Supplementary File 7: Narrative synthesis and additional Tables for Chapter 2, Results: Endopredict and EPClin

Study designs: EndoPredict and EPClin

Three data sets, all re-analyses of RCTs, have been used to validate the prognostic performance of EndoPredict (Table 1). Analysis of UK-based patients from the TransATAC trial was reported by Buus *et al*. (2016)[1] and updated data for 878 patients (used in this report) were provided via personal communication with the TransATAC team (Sestak, 2017).[2] Analysis of 1702 patients pooled from the Austrian Breast and Colorectal Cancer Study Group (ABCSG)-6 and ABCSG-8 trials was reported by Dubsky *et al* (2013a and 2013b) plus subgroup analyses provided to NICE by Myriad Genetics.[3-6] Finally, 555 patients from the Spanish GEICAM 9906 trial were analysed by Martin *et al*. (2014, 2016).[7, 8]

Patients: EndoPredict and EPClin

All three data sets either consisted of, or had analyses available for, ER+, HER2- patients. In terms of nodal status, two of the three data sets included LN0 patients (TransATAC[1, 2] and ABCSG-6+8[3-5]). All recruited only, or reported a subgroup of, patients who were ER+ HER2-. One reported on LN0 patients (N=680)[2, 9] and two[1, 2, 7, 8] on LN+ patients (total N=753; one[7, 8] included 36% patients with >3 positive nodes). One reported on patients unselected by LN status;[3, 4] additional analyses[5] were provided to the EAG as Commercial in Confidence data and cannot be reported here. Patients in all three analyses received 5 years of endocrine therapy. Patients in the GEICAM 9906 analysis[7, 8] also received adjuvant chemotherapy, while those in the other two analyses did not.

For TransATAC, two sets of data were presented in the analysis reported to the EAG via NICE.[2] The "full dataset" refers to data on all 878 patients with EndoPredict data available, while the "reduced dataset" refers to 774 patients with data for all four in-scope tests analysed in TransATAC. In this report, data for the "full dataset" is used where available; if not available then the "reduced dataset" is used. Both datasets gave very similar results.

Tests and comparators: EndoPredict and EPClin

All three data sets assessed the tests as marketed (though in TransATAC[1] a correction factor was applied to account for differences in RNA extraction methods), using qRT-PCR and standard cut-offs for risk groups (5 for EndoPredict and 3.3 for EPClin). The three data sets were also used to evaluate other in-scope tests as follows (see Appendix 5 on comparing tests). TransATAC was used to evaluate Oncotype DX, Prosigna and IHC4+C.[10-13] GEICAM 9906 was used to evaluate a "research-based" version of PAM50 ROR-PT. [7, 8] ABCSG-8 (but not ABCSG-6) was used to evaluate Prosigna.[14, 15]

Quality assessment: EndoPredict and EPClin

The EAG's assessment of study quality is provided in Table 2. All analyses excluded some original trial patients (or this was unclear), sometimes due to insufficient tumour sample which may introduce bias due to attrition of patients with smaller tumours. Blinding of test assessors to outcomes was reported in two analyses.[1, 7, 8] All used standardised outcomes.

Results: EndoPredict and EPClin

Chapter 2, Results: EndoPredict and EPClin of the main report, Table 3 and Table 20-21 of the main report present the data for all patients (mix of LN0 and LN+) and separate data for LN0 and LN+ patients.

Distribution of patients by risk group

The percentage of LN0 patients categorised as EPClin low-risk was 73% in TransATAC.[2] Far fewer LN+ patients were categorised as EPClin low-risk: 24% in TransATAC,[2] and 13% in GEICAM 9906[7, 8] (Table 20 of the main report).

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"

LN0: The analysis of LN0 patients (TransATAC[2]) showed that EPClin was statistically significantly prognostic for 10-year DRFS/DRFI. The proportion of patients with 10-year DRFS/DRFI in the EPClin low-risk groups was 94.1% [2] (Table 20 of the main report). HR for the low vs. high-risk group was 3.90 (95% CI: 2.33, 6.53, p=not reported).[2]

In terms of overall survival, EPClin was also statistically significantly prognostic for 10-year overall survival in the one study of LN0 patients reporting this outcome (TransATAC,[2] Table 3).

LN+: Both analyses of LN+ patients showed that EPClin was statistically significantly prognostic for 10-year DMFS/DRFS/DRFI. The proportion of patients with 10-year DMFS/DRFS/DRFI in the EPClin low-risk groups was 95.0% in TransATAC;[2] and 100% in GEICAM 9906[7, 8] (Table 20 of the main report). HRs for the low vs. high-risk groups were 6.77 (95% CI: 1.63, 28.07, p=not reported) in TransATAC;[2] and for GEICAM not estimable since there were no events in the low-risk group (p<0.0001).[7, 8] EPClin was also statistically significantly prognostic for 10-year overall survival in

TransATAC[2] (Table 3). However, as noted above, only a relatively small proportion of LN+ patients were classed as low-risk (13% to 24% across the two studies).[2, 7, 8]

Comparison to guidelines: In the ABCSG-6+8 analysis,[4] the hazard ratio for 10-year DRFI for low vs. intermediate/high-risk groups across all patients (two-thirds LN0) was higher for EPClin (HR 5.11, 95% CI: 3.48, 7.51, p<0.001) than when classifying patients as low/high risk according to any of three clinical guidelines: NCCN 2007 (HR 2.16, p=0.119), St Gallen 2011 (HR 2.78, p<0.001) or German S3 2008 guidelines (HR 2.20, p=0.014).

Patients at high clinical risk: The ABCSG-6+8 analysis[4] also reported results for patients classed as high or high/intermediate-risk via the three clinical guidelines: NCCN 2007, St Gallen 2011, and German S3 guidelines 2008. Around 60% were categorised as low-risk via EPClin. EPClin was statistically significantly prognostic for 10-year DRFI in these high-clinical-risk patients (Table 20 of the main report).

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.

Likelihood ratios: The TransATAC analysis[2] reports a reduced dataset of patients where data for all four in-scope tests are available. Additional prognostic value was assessed via increases in likelihood ratio χ^2 for 10-year DRFI, for EPClin plus NPI or CTS, over NPI or CTS alone (Table 21 of the main report). Increases in likelihood ratio χ^2 were statistically significant for LN0 patients: 15.22 (p<0.0001) over CTS and 17.00 (p<0.0001) over NPI, and also for LN+ patients: 7.36 (p=0.007) over CTS and 5.57 (p=0.02) over NPI.

C-indexes (AUC): In LN+ patients in GEICAM 9906, adding EndoPredict to a combination of clinicopathological variables increased the C-index from 0.654 to 0.672 (p=0.0018), while EPClin gave a higher C-index of 0.693 (p=NR; Table 21 of the main report).[8] In ABCSG-6+8 (two-thirds LN0), the C-index was only reported for years 5-10and 0-10 (no data for years 0-5).[3, 6] During both periods, the C-index increased when adding EndoPredict to a combination of clinical variables or to AOL (all p<0.001; Table 21 of the main report and data not shown).

Multivariable Cox models: Both ABCSG-6+8[3-5] (mix of LN0/LN+) and GEICAM 9906[7, 8] (LN+) used multivariable analyses to show that EndoPredict (no data reported for EPClin) was an

independent prognostic variable for 10-year DMFS/DRFI after adjustment for clinical variables ($p<0.001$;[3-5] $p=0.003$;[7, 8] Table 21 of the main report).

Table 1: Characteristics of prognostic studies: EndoPredict and EPClin

Reference(s)	Cohort(s)	N pts	Country	Study design	Test	Details of test	Cut-offs	Other tests	Population	Nodal status	Endo / chemo
Reanalyses of RCTs: LN status mixed											
100% ET monotherapy											
Sestak 2017 (data request),[2] Buus 2016[1]	**TransATAC**	878 (full dataset) 774 (reduced dataset)[a]	UK	R-RCT	EPClin	FFPE qRT-PCR, Sividon	3.3	O-DX ROR-PT IHC4+C	ER+ HER2- Postmeno 100% female	LN0, 77% LN1-3,23%	All ET 5yr No CT
Dubsky 2013a,[4] 2013b[3]	**ABCSG-6+8**	1702 (all)	Austria	R-RCT	EP EPClin	FFPE qRT-PCR	5 3.3	ROR-PT (ABSCG -8)	ER+ HER2- Postmeno Stage I-II 100% female	LN0, 68% LN1-3, 27% LN>3, 5%	All ET 5yr No CT
Reanalyses of RCTs: LN+											
100% CT&ET											
Martin 2016,[7] 2014[8]	**GEICAM 9906**	555	Spain	R-RCT	EP EPClin	FFPE qRT-PCR	5 3.3	ROR-PT	ER+ HER2- 46% postmeno Stage II-III 100% female	All N+ LN1-3, 64% LN>3, 36%	All ET 5yr All CT

ABCSG, Austrian Breast and Colorectal Cancer Study Group; 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; qRT-PCR, quantitative reverse transcription polymerase chain reaction; R-RCT, reanalysis of RCT

[a]Full dataset=all patients with EndoPredict data available; reduced dataset = patients with data for all four in-scope tests analysed in TransATAC

Table 2: Quality assessment of prognostic studies: EndoPredict and EPClin

Reference(s)	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?
Sestak 2017 (data request),[2] Buus 2016[1]	**TransATAC**	V	Y, R-RCT, no chemo	N, InT, MS, TP	Y	Y	Y	Y
Dubsky 2013a,[4] 2013b[3]	**ABCSG-6+8**	V	Y, R-RCT, no chemo	UC	UC	Y	N, (5% LN>3)	Y
Martin 2016,[7] 2014[8]	**GEICAM 9906**	V	N, R-RCT, adj chemo	N (reason NR)	Y	Y	N (36% LN>3)	N, Prosigna via qRT-PCR then microarray

Y, yes; N, no; UC, unclear
ABCSG, Austrian Breast and Colorectal Cancer Study Group; D, Development; InT, insufficient tissue; 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

Table 3: Prognostic performance of EndoPredict and EPClin: overall survival

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% pts per group		% OS risk: 0-5 yr		% OS risk: 0-10 yr		OS: HR (95% CI) 0-5 yr
						Low	High	Low	High	Low	High	
Reanalyses of RCTs: LN status mixed												
100% ET monotherapy												
Sestak 2017 (data request)[2] (reduced dataset)[b]	**TransATAC** R-RCT; UK	ER+ HER2- N=774	LN0, 76% LN1-3, 24%	All ET No CT	**EPClin**	61	39	-	-	-	-	**0-10 yr:** 2.15 (1.65, 2.80)
Reanalyses of RCTs: LN0												
100% ET monotherapy												
Sestak 2017 (data request)[2] (full dataset)[b]	**TransATAC** R-RCT; UK	ER+ HER2- N=680	LN0	All ET No CT	**EPClin**	73	27	93.6	90.0	80.0	63.4	**0-5yr:** 1.57 (0.88, 2.80) **0-10 yr:** 2.08 (1.51 2.87)
Reanalyses of RCTs: LN+												
100% ET monotherapy												

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% pts per group		% OS risk: 0-5 yr		% OS risk: 0-10 yr		OS: HR (95% CI) 0-5 yr
						Low	High	Low	High	Low	High	
Sestak 2017 (data request)[2] (full dataset)[b]	**TransATAC** R-RCT; UK	ER+ HER2- N=198	LN1-3	All ET No CT	**EPClin**	24	76	95.7	81.5	75.7	57.4	**0-5yr:** 4.68 (1.12, 19.66) **0-10 yr:** 2.24 (1.15, 4.37)
100% CT&ET												
Martin 2016,[7] 2014[8]	**GEICAM 9906** R-RCT; Spain	ER+ HER2- N=536	LN1-3, 64% LN>3, 36%	All ET All CT	**EP**	25	75	-	-	92	67[a]	**0-10 yr:** 3.9 (2.0, 7.5), p<0.0001
					EPClin	13	87	-	-	99 [a]	69 [a]	**0-10 yr:** 19.4 (2.7, 138.7), p<0.0001

-, not reported; ABCSG, Austrian Breast and Colorectal Cancer Study Group; CI, confidence interval; CT, chemotherapy; ET, endocrine therapy; ER, oestrogen receptor; HR, hazard ratio; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; OS, overall survival; R-RCT, reanalysis of RCT.

[a] Estimated off graph

[b]Full dataset=all patients with EndoPredict data available; reduced dataset = patients with data for all four in-scope tests analysed in TransATAC

REFERENCES

1. Buus R, Sestak I, Kronenwett R, Denkert C, Dubsky P, Krappmann K*, et al.* Comparison of EndoPredict and EPclin with Oncotype DX recurrence score for prediction of risk of distant recurrence after endocrine therapy. *J Natl Cancer Inst* 2016;108.
2. Sestak I, Dowsett M, Cuzick J. NICE request - TransATAC data analysis. In; 2017.
3. Dubsky P, Brase JC, Jakesz R, Rudas M, Singer CF, Greil R*, et al.* The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. *Br J Cancer* 2013;109:2959-64.
4. Dubsky P, Filipits M, Jakesz R, Rudas M, Singer CF, Greil R*, et al.* EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer. *Ann Oncol* 2013;24:640-7.
5. Myriad Genetics. EPclin in ABCSG-6+8: Subgroup analyses. 2015.
6. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF*, et al.* A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res* 2011;17:6012-20.
7. Martin M, Brase JC, Ruiz A, Prat A, Kronenwett R, Calvo L*, et al.* Prognostic ability of EndoPredict compared to research-based versions of the PAM50 risk of recurrence (ROR) scores in node-positive, estrogen receptor-positive, and HER2-negative breast cancer. A GEICAM/9906 sub-study. *Breast Cancer Res Treat* 2016;156:81-9.
8. Martin M, Brase JC, Calvo L, Krappmann K, Ruiz-Borrego M, Fisch K*, et al.* Clinical validation of the EndoPredict test in node-positive, chemotherapy-treated ER+/HER2- breast cancer patients: results from the GEICAM 9906 trial. *Breast Cancer Res* 2014;16:R38.
9. Sestak I, Buus R, Cuzick J, Dubsky P, Kronenwett R, Ferree S*, et al.* Comprehensive comparison of prognostic signatures for breast cancer recurrence in TransATAC. San Antonio Breast Cancer Symposium, abstract no. 5866.
10. Sestak I, Cuzick J, Dowsett M, Lopez-Knowles E, Filipits M, Dubsky P*, et al.* Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. *J Clin Oncol* 2015;33:916-22.
11. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW*, et al.* Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol* 2013;31:2783-90.
12. Sestak I, Dowsett M, Zabaglo L, Lopez-Knowles E, Ferree S, Cowens JW*, et al.* Factors predicting late recurrence for estrogen receptor-positive breast cancer. *J Natl Cancer Inst* 2013;105:1504-11.
13. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E*, et al.* Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol* 2011;29:4273-8.
14. Gnant M, Filipits M, Greil R, Stoeger H, Rudas M, Bago-Horvath Z*, et al.* Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Ann Oncol* 2014;25:339-45.
15. Filipits M, Nielsen TO, Rudas M, Greil R, Stoger H, Jakesz R*, et al.* The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. *Clin Cancer Res* 2014;20:1298-305.