Personalized Decision Making on Genomic Testing in Early Breast Cancer: Expanding the MINDACT Trial with Decision-Analytic Modeling

Ewout W. Steyerberg⬮**, Liesbeth C. de Wreede, David van Klaveren, and Patrick M. M. Bossuyt**

Background. Genomic tests may improve upon clinical risk estimation with traditional prognostic factors. We aimed to explore how evidence on the prognostic strength of a genomic signature (clinical validity) can contribute to individualized decision making on starting chemotherapy for women with breast cancer (clinical utility). **Methods.** The MINDACT trial was a randomized trial that enrolled 6693 women with early-stage breast cancer. A 70-gene signature (Mammaprint) was used to estimate genomic risk, and clinical risk was estimated by a dichotomized version of the Adjuvant!Online risk calculator. Women with discordant risk results were randomized to the use of chemotherapy. We simulated the full risk distribution of these women and estimated individual benefit, assuming a constant relative effect of chemotherapy. **Results.** The trial showed a prognostic effect of the genomic signature (adjusted hazard ratio 2.4). A decision-analytic modeling approach identified far fewer women as candidates for genetic testing (4% rather than 50%) and fewer benefiting from chemotherapy (3% rather than 27%) as compared with the MINDACT trial report. The selection of women benefitting from genetic testing and chemotherapy depended strongly on the required benefit from treatment and the assumed therapeutic effect of chemotherapy. **Conclusions.** A high-quality pragmatic trial was insufficient to directly inform clinical practice on the utility of a genomic test for individual women. The indication for genomic testing may be far more limited than suggested by the MINDACT trial.

Keywords
biomarker evaluation, breast cancer, clinical utility, genomic testing, risk stratification

New biomarkers and genomic tests hold substantial promise in improving upon clinical risk estimation with traditional prognostic factors.[1] Various genomic profiles are under development, in which genomic testing of tumors improves upon risk prediction with clinical and pathological characteristics in women with breast cancer.[2] Better risk prediction should lead to more targeted use of treatments, that is, that treatment is offered to those for whom the benefits outweigh the harms, such as the burden and costs of tests and treatments.

Ideally, we find predictive markers that relate to the mechanism of action of a treatment. There are at least 2 such success stories in breast cancer. Hormone receptors were discovered about 30 y ago and are key to guiding hormone therapies, which include tamoxifen and aromatase inhibitors.[3] The second success is the discovery of human epidermal growth factor receptor 2 (HER2), a

Corresponding Author
Ewout W. Steyerberg, Department of Biomedical Data Sciences, Leiden University Medical Center, LUMC, Albinusdreef 2, Rm S-5P, Leiden, 2333 ZA, The Netherlands (e.w.steyerberg@lumc.nl).

protein that appears on the surface of some breast cancer cells. For HER2-positive women, trastuzumab (Herceptin) is an effective treatment, a this drug directly targets the HER2 receptor.[4] There is an intensive search for "druggable mutations," with some success in other cancers (e.g., melanoma and lung cancer).[5]

Prognostic markers are defined as markers related to the outcome of a disease irrespective of the treatment. Far more examples can be found for prognostic factors than predictive factors such as HER2. Examples in oncology include the extent of disease (e.g., TNM or other staging systems) and functional status (e.g., indicated by comorbidity).[6] Genomic tests, based on the combination of multiple genetic variants, may also be primarily prognostic rather than predictive tests.[7,8] One such test is a 70-gene signature for women with breast cancer, the Mammaprint, which has been approved for the claim to have a prognostic effect and not for a predictive effect.[9] The Mammaprint was studied recently in a large randomized controlled trial: Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy (MINDACT).[10] The study goal was to help patients and clinicians avoid chemotherapy, which is the standard for these patients, by adding the genomic signature to the clinical evaluation of women. The MINDACT trial concluded that foregoing chemotherapy led to only a minor reduction in survival without distant metastasis among women with early-stage breast cancer who were at high clinical risk and low genomic risk. These women hence might not require chemotherapy. The study had sufficient statistical power to estimate the prognostic effect of the genomic signature but was too small to reliably estimate the effect of chemotherapy.[11] Moreover, clinical risk was categorized as either high or low. Contemporary risk calculators, such as Adjuvant!Online and PREDICT, provide continuous risk predictions.[12,13]

We aimed to explore how evidence on the prognostic strength of a gene signature can contribute to

individualized decision making on chemotherapy for women with early breast cancer. We hereto supplement the MINDACT analysis with a more refined decision-analytic approach, focused on estimating benefit for individual women. We find that the indication for testing and treatment needs to be restricted radically given plausible estimates of the decision threshold and the therapeutic effect of chemotherapy.

Methods

Design of the MINDACT Trial

We based our analyses on data reported from a large randomized controlled trial (MINDACT).[10] Genomic risk was evaluated in 6693 women using the Mammaprint, which had been validated before as a prognostic factor in several studies.[14] The genomic risk was categorized as low or high ("G-low" or "G-high"), excluding those in which the MammaPrint was not feasible, because of too few tumor cells, an inadequate/absent sample, or other reasons.

Clinical risk was evaluated using a modified version of Adjuvant!Online (version 8.0 with HER2 status). Adjuvant!Online has been promoted as a web tool for assessing the risk of breast cancer mortality for individual women. Prognostic factors in Adjuvant!Online include patient information (age, menopausal status, comorbidity) and tumor characteristics (tumor size, number of positive axillary nodes, hormone receptor status).[12]

Prognostic factors from Adjuvant!Online were categorized and linked to a binary classification as "low" or "high" clinical risk. Women at low clinical and low genomic risk did not receive chemotherapy (C-low/G-low), whereas those at high clinical and high genomic risk were recommended chemotherapy (C-high/G-high). Women with discordant risk results (i.e., C-low/G-high, or C-high/G-low) were randomly allocated to chemotherapy or no chemotherapy (discordant pair design). A secondary analysis assessed the effectiveness of chemotherapy, although the study was admittedly underpowered to reliably quantify a small treatment effect.[10] The primary research question in MINDACT focused on the C-high/G-low patients, to test whether the lower boundary of the 95% confidence interval for the 5-y distant metastasis–free survival (5y DMFS) was 92% or higher among the C-high/G-low patients who did not receive chemotherapy. The accompanying editorial specifies the underlying reasoning: if the 5y DMFS is already high, the benefit from chemotherapy cannot be that large that

Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands (EWS, LCdW); Department of Public Health, Erasmus MC, Rotterdam, The Netherlands (DvK); Predictive Analytics and Comparative Effectiveness (PACE) Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA (DvK); and Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, Amsterdam, The Netherlands (PMMB). The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors received no financial support for the research, authorship, and/or publication of this article.

it is worth the burden of treatment.[11] This reasoning is in line with classical decision theory on treatment choice.[15]

Design of the Decision Analysis

We set out to refine the indication for genomic testing and chemotherapy based on absolute benefit from adjuvant chemotherapy for individual women.[16] We hereto performed the following steps:

1. We started with estimating the continuous distributions of clinical and clinical plus genomic risks to closely match the observed risks with binary categorizations in the MINDACT trial.
2. We based the relative effect of chemotherapy on previous randomized trials for individualized estimates of benefit of treatment in terms of absolute 5y DMFS improvement.
3. We estimated the fractions of women who would be reclassified as benefiting from chemotherapy or not given a threshold for minimum required benefit.[16]
4. We then estimated the impact on fraction tested, receipt of chemotherapy, and prognosis and summarized the benefits of targeted treatment according to genomic risk in a summary measure for net benefit (NB).[17]
5. Sensitivity analyses were performed to assess key drivers of the robustness of our findings.

1. Risk distributions. We created a hypothetical MINDACT cohort 1000 times the size of the trial to mitigate the role of chance ($n = 6,693,000$ women, see R code in the Supplementary Appendix). We generated a random log-normal variable to simulate the distribution of clinical risk (defined as $1 - 5y$ DMFS), motivated by the fact that most risk distributions from clinical prediction models follow such a distribution when considered at the appropriate scale.[18] We added the genomic risk result as a correlated random binary variable, such that the higher frequency of women with high genomic risk among those at high clinical risk was closely approximated (Supplementary Table S1). Finally, a binary variable for chemotherapy was created to simulate randomization. We calibrated these risk distributions with the 2.41 relative risk of the genomic test result (details in the Supplementary Material).

2. Benefit of chemotherapy. In MINDACT, the pooled relative risk was 0.88 for chemotherapy, or a relative risk reduction of 12%, combining the results from the 2 discordant groups with inverse variance weighting of the log(HR).

In a base-case analysis, we assumed a relative effect for chemotherapy of 0.8, in line with effects in recent meta-analyses.[19] In sensitivity analyses, we considered relative effects of 0.9 and 0.5, corresponding to pessimistic and highly optimistic perspectives on the effect of chemotherapy, respectively. Individual benefit was the difference in calculated 5y DMFS risk with and without chemotherapy. Absolute benefit was estimated for each of the simulated patients. The median 5-y risks for simulated, MINDACT-like, patients were 3.2% and 4.0% with and without chemotherapy, respectively, or a median absolute benefit of 0.8% (Figure 1).

3. Reclassificaton by genomic risk. We followed clinical consensus in defining the threshold for the minimum benefit of chemotherapy as 3%.[20] Empirical studies eliciting preferences suggested a wide range of required benefits (medians 0.1%, 2%, 5%, and 7%).[21,22] This 3% threshold was also included in a American Society of Clinical Oncology guideline.[23] Here, the relative risk reduction by chemotherapy was assumed to be about 30%. With a baseline risk of 10%, a benefit of about 3% balances the harms of chemotherapy (2%–3% assumed).

4. NB of treatment. The decision-making strategies differed in the fraction tested, the fraction receiving chemotherapy, and the overall prognosis. To summarize strategies, we follow the decision-analytic principle of defining benefits and harms.[17] The increase in 5y DMFS was the benefit of testing, and the burden of chemotherapy is the harm. The decision threshold defines the relative weight w of the burden of chemotherapy to the gain in 5y DMFS. A 3% threshold implies that women are willing to receive up to $1/w = 1/0.03 = 33$ courses of chemotherapy to prevent one 5-y event. The NB is defined as:

$$\text{NB} = \Delta\text{Risk} - w \times \Delta\text{Treated}.$$

Here, ΔRisk is the reduction in 5y DMFS and $w \times \Delta$Treated is a weighted sum of treatments to achieve the reduction in risk, with $w =$ the decision threshold.[17] The NB of "treat none" is 0 (no reduction in risk, no treatment) and serves as a reference to other strategies. We may also state that this formulation of NB reflects the health system NB of treating women in whom the expected 5y DMFS benefit of chemotherapy exceeds 3.0%, which is the clinical expert consensus threshold for chemotherapy treatment.
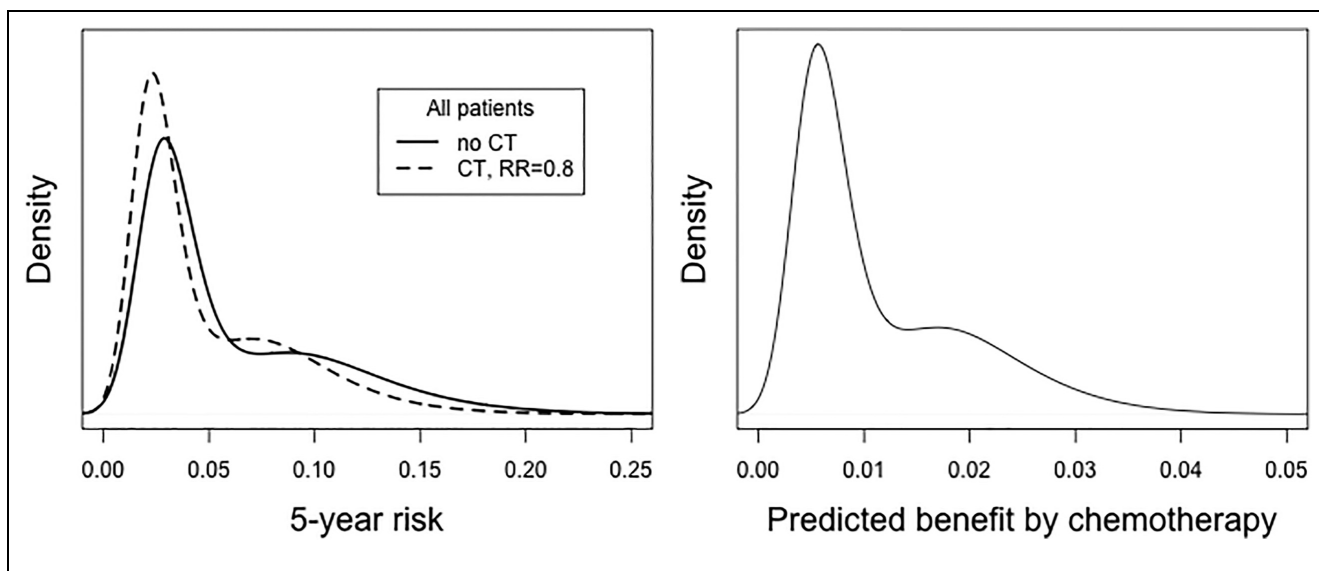
Figure 1 Simulated risk distribution of women enrolled in the MINDACT trial based on clinical and genomic risk, calibrated to MINDACT 5-y distant and metastasis free survival (DMFS), with and without chemotherapy (CT), assuming a relative risk reduction by CT of 20% (hazard ratio = 0.8). The median risks were 3.2% and 4.0% with and without chemotherapy, respectively (left panel), or a median benefit of 0.8% (right panel).

Sensitivity analyses. Key parameters were varied for NB of testing by the Mammaprint and targeted treatment by chemotherapy. Plausible ranges were considered as indicated above.

Results

Key Results from the MINDACT Trial

The 5y DMFS was estimated at 94.7% (95% confidence interval, 92.5%–96.2%) in the C-high/G-low women who did not receive chemotherapy and were in the primary test population.[10] Based on this result, the criterion for noninferiority was met, because the lower boundary of the 95% confidence interval exceeded 92%. In the C-high/G-low women who were randomized and in the intention-to-treat population, the 5y DMFS with chemotherapy was 95.9% (95% confidence interval [CI], 94.0 to 97.2) versus 94.4% (95% CI, 92.3 to 95.9) without chemotherapy. Hence, those with chemotherapy had a 1.5 percentage point higher 5y DMFS than those who did not receive chemotherapy. The adjusted hazard ratio for DMFS with chemotherapy versus no chemotherapy in this group was 0.78 (95% CI, 0.50 to 1.21; $P = 0.27$).[10]

There was no evidence in favor of directing chemotherapy on the basis of genomic risk among randomized patients in the C-low/G-high group. The 5y DMFS was 95.8% (95% CI, 92.9 to 97.6) among those allocated to chemotherapy and 95.0% (95% CI, 91.8 to 97.0%) among those allocated to no chemotherapy. The adjusted hazard ratio for DMFS with chemotherapy versus no chemotherapy in this group was 1.17 (95% CI, 0.59 to 2.28; $P = 0.66$).

The hazard ratio for the genomic test was 2.41 (1.79–3.26) in a Cox proportional hazards model, adjusting for clinical and pathological factors and treatment (Supplementary Table S10.[10] This is the key piece of evidence for the prognostic effect of the test (G-high v. G-low, Table 1).

Interpretation of the MINDACT Trial

The MINDACT investigators concluded that chemotherapy could be avoided in C-high/G-low women, at a cost of a slightly lower 5y DMFS. Women defined as C-high should be tested for genomic risk (3356/6693, 50%). Among these, 1550 had a low genomic risk (23% of the total group) and would not need chemotherapy, whereas 1806 (27% of the total group) had a high genomic risk and would be candidates for adjuvant chemotherapy.[10]
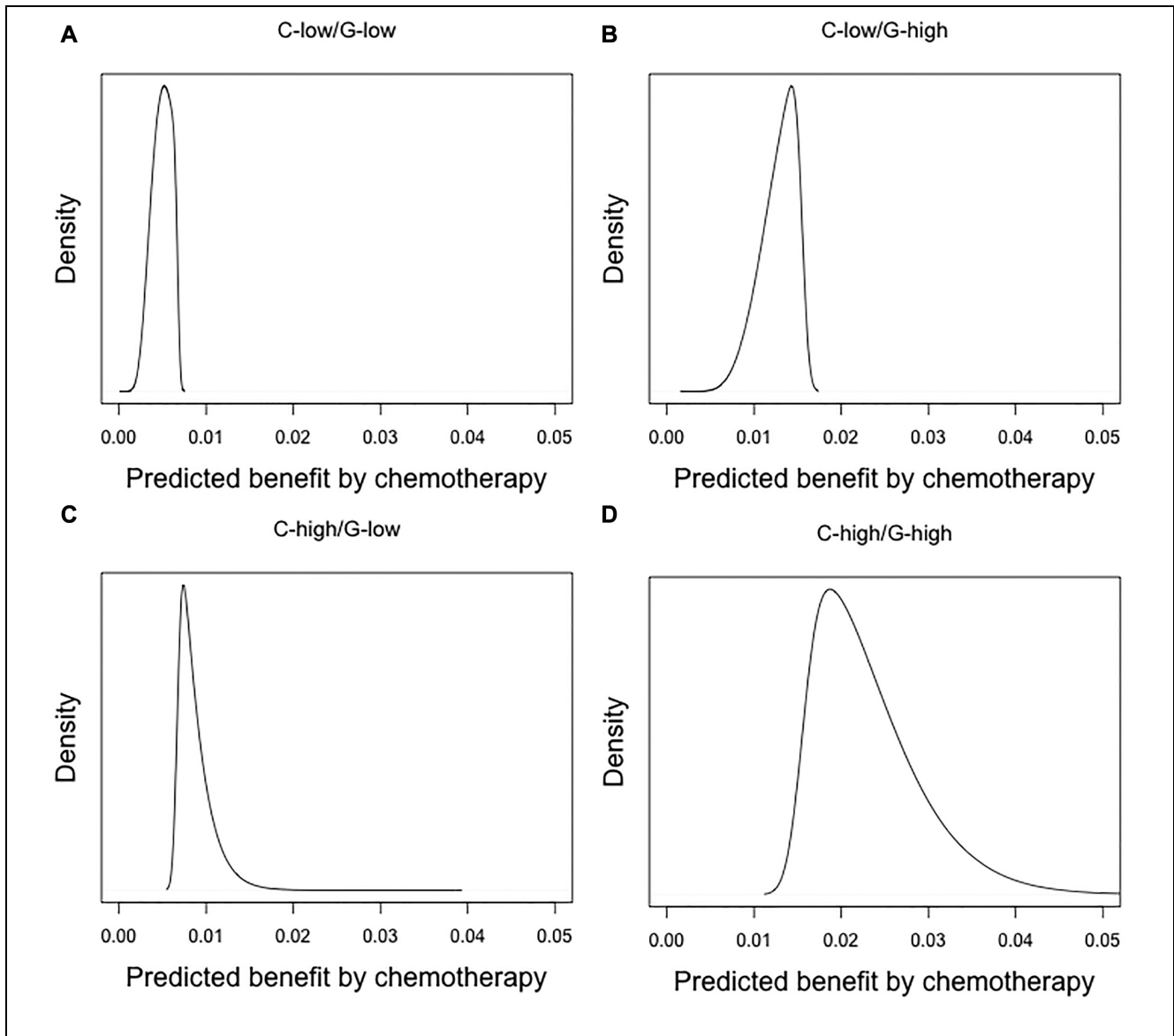
Figure 2 Distribution of predicted absolute benefit in terms of 5-y risk of mortality or distant metastases for women enrolled in MINDACT based on a relative risk reduction by chemotherapy of 20% (hazard ratio = 0.8). Subgroups were defined by combinations of clinical risk (C-low or C-high) and genomic risk (G-low or Ghigh). The median benefits were 0.4%, 1.1%, 0.8%, and 1.9%, respectively.

Results from Decision Analysis

Individual benefit was the difference in calculated 5y DMFS risk with and without chemotherapy (Figure 1). In line with the MINDACT findings, the predicted benefit from chemotherapy was limited for C-low women, irrespective of the genomic risk (Figure 2). The benefit distribution among those in the C-high/G-low group varied widely, with a median of 0.8% (2.5% and 97.5%

percentiles, p2.5 – p97.5: 0.7%–1.4%, Figure 2). The widest benefit distribution was among those in the C-high/G-high group: median 2.2% (p2.5–p97.5: 1.6%–3.8%). Remarkably, 87% of the women in this high-risk group were expected to have less than 3% benefit. Overall, only 3.3% of the women were expected to have more than 3% benefit from chemotherapy considering their risk estimate from the combination of clinical plus

genomic information. Thus, the MINDACT strategy would suggest chemotherapy for far too many women using a 3% threshold for the required benefit and assuming a 0.8 relative effect of chemotherapy.

With clinical risk at a continuous scale, we should administer chemotherapy if the predicted risk without chemotherapy exceeded 16%. Above this risk, the absolute benefit would exceed 3%, assuming a relative effect of chemotherapy of 0.8. Only 2.2% of women were in this category. With the genomic test, 0.3% would be reclassified to less than 3% expected benefit. On the other hand, the test would reclassify 1.3% of women from lower than 3% benefit according to clinical risk to a higher than 3% benefit according to clinical plus genomic risk estimation (Figure 3). Thus, the genomic risk would change decision making in 1.6%, leading to chemotherapy for 3.3% rather than 2.2% based on clinical risk alone.

Impact on Fraction Tested, Receipt of Chemotherapy, and Prognosis

Genomic testing was not indicated for those with predicted benefits less than 2.6% according to clinical risk, because a positive genomic test result ("G-high") would not lead to an indication for chemotherapy (Figure 3). Most women fell into this group (96%). Virtually no women would have such high clinical risk that chemotherapy would also be indicated with a G-low result. Thus, 4.0% of the women would be candidates for genomic testing.

Overall, decision making based on clinical risk would lead to an indication for chemotherapy in only 2.2% (as discussed above), with a 5-y risk of 5.89%. Decision making based on clinical plus genomic risk would lead to an indication for chemotherapy in 3.3% and a 5-y risk of 5.86%. Thus, in 1000 women, chemotherapy would be given to 11 more women (+1.1%), leading to 0.3 more with 5y DMFS (+0.03%). The simpler strategy based on the MINDACT dichotomization of clinical and genomic risk would lead to more testing (for 50% C-high women), more chemotherapy (for 27% in the C-high/G-low group), and a lower 5-y risk (5.36% overall; Table 2).

NB of Treatment

The NB of "treat none" is 0 (no reduction in risk, no treatment) and serves as a reference to other strategies. The NB of "treat all" is negative, since the expected overall risk reduction is below the 3% threshold. Treating only those in the C-high/G-high category implies 270
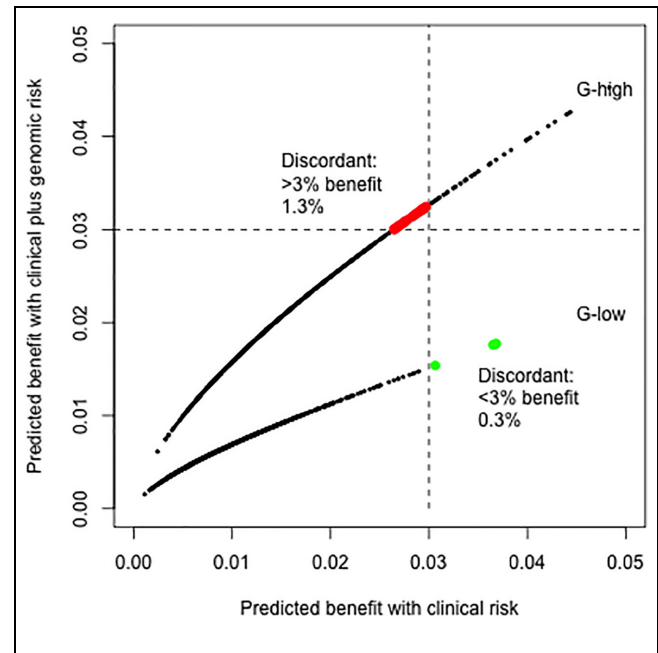

Figure 3 Reclassification by genomic risk. If a 3% threshold (indicated with dotted lines) is used to decide on the use of chemotherapy, 4% of women should be tested for genomic risk, with 1.3% receiving chemotherapy, whereas they would not if decision making were based on clinical risk alone (red dots). In contrast, 0.3% would not receive chemotherapy, although while they would if decision making were based on clinical risk alone (green dots). Thus, the genomic risk would change decision making in 1.6% of the patients. The dots are for a random sample of 2000 hypothetical patients.

women receive chemotherapy for 6.2 per 1000 more surviving without distant metastasis at 5 y (5.98%–5.36% = 0.62%). The NB = 6.2 – 3% × 270 = –1.9 per 1000. Treating the C-high category was also overall harmful (NB = –6.8). The decision-analytic strategies based on clinical risk and clinical plus genomic risk had small positive NBs (+0.17 and +0.18 per 1000 respectively; Table 2). In other words, 1 extra woman would survive at 5 y without distant metastasis per 100,000 for the same number of chemotherapy administrations, if decision making were based on the clinical plus genomic risk rather than clinical risk alone.

Sensitivity Analyses

If absolute 5y DMFS required to make the burden of chemotherapy acceptable were 1% rather than the 3%, the indication for testing and targeted chemotherapy would be much larger (70% and 39%, respectively), leading to lower 5-y risk (5.20% rather than 5.86%). Such a

Table 1 Results for the Discordant-Pairs Design as Applied in the MINDACT Trial versus Decision-Analytic Modeling of Individual Benefit

	Discordant Pairs: MINDACT RCT	**Decision-Analytic Modeling**
Clinical risk	Low vs high Adjuvant!Online 10-y risk; pooled 5y DMFS 97.1% v. 92.7%	Continuous Adjuvant!Online risk
Genomic risk	Nonparametric: subgroups (Supplementary Table S6) C-L/G-L, $n = 2745$: 97.6 [96.9–98.1] C-L/G-H, $n = 592$: 94.8 [92.4–96.4] C-H/G-L, $n = 1550$: 95.1 [93.8–96.2] C-H/G-H, $n = 1806$: 90.6 [89.0–92.0]	Parametric: Cox regression analysis, HR = 2.4
Treatment effect	Underpowered; inverse variance pooled estimate HR = 0.88	From meta-analysis: HR = 0.8
Interaction G × Tx	Nonparametric: subgroups C-H/G-L: HR = 0.78 [0.50–1.21]; ARR 1.5% at 5 y C-L/G-H: HR = 1.17 [0.59–2.28]; ARR 0.8% at 5 y	Assume constant relative effect
Indication for testing	C-high risk: 50% (3356/6693)	Threshold for benefit 3% implies C-risk >16.3%: 4%
Indication for treatment	27% (1806/6693)	If threshold 3%: 3.4%
Further studies	Larger trial Further analyses on risk × chemotherapy effect	Expand clinical risk calculator with genomic risk Further analyses on risk × chemotherapy effect

RCT, randomized controlled trial; C, clinical risk; G, genomic risk; ARR, absolute risk reduction; HR, hazard ratio.

reduction in decision threshold could arise if chemotherapy were better tolerated and significantly cheaper than currently. The associated marginal NB of genomic testing would increase markedly if chemotherapy had such a lower threshold level for DMFS, increasing from 0.01 per 1000 population (0.18 v. 0.17) to 0.70 (3.9 v. 3.2; Table 2).

In contrast, a 5% or higher threshold would imply that none of the patients should be tested nor receive chemotherapy. The MINDACT strategy would have a positive NB for thresholds below 2.3%. At a 1% threshold, the NB would be +0.3 net DMFS event prevented per 1000, as compared with decision making based on categorized clinical risk (NB 3.5 v. 3.2).

If chemotherapy would approximately halve the 5-y risks (HR = 0.5), this would have a major impact as compared with assuming HR = 0.8. Among 1000 patients, 667 would have an indication for testing, 350 would be treated, reducing the 5-y risk to 4.10%, for an NB of +83 (compared with +64 per 1000 for decision making on clinical risk alone). The combination of assuming an HR of 0.5 and a threshold for benefit between 1% and 3% would lead to a positive NB (Figure 4). Conversely, a modest treatment effect such as HR = 0.9 would make testing and treating moot (Table 2) unless a low threshold such as 1% was acceptable (Figure 4).

Assuming a larger prognostic effect of the genomic profile had minor impact. A stronger prognostic strength of the clinical risk model would imply somewhat more clinical value of the genomic test, with more testing, more treatment, and better survival (Table 2).

Discussion

This study evaluated the clinical utility of a genomic profile that is prognostic in women with early-stage breast cancer. The MINDACT trial confirmed the clinical validity of the genomic profile and suggested a simple approach for the clinical implementation: test all women who are at high clinical risk and treat only those with high genomic risk.[10] Unfortunately, this trial was underpowered to reliably answer the question about the impact of changing the indication for treatment directly. Moreover, the definition of "high risk" was arbitrary. Our more detailed decision-analytic modeling approach allowed for a more individualized assessment of the indication for testing, following basic decision-analytic principles.[16,24] We found that far fewer women would be candidates for genetic testing and far fewer women would be selected for chemotherapy compared with the trial recommendation, given the current evidence for the effectiveness of chemotherapy and the preferences of women for such treatment. In fact, following the

Table 2 Fraction Tested, Receiving Chemotherapy, overall 5-y risk of death or Distant Metastasis, and Net Benefit of treatment in a simulated MINDACT population[a]

				Net Benefit (per 1000 Women)	
Strategy	Tested	Chemotherapy	5-y Risk	C	C + G
Main analyses					
Reference strategies					
Treat none	0%	0%	5.98%	Reference	Reference
Treat all	0%	100%	4.84%	−18.6	−18.6
MINDACT strategies					
Treat C-high	0%	50%	5.16%	−6.8	
Treat C-high/G-high	50%	27%	5.36%		−1.9
Decision analytic					
Treat C if benefit >3%	0%	2.2%	5.89%	+0.17	
Treat C + G if benefit >3%	4%	3.3%	5.86%		+0.18
Sensitivity analyses[b]					
Decision threshold					
t = 1%	70%	39%	5.20%	3.2	3.9
t = 2%	25%	17%	5.54%	0.7	1.1
t = 3%	*4%*	*3.3%*	*5.86%*	*0.2*	*0.2*
t = 5%	0%	0%	5.98%	0	0
Chemotherapy effect					
HR = 0.9	0%	0%	5.98%	0	0
HR = 0.8	*4%*	*3.3%*	*5.86%*	*0.2*	*0.2*
HR = 0.5	67%	35%	4.10%	6.4	8.3
Genomic risk effect					
HR = 1.8, AUC + 0.015	2%	2%	5.92%	0.08	0.10
HR = 2.4, AUC + 0.027	*4%*	*3.3%*	*5.86%*	*0.2*	*0.2*
HR = 3.3, AUC + 0.045	7%	6%	5.78%	0.3	0.3
HR = 4.0, AUC + 0.054	8%	7%	5.74%	0.3	0.4
Reference model strength[c]					
AUC = 0.62	0%	0%	5.98%	0	0
AUC = 0.69	*4%*	*3.3%*	*5.86%*	*0.2*	*0.2*
AUC = 0.79	35%	8%	5.59%	0.9	1.1

[a]Results are shown for reference strategies, MINDACT strategies, and decision-analytic strategies, with various sensitivity analyses. NB = Benefit − Harms = ΔRisk $-w \times \Delta$Treated. Benefit is the difference in 5-y distant metastasis–free survival, and Harms is a weighted sum of treatments given to achieve the benefit. The weight w is the treatment threshold t. For example, a threshold of 3% implies that one 5-y event is worth 33 treatments.
[b]Parameters were varied for the "decision analytic, treat C + G if benefit >3%" strategy. The decision threshold is based on the clinical consensus for the increase in absolute distant metastatic–free survival required to make the burden of chemotherapy worthwhile. Numbers were rounded for ease of interpretation.
[c]The linear predictor in the reference model has a hazard ratio of 1 by definition; the prognostic effect could be halved for an area under the curve (AUC) of 0.62 rather than 0.69, or doubled, for an AUC of 0.79.

MINDACT strategy was harmful according to our model (Table 2). The proposed MINDACT strategy had an unfavorable balance of benefits in terms of improved 5-y disease and DMFS and harms in terms of the burden of chemotherapy. Among women in the high clinical and high genomic risk group, 87% did not meet the 3.0% absolute risk improvement in DMFS to justify use of chemotherapy. If we would add the genomic test to the risk estimation in an optimal way, the NB was close to null according to our decision analysis (1 extra net survivor at 5 y without distant metastasis per 100,000).

Limitations

The input for our analyses was from 3 sources, all of which have their limitations.

1. Clinical context: The design of the MINDACT trial specified the criterion of a 5y DMFS >92%. This definition of a threshold for absolute risk is uncommon but could be translated to a minimum required benefit of approximately 1.6% if the relative risk of chemotherapy were 0.8. Such a threshold of 1.6% defines the balance between the harms and benefits
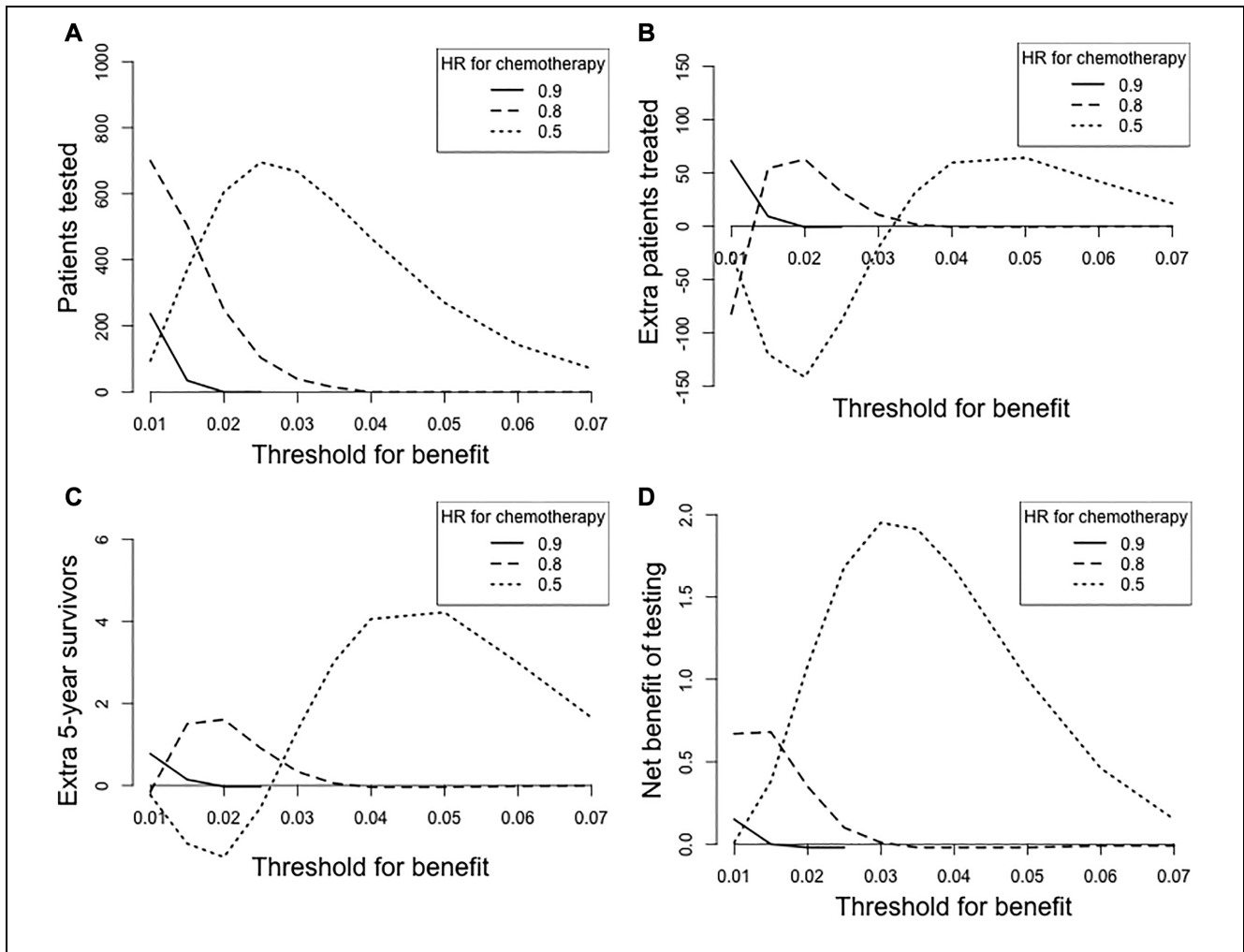
Figure 4 Sensitivity analyses for the "decision analytic, treat C + G if benefit >3%" strategy in Table 2. The threshold was varied between 1% and 7%, with a hazard ratio (HR) for chemotherapy of 0.9, 0.8, or 0.5 (relative risk reductions of approximately 10%, 20%, and 50%, respectively). All numbers are scaled per 1000 women with early breast cancer. For example, for a threshold of 3%, nearly no patients are tested or treated differently with an HR for chemotherapy of 0.8. With an HR for chemotherapy of 0.5, we should test 667 patients (with clinical risk between 1.6% and 5.5%), which would lead to treatment of 20 fewer patients (350 rather than 370), an improvement in 5y DMFS (95.8% to 95.9%, +1.3 per 1000), and a net benefit of 1.9 patients with DMFS (1.3% to 3% × 20 = 1.3 + 0.6 = 1.9 per 1000).

of treatment. This threshold is in contrast to the minimum benefit of 3% in 10-y survival set at the St Gallen convention,[20] which may translate to a 3% threshold for benefit in 5y DMFS. Preferences may vary widely according to various studies, up to a median required benefit of 7% in 10-y survival in one study.[22] The decision threshold is a key consideration in defining the indication for testing and candidates for chemotherapy (Figure 4).

2. Trials on treatment effect: The sample sizes in the discordant groups of the MINDACT trial were too small for reliable estimation of the treatment effect. The summary estimate for the 2 groups in MINDACT was 0.88, not far from estimates in meta-analyses of randomized controlled trials.[19] Ideally, the treatment effect would be based on the MINDACT trial, considering the specifics of the treatment regimens, potential historic changes in this therapeutic area, and application in relatively low-risk women. We note that the relative effect of chemotherapy was small in the recently completed TAILORx trial.[25] The overall effect was 1/1.08 =

0.93 in an intermediate-risk group according to the OncotypeDx 21-gene classifier.

3. MINDACT trial: the MINDACT trial reported on the 5-y risks in 4 distinct groups of patients, based on combinations of clinical risk (high v. low) and genomic risk (high v. low).[10] The relative risk of the genomic profile was studied with multivariable regression analysis (HR = 2.4). We could calibrate the relative risk of the clinical risk to the observed risks in the low and high groups with high accuracy. We hence approximated the continuous risk distribution from Adjuvant!Online, mitigating the categorization used in the MINDACT trial. Ideally, we would have had access to the individual patient data from the trial. Such access was sought and unfortunately denied despite repeated requests.

Sensitivity Analyses

We recognize that the decision threshold will differ between patients, which is why we varied it in sensitivity analyses. In addition, future chemotherapy might be better tolerated than currently available regimens. Moreover, the treatment effect may not be constant across risk groups, although this assumption may be reasonable in many medical domains.[18] In breast cancer, there is substantial interest in how the relative effect of treatment might differ by clinical profile. We might hypothesize that chemotherapy confers a constant absolute risk of harm to all women, irrespective of risk, which would imply a nonconstant relative effect. Our study illustrates the well-known phenomenon that the absolute benefit varies widely between patients, even if we assume that the relative effect is constant.[18,26] Our study also confirmed that the magnitude of the treatment effect was the key factor in clinical utility for patients and far more important than refining the indication for such therapy with a genomic profile with a relatively small prognostic effect. Remarkably, the impact of uncertainty of the prognostic effect of the genomic profile on the clinical utility was rather limited (Table 2). Having a better reference model would have more impact by increasing the number of women with risks close to the decision threshold and hence indicating genomic testing for more. Small improvements can easily be imagined for the current version of Adjuvant!Online, for example, including the rather cheap Ki-67 marker,[27] as included in the PREDICT model.[13] The PREDICT model is presented as an easy-to-use web-based decision support tool (https://breast.predict.nhs.uk/).

Further Research

We recognize that the clinical utility of a genomic test can be assessed in more detail according to analyses that focus on quality-adjusted life-years.[24] Utility estimates for the short-term and long-term burden of chemotherapy are then required, as well as estimates of short-term and long-term survival. In our modeling approach, we tried to stay close to the empirical data from a major study and used the threshold concept to define NB as the summary measure for clinical utility. NB combines the benefit of better DMFS versus the burden of chemotherapy in a single number. If we would perform a full cost-effectiveness analysis, we might also consider direct and indirect medical costs. The costs of the genomic test are nontrivial (more than 2500 Euro in Europe) but may be offset by reduced administration of chemotherapy when considered at a group level.[28]

An immediate improvement over the current MINDACT strategy is to extend current risk calculators with the MammaPrint as a prognostic factor (Table 1). Indeed, the PREDICT model is regularly updated, and work is under way to include the Mammaprint for refined decision making (https://www.health-holland.com/project/2019/improved-tools-for-breast-cancer-treatment-decisions). Given a clinical profile, the likelihood of a positive MammaPrint result should be estimated, and for both a negative and positive result the impact of treatment should be calculated. Such information will allow for shared decision making on ordering the test.[29,30]

In the Netherlands, recent clinical guidelines specify that ordering of the Mammaprint test needs to be discussed. Starting from a simple categorization as "you are at high clinical risk, and this test might reclassify you as low risk such that you don't need chemotherapy" is an oversimplification. The shared decision-making process should focus on the required benefit of chemotherapy. This requires elicitation of a woman's preferences at the individual level. The question then is whether this personal decision threshold is likely exceeded while considering the clinical risk of the individual woman and the revised risk estimate if the MammaPrint would be negative or positive. A threshold of about 3% implies that, for many women, chemotherapy will not be a viable treatment option and that there will be no value in testing. The benefit of chemotherapy will often be too low, even with a positive MammaPrint. Hence, the test often does not need to be ordered.

Further research is also desired on the effect of chemotherapy, both if assumed constant at the relative risk scale as in the current analysis or if assumed to be risk or subgroup specific.[18] Large-scale randomized trials are

required for such estimates, such as the recently completed TAILORx study.[25] An alternative design compared with MINDACT might randomize patients based on a range of expected benefit and allow for evaluation of statistical interaction (risk $\times$ treatment). A fuzzy regression discontinuity design could also be considered.[31,32] This design was followed in the TAILORx trial: low-risk women did not receive chemotherapy, whereas high-risk women did, and intermediate-risk women were randomized.[25]

Conclusions

A pragmatic trial, such as MINDACT, may provide high-quality evidence on the clinical validity of a genomic test. Estimates of the clinical utility depend on the required benefit of treatment. The estimated treatment effect may often be impossible to estimate reliably from a trial with randomization in relatively small discordant groups with few events. Our study illustrates that detailed risk estimation at a continuous rather than dichotomized scale and further decision-analytic modeling is essential to support optimal clinical implementation of trial results evaluating genomic markers.

ORCID iD

Ewout W. Steyerberg (iD) https://orcid.org/0000-0002-7787-0122

References

1. Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med*. 2010;363(4):301–4.
2. Yersal O, Barutca S. Biological subtypes of breast cancer: prognostic and therapeutic implications. *World J Clin Oncol*. 2014;5(3):412–24.
3. Davies C, Godwin J, Gray R, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet*. 2011;378(9793):771–84.
4. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med*. 2001;344(11):783–92.
5. Dietel M, Johrens K, Laffert MV, et al. A 2015 update on predictive molecular pathology and its role in targeted cancer therapy: a review focussing on clinical relevance. *Cancer Gene Ther*. 2015;22(9):417–30.
6. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010; 17(6):1471–4.
7. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Oncol Pract*. 2007;3(6):336–9.
8. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N Engl J Med*. 2009;360(8):790–800.
9. Slodkowska EA, Ross JS. MammaPrint 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Exp Rev Mol Diagn*. 2009;9(5):417–22.
10. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med*. 2016;375(8):717–29.
11. Hudis CA, Dickler M. Increasing precision in adjuvant therapy for breast cancer. *N Engl J Med*. 2016;375(8): 790–1.
12. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001;19(4):980–91.
13. Candido dos Reis FJ, Wishart GC, Dicks EM, et al. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. *Breast Cancer Res*. 2017;19(1):58.
14. Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst*. 2006; 98(17):1183–92.
15. Pauker SG, Kassirer JP. The toss-up. *N Engl J Med*. 1981;305(24):1467–9.
16. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med*. 1980;302(20):1109–17.
17. Vickers AJ, Kattan MW, and Daniel S. Method for evaluating prediction models that apply the results of randomized trials to individual patients. *Trials*. 2007;8:14.
18. Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG, Hayward RA. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *Int J Epidemiol*. 2016;45(6):2075–88.
19. Early Breast Cancer Trialists' Collaborative G. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet*. 2012;379(9814):432–44.
20. Rabaglio M, Aebi S, Castiglione-Gertsch M. Controversies of adjuvant endocrine treatment for breast cancer and recommendations of the 2007 St Gallen conference. *Lancet Oncol*. 2007;8(10):940–9.
21. Duric V, Stockler M. Patients' preferences for adjuvant chemotherapy in early breast cancer: a review of what makes it worthwhile. *Lancet Oncol*. 2001;2(11):691–7.
22. Hamelinck VC, Bastiaannet E, Pieterse AH, et al. Patients' preferences for surgical and adjuvant systemic treatment in

early breast cancer: a systematic review. *Cancer Treat Rev.* 2014;40(8):1005–18.

23. Harris LN, Ismaila N, McShane LM, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology clinical practice guideline. *J Clin Oncol.* 2016;34(10):1134–50.

24. Hunink MGM, Glasziou PP. *Decision Making in Health and Medicine: Integrating Evidence and Values*. Cambridge (UK): Cambridge University Press; 2014.

25. Sparano JA, Gray RJ, Makower DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med.* 2018;379:111–21.

26. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ.* 2013;346:e5793.

27. Jung S-Y, Han W, Lee JW, et al. Ki-67 Expression gives additional prognostic information on St. Gallen 2007 and Adjuvant! online risk categories in early breast cancer. *Ann Surg Oncol.* 2009;16(5):1112–21.

28. Rouzier R, Pronzato P, Chéreau E, Carlson J, Hunt B, Valentine WJ. Multigene assays and molecular markers in breast cancer: systematic review of health economic analyses. *Breast Cancer Res Treat.* 2013;139(3):621–37.

29. Pieterse AH, Baas-Thijssen MCM, Marijnen CAM, Stiggelbout AM. Clinician and cancer patient views on patient participation in treatment decision-making: a quantitative and qualitative exploration. *Br J Cancer.* 2008;99:875.

30. Oshima Lee E, Emanuel EJ. Shared decision making to improve care and reduce costs. *N Engl J Med.* 2013;368(1): 6–8.

31. van Leeuwen N, Lingsma HF, de Craen AJM, et al. Regression discontinuity design: simulation and application in two cardiovascular trials with continuous outcomes. *Epidemiology.* 2016;27(4):503–11.

32. van Leeuwen N, Lingsma HF, Mooijaart SP, Nieboer D, Trompet S, Steyerberg EW. Regression discontinuity was a valid design for dichotomous outcomes in three randomized trials. *J Clin Epidemiol.* 2018;98:70–9.