Summary of evidence relating to OncotypeDX reported in the Marchionni *et al.* systematic review[33]

Analytical validity	Clinical validity	Clinical utility
Reported in four studies.[39–42] Technical and operational aspects were reported in two studies[39,40] and test and assay variability and reproducibility were reported in three studies.[40–42] Conclusion: Preanalytic issues relating to sample storage and preparation appeared to play a larger role than within-laboratory variation Six studies reported overall success rate,[41–44,48,49] which ranged from 78.9% to 98.9% Not all of the studies provided detailed descriptions of the reasons for assay failure. When failures were reported they were mainly ascribed to an insufficient number of cancer cells in the specimens, poor RNA quality and, in a few cases, failure of the RT-PCR technique Systematic review conclusion: Evidence existed for some of the operational characteristics of this test but there was limited evidence for the reproducibility of the test. Reasonable reproducibility of the test across different samples of the same block, and samples from different blocks. No direct evidence was available about the effect of sample preparation. There was indirect evidence that the overall success rate of extracting analysable mRNA was fairly high. Centralisation was considered to be a current strength of OncotypeDX with regard to reproducibility	Reported in four studies in relation to the determination of recurrence risk (prognosis) Paik *et al.*[42] studied the prognostic validity of OncotypeDX in an independent tamoxifen-treated population. The RS was shown to be significantly correlated with DFS ($p<0.001$) and OS ($p<0.001$). RS alone was a better predictor of distant recurrence at 10 years than traditional clinicopathological predictors Esteva *et al.*[44] failed to find a correlation between RS and distant breast cancer recurrence in untreated node-negative (LN0) patients. In the reverse of what was expected, well-differentiated tumours were correlated with poorer survival than higher-grade tumours Cobleigh *et al.*[43] reported that the RS score was significantly correlated with DRFS in a training set of LN0 patients. As this data set related to training and not validation, it was considered to present minimal evidential value Habel *et al.*[41] assessed the risk of breast cancer-specific mortality among women in a large case–control study of ER+, LN0 breast cancer patients treated with tamoxifen. The 10-year risk of death from breast cancer was 3% for patients with a low RS, 12% for patients with an intermediate RS and 27% for patients with a high RS. Multivariate analysis showed that RS and tumour size were independent risk predictors of breast cancer death in ER+, tamoxifen-treated patients (RR (relative risk) for RS (risk score) per 50 units = 7.6, $p<0.001$) and untreated patients (RR (relative risk) for RS (risk score) per 50 units = 4.1, $p<0.001$). The RS score also showed some prognostic value in ER– patients Three posters describing studies that compared risk predictions provided by OncotypeDX assays and standard risk classification methods were reported.[45,46,47] The data presented in these posters suggested that optimal predictions may come from a combination of gene expression tests and standard risk assessment methods Systematic review conclusion: Fairly strong support for the clinical validity of the OncotypeDX test over and above standard clinical predictors in ER+, LN0 and tamoxifen-treated patients with a clear treatment indication for adjuvant chemotherapy. The authors noted, however, that it was not clear (1) how much the test added to the management of patients, (2) what proportion of patients would benefit from the use of the OncotypeDX test and (3) the stability of the observed risk categories in other populations, particularly those treated with current therapies	No published studies reported demonstrating clinical utility (direct evidence) Two studies reported that provided preliminary evidence of the potential predictive power of OncotypeDX (indirect evidence) Paik *et al.*,[49] using specimens and data from an existing trial (NSABP B20), addressed the potential value of the RS in predicting chemotherapy benefit in a population of ER+, LN0 patients. This study compared a group of patients treated with tamoxifen and chemotherapy with a group of patients who were randomised to tamoxifen only. The RS was found to be correlated with chemotherapy benefit, defined in terms of 10-year DRFS, with a significant benefit from the use of chemotherapy in the high RS group ($p=0.001$). However, in a multivariate analysis the benefit from chemotherapy was unclear because of large CIs in the low and intermediate RS risk groups Oratz *et al.*[48] reported that knowledge of the RS changed the clinicians' treatment recommendations for 21% of patients and the actual administered treatment for 25% of patients. They did not report what the patients (or doctors) were told or understood about the risk of recurrence Systematic review conclusion: the Paik *et al.* study[49] represented the strongest evidence derived from already existing data regarding the clinical utility of the OncotypeDX test. This study also noted that, although prospective confirmation of these findings was required, the evidence provided reasonable justification in the interim for the use of the test by women in this specific population

Systematic review summary

The studies assessed in this review were heterogeneous in focus and quality. Few of the publications addressed technical aspects of the tests. A number of the reports focused on prognostic prediction. Only one study examined the prediction of treatment benefit. Most of the published evidence available for OncotypeDX was obtained using the marketed assay. Overall, the evidence presented for the clinical validity of OncotypeDX/21-gene signature in the systematic review was considered to have provided fairly strong support for the clinical performance of the test compared with standard predictors in a well-defined population (ER+, LN0, tamoxifen-treated women). It was considered that there was strong enough evidence of the clinical utility of the test in retrospectively collected data from one large clinical trial to provide reasonable justification for the interim use of the test in women in the same population group as the trial patients. There was little information about the impact of the test on clinical decision-making

Summary of evidence relating to OncotypeDX reported in the Smartt systematic review[34]

Clinical validity	Clinical utility
Two studies reported on the clinical validity of the test[50,51] The purpose of the Goldstein *et al.* study[49] was to evaluate the prognostic value of OncotypeDX in hormone receptor-positive, LN0 or LN+ patients and to determine whether or not it could better predict outcome at 5 years than a modified Adjuvant! Online algorithm. The 21-gene assay was a more accurate predictor of relapse than standard clinical features for individual patients with hormone receptor-positive operable breast cancer treated with chemotherapy/hormonal therapy and provides information that is complementary to features typically used in anatomic staging, such as tumour size and LN involvement. The 21-gene assay may be used to select low-risk patients for abbreviated chemotherapy regimens similar to those used in our study or high-risk patients for more aggressive regimens or for clinical trials In the Wolf *et al.* study[51] the authors sought to assess the correlation between standard clinical and pathological breast cancer characteristics and the RS in a cohort of Israeli breast cancer patients and to compare the stratification of patients using RS with that of commonly used clinical guidelines. High tumour grade, low PR expression, infiltrating ductal histology and HER2 overexpression were found to be associated with a high RS. Patient age, tumour size, ER expression, and LN micrometastasis were found to correlate poorly with the RS. The ability of any of these variables, either alone or in combination, to predict the RS was limited. Similarly, none of the guidelines nor the Adjuvant! Online software could predict the RS. This study reported on a selected population of patients who were referred to undergo the OncotypeDX test. No association was noted between the RS and patient age or ER intensity and only a modest association was noted between the RS and tumour size. The clinical utility of these comparisons was not made clear Summary of reported conference abstracts: Shak *et al.*[59] demonstrated that the distribution of RS was similar for men and women with breast cancer	Four studies reported on the clinical utility (indirect evidence) of the test[52–55] The purpose of the Asad *et al.* study[52] was to determine whether or not the results of OncotypeDX influence the decision to administer chemotherapy. The OncotypeDX results influenced the decision for chemotherapy in 37 (44%) patients; four patients classified as low risk by the NCCN guidelines[129] (tumours <1 cm) were advised to have chemotherapy and 33 patients classified as high risk by the NCCN guidelines (tumours ≥ 1 cm) were advised to undergo hormone treatment only. The authors concluded that the OncotypeDX RS is significantly related to tumour grade and HER2/neu status. **Comment:** There was no evidence that OncotypeDX changed clinical outcomes The Henry *et al.* study[53] reported on the functional and clinically relevant impact of the RS on the adjuvant therapy administered to 29 patients with ER+, LN0 breast cancer, as well as its influence on a panel of five expert breast oncologists. They concluded that the RS contributed to chemotherapy changes in 31% of patients, with more changes made against than for adjuvant chemotherapy. The RS increased consensus recommendations by 10% but did not appear to increase the reported strength of panellists' recommendations. **Limitations:** The small sample size increased the likelihood of a type 2 error (false-negative result) and the study lacked statistical power to draw definitive conclusions. Determination of therapy received was retrospective and may have been subject to the well-established biases (e.g. selection bias, information bias) associated with this methodology. Panellists were the same medical oncologists who administered chemotherapy and panellists may have remembered their recommendations from when they were actually managing these patients. The 2-month washout period may have been insufficient to erase all recollections of previous recommendations (recall bias). Although the RS predicts only distant relapse, Adjuvant! Online includes distant and local relapse, thus the estimate of recurrence for Adjuvant! Online was much higher (90%) than that for the RS and the chemotherapy decision for 54% of patients was changed with RS information. One patient was male Li *et al.*[54] hypothesised that an integrated gene expression profile could a predict patient's response to chemotherapy. The main purpose of this study was the validation of a new gene signature, which overlapped in part with OncotypeDX and the 70-gene signature. The authors reported that their integrated signature was a stronger prediction of chemotherapy outcome than the single signatures (OncotypeDX and the 70-gene signature). **Comment:** Neither OncotypeDX nor the 70-gene signature formed the main focus of this study. Both signatures were used in populations that were very different from those that the tests were validated for. The follow-up was short The purpose of the Rayhanabad *et al.* study[55] was to examine the utility of OncotypeDX in the prediction of recurrence and the degree of benefit from chemotherapy. Treatment received after OncotypeDX testing was compared with treatment based on NCCN guidelines.[129] A total of 13 out of 18 high-risk NCCN, low-risk RS patients did not receive chemotherapy ($p<0.001$); 11 patients with an intermediate RS received chemotherapy. OncotypeDX results changed management in 15 (26%) patients ($p=0.05$). The authors concluded that the use of gene assays altered recurrence risk stratification and the decision for chemotherapy in a significant number of patients. This allowed better individualised treatment for patients, reserving chemotherapy for those at high risk of recurrence, whereas low-risk patients were spared the morbidity associated with chemotherapy. However, Smartt reported that there were a number of serious limitations in this study, which threaten the validity of the reported results Summary of reported conference abstracts: Most studies reported examined or modelled the impact of the RS on clinical decision-making in relation to adjuvant chemotherapy. Erb *et al.*[56] reported a significant decline in the use of adjuvant chemotherapy after the introduction of the test in the authors' institution. Gold *et al.*,[57] reporting on how clinicians integrated RS into their decision-making, found that RS, tumour grade and size were all independent predictors of chemotherapy administration. Lo *et al.*[58] examined the effect of knowledge of the RS on both patients and medical oncologists in relation to their adjuvant therapy choice. In total, 22% of oncologists and 10% of patients changed from chemotherapy to hormone therapy. The change in the other direction (i.e. from hormone therapy to chemotherapy) occurred in 3% and 8% respectively

Systematic review summary

There were no additional studies reporting on the analytical validity of the test and this remains an area of weakness in the evidence story to date. In contrast to the studies reported in the original systematic review,[33] the majority of these studies primarily addressed questions relating to the clinical utility of OncotypeDX, some reported further on the clinical validity or validity and utility of the test and one study reported, for the first time, on the use of the test in male breast cancer. The additional studies reporting on the clinical validity of OncotypeDX further endorsed the advantages of the test compared with standard clinicopathological assessment of risk and extended the examination of its prognostic value beyond clinical trial populations to a general population, as well as the cohort of male breast cancer patients. The studies reporting on the clinical utility of the test examined its ability to predict response to treatment or its impact on clinical decision-making. The latter studies all reported a positive impact of the test on clinical decision-making and generally claimed that there was a reduction in the number of patients who were or would have been considered for chemotherapy. However, the studies generally had methodological weaknesses that were likely to have overestimated the effect/influence of the test and were not designed to assess the effect of the test on clinical outcomes. Studies examining the ability of OncotypeDX to predict response to adjuvant and neoadjuvant endocrine therapy and chemotherapy generally reported that OncotypeDX was predictive, to a greater or lesser extent, of response to therapy; however, as the design of the studies precluded any firm conclusions about the ability of the test to predict response to therapy, these studies did not materially add to the body of evidence in this area