B-14:** LN0 **TransATAC:** LN+/- **B-20:** LN0	**B-14:** 100% ET **TransATAC:** 100% ET **B-20:** 36% ET; 64% CT&ET

O-DX, Oncotype DX; MMP, MammaPrint; FFPE, formalin fixed, paraffin embedded tumour samples; SYSMH, Sun Yat-sen Memorial Hospital; CCSYSU, Cancer Centre of Sun Yat-sen University; 3rdHNC, Third Hospital of Nanchang City

[a] TransATAC is reported across several publications, each with a different aim and/or reporting results of different tests. Data was also made available to the EAG via NICE which included only patients with HR+, HER2- disease with LN0-3; [b] from Paik 2006, about Paik 2004 "a preliminary version of the RT-PCR assay (lacking standardized reagents, calibrators, and controls)"; [c] Note data for B-14 also reported in this article, but reported here under Paik 2004;[d] Note a second abstract (Mamounas 2012)[19] presented data for the same cohort, but split by chemotherapy treatment group, and has been excluded as it added no new data to Mamounas 2012[9]

Table 2: Quality assessment of Oncotype DX prognostic performance studies

Reference: N	Cohorts	Derivation or validation?	Study design appropriate?	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?
Albain 2010[5]	SWOG-8814	V	Y, reanalysis of endocrine only arm of RCT	N InT; TF	Y	Y	No, InT, TF, >20% LN>3. However, adjustments applied in several analyses	Y
Goldstein 2008 (5 year);[7] Sparano 2012[12] (10-year)	E2197 (ECOG trial)	V	N- authors identify possible bias; all CT	UC	Y	Y	No, >20% HER2+	Y
Gong 2016[8]	SYSMH; CCSYU; 3rdHNC	V	N, cohort study, some CT	N InT; MD	UC	Y	N, InT, MD	N – Oncotype DX algorithm, but used Surexam, Guangzhou, China assay.
Paik 2004[1]	NSABP B-14	V	Y, reanalysis of RCT; endocrine only	N InT	UC	Y	N, InT, %HER2- NR	UC
Paik 2006[10]	NSABP B-20	V (ET&CT arm)	N, reanalysis of RCT; CT	N InT	UC	Y	N, InT, %HER2- NR	Y
Penault-Llorca 2014[11]	PACS01	V	N, reanalysis of RCT; some CT	N InT	UC	Y	N InT	UC[a]
Russell 2016[16]	University of South Florida; Morton Plan Hospital	V	N, cohort study, usual practice (some CT)	N InT, SfT	UC	Y	N InT	Y
Sun 2011[13] N=93	Hospital Affiliated Academy of Military Medical Science (HAAMMS), Beijing	V	N, cohort study (retrospective) some CT	N InT; MD	UC	Y	N InT, MD, 18% LN>3	N Oncotype DX algorithm, but assay not Genomic Health
Toi 2010[14]	8 Japanese hospitals (unnamed)	V	UC, cohort study (retrospective), %CT NR	N InT; MD; FT	UC	Y	N InT, MD, FT, HER2 NR	Y

Reference: N	Cohorts	Derivation or validation?	Study design appropriate?	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?
Wolmark 2016[15] Mamounas 2012[9] N=1065	NSABP B-28	V	N, reanalysis of RCT; all CT	N InT; FT	UC	Y	N InT; FT	Y
Sestak 2017 (data request)[17] Dowsett 2010[6]	TransATAC	V	Y, reanalysis of RCT, ET monotherapy	N InT; FT	Y	Y	N InT, FT	Y
RSPC								
Tang 2011b[18]	NSABP B-14 & TransATAC meta-analysed NSABP B-20	D, V	Y, reanalysis of RCT N, B-20 some CT	N, InT; ER+ by RS[b]	UC	Y	UC %HER- NR	Y

CT, chemotherapy; TF; InT, insufficient tissue; test failure; MD, missing data; ER, oestrogen receptor status; RS, recurrence score; SfT, sent for test; SYSMH, Sun Yat-sen Memorial Hospital; CCSYSU, Cancer Centre of Sun Yat-sen University; 3rdHNC, Third Hospital of Nanchang City

[a] In this analysis, patients were classed as ER+ using RS rather than histology, which does not reflect clinical practice as patients would be selected for RS testing using histology

[b] from Paik 2006, about Paik 2004, "a preliminary version of the RT-PCR assay (lacking standardized reagents, calibrators, and controls)" suggests the assay used was somewhat different to the commercial version now available.

Table 3: Oncotype DX prognostic performance, OS & BCSS

Reference; N	Cohorts	Population	Nodal status	ET/CT	outcome	% pts per group			% risk: 0-5 yr (95% CI)			% risk: 0-10 yr (95% CI)			OS: HR (95% CI)		
						Low	Inter	High	Low	Inter	High	Low	Inter	High	0-5 years	0-10 years	Other
LN0/+, variable ET&CT																	
Sun 2011[13] N=93	HAAMMS	100% HR+ 86% HER2- (7.5% unclear)	LN+/-	75.3% ET 80.6% CT	BCSS[a]	37	31	32	RS as categorical or continuous variable p=0.553.								
LN0, 100% ET monotherapy																	
Toi 2010[14] N=200	8 Japanese hospitals (unnamed)	100% ER+ HER2 NR	LN0	100% ET	OS	48	20	33				93.6 (86.4, 97.1)	97.4 (83.2, 99.6)	80.9 (68.7, 88.7)			
												p=0.008 log rank test high vs low					
Sestak 2017 (data request)[17] Dowsett 2010[6] N=829	TransATAC	100% HR+ 100% HER2- Postmeno	LN0	100% ET	OS	65	27	9	95.0	90.9	84.9	81.2	73.7	60.2	**Inter vs Low:** 1.82 (1.02, 3.24) **High vs Low:** 3.16 (1.57, 6.38)	**Inter vs Low:** 1.46 (1.04, 2.04) **High vs Low:** 2.54 (1.65, 3.89)	
LN+, 100% ET monotherapy																	
Albain 2010[5] N=148	SWOG-8814	100% HR+ 91% HER2- Postmeno	LN+, 100% LN>3, 37%	100% ET	OS	37	31	32				77	68	51		RS 50 point difference: 4.42 (1.96, 9.97, p=0.0006)	RS risk categories: log rank p=0.003 Proportional hazards not met

Reference; N	Cohorts	Population	Nodal status	ET/CT	outcome	% pts per group			% risk: 0-5 yr (95% CI)			% risk: 0-10 yr (95% CI)			OS: HR (95% CI)		
						Low	Inter	High	Low	Inter	High	Low	Inter	High	0-5 years	0-10 years	Other
Sestak 2017 (data request)[17] Dowsett 2010[6] N=219	TransATAC	100% HR+ 100% HER2- Postmeno	LN1-3	100% ET	OS	57	32	11	92.0	79.7	76.0	66.9	61.7	47.7	**Inter vs Low:** 2.70 (1.20, 6.07) **High vs Low:** 3.35 (1.22, 9.21)	**Inter vs Low:** 1.36 (0.82, 2.24) **High vs Low:** 2.19 (1.17, 4.11)	
LN+, variable ET&CT																	
Penault-Llorca 2014[11] N=530	PACS01	100% HR+	LN+	100% CT 74.2% ET	OS	39	30	31	99.0 (96.2, 99.8)	95.6 (90.9, 97.9)	85.6 (79.1, 90.2)					7.7yr median FU, RS 50 point difference: 5.0 (CI NR), p<0.001	
									p<0.001								
Wolmark 2016[15] Mamounas 2012[9] N=1065	NSABP-28	100% ER+ HER2 NR Meno NR Female NR	LN+	100% CT & ET	OS	36	34	30				90.0 (86.4, 92.6)	74.7 (69.8, 78.9)	63.0 (57.4, 68.2)			
												p<0.001					

HAAMMS, Hospital Affiliated Academy of Military Medical Science, Beijing; N, number of patient; ET, endocrine therapy; CT, chemotherapy; pts, patients; Inter, intermediate group; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Fu, follow-up; RS, Oncotype DX recurrence score;

[a] Called "overall survival" in the publication, but defined as only breast-cancer deaths

Table 4: Oncotype DX prognostic performance, RFI

Reference; N	Cohorts	Population	Nodal status	ET/CT	% pts per group			% RFI risk: 0-5 yr			% RFI risk: 0-10 yr			Other
					Low	Inter	High	Low	Inter	High	Low	Inter	High	
LN0, 100% ET monotherapy														
Toi 2010[14] N=200	8 Japanese hospitals (unnamed)	100% ER+ HER2 NR Meno NR % Female NR T1-T2	LN0	100% ET	48	20	33				94.5 (87.2, 97.7)	97.5 (83.5, 99.6)	75.4 (62.4, 84.4)	
											High vs Low: p<0.05			
LN0, 100% ET&CT														
Goldstein 2008;[7] Sparano 2012[12] N=233	E2197 (ECOG trial)	100% HR+ 44% HER2- Pre/post-meno	LN0	100% ET&CT	-	-	-	96[a]	86[a]	87[a]	93[a]	76[a]	81[a]	
LN+/-, 100% ET&CT														
Goldstein 2008;[7] Sparano 2012[12] N=465	E2197 (ECOG trial)	100% HR+ 44% HER2- Pre/post-meno	LN0, 56.5% LN1-3 43.5%	100% ET&CT	46	30	24	96[a]	87[a]	83[a]	92[a]	77[a]	75[a]	C-index (AUC) 0.69 at 0-5yr
LN+, 100% ET&CT														
Goldstein 2008;[7] Sparano 2012[12] N=232	E2197 (ECOG trial)	100% HR+ 44% HER2- Pre/post-meno	LN1 (N=123)	100% ET&CT	-	-	-	98[a]	90[a]	82[a]	93.5[a]	85[a]	62.5[a]	
			LN2-3 (N=109)		-	-	-	92[a]	84[a]	67[a]	88[a]	76[a]	63[a]	

RFI, recurrence-free interval; N, number of patient; ET, endocrine therapy; CT, chemotherapy; pts, patients; Inter, intermediate group; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; AUC, area under the curve;

[a] Read off graph, RFI from recurrence rates

Table 5: Oncotype DX prognostic performance, RFS

Reference; N	Cohorts		Population	Nodal status	ET/CT	% pts per group			% RFS risk: 0-10 yr			HR
						Low	Inter	High	Low	Inter	High	10 year
	LN0, 100% ET monotherapy											
Toi 2010[14] N=200	8 Japanese hospitals		100% ER+ HER2 NR	LN0	100% ET	48	20	33	90.4 (82.4, 94.9)	94.9 (81.2, 98.7)	76.6 (64.1, 85.2)	
									High vs Low: p<0.05			
	RFS, relapse-free survival (events include locoregional or distant recurrence or death from any cause; censored are contralateral disease, new cancer, deaths before recurrence)											

REFERENCES

1. 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.
2. Cobleigh M, Bitterman P, Baker J, Cronin M, Liu M, Borchik R, *et al.* Tumor gene expression predicts distant disease-free survival (DDFS) in breast cancer patients with 10 or more positive nodes: High throughput RT-PCR assay of paraffin-embedded tumor tissues. Proc Am Soc Clin Oncol, abstract no. 6402, p. 2003.
3. Esteban J, Baker J, Cronin M, Liu M, Llamas M, Walker M, *et al.* Tumor gene expression and prognosis in breast cancer: multi-gene RT-PCR assay of paraffin-embedded tissue. Proc Am Soc Clin Oncol, abstract no. 6400, p. 2003.
4. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, *et al.* Multi-gene RT-PCR assay for predicting recurrence in node negative breast cancer patients-NSABP studies B-20 and B-14. *Breast Cancer Res Treat* 2003;82:S10-S1.
5. 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.
6. Dowsett M, Cuzick J, Wale C, Forbes J, Mallon EA, Salter J, *et al.* Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: A TransATAC study. *J Clin Oncol* 2010;28:1829-34.
7. Goldstein LJ, Gray R, Badve S, Childs BH, Yoshizawa C, Rowley S, *et al.* Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features. *J Clin Oncol* 2008;26:4063-71.
8. Gong C, Tan W, Chen K, You N, Zhu S, Liang G, *et al.* Prognostic value of a BCSC-associated microRNA signature in hormone receptor-positive HER2-negative breast cancer. *EBioMedicine* 2016;11:199-209.
9. Mamounas EP, Tang G, Paik S, Baehner FL, Liu Q, Jeong JH, *et al.* Prognostic impact of the 21-gene recurrence score (RS) on disease-free and overall survival of node-positive, ER-positive breast cancer patients (pts) treated with adjuvant chemotherapy: Results from NSABP B-28. *J Clin Oncol Conf* 2012;30.
10. 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. *J Clin Oncol* 2006;24:3726-34.
11. Penault-Llorca FM, Filleron T, Asselain B, Baehner FL, Fumoleau P, Lacroix-Triki M, *et al.* Prediction of recurrence with the Oncotype DX recurrence score in node-positive, HR-positive, breast cancer patients treated with adjuvant chemotherapy: Results from PACS01 trial. *J Clin Oncol Conf* 2014;32.
12. 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.
13. Sun B, Zhang F, Wu SK, Guo X, Zhang LL, Jiang ZF, *et al.* Gene expression profiling for breast cancer prognosis in Chinese populations. *Breast J* 2011;17:172-9.
14. Toi M, Iwata H, Yamanaka T, Masuda N, Ohno S, Nakamura S, *et al.* Clinical significance of the 21-gene signature (Oncotype DX) in hormone receptor-positive early stage primary breast cancer in the Japanese population. *Cancer* 2010;116:3112-8.
15. Wolmark N, Mamounas EP, Baehner FL, Butler SM, Tang G, Jamshidian F, *et al.* Prognostic impact of the combination of recurrence score and quantitative estrogen receptor expression (ESR1) on predicting late distant recurrence risk in estrogen receptor-positive breast cancer after 5 years of tamoxifen: Results from NRG Oncology/National Surgical Adjuvant Breast and Bowel Project B-28 and B-14. *J Clin Oncol* 2016;34:2350-8.

16. Russell S, Shivers SC, Blumencranz L. Long-term follow-up of early stage breast cancer patients with results of MammaPrint, Oncotype DXand MammoStrat risk classif ication assays. *SABCS* 2016.
17. Sestak I, Dowsett M, Cuzick J. NICE request - TransATAC data analysis. In; 2017.
18. 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.
19. Mamounas EP, Tang G, Paik S, Baehner FL, Liu Q, Jeong JH*, et al.* Association between the 21-gene recurrence score (RS) and benefit from adjuvant paclitaxel (Pac) in node-positive (N+), ER-positive breast cancer patients (pts): Results from NSABP B-28. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72.