Supplementary File 5: Narrative synthesis and additional Tables from Chapter 2, Results: MammaPrint

Developmen of MammaPrint

Derivation of the 70-gene signature used a case-control design with 78 node-negative (LN0) patients aged under 55 years: 34 patients with and 44 patients without distant metastases within 5 years (van 't Veer *et al*., 2002).[1] Validation in an additional 19 patients is described within the same article; these patients were also young and LN0, 12 with and 7 without distant metastases within 5 years. Derivation of the 70-gene signature used a DNA microarray containing approximately 25,000 genes.

The first main validation study of the 70-gene signature used a retrospective consecutive series of 295 patients (151 LN0 and 144 LN+ patients) from the Netherlands Cancer Institute (NKI), described by van de Vijver *et al*.[2] (2002). Of these, 61 patients (21%) were also part of the derivation set.[1] Again a 25,000-gene microarray was used to identify the 70-gene signature. This study showed that the signature was significantly prognostic for 5-year DMFS and OS in LN0 and LN+ patients. Updated results for this cohort have since been reported and are presented below.

Threshold: MammaPrint

MammaPrint classifies patients as low-risk (good prognosis) and high-risk (poor prognosis). Correlation coefficients are calculated for the expression level of the 70 genes between individual patients and an "average" good prognosis profile based on the derivation study by van 't Veer *et al*..[1] In the first version of MammaPrint, samples were classified as low-risk if the correlation coefficient was greater than 0.4 and high-risk if less than 0.4 (van 't Veer *et al*.).[1] In a later version of MammaPrint, this threshold was mathematically adjusted to 0 so that low-risk samples are greater than 0 and high-risk samples are ≤0. Both thresholds are the same apart from the adjustment to zero. The same threshold is used in all clinical studies (personal communication with manufacturer).

Prognostic performance in derivation and first validation cohorts

In the derivation cohort (n=78),[1] the test incorrectly identified 3/34 patients who recurred as good prognosis and 12/44 patients who did not recur as poor prognosis. The initial validation cohort in the same article (n=19)[1] incorrectly identified 2/19 patients (whether these were recurrences or non-recurrences was not reported). A multivariable logistic regression analysis that included "classical prognostic factors" (variables not reported) reported an odds ratio for distant metastasis of 18 (95% CI 3.3 to 94) for low- compared with high-risk patients in the derivation cohort (n=78), and a likelihood ratio p-value of 0.0001, though it was unclear whether the patients included were from the derivation cohort or the validation cohort.

Equivalence of different test methods: MammaPrint

Following development of the MammaPrint mini-array specific to the 70 genes, Glas *et al.*[3] (2006) demonstrated that the 70-gene MammaPrint microarray provided very similar results to the 25,000-gene microarray. Within the 78 patients from the derivation set,[1] risk group classification was very similar between the 25,000-gene array and the MammaPrint 70-gene array (Pearson correlation 0.92). For 145 of 151 LN0 patients from the van de Vijver *et al.* validation study,[2] HRs for low vs. high-risk for DMFS over all follow-up were very similar for the two array types: HR 5.5 (95% CI: 2.5, 12.2) for the 25,000-gene array, and HR 5.6 (95% CI: 2.4, 7.3) for the 70-gene array.[3]

Beumer *et al*. (2016)[4] showed that fresh-frozen and FFPE paired samples give very similar results (Pearson correlation 0.93); that the MammaPrint 70-gene mini-array and whole-genome 25,000-gene array give near-perfect correlation (Pearson correlation 0.99); that samples repeated over 10 years give an overall reproducibility of 97%; and that precision and repeatability (using repeated measurements) are both 98% overall.

Prognostic performance: MammaPrint

Study designs and patients: MammaPrint prognostic performance

Several publications describe validation of the prognostic value of MammaPrint. Many include overlapping cohorts of patients, sometimes pooled with other cohorts, sometimes focussing on patient subgroups (e.g. ER+ or LN0/LN+), sometimes updating the data with longer follow-up, and reporting a range of different outcomes. Therefore, it should be noted that there is some overlap between patient cohorts within the references included here. Table 1 shows both the study reference(s) (column 1) and the cohort(s) (column 2) used for each analysis.

Prognostic data on MammaPrint mainly consists of retrospective analyses of consecutive patient series, many from the Netherlands plus some from other countries. The main nine cohorts are listed below (and in Table 1). Five cohorts consisted of LN0 patients,[5-9] one of LN+,[10] and three included a mix of LN0 and LN+ patients.[2, 11, 12] Three cohorts did not receive adjuvant chemotherapy,[6, 7, 11] while in the other six a subset received chemotherapy,[2, 5, 8-10, 12] though treatment was not influenced by the MammaPrint test since this was performed later on stored tumour samples. In the majority of these series, around 70-80% of patients were ER+, while HER2 was not well reported (Table 1). The nine cohorts are:

- van de Vijver 2002:[2] Retrospective analysis of consecutive series from the Netherlands Cancer Institute (NKI, 1984-95), age ≤52 years; 51% LN0. Updated data are presented in subsequent articles, the most recent being Drukker *et al.*[13] (2014) Independent data from the same centre are reported in Mook 2010[6] (ages 55-71, LN0) and Bueno-de-Mesquita 2009,[5] and there may be some overlap with Mook 2009[10] (1994-2001) and Kok 2009 (1985-94).[11]

- Bueno-de-Mesquita 2009:[5] Retrospective analysis of consecutive series from two Dutch hospitals (NKI and Reinier de Graaf Hospital, 1996-99), all LN0.
- Mook 2010:[6] Retrospective analysis of consecutive series from NKI (1984-96), age 55-71 years, all LN0.
- Mook 2009:[10] Retrospective analysis of consecutive series from NKI and Italy (1994-2001), all LN+ (LN1-3).
- Kok 2009/2012:[11] Retrospective analysis of consecutive series from NKI (1985-94), 82% LN+.
- Buyse 2006[7] (TRANSBIG): Retrospective cohort from the UK, France and Sweden (1980-1999); all LN0.
- Yao 2015:[12] Retrospective analysis of consecutive series from two US centres (1992-2010); mix of LN0 and LN+.
- Wittner 2008:[9] Retrospective analysis of consecutive series from one US centre (1985-1997); all LN0.
- Ishitobi 2010:[8] Retrospective analysis of cases from Osaka Medical Centre, Japan (1998-2001); all LN0.

In addition, there is one retrospective analysis of an RCT:

- Stockholm Tamoxifen (STO-3) trial (Esserman 2016,[14] Lindstrom 2015,[15] van 't Veer 2017,[16] company submission[17]): LN0 patients receiving no chemotherapy.

A number of additional analyses pooled data on patients with specific characteristics from two or more of the above cohorts, as follows:

- Mook *et al.* (2010)[18] pooled 964 patients from seven series and reports prognostic performance;[2, 5-7, 10, 11, 19] patients are a mix of LN0 and LN+, and the analysis is restricted to T1 patients (tumour ≤2cm) which means that a higher proportion of patients are ER+ than in the original analyses. The analysis included six series in which MammaPrint did not influence treatment, plus one study (RASTER)[19-21] in which patients were treated according to usual practice plus MammaPrint.
- Knauer *et al.*[22] (2010) pooled 541 patients from six of seven series above (LN0 or LN1-3) and reports whether MammaPrint predicts benefit from chemotherapy (Chapter 2, Chemotherapy benefit: MammaPrint). Again, this analysis included the RASTER observational study.[19-21]
- Bueno-de-Mesquita *et al.*[23] (2011) pooled 139 ER+ LN0 untreated patients from two series[2, 5]
- Beumer *et al.*[24] (2016) pooled patients with lobular breast cancer from five series.[2, 11, 19, 25]

Tests and comparators: MammaPrint prognostic performance

All prognostic studies used the MammaPrint 70-gene microarray. The majority used frozen tumour samples, while FFPE samples were used in the STO-3 trial,[14-16] and both frozen and FFPE samples were used in the USA series (Yao *et al*., 2015[12]) (Table 1). Patients were categorised as low-risk (or good prognosis) and high-risk (or poor prognosis).

None of the MammaPrint analyses included other in-scope tests (except for some of the whole-transcriptome microarray studies; see Appendix 5). Comparators for prognostic studies included AOL and NPI.

Quality assessment: MammaPrint prognostic performance

All data sets included for prognostic performance were validation studies (Table 2), though the Van de Vijver 2002[2] cohort included a small proportion of patients from the derivation set (Van 't Veer *et al*. 2002) which may lead to overestimation of prognostic performance, though a "leave-one-out" analysis was used to mitigate to some extent this problem.[1] Most analyses excluded some patients recruited to the original trial or cohort, or this was unclear. Blinding of test assessors to outcomes was reported in around half the studies. Outcomes did not always match standardised defintions; several described analyses of distant metastases but were not clear whether all deaths and breast cancer deaths were counted as events or were censored, which makes it difficult to know whether the analyses were of DRFS or DRFI.[5, 7, 8, 16, 22-24] As noted above, many studies were retrospective analyses of patient series of whom some received chemotherapy in accordance with usual practice; the corresponding different levels of chemotherapy use in the high- and low-risk groups may confound results. Additionally, retrospective selection of cohorts who did (or did not) have chemotherapy may introduce spectrum bias since these patients may be systematically different to the whole population. In addition, many studies included a proportion of patients who were out of scope (ER- and/or HER2+ and/or >3 positive nodes).

Results: MammaPrint prognostic performance

Prognostic data for MammaPrint is provided in Tables 10 and 11 of the main report.

Distribution of patients by risk group

For LN0 patients, the percentage of patients categorised as low-risk varied widely: 20% to 71% across seven analyses[5-9, 13, 14, 16] (Table 10 of the main report). For LN+ patients, 38% and 41% were categorised as low-risk in two analyses.[10, 13] A further analysis of LN0 patients showed that, of those who were low clinical risk (via three tools: AOL, NPI and St Gallen), 77% were MammaPrint low-risk; conversely, of those at high clinical risk, only 27% were MammaPrint low-risk.[23]

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"

Mix of LN0/+ patients with varying endocrine and chemotherapy use: Two unadjusted analyses pooled six or seven European validation series; both showed MammaPrint to be significantly prognostic for DRFS/DRFI and BCSS. Mook *et al.*[18] (2010) pooled 964 patients from seven series[2, 5-7, 10, 11, 19] (84% ER+, varying levels of chemotherapy and endocrine therapy). MammaPrint was significantly prognostic for 10-year DRFS (HR 2.70 (95% CI 1.88 to 3.88, p<.0001); Table 10 of the main report) and BCSS (HR 4.22 (95% CI 2.70 to 6.60, p<0.001); Table 3), with 10-year DRFS rates in the low-risk group of 87% at 10 years (Table 10 of the main report). Knauer *et al.*[22] (2010) pooled 541 patients from six of these series (restricted to LN0-3 patients, all had endocrine therapy and 42% chemotherapy). MammaPrint was again significantly prognostic for 5-year DRFS and BCSS, with 95% DRFS in the low-risk group at 5 years (no data for later follow-up). Separate results for ER+ patients from three of the above series were reported by Kok *et al.* (2012);[11] MammaPrint was significantly prognostic for 10-year BCSS among patients pooled from two series[2, 6] (all ER+, 91% LN0, no adjuvant treatment, HR 4.52 (95% CI 2.01 to 10.2, p<0.001)) and also from NKI patients[11] (all ER+, 82% LN+, all endocrine-treated, HR 2.78 (95% CI 1.30 to 5.94, p=0.008)) (Table 4).

In terms of longer follow-up, 25-year follow-up[13] of the initial van de Vijver (2002)[2] cohort (51% LN0; 37% had chemotherapy and 14% endocrine therapy) reported that MammaPrint was statistically significantly prognostic for unadjusted analyses of DRFS for the whole 0-25 year period (HR 3.1 (95% CI 2.02 to 4.86, p<0.001)); however, most of this difference was seen in the first 5 years (HR 9.6, 95% CI 4.2 to 22.1), with subsequent individual 5-year bands from 5-10 years to 20-25 years not showing a statistically significant difference in DRFS between risk groups (Table 10 of the main report). Results for OS showed a similar pattern, with a statistically significant prognostic effect for years 0-5 and 0-25 (p<0.0001); there was also a statistically significant difference in years 5-10 for OS (p=not reported; Table 3). A separate USA series (Yao *et al.* 2015,[12] 72% LN0, 43% had chemotherapy and 87% endocrine therapy) also showed statistically significant prognostic ability for DRFS at 10 years (HR 2.91 (95% CI 0.97 to 8.68), p=0.045, Table 10 of the main report) with DRFS rates in the low-risk group of 96% at 10 years; results were similar (low-risk 10-year DRFS 98%) in a subset with no chemotherapy.

LN0: Four of five retrospective LN0 cohorts (all having varying levels of endocrine and chemotherapy) assessing the prognostic ability of MammaPrint reported statistically significant prognostic performance in unadjusted analyses.[2, 5-7] The exception was one study of 100 US patients

(Wittner, 2008)[9] in which MammaPrint was not statistically significantly prognostic for DRFI (p=0.330 at 10 years; HR NR). In the van de Vijver 2002[2] cohort (age ≤52 years), MammaPrint was statistically significantly prognostic for DRFS (Table 10 of the main report) and OS (Table 3) over years 0-10[5] and years 0-25[13] (HRs range from 4.6 to 10.7, all p<0.001). In the Bueno-de-Mesquita 2009 cohort (age <55 years),[5] MammaPrint was statistically significantly prognostic for DRFS (HR 5.7 (95% CI 1.6 to 20, p=0.007)) and OS (HR 3.4 (95% CI 1.2 to 9.6, p=0.021)) at 5 years. In Mook 2010 (age 55-71 years),[6] MammaPrint was statistically significantly prognostic for 5-year DRFS (4.6 (95% CI 1.8 to 12.0, p=0.01) and BCSS (HR 19.1 (95%CI 2.5 to 148, p=0.005), though 10 year outcome data were available but no statistical significance levels were reported (Table 10 of the main report; Table 4). In TRANSBIG (Buyse 2006[7]), for all follow-up (median 13.6 years), MammaPrint was statistically significantly prognostic for DRFI (HR 2.32 (95% CI 1.35 to 4.00, p=0.002), OS (HR 2.79 (95% CI 1.60 to 4.87, p<0.001) and BCSS (HR 1.50 (95% CI 1.04 to 2.16, p=0.032). In addition, the STO-3 trial (van 't Veer 2017[16]) reported 10-year DRFS rates (93% in low-risk; 85% in high-risk; Table 10 of the main report) but no statistical significance levels were reported.[16] An additional analysis was provided by the company in confidence and cannot be reported here. 5-year DRFS was also reported for a Japanese cohort (Ishitobi *et al*., 2010[8]), with 5-year DRFS of 100% for low-risk patients and 94% for high-risk; however, no statistical significance levels were reported (Table 10 of the main report).

Patient outcomes may vary by receipt of chemotherapy and endocrine therapy. In low-risk patients, 10-year DRFS rates were 88% in a pooled analysis of patients receiving no chemotherapy or endocrine therapy from the van de Vijver[2] and Bueno-de-Mesquita[5] cohorts (Bueno-de-Mesquita 2011[23]); 86% in van de Vijver 2002[5] (4% chemotherapy, 4% endocrine therapy); 80% in Mook 2010[6] (no chemotherapy, 18% endocrine therapy); and in the STO-3 trial (van 't Veer 2017,[16] ER+ patients), 10-year DRFS was 93% with endocrine monotherapy and 83% without endocrine or chemotherapy, while 10-year DRFI was 90% in TRANSBIG (no chemotherapy or endocrine therapy).[7]

Three LN0 cohorts included comparisons to clinical risk tools (AOL and NPI), which appeared to have less prognostic value than MammaPrint, though there were no comparisons available for some in-scope comparators (such as PREDICT or modified AOL). NPI was statistically significantly prognostic for 10-year DRFS and OS (both p<0.001)) in the van de Vijver 2002 cohort,[2, 5] but was not statistically significantly prognostic for 5-year DRFS (p=0.14)) and borderline non-significant for 5-year OS (p=0.053) in the Bueno-de-Mesquita 2009 cohort,[5] and was statistically significantly prognostic for DRFI (p=0.043) but not OS (p=0.092) or DFS (p=0.58) in TRANSBIG[7] (all follow-up; Table 10 of the main report, Table 3 and Table 4). AOL was statistically significantly prognostic for 10-year OS (p=0.017) but not DRFS (p=0.14) in the van de Vijver 2002 cohort,[2, 5] but was not

statistically significantly prognostic for 5-year DRFS (p=0.14) or OS (p=0.22) in the Bueno-de-Mesquita 2009 cohort,[5] nor for DRFI (p=0.092), OS (p=0.085) or BCSS (p=0.092) in TRANSBIG.[7]

LN+: Two cohorts reported separate results for LN+ patients, both with varying endocrine and chemotherapy use; both showed statistically significant prognostic performance of MammaPrint.[2, 10] In the van de Vijver 2002[2] cohort (in which a quarter had more than 3 positive nodes), MammaPrint was statistically significantly prognostic for DRFS (HR 2.24 (95% CI 1.25 to 4.00, p=0.01)) and OS (HR 1.83 (95% CI 1.07 to 3.11, p=0.03) over years 0-25[13] and for 10-year BCSS[10] (HR 6.60 (95% CI 1.97 to 22.10, p=0.002)) (Table 10 of the main report, Table 3 and Table 4). In the Mook 2009[10] cohort (all LN1-3), MammaPrint was again statistically significantly prognostic for DRFS (HR 4.13 (95% CI 1.72 to 9.96), p=0.002), OS (HR 5.40 (95% CI 2.11 to 13.80, p<0.001)) and BCSS (HR 5.70 (95% CI 2.01 to 16.23, p=0.001)) over 0-10 years. In both cohorts, some patients received chemotherapy, though results remained statistically significant in a subgroup of patients not receiving chemotherapy in Mook 2009[10] (only reported for BCSS, HR 7.33 (95% CI 1.61 to 33.49, p=0.01); Table 4). In low-risk patients, 10-year DRFS rates were 79% in van de Vijver 2002[5] (rates of adjuvant treatment not reported) and 91% in Mook 2009[10] (56% chemotherapy, 73% endocrine therapy).

Low or high clinical risk: Patients at low- or high-risk via three clinical tools (AOL, NPI and St Gallen) were assessed in a pooled analysis of LN0 untreated patients from two series[2, 5] (Bueno-de-Mesquita *et al.*,[23] 2011; Table 11 of the main report). Patients with all-low clinical risk according to all three clinical tools showed a statistically significant prognostic effect of MammaPrint on 10-year OS (HR NR, p=0.016) but not DRFI (HR NR, p=0.19), though 10-year DRFI was numerically more favourable in the MammaPrint low-risk group (87%) than in the high-risk group (70%). Patients with all-high clinical risk did not show a statistically significant effect on either OS (HR NR, p=0.17) or DRFI (HR NR, p=0.19), and had relatively poor 10-year DRFI even in the MammaPrint low-risk group (77%) though this was numerically more favourable than in the high-risk group (45%). In a separate analysis, LN+ patients (LN1-3) at high clinical risk via AOL in Mook 2009[10] showed a statistically significant prognostic effect of MammaPrint on 10-year BCSS (HR 4.12 (95%IC 1.45 to 11.76, p=0.008)); Table 11 of the main report). Statistical significance levels in this analysis may have been affected by the small sample sizes per subgroup.

Lobular breast cancer: A pooled analysis of patients with invasive lobular breast cancer from five series[2, 11, 19, 25] (Beumer *et al.*,[24] 2016) showed that MammaPrint was statistically significantly prognostic for 10-year DRFS (HR : 3.31 (95%CI 1.79 to 6.12, p<0.001)) and OS (HR 3.58 (95% CI 1.84 to 6.95, p<0.001)) in all patients (34% LN+) and in a sub-analysis of LN0 patients (DRFS HR 7.81 (95% CI 2.89 to 21.07, p<0.001); OS HR 7.47 (95% CI 2.58 to 21.58, p<0.001)), Table 10 of the main report and Table 3).

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.

Among mixed LN0/+ cohorts, the van de Vijver 2002 cohort reported that MammaPrint was statistically significantly prognostic for 10-year DRFS (HR 4.6 (95% CI 2.3–9.2, p<0.001) in a multivariable analysis which included age, lymph node status, tumour size, grade, vascular invasion, ER status, surgery type, chemotherapy and endocrine therapy. In the pooled analysis of seven series by Mook *et al.*[18] (2010), which incorporated some or all of the van de Vijver 2002[2] cohort, MammaPrint was also statistically significantly prognostic for 10-year DRFS (HR 2.43 (95% CI 1.56 to 3.77, p<0.001) and BCSS in a multivariable analysis adjusted for age, tumour size, nodal status, grade, ER, HER2, surgery, endocrine therapy and chemotherapy (p<0.001; Table 12 of the main report and Table 6). However, in the USA series (Yao *et al.* 2015,[12]), MammaPrint prognostic value for 10-year DRFS was borderline statistically significant in the unadjusted analysis (p=0.045, Table 12 of the main report) and borderline non-statistically significant in a multivariable analysis (HR 3.01 (95% CI 0.88 to 10.33, p=0.08, Table 12 of the main report).

Among LN0 patients, MammaPrint remained statistically significantly prognostic for distant recurrence when adjusted for either AOL or NPI in three cohorts: for 10-year DRFI in van de Vijver 2002 (p=0.001),[2, 5, 7] for 5-year DRFI in Bueno-de-Mesquita 2009[5] (p=0.02), and for DRFI (all follow-up) in TRANSBIG[7] (p=not reported) (Table 12 of the main report). C-indices (reported as AUC) were reported by Bueno-de-Mesquita 2009[5] for both cohorts (Bueno-de-Mesquita[5]; van de Vijver 2002[2]) and showed a higher value (0.75 (95% CI 0.61 to 0.89) and 0.76 (95% CI 0.68 to 0.85) respectively) for MammaPrint and clinicopathological factors (age, tumour size, grade, ER, PR, HER2) than for either the factors on their own, or MammaPrint on its own, though differences were not statistically compared (Table 12 of the main report). For OS (Table 5), MammaPrint remained statistically significantly prognostic in van de Vijver 2002[2, 5, 7] at 10-year when adjusted for AOL or NPI (p<0.001), in TRANSBIG[7] (all follow-up) when adjusted for AOL or NPI (p=not reported), and in Bueno-de-Mesquita 2009[5] at 5-year when adjusted for AOL (p=0.044), but not NPI (p=0.086). C-indicies reported by Bueno-de-Mesquita 2009[5] for OS showed the same trends as for DRFI (data not shown). For other outcomes, MammaPrint remained statistically significantly prognostic for 5-year BCSS in Mook 2010[6] when adjusted for AOL (p=0.01) and for DFS (all follow-up) in van de Vijver 2002[2, 7] when adjusted for AOL (HR 4.80 (95%CI 2.37 to 9.71, p not reported)) but not for DFS in TRANSBIG[7] when adjusted for AOL or NPI (p=not reported) (Table 6).

Among LN+ patients, MammaPrint was statistically significantly prognostic for 10-year BCSS (HR 7.17 (95% CI 1.81 to 28.43, p=0.005), Table 6) but borderline significant for 10-year DRFS (2.99 (95% CI 0.996 to 8.99, p=0.051), Table 12 of the main report) in Mook 2009[10] when adjusted for age, tumour size, nodal status, grade, ER, HER2, surgery, endocrine therapy, chemotherapy. MammaPrint was borderline non-statistically significantly prognostic for 10-year BCSS in van de Vijver 2002[2, 10] (HR 3.63 (95% CI 0.88 to14.96, p=0.07)) when adjusted for the same variables.

Among lobular breast cancer patients, MammaPrint was statistically significantly prognostic for 10-year DRFS (p=0.037 in all patients and p=0.001 in LN0; Table 12 of the main report) but not statistically significant for 10-year OS (p=0.070 in all patients and p=0.008 in LN0) when adjusted for age, nodal status, grade, ER, HER2, and chemotherapy.

Chemotherapy benefit: MammaPrint

Study designs and patients

Two references have reported the ability of MammaPrint to predict the benefit of chemotherapy, i.e. whether the relative effect of chemotherapy differs between MammaPrint risk groups. The article by Knauer *et al.*[22] (2010) reported a pooled analysis of 541 patients, of whom 100% received endocrine therapy and 42% recevied chemotherapy, from six consecutive patient series as detailed in Table 7. Overall, 90% were ER+, 89% HER2-, and half were LN0 while half had 1-3 positive nodes (LN1-3). This publication did not report separate analyses for LN0 and LN+ groups.

Additionally, the article by Mook *et al.*[10] (2009) reported a pooled analysis of two of the six patient series from Knauer *et al.*[22] (Table 7), with an extended follow-up (10 years), but restricted to LN1-3 patients (including micrometastases).

Quality assessment

Table 8 presents the quality assessment of studies assessing MammaPrint prediction of chemotherapy benefit. There were no reanalyses of RCTs assessing chemotherapy benefit. Both studies used pooled retrospective cohorts, where patients were treated according to usual practice (in addition, one of the six cohorts in Knauer[22] was the prospective RASTER study[19] where patients were treated according to usual practice plus MammaPrint). As such, those who received chemotherapy are likely to be systematically different in terms of known (and potentially unknown) prognostic and clinical factors (e.g. age, nodal status) to those who did not, leading to a high risk of confounding. Both studies blinded the test assessors to clinical outcomes, and both used standard outcome definitions. Both studies included a proportion of patients outside the scope (ER- and/or HER2+).

Results

The pooled analysis of six consecutive series by Knauer *et al.*[22] (2010) reported that at 5 years, there was a statistically significant effect of chemotherapy in the MammaPrint high-risk group but no statistically significant effect in the low-risk group, though HRs favoured chemotherapy in both groups (Table 13 of the main report). Unadjusted HRs for DRFS (for no chemotherapy vs. chemotherapy) were 0.26 (95% CI: 0.03, 2.02, p=0.20) in the low-risk group and 0.35 (95% CI: 0.17, 0.71, p<0.01) in the high-risk group, while unadjusted HRs for BCSS were 0.58 (95% CI: 0.07, 4.98, p=0.62) in the low-risk group and 0.21 (95% CI: 0.07, 0.59, p<0.01) in the high-risk group. Multivarible analyses of the effect of chemotherapy on 5 year BCSS were again statistically significant in the high-risk group (HR 0.21, 95% CI: 0.06, 0.80, p=0.02) but not the low-risk group (HR not estimable, p=0.98) (Table 13 of the main report). However, the interaction test for chemotherapy treatment and risk group was not statistically significant (p=0.45; the interaction test appears to relate to 5-year BCSS as opposed to DRFS but this is unclear in the publication). This indicates that the effect of chemotherapy versus no chemotherapy on 5-year BCSS was not statistically significantly different between risk groups. It is unclear whether this interaction test relates to the adjusted or unadjusted analysis.

For the two pooled LNmicro-3 cohorts reported by Mook *et al.*, 2009[10] (these were subsets of two of the six cohorts pooled in Knauer *et al.*[22]), the only evidence relating to prediction of chemotherapy benefit was a test of the interaction between chemotherapy treatment and risk group (within a multivariable analysis of 10-year BCSS), which was not statistically significant (p=0.95, Table 13 of the main report).

Discussion: MammaPrint chemotherapy benefit

Prediction of chemotherapy benefit for MammaPrint was reported within a pooled analysis of 541 patients across six patient series (half LN0, half LN1-3).[22] The effect of chemotherapy versus no chemotherapy on 5-year DRFS and BCSS was statistically significant in the MammaPrint high-risk group but not in the low-risk group in unadjusted analyses for 5-year DRFS and BCSS and in analyses for 5-year BCSS adjusted for clinicopathological variables (not reported for DRFS). However, the interaction test for chemotherapy treatment and risk group (for 5 year BCSS) was non-significant (p=0.45). A further pooled analysis of two of the above series, with follow-up to 10 years but restricted to LN1-3 patients, also reported a statistically non-significant interaction between chemotherapy treatment and risk group for 10-year BCSS (p=0.95).[10]

Both studies used pooled retrospective cohorts where patients were treated according to usual practice (or usual practice plus MammaPrint within RASTER,[19] one of the six pooled cohorts). As such, those who received chemotherapy are likely to be systematically different in terms of known (and

unknown) prognostic and clinical factors to those who did not, leading to a high risk of confounding. In the analysis of six series,[22] it was unclear whether the interaction test was unadjusted or adjusted, and if so for which factors. In the analysis of LN1-3 patients from two series,[10] the interaction test was conducted within a multivariable analysis adjusted for clinicopathological variables.

Conclusions: MammaPrint chemotherapy benefit

Prediction of chemotherapy benefit for MammaPrint was reported within a pooled analysis of 541 patients within six non-randomised patient series (half LN0, half LN1-3) in which patients were treated according to usual practice. The effect of chemotherapy versus no chemotherapy on 5-year DRFS and BCSS was statistically significant in the MammaPrint high-risk group but not in the low-risk group in unadjusted analyses for 5-year DRFS and BCSS and in adjusted analyses for 5-year BCSS. However, the interaction test for chemotherapy treatment and risk group (for 5 year BCSS) was non-significant (p=0.45). A further pooled analysis of two of the above series, restricted to LN1-3 patients, also reported a statistically non-significant interaction between chemotherapy treatment and risk group for 10-year BCSS (p=0.95). The evidence for the ability of MammaPrint to predict chemotherapy benefit is extremely limited; although unadjusted analyses suggest a greater effect of chemotherapy in high-risk groups, adjusted analyses were only reported for one outcome, and the non-significant interaction tests suggest there was no statistically significant difference in effect of chemotherapy between risk groups.

Clinical Utility: MammaPrint

Overview

Two studies reported evidence relating to clinical utility of MammaPrint (the impact of prospective use of the test on clinical outcomes). MINDACT is an RCT of MammaPrint versus clinical practice.[26] RASTER[19-21] is a prospective observational study in which patients were treated according to usual practice plus MammaPrint. As these two studies are very different in design, they are reported separately below.

Clinical utility RCT: MINDACT

Study design

MINDACT (Cardoso *et al*., 2016)[26] is a partially-randomised prospective study of MammaPrint versus clinical practice. Patients with discordant risk scores (high/low or low/high) according to MammaPrint and modified AOL (mAOL; included HER2 status) were randomised to chemotherapy or no chemotherapy; importantly, this also means discordant-risk patients were randomised to treatment determined by MammaPrint or treatment determined by mAOL.

Patients with concordant risk were not randomised, but were followed as prospective cohorts. High/high-risk patients (via both MammaPrint and mAOL) were all recommended to receive chemotherapy, while low/low-risk patients were all recommended no chemotherapy.

The primary aim was to determine whether patients who were high-clinical and low-MammaPrint risk could avoid chemotherapy by comparing outcomes for patients randomised to chemotherapy or no chemotherapy. Results were also presented for low-clinical high-MammaPrint patients. Secondary analyses included an analysis of discordant patients according to treatment group (chemotherapy versus no chemotherapy), as well as for all patients when chemotherapy was recommended according to clinical risk or to MammaPrint risk. The percentage of patients assigned to chemotherapy with each strategy was also reported.

Patients and tests

MINDACT enrolled 6693 patients from nine European countries (Table 9). Of these, using ITT analyses, 2634 (39%) were low clinical, low MammaPrint risk and were assigned to no chemotherapy; 1873 (28%) were high clinical, high MammaPrint risk and were assigned to chemotherapy; 1497 (22%) were high clinical, low MammaPrint risk and were randomised to chemotherapy or no chemotherapy; and 690 (10%) were low clinical, high MammaPrint risk and were again randomised.

Of all 6693 patients, 88% were hormone-receptor-positive (HR+) and 90% HER2-. In terms of nodal status, overall 79% were LN0 and 21% LN1-3. However, this varied by group: in the discordant groups, only 52% were LN0 among high clinical, low MammaPrint patients, while 98% were LN0 among low clinical, high MammaPrint patients; in the concordant groups 94% were LN0 in the low-risk concordant group and 74% were LN0 in the high-risk concordant group.

Frozen tumour samples were used, and the MammaPrint 70-gene test was conducted using an FDA-approved MammaPrint whole-transcriptome microarray. Cut-offs were not reported, but were assumed by the EAG to be the same as in previous studies.

Quality assessment

Discordant-risk patients were randomised centrally and randomisation was stratified by institution, risk group, ER, PR, nodal status, age, HER2, axillary treatment, and type of surgery; hence, randomisation sequence and allocation concealment were judged to be low risk of bias. No details of blinding were reported (Table 10).

Intention-to-treat (ITT) and per-protocol analyses were reported. Some patients did not adhere to their recommended chemotherapy or no chemotherapy allocation. Other patients had a change in clinical risk group due to initial incorrect reporting of clinical characteristics, or a change in MammaPrint risk group due to a change in the RNA-extraction solution which affected the calculation of risk group. For ITT, patients were analysed in their originally-allocated clinical/MammaPrint risk groups and in their randomised treatment groups. Per protocol analysis excluded patients who were ineligible, or were non-adherent to chemotherapy recommendations, or had a change in their clinical or MammaPrint risk group. This report uses ITT results (where available).

Results

Adherence to recommended treatment

In the discordant-risk groups, overall adherence to chemotherapy assignment was 86%. Among high clinical, low MammaPrint risk patients, adherence was 85% for chemotherapy and 89% for no-chemotherapy. Among low clinical, high MammaPrint risk patients, adherence was 80% for chemotherapy and 88% for no-chemotherapy. However, results presented here are for the ITT analyses which analyse patients within their allocated groups regardless of adherence.

High clinical, low MammaPrint group

The primary aim was to assess whether patients who were high-clinical (mAOL) but low-MammaPrint risk could avoid chemotherapy, i.e. whether outcomes were similar for chemotherapy versus no chemotherapy. In this group (N=1497; 52% LN0), using ITT analyses, 5-year DMFS was 95.9% (95% CI: 94.0, 97.2) with chemotherapy and 94.4% (95% CI: 92.3, 95.9) without chemotherapy, an absolute difference of 1.5% favouring chemotherapy, though the HR was not statistically significant (adjusted HR 0.78, 95% CI 0.50, 1.21, p=0.267). Similar differences between chemotherapy and no chemotherapy were reported for 5-year DMFI, DFS and OS, as well as among both LN0 and LN1-3 patients and a LN0 HR+ HER2- subgroup (Table 14 of the main report).

This finding was interpreted by the authors as showing little difference in outcomes for chemotherapy versus no chemotherapy, implying that patients who were high-clinical but low-MammaPrint risk could potentially avoid chemotherapy. Statistically, this met the primary objective in that the lower bound of the 95% CI for 5-year DMFS in the no-chemotherapy group was at least 92% (this lower bound was 92.3% in the ITT analysis and 92.5% in the per protocol analysis).

Low clinical, high MammaPrint group

Results were also presented for the low-clinical (mAOL) high-MammaPrint risk group (Among these patients (N=690; 98% LN0), again using ITT data, 5-year DMFS was 95.8% (95% CI: 92.9, 97.6) with chemotherapy and 95.0% (95% CI: 91.8, 97.0) without chemotherapy, an absolute difference of

0.8% (adjusted HR 1.17, 95% CI: 0.59, 2.28, p=0.657). This finding, though again showing little difference in outcomes between chemotherapy and no chemotherapy, has quite a different interpretation. Given that low clinical risk patients could be assumed (in general) to not be recommended chemotherapy in current practice, these results imply that low-clinical risk patients with a high-risk MammaPrint result still have little benefit from chemotherapy, implying that MammaPrint should not be used to guide treatment in low clinical risk patients as it would result in patients receiving chemotherapy but not gaining any benefit.

Non-randomised concordant-risk groups

In terms of outcomes for the non-randomised groups, patients with low/low-risk (recommended no chemotherapy) had a 5-year DMFS of 97.6% (95% CI: 96.9, 98.1), i.e. slightly more favourable than the discordant groups. Conversely, patients with high/high-risk (recommended chemotherapy) had a 5-year DMFS of 90.6 (95% CI: 89.0, 92.0), i.e. slightly less favourable than the discordant groups. Results for DFS and OS followed a similar pattern (Table 14 of the main report).

Estimated outcomes according to clinical and MammaPrint treatment strategies

Results were also reported for analyses, firstly assuming that chemotherapy recommendations were determined by clinical risk, and secondly by MammaPrint risk (Table 11). Both these analysis included all concordant-risk patients (low/low, recommended no chemotherapy, and high/high, recommended chemotherapy). Of the discordant-risk patients, the clinical strategy only included the clinical high, MammaPrint low patients who were randomised to chemotherapy and the clinical low, MammaPrint high patients who were randomised to no chemotherapy (and vice versa for the MammaPrint strategy; see Table 11). Since half of randomised patients were excluded from each analysis, the remaining discordant patients were double-weighted; the outcomes are therefore described as “estimated”.

The 5-year DMFS for both strategies were very similar: 95.0% for the clinical strategy and 94.7% for the MammaPrint strategy (95% CIs not reported). This was interpreted as the MammaPrint strategy leading to little difference in outcomes even though fewer patients had chemotherapy (see below). However, any potential difference between treatment according to the MammaPrint or clinical strategy in the discordant group could be considered to be “diluted” by the concordant-risk groups who had the same treatment and outcomes with either strategy. This analysis also assumes that in the MammaPrint strategy, all patients would be treated according to MammaPrint, whereas the results above indicate this may not be justified for low-clinical high-MammaPrint patients.

Reclassification of patients via clinical or MammaPrint risk (and implications for chemotherapy)

Of all 6693 patients, 3356 (50%) overall were high clinical risk via mAOL, while 2398 (36%) were high MammaPrint risk (Table 9). Therefore, overall, 14% fewer (958/6693) were categorised as high-risk via MammaPrint than mAOL. Of those at high clinical risk, 46% (1550/3356) could be reclassified to low-risk by MammaPrint.

Multivariable analysis

In a multivariable analysis adjusted for chemotherapy use, clinical risk, and patient and tumour characteristics, MammaPrint low/high-risk grouping was statistically significantly associated with 5-year DMFS (HR for high vs low-risk 2.41, 95% CI: 1.79, 3.26, p<0.001).

Discussion: RCT of clinical utility for MammaPrint (MINDACT)

One RCT assessed the clinical utility of MammaPrint. In MINDACT (total N=6693),[26] patients with discordant risk scores via MammaPrint and mAOL were randomised to chemotherapy or no chemotherapy, while patients with concordant high-risk were recommended chemotherapy and those with concordant low-risk were recommended no chemotherapy. The primary aim was to determine whether patients who were high-clinical but low-MammaPrint risk could avoid chemotherapy. In this group (N=1550; 52% LN0), 5-year DMFS was 95.9% (95% CI: 94.0, 97.2) with chemotherapy and 94.4% (95% CI: 92.3, 95.9) without chemotherapy, an absolute difference of 1.5% (adjusted HR 0.78, 95% CI 0.50, 1.21, p=0.267). This finding was interpreted by the authors as suggesting that these patients could avoid chemotherapy. Clinical advice to the EAG suggests that chemotherapy would usually only be indicated where it is likely to provide an absolute improvement in 5-year DRFS of 2%-3%, which suggest that it may be reasonable to withhold chemotherapy in patients with high-clinical low-MammaPrint risk given the above absolute difference in 5-year DRFS of 1.5% for chemotherapy vs. no chemotherapy.

In patients who were low-clinical but high-MammaPrint risk (N=592; 98% LN0), 5-year DMFS was 95.8% (95% CI: 92.9, 97.6) with chemotherapy and 95.0% (95% CI: 91.8, 97.0) without chemotherapy, an absolute difference of 0.8% (adjusted HR 1.17, 95% CI: 0.59, 2.28, p=0.657). This finding could be interpreted as showing that use of MammaPrint in low clinical risk patients could lead to more patients being prescribed chemotherapy, but not receiving a survival benefit from treatment. Additional analyses assessed strategies in which chemotherapy recommendations for all patients were determined by either clinical risk or MammaPrint risk. These included concordant (non-randomised) and discordant (randomised) patients who had treatment that matched either their clinical risk (treatment determined by clinical risk group) or MammaPrint risk (treatment determined by MammaPrint risk group). The 5-year DMFS was very similar: 95.0% for clinical strategy and 94.7% for MammaPrint strategy. This was interpreted as the MammaPrint strategy leading to little difference

in outcomes while sparing many patients from chemotherapy (of those at high clinical risk, 46% were MammaPrint low-risk and could potentially be spared chemotherapy). Given the results in the low clinical risk group (where treatment according to MammaPrint risk groups would result in more patients receiving chemotherapy but with no DMFS advantage), the most advantageous strategy may be to only test clinical high-risk patients with MammaPrint. However, the comparator in this study was mAOL, and it is unclear whether the same would be true for other clinical risk scores.

Conclusions: RCT of clinical utility for MammaPrint (MINDACT)

MINDACT randomised patients with discordant MammaPrint and mAOL risks to chemotherapy or no chemotherapy. For patients who were high-clinical, low-MammaPrint risk, 5-year DMFS was 95.9% with chemotherapy and 94.4% without chemotherapy, an absolute difference of 1.5%. This raises the possibility of avoiding chemotherapy in these patients. In patients who were low-clinical, high-MammaPrint risk, 5-year DMFS was 95.8% with chemotherapy and 95.0% without chemotherapy, an absolute difference of 0.8%. This could be interpreted as showing that MammaPrint may not be useful in this group as it would increase chemotherapy rates without improving outcomes. However, the comparator was mAOL, and it is unclear whether the same would be true for other clinical risk scores.

Clinical utility observational study: RASTER

Study design

RASTER (Drukker *et al*., 2013;[20] Drukker *et al*., 2014;[21] Bueno-de-Mesquita *et al*., 2007;[19] Vliek 2017a[27]) is a prospective observational study in which LN0 patients in the Netherlands were treated according to MammaPrint plus usual clinical practice (Dutch Institute of Healthcare Improvement (CBO) guidelines of 2004[28] and clinician and patient preference). The aims were to assess the impact of MammaPrint on treatment decisions and to prospectively record outcomes for patients categorised as high or low-risk via MammaPrint, via clinical risk tools, and for various combinations of MammaPrint risk and clinical risk. An additional analysis conducted retrospectively in LN+ patients was reported separately (Vliek 2017b;[27]).

In the prospective observational study of LN0 patients, receipt of chemotherapy was guided by MammaPrint in combination with the Dutch Institute of Healthcare Improvement (CBO) guidelines of 2004[28] and clinician and patient preference. As such, estimates of prognostic performance (HRs between groups; c-indices) are confounded by the differing rates of chemotherapy in different risk groups (usually more chemotherapy in the high-risk group compared with the low-risk group). Estimates of the impact of the test on clinical outcomes (DRFI, DRFS and OS rates) and chemotherapy use for MammaPrint reflect the use of MammaPrint in routine clinical practice in conjunction with the CBO guidelines, rather than MammaPrint on its own. Conversely, estimates for

other risk tools (NPI, Predict, AOL) are confounded by differential rates of chemotherapy in each risk group, and cannot be used to estimate the impact of those tests on clinical outcomes, but can provide some estimate of prognostic performance, albeit confounded by treatment.

Patients and tests

RASTER assessed prospective use of MammaPrint in 427 LN0 patients, age <61 years, of whom 80% were ER+ and 84% HER2- (Table 13).[20, 21] In addition, MammaPrint was conducted retrospectively for 164 LN+ patients (Vliek 2017b;[27]). Frozen tumour samples were used[20] (except in the retrospective analysis of LN+ patients where FFPE samples were used[27]). The MammaPrint 70-gene microarray was used, stating that cut-offs were the same as in previous studies.[20]

Quality assessment

Since RASTER was not an RCT, it was judged to be at a high risk of bias using standard RCT criteria (

Table **14**).

Results for LN0 patients

Results for MammaPrint (in conjunction with CBO guidelines and patient/clinician preference): Of all 427 LN0 patients in RASTER, MammaPrint was low-risk in 51% (of whom 15% received chemotherapy) and high-risk in 49% (of whom 81% received chemotherapy). At 5 years, DRFI was 97.0% for low-risk and 91.7% for high-risk (p=0.03 between groups, HR NR; Table 15 of the main report).[20, 21] The 10-year DRFI was 93.7% for low-risk and 86.8% for high-risk patients; HR 1.4 (95% CI: 1.0, 1.9). Results at 10 years were similar within the 342 ER+ patients, though not statistically significant (conference abstract by Vliek 2017a;[27] Table 15 of the main report). 5-year overall survival was not statistically significantly different between MammaPrint groups (p=0.35, HR NR; Table 15).[20, 21]

Results for clinical risk tools: MammaPrint results were compared against various clinical risk tools applied retrospectively to the data (Table 15 and Table 16 of the main report). Both NPI and PREDICT Plus categorised approximately the same number of patients into the high-risk groups (42% and 47% respectively) as did MammaPrint (49%), and chemotherapy rates in high-risk groups for NPI and PREDICT Plus (84% and 78% respectively) were similar to MammaPrint (81%). Likewise, 5-year DRFI rates in the low-risk groups for NPI and PREDICT Plus (96.7% and 96.8% respectively) were similar to MammaPrint (97.0%), and likewise 5-year DRFI rates in the high-risk groups for NPI and PREDICT Plus (91.3% and 91.7% respectively) were similar to MammaPrint (91.7%). Both NPI and PREDICT Plus showed a significant difference between groups (p=0.03 and p=0.004).[20, 21]

Conversely, AOL categorised more patients as high-risk (69%) than did MammaPrint, NPI or PREDICT Plus, and high-risk AOL patients had a lower chemotherapy rate (60%). 5-year DRFI was similar for the low-risk group (96.7%) but not so much reduced in the high-risk group (93.4%) as for MammaPrint, NPI or PREDICT Plus, and the difference between groups for AOL was not statistically significant (p=0.24; Table 15 of the main report).[20, 21] Interestingly, mAOL categorised similar numbers of patients as high/low-risk as did MammaPrint, NPI and PREDICT Plus. The 10-year DRFI for mAOL was more favourable for the low-risk than the high-risk group, but this was not statistically significant (HR 1.4, 95% CI: 0.8, 2.6) and any difference was lost when restricting to ER+ patients (Table 15 of the main report; Vliek 2017a).[27]

MammaPrint results for patients at high/low clinical risk: Also presented were results by MammaPrint risk group for patients at a high or low clinical risk. The high clinical risk group is particularly of interest to determine whether patients with a high-clinical low-MammaPrint result could safely avoid chemotherapy. Within patients who were high-risk via NPI or PREDICT Plus or modified AOL,[27] 25% of each (same percentage in all cases) were MammaPrint low-risk (of whom 41-57% received chemotherapy) while 75% were MammaPrint high-risk (of whom 91-93% received chemotherapy). Within NPI and PREDICT Plus high-risk patients, 5-year DRFI for MammaPrint low-risk was 95.5% and 93.9%, while for MammaPrint high-risk it was 89.9% and 91.0%, respectively (Table 15 of the main report; no p-values reported).[20, 21]

Conversely, AOL categorised more patients as high-risk than did NPI or PREDICT Plus. Of these, a higher proportion fell into the MammaPrint low-risk group (42%), in which chemotherapy rates were lower (24%) and 5-year DRFI higher (98.4%). Of 117 AOL-high-risk patients who received no chemotherapy, 80% were MammaPrint low-risk, and 5-year DRFI for these MammaPrint low-risk patients was 98.9%.[20, 21] However, no such data are reported for NPI or PREDICT Plus, which categorise fewer patients as high-risk.

Of patients at low clinical risk, 5-year DRFI for MammaPrint low-risk patients ranged from 95.3% to 98.0% (Table 15 of the main report),[20, 21] whilst for MammaPrint high-risk patients 5-year DRFI ranged from 93.9% to 100%, though it should be noted that high-risk patients had more chemotherapy (57-59%) than low-risk patients (3-8%).

Additional prognostic value of MammaPrint:

Table 16 of the main report shows C-indexes (AUC) for clinical risk tools alone and in addition to MammaPrint. The addition of MammaPrint to AOL or NPI statistically significantly increased the C-index (AUC) (p=0.03 and p=0.05 respectively), while the addition of MammaPrint to PREDICT Plus did not statistically significantly increase the C-index (AUC) (p=0.27; Table 16 of the main report).[21]

Results for LN+ patients

A conference poster by Vliek *et al*.[27] (2017b) reported results for 164 LN+ patients followed up in RASTER, for whom MammaPrint was retrospectively conducted (Table 17 of the main report). Over 95% of patients received chemotherapy. MammaPrint categorised 48% of LN1-3 patients as low-risk. The 5-year DRFI was 98.4% for low-risk and 86.9% for high-risk patients, while 10-year DRFI was 94.9% for low-risk and 80.7% for high-risk patients, showing a statistically significant difference between groups (HR 4.7; 95% CI: 1.3, 16.2). A comparison was made to modified AOL, though this analysis included 30 additional patients with LN>3 who were automatically classed as high-risk. Modified AOL categorised only 14% as low-risk; 10-year DRFI was 94.4% for low-risk and 85.8% for high-risk, which was not statistically significantly different (HR 3.7; 95% CI: 0.5, 28.5). Within mAOL high-risk patients, 10-year DRFI was statistically significantly better in MammaPrint low-risk (95.2%) than high-risk (79.6%) patients (HR 4.8; 95% CI: 1.1, 21.4).[27, 29]

Table 1: Characteristics of prognostic studies: MammaPrint

Reference(s)	Cohort(s)	N	Country	Study design	Test	Test details	Cut-offs	Population	Nodal status	ET / CT
Pooled analyses of patient cohorts: LN status mixed										
Variable ET&CT										
Beumer 2016[24] **Lobular cancer**	Lobular cancers, 5 pooled series: - van de Vijver 2002[2] - Bueno-de-Mesquita 2007[19] (RASTER) - Kok 2012[11] - Michaut 2016[25] (RATHER; NKI, UK) - North Shore & Fox Chase, US	217	Neths, US, UK	Pooled cohorts	**MMP**	Sample type NR MMP microarray	Low, high; details NR	Invasive lobular breast cancer 94% ER+ 92% HER2- % female NR	LN0, 66% LN1-3, 24% LN>3, 9%	59% ET (low 58%, high 62%) 22% CT (low 19%, high 33%)
Knauer 2010[22]	Pooled 6 series: - van de Vijver 2002[2] - Bueno-de-Mesquita 2009[5] - Mook 2009[10] (LN1-3) - Mook 2010[6] (age 55-71) - Bueno-de-Mesquita 2007[19] (RASTER) - Kok (personal com.)	541	Various	Pooling of 6 consecutive cohorts	**MMP**	Frozen MMP microarray	Low, high (details NR)	90% ER+ 89% HER2- Pre/post-meno % female NR pT1-3	LN0, 49% LN1-3, 51%	All ET 42% CT
Mook 2010[18]	Pooled 7 series: - van de Vijver 2002[2] (NKI 84-95) - Bueno-de-Mesquita 2009[5] (NKI+RdGG) - Mook 2009[10] (LN1-3, NKI+Italy) - Mook 2010[6] (age 55-71, NKI) - Bueno-de-Mesquita 2007[19] (RASTER) - Kok 2012[11] (NKI 1985-94) - Buyse 2006[7] (TRANSBIG)	964	Various	Pooling of 7 consecutive cohorts	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	84% ER+ 68% HER2- (23% missing) Pre/post-meno % female NR pT1 (≤2cm)	LN0, 72% LN+, 27% (% LN>3 NR)	32% ET (low 27%, high 38%) 22% CT (low 10%, high 37%)
No ET&CT										
Kok 2012[11]	Pooled 2 series: - van de Vijver 2002[2] - Mook 2010[6] age 55-71	100 + 51	Neths	Two pooled cohorts	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	All ER+ HER2 NR Pre/post-meno % female NR	LN0, 91% LN1-3, 7% LN>3, 2%	No ET/CT

Reference(s)	Cohort(s)	N	Country	Study design	Test	Test details	Cut-offs	Population	Nodal status	ET / CT
Retrospective studies: LN status mixed										
100% ET monotherapy										
Kok 2012[11]	Kok 2009[30] (NKI 1985-94)	121	Neths	1 cohort	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	All ER+ HER2 NR Pre/post-meno % female NR	LN0, 18% LN1-3, 65% LN>3, 18%	All ET, no CT
Variable ET&CT										
Drukker 2014[13] van de Vijver 2002[2]	- van de Vijver 2002[2]	295	Neths	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low >0.4, high <0.4	77% ER+ HER2 NR Age ≤52 100% female	LN0, 51% LN1-3, 36% LN>3, 13%	14% ET (low 15%, high 13%) 37% CT (low 38%, high 37%)
Yao 2015[12]	NorthShore University Health System & Fox Chase Cancer Center (1992-2010)	373 (all) 238 (subgrp)	USA	Retrospective, consecutive	**MMP**	Frozen or FFPE MMP microarray	Low, high; details NR	**All:** 74% ER+ 83% HER2- Stage 1-2b **Subgrp:** All HR+ all HER2- 100% female	LN0, 72% LN1-3, 25% LN>3, 5%	**Subgrp:** 87% ET (low 92%, high 79%) 43% CT (low 37%, high 53%)
Reanalyses of RCTs: LN0										
100% ET monotherapy OR No ET&CT										
van 't Veer 2017[16] Esserman 2016[14] Lindstrom 2015[15]	Stockholm Tamoxifen (STO-3) trial: ER+ subgroup	538	Sweden	Reanalysis of RCT	**MMP**	FFPE MMP microarray	Low >0, high <0	All ER+ 96% HER2- Post-meno % female NR Tumours <30mm	LN0	Analysis 1: All ET, no CT Analysis 2: No ET, no CT
Pooled analyses of patient cohorts: LN0										
No ET&CT										
Bueno-de-Mesquita 2011[23]	Pooled 2 series: - van de Vijver 2002 (NKI, 84-95)[2] - Bueno-de-Mesquita 2009[5] (NKI 96-99)	186	Neths	Pooling of 2 cohorts to form 1 consecutive series	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	76% ER+ 76% HER2- Pre/post-meno 100% female	LN0	No ET No CT

Reference(s)	Cohort(s)	N	Country	Study design	Test	Test details	Cut-offs	Population	Nodal status	ET / CT
Retrospective studies: LN0										
Variable ET&CT										
Bueno-de-Mesquita 2009[5]	1) Bueno-de-Mesquita 2009[5] (NKI+RdGG 1996-99) 2) van de Vijver 2002[2]	1) 123 2) 151	Neths	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	1) 76% ER+ 93% HER2- pT1–2, <55 yr 2) 72% ER+ HER2 NR, pT1–2, age ≤52	LN0	1) 22% ET (low 28%, high 15%); 25% CT (low 16%, high 36%) 2) 4% ET (low 5%, high 3%); 4% CT (low 3%, high 4%)
Buyse 2006[7]	1) TRANSBIG (1980-1999)[7] 2) van de Vijver 2002[2]	1) 302 2) 151	1) France, Sweden, UK 2) Neths	Retrospective cohorts	**MMP**	Frozen MMP microarray	Correlation coeff. low >0.4, high <0.4	1) 70% ER+ HER2 NR, <61yr T1-2 (≤5cm) % female NR 2) 72% ER+ HER2 NR, pT1–2, age ≤52	LN0	1) No ET/CT 2) Some ET/CT
Ishitobi 2010[8]	Osaka Medical Centre (1998-2001)	102	Japan	Retrospective analysis of cases	**MMP**	Frozen MMP microarray	Good (low, if above threshold) or poor (high)	51% ER+ HER2 NR ≤70yrs, T1-3 100% female	LN0	73% ET (low 85%, high 70%) 28% CT (low 10%, high 33%)
Mook 2010[6]	NKI 1984-96[6] (55-71yr)	148	Neths	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	78% ER+ HER2 NR Post-meno, T1-2 100% female	LN0	18% ET No CT
Wittner 2008[9] N=100	Massachusetts General Hospital (1985-1997)	100	USA	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low (good) >0.4, high (poor) <0.4	80% ER+ HER2 NR Pre/post-meno 100% female	LN0	24% ET 21% CT

Reference(s)	Cohort(s)	N	Country	Study design	Test	Test details	Cut-offs	Population	Nodal status	ET / CT
Retrospective studies: LN+										
Variable ET&CT										
Mook 2009[10]	1) NKI+Italy 1994-2001[10] 2) van de Vijver 2002[2]	1) 241 2) 106	Neths, Italy	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	1) 79% ER+ 84% HER2- 2) 82% ER+ 84% HER2- All: Pre/post-meno, age ≤70 % female NR T1-3	LN1-3 (inc. micromets)	1) 73% ET (low 82%, high 65%); 56% CT (low 41%, high 67%) 2) 23% ET (low 26%, high 21%); 70% CT (low 77%, high 65%)

CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; HR+, hormone-receptor positive; LN, number of positive nodes; MMP, MammaPrint; NKI, Netherlands Cancer Institute; NR, not reported; qRT-PCR, quantitative reverse transcription polymerase chain reaction; RASTER, MicroarRAy PrognoSTics in Breast CancER study; RATHER, RAtional THerapy for breast cancer study; RdGG, Reinier de Graaf Hospital; R-RCT, reanalysis of RCT

Table 2: Quality assessment of prognostic studies: MammaPrint

Reference(s); N	Cohort(s)	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)?	Outcome definition standardised or *a priori*?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
Beumer 2016[24] N=217 Lobular cancer	Lobular cancers, 5 **pooled** series[2, 11, 19, 25]	V	N, cohorts, some CT	UC	UC	Y	Most (6% ER-, 8% HER2+, 9% LN>3)	Y
Bueno-de-Mesquita 2011[23] N=139	**Pooled** 2 series: van de Vijver 2002;[2] Bueno-de-Mesquita 2009[5]	V[a]	Y, consecutive cohorts, no CT	UC	Y	Y	Most (all ER+, all LN0, 86% HER2-)	Y
Bueno-de-Mesquita 2009[5] N=123+151	1) Bueno-de-Mesquita 2009[5] 2) van de Vijver 2002[2]	V[a]	N, consecutive cohorts, some CT	N InT	Y	Y	N (24%+7% ER-, 7% HER2 or NR)	Y
Buyse 2006[7] N=302+151	1) TRANSBIG[7] 2) van de Vijver 2002[2]	V[a]	Y, retrospective cohort, no CT	N RNA qual, missing data	UC	Y	N (ER- 30%, HER2 NR)	Y
Drukker 2014[13] van de Vijver 2002[2] N=295	- van de Vijver 2002[2]	V[a] (21% also in derivation set)	N, retrospective, some CT	Y	UC	Y	N (23% ER-, HER2 NR, 13% LN>3)	Y
van 't Veer 2017[16] Esserman 2016[14] Lindstrom 2015[15]	Stockholm Tamoxifen (STO-3) trial: ER+ subgroup	V	Y, reanalysis of RCT, no CT	N InT, TF	UC	Y	Most (HER2 NR)	Y
Ishitobi 2010[8]	Osaka Medical Centre	V	N, case series, some CT	N Lack of RNA, TF	Y	Y	N (49% ER-, HER2 NR)	Y
Knauer 2010[22] N=541	**Pooled** 6 series[2, 5, 6, 10, 11, 19]	V[a]	N, cohorts, some CT	UC	Y	Y	Most (10% ER-, 11% HER2+)	Y
Kok 2012[11] 1) N=121 2) N=100+51	1) Kok 2009[30] 2) **Pooled** 2 series: - van de Vijver 2002[2] - Mook 2010[6] (55-71)	V[a]	Y, consecutive cohorts, no CT	UC	UC	Y	Most (HER2 NR; LN>3 18% (1) and 2% (2))	Y
Mook 2010[18] N=964	**Pooled** 7 series[2, 5-7, 10, 11, 19]	V[a]	N, cohorts, some CT	N TF, MD	Y	Y	UC (16% ER-, 9% HER2+, 23% HER2	Y

Reference(s); N	Cohort(s)	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)?	Outcome definition standardised or *a priori*?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
							unknown, LN>3 % NR)	
Mook 2010[6] N=148	NKI 1984-96[6]	V	Y, consecutive cohort, no CT	N InT, RNA qual, MD	Y	Y	N (22% ER-, HER2 NR)	Y
Mook 2009[10] N=241+106	1) NKI+Italy[10] 2) van de Vijver 2002[2]	V[a]	N, retrospective, 56% + 70% CT	N InT, RNA qual	Y	Y	N (21%+18% ER-, 16% HER2+)	Y
Wittner 2008[9] N=100	Massachusetts, USA	V	N, retrospective, some CT	UC	UC	Y	N (20% ER-, HER2 NR)	Y
Yao 2015[12] N=238	NorthShore & Fox Chase	V	N, retrospective, some CT	UC	Y	Y	Most (for HR+ HER2- subgroup; LN NR)	Y

Y, yes; N, no; UC, unclear

D, Development; InT, insufficient tissue; MD, missing data; MS, missing samples; LN, number of positive nodes; qRT-PCR, quantitative reverse transcription polymerase chain reaction; R-RCT, reanalysis of RCT; TF, test failure; V, validation

[a]van de Vijver 2002[2] included 61 patients from the derivation set

Table 3: Prognostic performance of MammaPrint: Overall survival

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comp.	% pts per group		% OS risk: 0-5 yr		% OS risk: 0-10 yr		OS: HR (95% CI) (unless stated otherwise)
						Low	High	Low	High	Low	High	
Pooled analyses of patient cohorts: LN status mixed												
Variable ET&CT												
Beumer 2016[24] N=217 **Lobular cancer**	Lobular cancers, 5 pooled series[2, 11, 19, 25]	94% ER+ 92% HER2-	LN0, 66% LN+, 34%	59% ET 22% CT	**MMP**	76	24	-	-	-	-	**0-10yr:** 3.58 (1.84, 6.95), p<0.001
		93% ER+ 93% HER2-	LN0	51% ET 12% CT	**MMP**	82	18	-	-	-	-	**0-10yr:** 7.47 (2.58, 21.58), p<0.001
Retrospective studies: LN status mixed												
Variable ET&CT												
Drukker 2014[13] N=295	- van de Vijver 2002[2]	77% ER+ HER2 NR	LN0, 51% LN1-3, 36% LN>3, 13%	14% ET 37% CT	**MMP**	39	61	97.4	74.0	92.8	55.7	**0-5yr:** 11.3 (3.5, 36.4), p=NR **5-10yr:** 6.1 (2.4, 15.6), p=NR **10-15yr:** 1.5 (0.6, 3.5), p=NR **15-20yr:** 0.6 (0.2, 1.7), p=NR **20-25yr:** 0.2 (0, 2.1), p=NR **0-25yr:** 2.9 (1.90, 4.28), p<0.0001
Pooled analyses of patient cohorts: LN0												
No ET&CT												
Bueno-de-Mesquita 2011[23]	Pooled[2, 5] N=186	76% ER+ 76% HER2-	LN0	No ET/CT	**MMP**	45	55	-	-	91	56	
Retrospective studies: LN0												
Variable ET&CT												
Bueno-de-Mesquita 2009[5] N=123	- Bueno-de-Mesquita 2009[5]	76% ER+ 93% HER2-	LN0	22% ET 25% CT	**MMP**	52	48	97	82	-	-	**0-5yr:** 3.4 (1.2, 9.6), p=0.021
					AOL	-	-	-	-	-	-	**0-5yr:** 2.5 (0.59, 11), p=0.22
					NPI	-	-	-	-	-	-	**0-5yr:** 2.8 (0.99. 7.8), p=0.053
Buyse 2006[7] N=302	- TRANSBIG[7]	70% ER+ HER2 NR	LN0	No ET/CT	**MMP**	37	63	-	-	-	-	**All (med 13.6yr):** 2.79 (1.60, 4.87), p<0.001 **C-index (AUC)** 0.648
					AOL							**All (med 13.6yr):** 1.67 (0.93, 2.98), p=0.085 **C-index (AUC)** 0.576
					NPI							**All (med 13.6yr):** 1.49 (0.94, 2.36), p=0.092
Drukker 2014[13] Bueno-de-	- van de Vijver 2002[2]	72% ER+ HER2 NR	LN0	4% ET 4% CT	**MMP**	40	60	96.7[13]	71.1[13]	94[5]	51[5]	**0-10yr:** 10.7 (3.9., 30), p<0.001[5] **0-25yr:** 4.73 (2.46, 9.07); p<0.0001[13]

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comp.	% pts per group		% OS risk: 0-5 yr		% OS risk: 0-10 yr		OS: HR (95% CI) (unless stated otherwise)
						Low	High	Low	High	Low	High	
Mesquita 2009[5] N=151					**AOL**	-	-	-	-	-	-	**0-10yr:** 2.8 (1.2, 6.6), p=0.017[5]
					NPI	-	-	-	-	-	-	**0-10yr:** 3.4 (1.8, 6.6), p<0.001[5]
Retrospective studies: LN+												
Variable ET&CT												
Drukker 2014[13] N=144	- van de Vijver 2002[2]	ER+/- HER2 NR	LN1-3, 74% LN>3, 26%	Some ET/CT	**MMP**	38	62	98.2	76.9	92.5	58.7	**0-5yr and 0-10yr:** HRs not reported **0-25yr:** 1.83 (1.07, 3.11); p=0.03
Mook 2009[10] N=241	- NKI+Italy[10]	79% ER+, 84% HER2-	LN1-3	73% ET 56% CT	**MMP**	41	59	-	-	-	-	**0-10yr:** 5.40 (2.11, 13.80), p<0.001

-, not reported; AOL, Adjuvant! Online; CI, confidence interval; comp, comparator; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; MMP, MammaPrint; NPI, Nottingham Prognostic Index; OS, overall survival.

Table 4: Prognostic performance of MammaPrint: Other outcomes

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or comp.	% pts per group		% risk of outcome: 0-5yr		% risk of outcome: 0-10yr		HR (95% CI)
							Low	High	Low	High	Low	High	
Pooled analyses of patient cohorts: LN status mixed													
Variable ET&CT													
Knauer 2010[22] N=541	Pooled 6 series[2, 5, 6, 10, 11, 19]	90% ER+ 89% HER2-	LN0, 49% LN1-3, 51%	All ET 42% CT	**BCSS**	**MMP**	47	53	97	87	-	-	**0-5 yr:** 4.81 (1.98, 11.67), p<0.01
Mook 2010[18] N=964	Pooled 7 series[2, 5-7, 10, 11, 19]	84% ER+ 68% HER2-	LN0, 72% LN+, 27%	32% ET 22% CT	**BCSS**	**MMP**	54	46	99	88	91	72	**0-10 yr:** 4.22 (2.70, 6.60), p<0.001
		N=552	LN+/-	No ET/CT					99	85	91	69	**0-10 yr:** 4.67 (2.67, 8.18), p<0.001
No ET&CT													
Kok 2012[11] N=100+51	Pooled 2 series: van de Vijver[2] + Mook 2010[6]	ER+ HER2 NR	LN0, 91% LN+, 9%	No ET/CT	**BCSS**	**MMP**	56	44	97.6	80.9	90.2	63.3	**0-10 yr:** 4.52 (2.01, 10.2), p<0.001
Retrospective studies: LN status mixed													
100% ET monotherapy													
Kok 2012[11] N=121	NKI 1985-94[30]	ER+ HER2 NR	LN0, 18% LN+, 82%	All ET No CT	**BCSS**	**MMP**	69	31	96.2	72.5	80.6	63.4	**0-10 yr:** 2.78 (1.30, 5.94), p=0.008
Reanalyses of RCTs: LN0													
Variable ET&CT													
van 't Veer 2017[16] Esserman[14]	STO-3 trial: ER+ analysis N=538	All ER+ HER2 NR	LN0	52% ET No CT	**BCSS**	**MMP**	69	31	-	-	-	-	**0-20 yrs:** p<0.0001
Retrospective studies: LN0													
Variable ET&CT													
Mook 2010[6] N=148	- NKI 1984-96[6] (55-71yr)	78% ER+ HER2 NR	LN0	18% ET No CT	**BCSS**	**MMP**	61	39	99	80	90	69	**0-5 yr:** 19.1 (2.5, 148), p=0.005 **0-10 yr:** 3.9 (CI NR), p=NR
						AOL	50	50	-	-	-	-	**0-5 yr:** 5.3 (CI NR) **0-10 yr:** 6.2 (2.1, 18.0), p=0.001

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or comp.	% pts per group		% risk of outcome: 0-5yr		% risk of outcome: 0-10yr		HR (95% CI)
							Low	High	Low	High	Low	High	
No ET&CT													
Buyse 2006[7] N=302	- TRANSBIG[7]	70% ER+ HER2 NR	LN0	No ET/CT	**DFS**	**MMP**	37	63	-	-	-	-	**All FU (med 13.6yr):** 1.50 (1.04, 2.16), p=0.032
						AOL							1.30 (0.86, 1.95), p=0.21
						NPI							1.10 (0.78, 1.56), p=0.58
Retrospective studies: LN+													
Variable ET&CT													
Mook 2009[10] N=241	- NKI+Italy[10]	79% ER+ 84% HER2-	LN1-3	73% ET 56% CT	**BCSS**	**MMP**	41	59	99	88	96	76	**0-10 yr:** 5.70 (2.01, 16.23), p=0.001
		All ER+ N=191	LN1-3	Some ET/CT			NR	NR	-	-	-	-	**0-10 yr:** 9.75 (2.26, 42.01), p=0.002
		N=101	LN1-3	No CT			NR	NR	-	-	-	-	**0-10 yr:** 7.33 (1.61, 33.49), p=0.01
		N=166	LN1-3	All ET			NR	NR	-	-	-	-	**0-10 yr:** 3.63 (1.21, 10.94), p=0.02
van de Vijver 2002;[2] Mook 2009[10] N=106	- van de Vijver 2002[2]	82% ER+ 84% HER2-	LN1-3	23% ET 70% CT	**BCSS**	**MMP**	41	59	-	-	98	64	**0-10 yr:** 6.60 (1.97, 22.10), p=0.002

-, not reported; AOL, Adjuvant! Online; BCSS, breast cancer-specific survival; CI, confidence interval; CT, chemotherapy; DFS, disease-free survival; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; MMP, MammaPrint.

Table 5: Additional prognostic value for overall survival: MammaPrint

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Multivariable model adj. for CP factors[a], AOL[b] or NPI[c] : HR (95% CI)
Pooled analyses of patient cohorts: LN status mixed								
Variable ET&CT								
Beumer 2016[24] N=217 **Lobular cancer**	Lobular cancers, 5 pooled series[2, 11, 19, 25]	94% ER+ 92% HER2-	LN0, 66% LN+, 34%	59% ET 22% CT	**MMP**			**10yr:** 2.02 (0.94, 4.30), p=0.070[a]
		93% ER+ 93% HER2-	LN0	51% ET 12% CT	**MMP**			**10yr:** 5.10 (1.52, 17.17), p=0.008[a]
Pooled analyses of patient cohorts: LN0								
No ET&CT								
Bueno-de-Mesquita 2011[23] N=186	Pooled[2, 5]	76% ER+ 76% HER2-	LN0	No ET/CT	**MMP**		Change log likelihood p=0.005	
Retrospective studies: LN0								
Variable ET&CT								
Bueno-de-Mesquita 2009[5] N=123	- Bueno-de-Mesquita[5]	76% ER+ 93% HER2-	LN0	22% ET 25% CT				**5yr:** 3.0 (1.0, 8.9), p=0.044[b] 2.7 (0.87, 8.1), p=0.086[c]
van de Vijver 2002;[2] Bueno-de-Mesquita 2009;[5] Buyse 2006[7] N=151	- van de Vijver 2002[2]	72% ER+ HER2 NR	LN0	4% ET 4% CT	**MMP**		Change log likelihood 19.7, p<0.01	**10yr:**[5] 9.6 (3.4, 27), p<0.001[b] 8.5 (2.9, 25), p<0.001[c] **All FU (med 6.7yr):** [7] 17.46 (4.12, 74.00)[b]
No ET&CT								
Buyse 2006[7] N=302	- TRANSBIG[7]	70% ER+ HER2 NR	LN0	No ET/CT	**MMP**			**5yr:** 16.99 (CI NR)[b] **10yr:** 3.46 (CI NR)[b] **All (med 13.6yr):** 2.63 (1.45, 4.79);[b] 2.89 (1.58, 5.29)[c]

AOL, Adjuvant! Online; CI, confidence interval; CP, clinical/pathological; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; LR, likelihood ratio; MMP, MammaPrint; NPI, Nottingham Prognostic Index; OS, overall survival.
[a]Adjusted for: Bueno-de-Mesquita 2011+2009: age, tumour size, grade, ER, PR, HER2; Beumer 2016: age, nodal status, grade, ER, HER2, chemotherapy (similar results when only adjusting for CP factors associated with MMP outcome). [b]Adjusted for AOL. [c]Adjusted for NPI.

Table 6: Additional prognostic value for other outcomes: MammaPrint

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Multivariable model adj. for CP factors[a], AOL[b] or NPI[c]: HR (95% CI)
Pooled analyses of patient cohorts: LN status mixed									
Variable ET&CT									
Mook 2010[18] N=964	Pooled 7 series[2, 5-7, 10, 11, 19]	84% ER+ 68% HER2-	LN0, 72% LN+, 27%	32% ET 22% CT	**BCSS 10yr**	**MMP**			HR 3.25 (1.92, 5.51), p<0.001[a]
		All ER+ (n=788)	LN+/-		**BCSS 10yr**	**MMP**			3.43 (1.98, 5.95), p<0.001[a]
			LN+/-	No ET/CT	**BCSS 10yr**	**MMP**			3.47 (1.83, 6.60), p<0.001[a]
No ET&CT									
Kok 2012[11] N=100+51	Pooled 2 series: van de Vijver[2] + Mook 2010[6]	ER+ HER2 NR	LN0, 91% LN+, 9%	No ET/CT	**BCSS 10yr**	**MMP**			2.56 (0.91, 7.17), p=0.074
Retrospective studies: LN status mixed									
100% ET monotherapy									
Kok 2012[11] N=121	NKI 1985-94[30]	ER+ HER2 NR	LN0, 18% LN+, 82%	All ET No CT	**BCSS 10yr**	**MMP**			1.88 (0.77, 4.61), p=0.17
Retrospective studies: LN0									
Variable ET&CT									
Mook 2010[6] N=148	- NKI 1984-96[6] (55-71yr)	78% ER+ HER2 NR	LN0	18% ET No CT	**BCSS**	**MMP**			**5yr:** 14.4 (1.7, 122), p=0.01[b] **10yr:** 2.2 (CI NR)[b]
van de Vijver 2002;[2] Bueno-de-Mesquita 2009;[5] Buyse 2006[7]	- van de Vijver 2002[2] N=151	72% ER+ HER2 NR	LN0	4% ET 4% CT	**DFS**	**MMP**			**All FU (med 6.7yr):** [7] 4.80 (2.37, 9.71)[b]
No ET&CT									
Buyse 2006[7] N=302	- TRANSBIG[7]	70% ER+ HER2 NR	LN0	No ET/CT	**DFS**	**MMP**			**5yr:** 2.16 (CI NR)[b] **10yr:** 1.66 (CI NR)[b] **All (med 13.6yr):** 1.36 (0.91, 2.03);[b] 1.45 (0.97, 2.16)[c]

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Multivariable model adj. for CP factors[a], AOL[b] or NPI[c]: HR (95% CI)
Retrospective studies: LN+									
Variable ET&CT									
Mook 2009[10] N=241	- NKI+Italy[10]	79% ER+ 84% HER2-	LN1-3	73% ET 56% CT	**BCSS 10yr**	**MMP**			7.17 (1.81, 28.43), p=0.005[a]
van de Vijver 2002;[2] Mook 2009[10] N=106	- van de Vijver 2002[2]	82% ER+ 84% HER2-	LN1-3	23% ET 70% CT	**BCSS 10yr**	**MMP**			3.63 (0.88, 14.96), p=0.07[a]

AOL, Adjuvant! Online; BCSS, breast cancer-specific survival; CI, confidence interval; CP, clinical/pathological; CT, chemotherapy; DFS, disease-free survival; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; LR, likelihood ratio; MMP, MammaPrint; NPI, Nottingham Prognostic Index.
[a]Adjusted for: Mook 2009 (LN1-3) + Mook 2010 (pooled): age, tumour size, nodal status, grade, ER, HER2, surgery, endocrine therapy, chemotherapy. [b]Adjusted for AOL. [c]Adjusted for NPI.

Table 7: Characteristics of chemotherapy benefit studies: MammaPrint

Reference(s)	Cohort(s)	N	Country	Study design	Test	Details of test	Cut-offs	Population	Nodal status	ET / CT
Knauer 2010[22]	Pooled 6 series: - van de Vijver 2002[2] - Bueno-de-Mesquita 2009[5] - Mook 2009[10] (LN1-3) - Mook 2010[6] (age 55-71) - Bueno-de-Mesquita 2007[19] (RASTER) - Kok (personal com.)	541	Various	Pooling of 6 consecutive cohorts	**MMP**	Frozen MMP microarray	Low, high (details NR)	90% ER+ 89% HER2- Pre/post-meno % female NR pT1-3	LN0, 49% LN1-3, 51%	All ET 42% CT
Mook 2009[10]	1) NKI+Italy 1994-2001[10] 2) van de Vijver 2002[2]	1) 241 2) 106	Neths, Italy	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	1) 79% ER+ 84% HER2- 2) 82% ER+ 84% HER2- All: Pre/post-meno, age ≤70 % female NR	LN1-3 (inc. micromets)	1) 73% ET (low 82%, high 65%); 56% CT (low 41%, high 67%) 2) 23% ET (low 26%, high 21%); 70% CT (low 77%, high 65%)

CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR+, hormone-receptor positive; LN, number of positive nodes; MMP, MammaPrint; NKI, Netherlands Cancer Institute; NR, not reported; qRT-PCR, quantitative reverse transcription polymerase chain reaction; RASTER, MicroarRAy PrognoSTics in Breast CancER study; RdGG, Reinier de Graaf Hospital

Table 8: Quality assessment of studies predicting chemotherapy responsiveness: MammaPrint

Reference(s)	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?
Knauer 2010[22]	Pooled 6 series: - NKI, van de Vijver 2002[2] - Bueno-de-Mesquita 2009[5] - Mook 2009[10] (LN1-3), NKI+EIO (Italy) - Mook 2010[6] (age 55-71) - Bueno-de-Mesquita 2007[19] (RASTER) - Kok (personal com.)	V	N, not RCT data, pooled cohorts,	NR	Y	Y	Most (10% ER-, 11% HER2+)	Y
Mook 2009[10] (LN1-3)	- NKI+EIO (Italy)[10] - NKI, van de Vijver 2002[2]	V	N, not RCT data,retrospective cohort,	N InT, RNA quality	Y	Y	N (21%+18% ER-, 16% HER2+)	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; CT, chemotherapy; NKI, Netherlands Cancer Institute-Antoni van Leeuwenhoek hospital, Amsterdam; European Institute of Oncology, Italy

Table 9: Study and patient characteristics: MINDACT (clinical utility RCT)

Reference	Cohort; N		Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Cardoso 2016[26]	MINDACT RCT	6693 total (see below)	9 European countries	RCT and prospective cohort (see below)	Frozen MMP whole-transcriptome microarray	Low, high (no details)	88% ER+/PR+ 90% HER2-	[a]LN0, 79% LN1-3, 21%	Some ET (% NR) CT according to clinical / MMP risk
		1497 high clin, low MMP		RCT			98% HR+ 92% HER2-	LN0, 52% LN1-3, 48%	Randomised to CT or no CT
		690 low clin, high MMP		RCT			90% HR+ 88% HER2-	LN0, 98% LN1-3, 2%	Randomised to CT or no CT
		2634 low clin, low MMP		Prospective cohort			100% HR+ 96% HER2-	LN0, 94% LN1-3, 6%	No CT recommended
		1873 high clin, high MMP		Prospective cohort			62% HR+ 81% HER2-	LN0, 74% LN1-3, 26%	CT recommended

-, not reported; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; MMP, MammaPrint; NR, not reported.
[a]Micrometastases 0.2-2mm considered LN+; isolated tumour cells considered LN0

Table 10: Quality assessment: MINDACT (clinical utility RCT)

	Random sequence generation	Allocation concealment	Blinding (participants and personnel)	Blinding (outcome assessment)	Incomplete outcome data	Selective reporting
Cardoso 2016[26] MINDACT RCT	Low (stratified)	Low (centrally randomised)	Unclear	Unclear	Low	Low

High/low/unclear relates to risk of bias on each criterion.

Table 11: Clinical utility of MammaPrint (MINDACT): Estimated outcomes according to clinical and MMP treatment strategies (ITT)

Study	Subgroup	Patients & treatment	N	DMFS 5yr: % estimated risk
Cardoso 2016[26]	Clinical strategy: CT recommended according to clinical (mAOL) risk	Clin low MMP low: no CT Clin low MMP high: no CT Clin high MMP low: CT Clin high MM high: CT (Excludes: Clin low MMP high: CT) (Excludes: Clin high MMP low: no CT)	6698 (discordant patients double weighted since under-represented)	95.0 (CI NR)
	MMP strategy: CT recommended according to MMP risk	Clin low MMP low: no CT Clin low MMP high: CT Clin high MMP low: no CT Clin high MM high: CT (Excludes: Clin low MMP high: no CT) (Excludes: Clin high MMP low: CT)	6690 (discordant patients double weighted since under-represented)	94.7 (CI NR)
CI, confidence interval; CT, chemotherapy; DMFS, distant metastasis-free survival; mAOL, modified Adjuvant! Online; MMP, MammaPrint; NR, not reported.				

Table 12: Clinical utility of MammaPrint (MINDACT): Estimated outcomes according to clinical and MMP treatment strategies (ITT)

Study	Subgroup	Patients & treatment	N	DMFS 5yr: % estimated risk
Cardoso 2016[26]	Clinical strategy: CT recommended according to clinical (mAOL) risk	Clin low MMP low: no CT Clin low MMP high: no CT Clin high MMP low: CT Clin high MM high: CT (Excludes: Clin low MMP high: CT) (Excludes: Clin high MMP low: no CT)	6698 (discordant patients double weighted since under-represented)	95.0 (CI NR)

Study	Subgroup	Patients & treatment	N	DMFS 5yr: % estimated risk
	MMP strategy: CT recommended according to MMP risk	Clin low MMP low: no CT Clin low MMP high: CT Clin high MMP low: no CT Clin high MM high: CT (Excludes: Clin low MMP high: no CT) (Excludes: Clin high MMP low: CT)	6690 (discordant patients double weighted since under-represented)	94.7 (CI NR)
CI, confidence interval; CT, chemotherapy; DMFS, distant metastasis-free survival; mAOL, modified Adjuvant! Online; MMP, MammaPrint; NR, not reported.				

Table 13: Study and patient characteristics: RASTER (clinical utility observational study)

Reference	Cohort; N		Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Drukker 2013[20] Drukker 2014[21] Bueno-de-Mesquita 2007[19] Vliek 2017a[27]	RASTER node-negative 16 community hospitals	427	Neths	Prospective observational; treatment influenced by MMP result	Frozen MMP microarray	Low (good), high (poor); cut-offs as in previous studies	80% ER+ 84% HER2- Age <61 100% female	LN0	43% ET (low 27%, high 59%) 47% CT (low 15%, high 81%)
Vliek 2017b,[27]	RASTER node-positive 16 community hospitals	164	Neths	Prospective observational; treatment NOT influenced by MMP result (test retrospective)	FFPE MMP microarray	Low, high (no details)	83% ER+ HER2 NR Age <61 100% female	LN1-3, 82% LN>3, 18% (some analyses LN1-3 only)	ET NR 95% CT (low 92%, high 97%)
-, not reported; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; MMP, MammaPrint; NR, not reported; RASTER, MicroarRAy PrognoSTics in Breast CancER study.									

Table 14: Quality assessment: RASTER (clinical utility observational study)

	Random sequence generation	Allocation concealment	Blinding (participants and personnel)	Blinding (outcome assessment)	Incomplete outcome data	Selective reporting
Drukker 2013[20]	High	High	High	UC	High	Unclear

Drukker 2014[21] Bueno-de-Mesquita 2007[19] Vliek 2017a[27]Cohort study						
High/low/unclear relates to risk of bias on each criterion. RASTER, MicroarRAy PrognoSTics in Breast CancER study.						

Table 15: Clinical utility of MammaPrint (RASTER study): overall survival in node-negative patients

Study	Subgroup N	Population	Nodal status	Endo / chemo	Outcome	Test or comparator	% pts per group		% CT per group		% OS risk per group		HR (95% CI), p-value
							Low	High	Low	High	Low	High	0-5 yr
Node-negative													
RASTER[20, 27, 31]	All patients N=427	80% ER+ 84% HER2-	LN0	43% ET 47% CT	**OS 5yr**	**MMP**	51	49	15	81	98.3	96.9	p=0.35
						AOL	31	69	18	60	100.0	96.5	p=0.02

AOL, Adjuvant! Online; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; MMP, MammaPrint; OS, overall survival; RASTER, MicroarRAy PrognoSTics in Breast CancER study.

REFERENCES

1. van de Vijver MJ, He Y, van't Veer LJ, Dai H, Hart AA, Voskuil DW*, et al.* A gene-expression signature as a predictor of survival in breast cancer. 347 (25), 1999-2009. *New Engl J Med* 2002;347:1999-2009.
2. van de Vijver MJ, He Y, van't Veer LJ, Dai H, Hart AA, Voskuil DW*, et al.* A gene-expression signature as a predictor of survival in breast cancer. *New Engl J Med* 2002;347:1999-2009.
3. Glas A, Floore A, Delahaye L, Witteveen A, Pover R, Bakx N*, et al.* Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006;7:278-.
4. Beumer I, Witteveen A, Delahaye L, Wehkamp D, Snel M, Dreezen C*, et al.* Equivalence of MammaPrint array types in clinical trials and diagnostics. *Breast Cancer Research & Treatment* 2016;156:279-87.
5. Bueno-De-Mesquita JM, Linn SC, Keijzer R, Wesseling J, Nuyten DSA, van Krimpen C*, et al.* Validation of 70-gene prognosis signature in node-negative breast cancer. *Breast Cancer Research & Treatment* 2009;117:483-95.
6. Mook S. The 70-gene prognosis signature predicts early metastasis in breast cancer patients between 55 70 years of age. *Annals of Oncology* 2010;21:717-22.
7. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM*, et al.* Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *Journal of the National Cancer Institute* 2006;98:1183-92.
8. Ishitobi M, Goranova TE, Komoike Y, Motomura K, Koyama H, Glas AM*, et al.* Clinical utility of the 70-gene MammaPrint profile in a Japanese population. *Japanese Journal of Clinical Oncology* 2010;40:508-12.
9. Wittner BS, Sgroi DC, Ryan PD, Bruinsma TJ, Glas AM, Male A*, et al.* Analysis of the MammaPrint breast cancer assay in a predominantly postmenopausal cohort. *Clinical Cancer Research* 2008;14:2988-93.
10. Mook S, Schmidt MK, Viale G, Pruneri G, Eekhout I, Floore A*, et al.* The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. *Breast Cancer Res Treat* 2009;116:295-302.
11. Kok M, Koornstra RH, Mook S, Hauptmann M, Fles R, Jansen MP*, et al.* Additional value of the 70-gene signature and levels of ER and PR for the prediction of outcome in tamoxifen-treated ER-positive breast cancer. *Breast* 2012;21:769-78.
12. Yao K, Goldschmidt R, Turk M, Wesseling J, Stork-Sloots L, de Snoo F*, et al.* Molecular subtyping improves diagnostic stratification of patients with primary breast cancer into prognostically defined risk groups. *Breast Cancer Res Treat* 2015;154:81-8.
13. Drukker CA, van Tinteren H, Schmidt MK, Rutgers EJ, Bernards R, van de Vijver MJ*, et al.* Long-term impact of the 70-gene signature on breast cancer outcome. *Breast Cancer Res Treat* 2014;143:587-92.
14. Esserman LJ, Thompson CK, Yau C, Van't Veer LJ, Borowsky AD, Tobin NP*, et al.* Identification of tumors with an indolent disease course: MammaPrint ultralow signature validation in a retrospective analysis of a Swedish randomized tamoxifen trial. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76.
15. Lindstrom LS, Benz CC, Yau C, van't Veer LJ, Thompson CK, Esserman LJ. MammaPrint accurately predicts long-term survival (25 years) and adjuvant tamoxifen therapy benefit in lymph node negative patients. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75.
16. van 't Veer LJ, Yau C, Yu NY, Benz CC, Nordenskjold B, Fornander T*, et al.* Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. *Breast Cancer Res Treat* 2017; 10.1007/s10549-017-4428-9.
17. Agendia. MammaPrint - evidence provided to NICE. 2017.

18. Mook S, Knauer M, Bueno-De-Mesquita JM, Retel VP, Wesseling J, Linn SC*, et al.* Metastatic potential of T1 breast cancer can be predicted by the 70-gene MammaPrint signature. *Annals of Surgical Oncology* 2010;17:1406-13.
19. Bueno-de-Mesquita JM, van Harten WH, Retel VP, van 't Veer LJ, van Dam FS, Karsenberg K*, et al.* Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER). *Lancet Oncol* 2007;8:1079-87.
20. Drukker CA, Bueno-de-Mesquita JM, Retel VP, van Harten WH, van Tinteren H, Wesseling J*, et al.* A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. *Int J Cancer* 2013;133:929-36.
21. Drukker CA, Nijenhuis MV, Bueno-de-Mesquita JM, Retel VP, van Harten WH, van Tinteren H*, et al.* Optimized outcome prediction in breast cancer by combining the 70-gene signature with clinical risk prediction algorithms. *Breast Cancer Research & Treatment* 2014;145:697-705.
22. Knauer M, Mook S, Rutgers EJT, Bender RA, Hauptmann M, van de Vijver MJ*, et al.* The predictive value of the 70-gene signature for adjuvant chemotherapy in early breast cancer. *Breast Cancer Research & Treatment* 2010;120:655-61.
23. Bueno-de-Mesquita JM, Sonke GS, van de Vijver MJ, Linn SC. Additional value and potential use of the 70-gene prognosis signature in node-negative breast cancer in daily clinical practice. *Ann Oncol* 2011;22:2021-30.
24. Beumer IJ, Persoon M, Witteveen A, Dreezen C, Chin SF, Sammut SJ*, et al.* Prognostic value of MammaPrint in invasive lobular breast cancer. *Biomark Insights* 2016;11:139-46.
25. Michaut M, Chin SF, Majewski I, Severson TM, Bismeijer T, de Koning L*, et al.* Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. *Sci Rep* 2016;6:18517.
26. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S*, et al.* 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *New Engl J Med* 2016;375:717-29.
27. Vliek SB, Retel V, Drukker C, Rutgers E, van Tinteren H, Van de Vijver MJ*, et al.* 10 years follow up of the RASTER study; implementing a genomic signature in daily practice (poster). *Ann Oncol* 2017a;28 (suppl_5):v43-v67.
28. VvlK. KvdGC. Adjuvante Systemische Therapie voor het Operabel Mammacarcinoom. Behandeling van het Mammacarcinoom.; 2004.
29. Linn SC, Retel V, Drukker C, Bueno-de-Mesquita JM, Rutgers E, van Tinteren H. 10 years follow up of the RASTER study; implementing a genomic signature in daily practice (abstract). *ESMO conference* 2017.
30. Kok M, Koornstra RH, Margarido TC, Fles R, Armstrong NJ, Linn SC*, et al.* Mammosphere-derived gene set predicts outcome in patients with ER-positive breast cancer. *J Pathol* 2009;218:316-26.
31. Drukker CA, Nijenhuis MV, Bueno-de-Mesquita JM, Retel VP, van Harten WH, van Tinteren H*, et al.* Optimized outcome prediction in breast cancer by combining the 70-gene signature with clinical risk prediction algorithms. *Breast Cancer Res Treat* 2014;145:697-705.