OPEN

The Story of the Magee Equations: The Ultimate in Applied Immunohistochemistry

Rohit Bhargava, MBBS and David J. Dabbs, MD†*

Abstract: Magee equations (MEs) are a set of multivariable models that were developed to estimate the actual Onco*type* DX (ODX) recurrence score in invasive breast cancer. The equations were derived from standard histopathologic factors and semiquantitative immunohistochemical scores of routinely used biomarkers. The 3 equations use slightly different parameters but provide similar results. ME1 uses Nottingham score, tumor size, and semiquantitative results for estrogen receptor (ER), progesterone receptor, HER2, and Ki-67. ME2 is similar to ME1 but does not require Ki-67. ME3 includes only semiquantitative immunohistochemical expression levels for ER, progesterone receptor, HER2, and Ki-67. Several studies have validated the clinical usefulness of MEs in routine clinical practice. The new cut-off for ODX recurrence score, as reported in the Trial Assigning IndividuaLized Options for Treatment trial, necessitated the development of Magee Decision Algorithm (MDA). MEs, along with mitotic activity score can now be used algorithmically to safely forgo ODX testing. MDA can be used to triage cases for molecular testing and has the potential to save an estimated $300,000 per 100 clinical requests. Another potential use of MEs is in the neoadjuvant setting to appropriately select patients for chemotherapy. Both single and multi-institutional studies have shown that the rate of pathologic complete response (pCR) to neoadjuvant chemotherapy in ER + /HER2-negative patients can be predicted by ME3 scores. The estimated pCR rates are 0%, <5%, 14%, and 35 to 40% for ME3 score <18, 18 to 25, >25 to <31, and 31 or higher, respectively. This information is similar to or better than currently available molecular tests. MEs and MDA provide valuable information in a time-efficient manner and are available free of cost for anyone to use. The latter is certainly important for institutions in resource-poor settings but is also valuable for large institutions and integrated health systems.

Key Words: magee equations™, magee decision algorithm™, onco*type* DX®, breast cancer, molecular testing, Immunohistochemistry, ER, PR, HER2, Ki-67

(*Appl Immunohistochem Mol Morphol* 2023;31:490–499)

Breast cancer is a systemic disease and is treated using multimodality therapy. Surgery and radiation therapy are used for local control of the disease, and subsequently, the majority of breast cancer patients are also offered some form of systemic therapy. Almost all patients with estrogen receptor (ER) positive tumors are offered hormonal therapy, and patients with ER-negative tumors are generally offered chemotherapy. A small percentage of patients with ER+ tumors are also treated with chemotherapy. The tumor size, lymph node status, and the overall receptor profile are critical in deciding the type of systemic therapy. Although patients with multiple positive nodes and higher-stage disease (regardless of receptor status) often receive chemotherapy, the type of systemic therapy offered to patients with lymph node-negative or early-stage disease is often based on receptor profile. The approximate percentage of breast cancer immunohistochemical (IHC) subtypes and the type of systemic therapy offered based on receptor profile is shown in Figure 1. It is evident that ~70% of breast cancers are ER + /HER2-negative and almost all such patients are offered hormonal therapy, but only a small percentage receive and benefit from additional chemotherapy. Which ER + /HER2-negative breast cancer patients will benefit and shall receive chemotherapy is the crux of molecular testing in breast cancer. To understand this dilemma, one has to consider some historical background of chemotherapy use in breast cancer.

An important document that standardized the use of systemic therapy was the National Institute of Health consensus statement of 2000 on adjuvant therapy use in breast cancer.[1] The document stated that it is "accepted practice to offer chemotherapy to most women with lymph node metastases or with primary breast cancers larger than 1 cm (both node-negative and node-positive)" and "decision to consider chemotherapy in node-negative cancers less than 1 cm should be individualized." These statements essentially meant that chemotherapy can be offered to almost any patient with a breast tumor size larger than 1 cm. Importantly, no consideration was given to tumor grade even though the Nottingham group has clearly shown the prognostic value of tumor grading

Received for publication February 24, 2022; accepted August 19, 2022.
From the *Department of Pathology, UPMC Magee-Womens Hospital, Pittsburgh, PA; and †PreludeDx, Laguna Hills, CA.
R.B., declares receiving consultation fee from Agilent Technologies and ImmunoGen. D.J.D. author declare no conflict of interest.
Reprints: Rohit Bhargava, MBBS, Department of Pathology, Magee-Womens Hospital of UPMC 300 Halket Street, Pittsburgh, PA 15213 (e-mail: bharrx@upmc.edu).

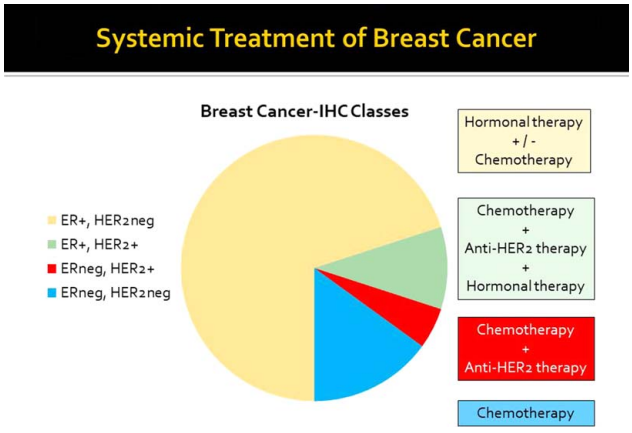

FIGURE 1. Immunohistochemical subclasses of invasive breast carcinoma and their general management.

in breast cancer a decade earlier.[2] Tumor receptor status was also not taken into account; however, one could argue that receptor assays were still in their infancy in the late 90s/early 2000. Although the intention of the National Institute of Health consensus statement was not to leave any patient out of the benefits of chemotherapy, it resulted in gross overtreatment. Around the same time, advancement in genomics was being applied to cancers, starting first with hematologic malignancies and soon thereafter to solid cancers.[3–5] Breast cancer was the first solid cancer to be profiled by gene expression analysis, which further enhanced our understanding of biology and its impact on clinical behavior.[5–7] Breast cancer molecular portraits were described using an "intrinsic" gene set, which essentially correlated to the tumor estrogen receptor (ER) profile. Several other groups independently used gene expression profiling to describe different prognostic groups in breast cancer.[8,9] New tests were designed based on these prognostic classifiers and have since been commercialized. All of these assays were initially devised for prognostic use, but are now routinely used for making therapy decisions in estrogen receptor (ER) positive breast cancers.[8–14] Although these tests appear to be comparable, the results may vary significantly on an individual case level. In a randomized feasibility study called the "optimal personalized treatment of early breast cancer using multiparameter analysis (OPTIMA prelim)", several molecular assays categorized comparable numbers of tumors into low-risk or high-risk groups and/or equivalent molecular subtypes but there was only moderate agreement between tests at an individual tumor level (kappa ranges 0.33 to 0.60 and 0.39 to 0.55 for tests providing risks and subtypes, respectively).[15] However, these results have not dampened the enthusiasm among medical oncologists to frequently request molecular testing in ER+ breast cancer to make chemotherapy decisions, given that this decision may be different depending on the type of test ordered. One of the most commercially successful tests used for this purpose is the Onco*type* DX (ODX; Redwood City, CA, a subsidiary of Exact Sciences Corp., Madison, WI).

ODX is also known as the 21 gene assay (16 cancer-related and 5 housekeeping genes) and was designed to estimate the risk of distant recurrence for patients with ER-positive, lymph node-negative breast cancers when treated with endocrine therapy, specifically tamoxifen. The test is reported as a numerical "recurrence score (RS)" which is calculated using a formula based on gene expression levels of the 16 cancer-related genes and ranges from 0 to 100. Based on earlier studies that utilized tissue blocks from National Surgical Adjuvant Breast and Bowel Project B-14 and B-20 clinical trials, ODX RS categories were defined as low-risk (score 0 to <18, average risk of 7% assuming patient receives tamoxifen for 5 y), intermediate-risk (score 18 to 30, average risk 14%), and high-risk ($\geq$31, average risk approaching 30%).[8,16] These retrospective studies showed the benefit of chemotherapy only in the high-risk group with no benefit in low-risk and negligible benefit in the intermediate-risk group. The test became commercially available in the later part of 2004 and became increasingly popular over the years. It was also endorsed by national societies long before any prospective clinical trial data became available. The prospective trial data was reported in 2018.[17] The Trial Assigning IndividuaLized Options for Treatment was designed to assess the usefulness of ODX testing, which arbitrarily redefined the intermediate-risk group as score 11 to 25. Consequently, patients with scores 0 to 10 received only endocrine therapy, and patients with scores >25 received both endocrine and chemotherapy. Patients with ODX RS 11 to 25 were randomized to receive either endocrine therapy alone or both endocrine and chemotherapy. After 9 years of average follow-up, the recurrence rate and survival were similar between the endocrine-only group and the chemo-endocrine group, concluding that there is a lack of chemotherapy benefit in patients with RS 11 to 25. Since then, the new cut-off for chemotherapy consideration in postmenopausal patients has been 25.

DEVELOPMENT OF MAGEE EQUATIONS

We have followed the course of ODX and other similar testing closely since its inception. We were always intrigued by the similarity of the genes (or the genetic pathways) examined by the molecular tests and the routinely examined pathology prognostic/predictive markers. The first proof of concept was established in 2006 when we examined 42 invasive breast cancers that had available ODX RS (first presented at the 2006 International Academy of Pathology meeting in Montreal, Canada). We found that ODX RS correlated with tumor nuclear grade, mitotic activity, HER2 status, ER, and progesterone receptor (PR) histochemical scores (H-scores).[18] Although this preliminary study was based on a very small number of cases, it allowed a detailed review of cases to provide insight on cases that showed discordant results, viz. grade I tumors with intermediate RS showed lower ER and PR expression, grade II tumors with low RS showed high ER and PR expression, grade II tumors

with high RS showed lower ER/PR expression and/or HER2 overexpression, and grade III tumors with intermediate RS showed high ER/PR expression. Interestingly, none of the grade I tumors showed high RS and none of the grade III tumors showed low RS. These observations were appropriately captured in the regression equation RS = 13.424+5.420 (nuclear grade) +5.538 (mitotic count) -0.045 (ER H-score) –0.030 (PR H-score) +9.486 (HER2 status), which predicted the RS with an R^2 of 0.66, that is the full model accounted for 66% of the data variability.[18]

The original equation was based only on 42 cases. Subsequently, ODX testing was increasingly requested by clinicians on ER+ breast cancers. Our ongoing quality assurance program allowed us to accumulate over 800 cases with histopathologic data and ODX RS between 2004 and 2009. We used these cases to build new models depending on different hypotheses and data availability. Of note, the Ki-67 proliferation index was not available in cases before 2007. The data from these cases resulted in 3 different models, which we named the new Magee Equations (MEs), ME1, ME2, and ME3.[19] ME1 uses Nottingham score, tumor size, and semiquantitative results for ER, PR, HER2, and Ki-67. ME2 is similar to ME1 but does not require Ki-67. ME3 includes only semiquantitative immunohistochemical expression levels for ER, PR, HER2, and Ki-67. An example of how MEs can be calculated is shown in Figure 2. The 3 new equations and the original equation were tested on a separate set of 255 cases (validation set) sent for clinical testing in 2010 and the earlier part of 2011. Case categorization was performed according to ODX risk categories used clinically at that time (< 18, 18 to 30, and ≥ 31). Concordance statistics were performed between actual RS and RS estimated from MEs. Pearson correlation coefficients were also calculated. The results showed the 2-step discordance rate that is, estimated low-risk by MEs with high-risk actual RS or estimated high-risk by MEs with low-risk actual RS was negligible (ranged from 0% to <1%). If the results were predicted to be "intermediate-risk" (score 18 to 30) by MEs, the actual RS was low-risk 85% of the time. The overall correlation between MEs and actual RS was good but less than perfect. There are several reasons for this less-than-perfect correlation/discordance. ODX measures mRNA levels of 16 genes, and MEs use morphologic and IHC data on 4 important proteins highlighting the fact that these are different technologies and some variability is expected. There can be variability in grading and/or semi-quantification of receptor results. However, the problems intrinsic to mRNA extraction are often ignored. Some of this could be operator-related (suboptimal gross dissection of invasive carcinoma), and in other cases, it may be impossible to separate invasive carcinoma with benign ducts/lobules, in-situ disease, inflammatory infiltrate, and biopsy site (Fig. 3). This "contamination" from noninvasive components can sometimes contribute to erroneous ODX and single gene results with grave clinical consequences (Fig. 4).[20,21]

VALIDATION AND MAGEE DECISION ALGORITHM

Soon after the validation, the MEs were made publicly available for anyone to use for free (https://path.upmc.edu/onlineTools/mageeequations.html). Now an app is also available for iOS devices (UPMC Magee Equations on the App Store). Since those original publications, other investigators have examined MEs in their practice and have found them useful.[22–35] We have utilized MEs to further validate and create a decision algorithm for triaging breast cancer cases and also tested its usefulness in the neoadjuvant setting. A brief overview of these internal studies is provided below.

First, we performed a detailed review of cases that showed ODX RS of ≤ 25 with a special focus on cases with RS of ≤ 10. We identified that up to a third of cancers with RS of ≤ 10 are of low-grade special subtype (such as tubular, cribriform, mucinous, classic lobular), and most others are enriched in high hormone receptor content and almost always have a mitosis score of 1.[36] This helped us design an algorithm based on MEs and mitosis score, which we prospectively applied to cases requested for clinical testing. The prospective value study included all clinical requests for ODX testing from the latter half of 2016 to the first quarter of 2018 at our center, resulting in a total of 205 cases.[37] Before sending the tissue block for ODX testing, the case was labeled as "do not send" (DNS) or "send" (SND) based on prespecified rules. The cases were designated as DNS if all 3 MEs showed scores less than 18 (ie, clearly low-risk) or all 3 equations showed scores 31 or higher (ie, clearly high-risk). If any of the equations showed scores 18 to 25, the mitosis score (one of the 3 components of Nottingham grading) was taken into consideration. If the mitosis score was 1, the case was again labeled as DNS. The remaining cases were labeled as SND, that is, cases with any MEs score more than 25 to less than 31 and MEs scores from 18 to 25 but mitosis score of 2 or 3. The algorithm classified 71% of the cases (n = 146) as DNS and only 29% (n = 59) as SND. Of the 146 DNS cases, 143 (98%) were accurately predicted, with the majority being "DNS, expect low-risk" ODX RS. We reviewed the 3 "discordant" cases (ie, predicted low-risk by the algorithm but high-risk on actual ODX testing) and found that the cases had noninvasive tumor factors that may have contributed to increased ODX RS, such as HER2 positivity of in-situ carcinoma and increased Ki-67 proliferation index of biopsy site stroma.[37] On the basis of the results of this prospective value study, we named the algorithm as Magee Decision Algorithm (MDA) and made it available on the ME website (Fig. 5). Since the prospective value study had only a limited number of cases, we continued our prospective evaluation and also decided to retrospectively apply the MDA to our archival cases resulting in a total of 2196 cases.[38] All of the cases have been sent for clinical ODX testing as per oncologists' requests over the years. This prospective-retrospective cohort included patients from 26 to 87 years of age, with a median age of 59 years. Most were early-stage breast cancers. The

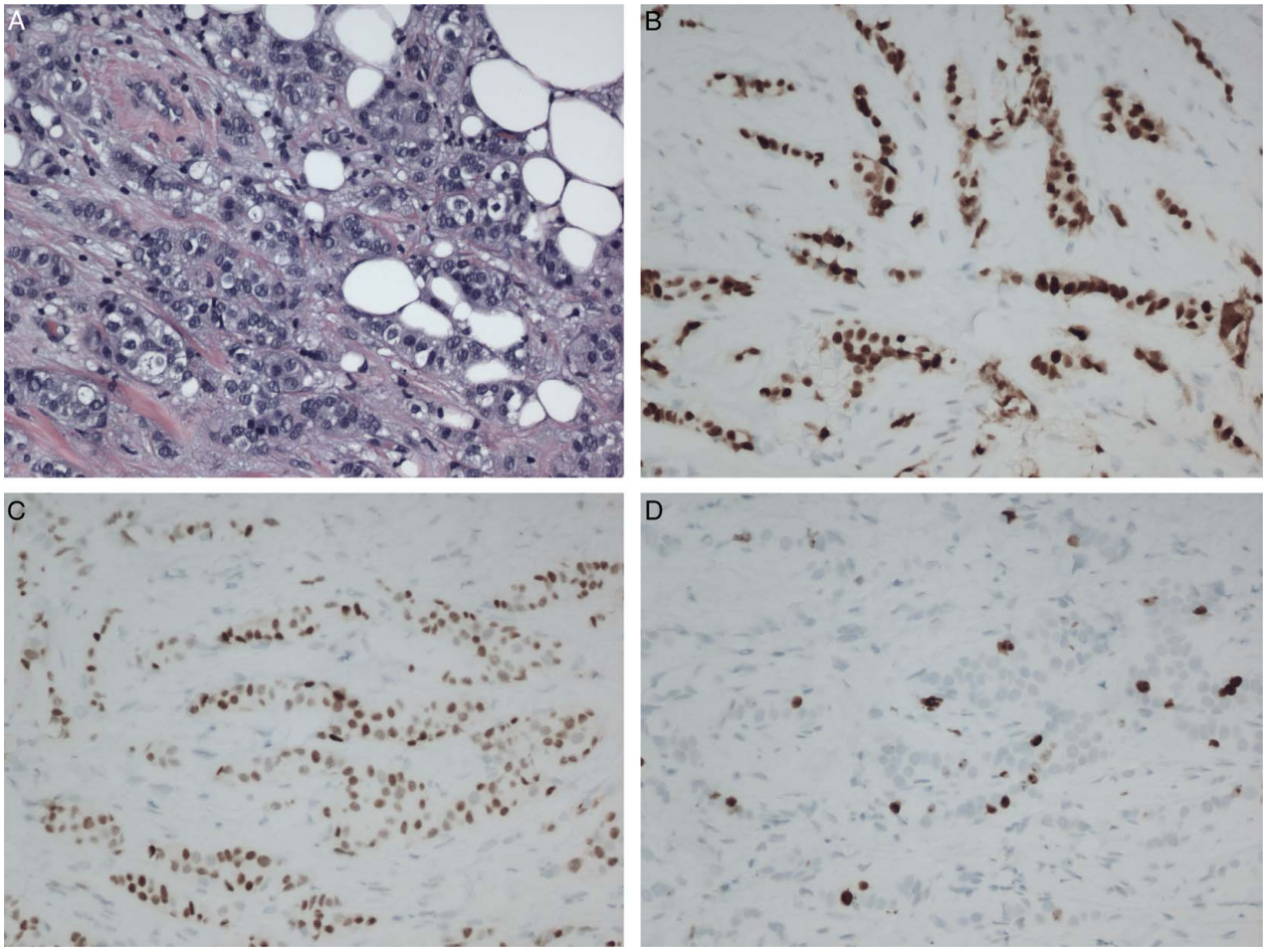

FIGURE 2. An example of how Magee Equations can be calculated. A 2 cm grade II (Nottingham score of 6) ER + HER2-negative invasive breast cancer (A) shows strong ER reactivity (B) with an H-score of 290 (percentage of cells with no staining: 0%, weak 1+ staining: 0%, moderate 2+ staining: 10%, and strong 3+ staining: 90%), moderate PR reactivity (C) with an H-score of 195 (percentage of cells with no staining: 5%, weak 1+ staining: 25%, moderate 2+ staining: 40%, and strong 3+ staining: 30%), and a Ki-67 proliferation index of 10% (D). Based on the morpho-immunohistologic results, the ME1 score is 15.4, ME2 score is 16.1, and ME3 score is 14.2. With all 3 equation scores of less than 18, the likelihood of an actual ODX recurrence score coming back as ≤ 25 is $\geq 95\%$.

median tumor size was 1.6 cm. Of the 2196 cases, 1879 (86%) were lymph node-negative. The 2196 cases included 503 grade 1 (23%), 1352 grade 2 (61%), and 36 grade 3 (16%) tumors. A higher number of grade 2 tumors indicate the selection bias for requesting clinical ODX testing. All cases were ER-positive and 2018 (92%) were PR positive. Unequivocal HER2-positive cases were excluded from this cohort. Interestingly, the results were very similar to the smaller prospective value study, ie, 1538 (70%) of the cases were classified as DNS and 658 (30%) were classified as SND.[37,38] The classification accuracy of the DNS group was 95%, ie, 1462 of 1538 DNS cases predicted accurately. The "discordant" cases (expected ≤ 25 but actual RS > 25; $n = 75$) were evaluated for clinical outcomes. Of these 75 cases, 41 received chemo-endocrine therapy, 2 received chemotherapy only, 26 had endocrine therapy alone (mostly an aromatase inhibitor), and 6 did not receive any systemic therapy. With an average follow-up of 71 months, the distant recurrence-free survival was not worse for patients who received endocrine therapy alone. There were 3 distant recurrences, 2 in patients that received chemo-endocrine therapy and 1 in a patient who did not receive any systemic therapy. No distant recurrences were recorded in the group that received hormonal therapy alone.[38] This indicated that there is no harm to patient care when MDA is used to triage cases for molecular testing.

Since MEs were designed to estimate the ODX RS, it is of no surprise that the equations used either alone or along with the decision algorithm performs very well in ODX RS estimation and triaging patients for further testing. Although we are convinced about the predictive power of MEs as a standalone test based on our routine practice, it is a bit difficult to study systematically because it requires long-term follow-up of patients who have been homogeneously treated. This data is difficult to obtain outside of a clinical trial. One alternative is to apply MEs in the neoadjuvant setting.

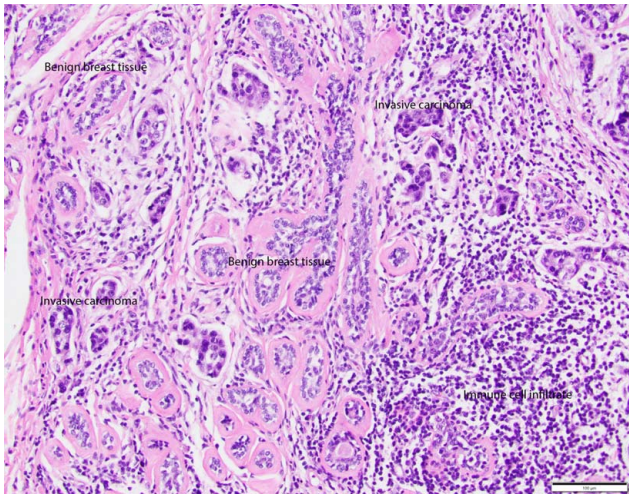

FIGURE 3. An invasive breast carcinoma is often admixed with noninvasive components, which can contribute to quantitative gene expression levels if the invasive component is not carefully micro-dissected.

NEOADJUVANT CHEMOTHERAPY AND ME3

The use of neoadjuvant chemotherapy (NACT) in breast cancer has evolved in the last decade. Once used for inoperable disease, it is now frequently used in early-stage breast cancer. The primary goal is to shrink the tumor significantly so that a tumor that would have required a larger resection (or total mastectomy) would require only a smaller resection (or segmental resection) after chemotherapy.

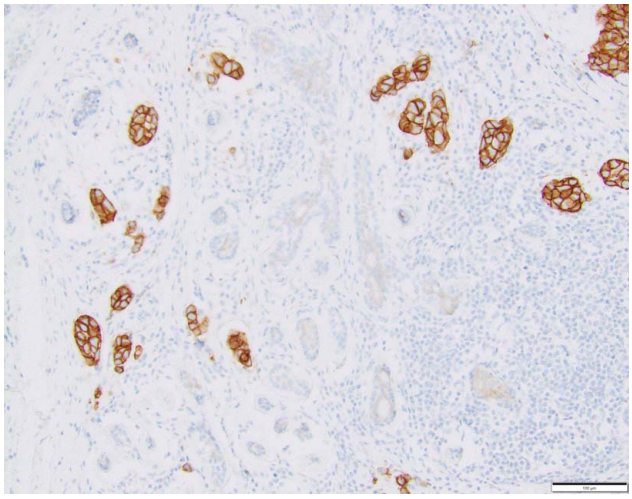

FIGURE 4. Same cases as shown in Fig. 3. An invasive breast cancer with HER2 IHC score of 3+ (positive for HER2 overexpression and eligible for anti-HER2 therapy). This case was also highly amplified by in-situ hybridization (not shown). Although HER2+ cases should not be sent for ODX testing, this HER2-positive case was sent during the initial phase of ODX testing. Single gene score for HER2 was erroneously reported as "equivocal" by the company, likely due to dilution of tumor mRNA with other nontumor components (lymphocytes in this case).

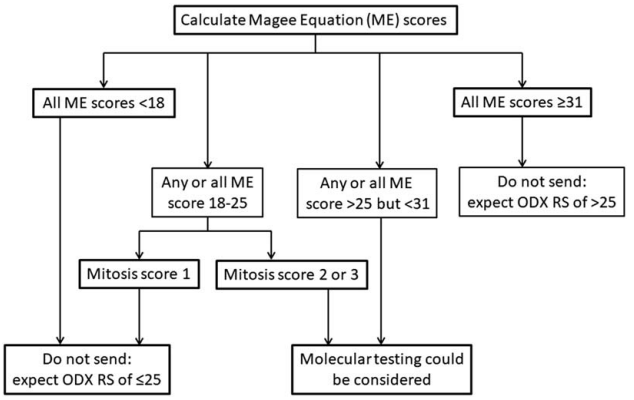

FIGURE 5. Magee Decision Algorithm can be used to triage cases for molecular testing.

Apart from tumor size and lymph node status, tumor receptor status is an important variable in making this decision. There is a selection bias for ER-negative and HER2-positive disease to be considered for NACT. However, the decision to use NACT in ER+/HER2-negative tumors is quite difficult. Often the clinicians have resorted to molecular tests (such as ODX or MammaPrint) in recent years, but the available evidence for such an approach is quite limited. Moreover, these assays are expensive, subject to inconsistent reimbursement in the preoperative setting, may result in a delay in initiating care, may lead to indeterminate results due to limited tissue in core biopsy samples, are not available in resource-poor locations, and are sometimes subject to inaccurate results due to suboptimal microdissection. A readily available alternative in such situations is the application of MEs. Of the 3 equations, ME3 is more widely applicable as it does not require pathologic tumor size and grade information, which is difficult to obtain at the time of pretherapy core biopsy. We studied 237 cases subjected to neoadjuvant therapy at our institution.[39] The cases were divided into a low score (0 to <18), intermediate score (18 to <31), and high score ($\geq$31). The distribution of ME3 scores was low in 65 (27%), intermediate in 116 (49%), and high in 56 (24%). After excluding 46 patients who received neoadjuvant hormonal therapy only, the pathologic complete response (pCR) rate in 191 cases subjected to NACT was 0% with a low score, 5% with an intermediate score, and 36% with a high score, a difference that was statistically significant ($P < 0.0001$). Increasing ME3 score also correlated with an increase in the estimated tumor size/volume reduction in the breast. Average tumor size/volume reduction in low, intermediate, and high ME3 score categories was 46%, 59%, and 77% respectively. Fifty-six percent of patients in the low/intermediate ME3 score category showed >50% tumor size/volume reduction, compared with 80% of the patients in the high ME3 score category ($P = 0.0024$). Moreover, the ME3 scores on pretherapy samples were prognostic. Patients with high ME3 scores had an ~3-fold increased risk of overall recurrence (95% CI 3.26-28.87, $P < 0.0001$) and a 10-fold increased risk of death [95% confidence interval (CI) 1.57-5.25, $P = 0.0016$] compared with those with ME3 scores less

than 31.[39] A subsequent multi-institutional study was undertaken to examine the validity of MEs outside of Magee.[40] In this multi-institutional study, 7 different academic and nonacademic centers participated. The combined dataset of 166 cases showed essentially similar results to the previously published single institutional study. The pCR rate according to ME3 scores was 0% (0 of 64) in ME3 < 18, 0% (0 of 46) in ME3 18 to 25, 14% (3 of 21) in ME3 > 25 to <31, and 40% (14 of 35) in ME3 score 31 or higher (*P*-value: <0.0001). There were no distant recurrences and no deaths in the 17 patients with pCR. In the remaining 149 cases with residual disease, the ME3 score of >25 was significantly associated with shorter distant recurrence-free survival and showed a trend for shorter breast cancer-specific survival.[40] The ME3 results in the neoadjuvant setting are comparable with or even better than many different molecular tests that are used for predicting chemotherapy benefits in ER+ breast cancers (see Table 1).[39–48]

WHY MAGEE EQUATIONS?

Although MEs were the first multivariable model described to estimate the ODX RS, it is not the only model. Since then, several rules, models, and nomograms have been described and published. Most provide a reasonable estimate of ODX RS; however, the accuracy and simplicity of MEs make it easier to use in routine practice to make confident clinical decisions. MEs require and also provide more granular data to make therapy decisions compared with other published models. For example, the University of Tennessee Medical Center (UTMC) nomogram provides the likelihood of a low-risk (≤25) and a high-risk (>25) result rather than an actual number.[49,50] It works well at the extremes of scenarios, but it is difficult to make decisions on the majority of cases in routine clinical practice. This can be illustrated by using a hypothetical case of a 55-year-old woman with a 2.0 cm, grade II (Nottingham score 6, with mitosis score of 1), lymph node-negative, ER+ (H-score of 300), PR negative (H-score 0), HER2-negative, Ki-67 proliferation index of 15% invasive ductal carcinoma. The probability of a low-risk ODX is 53% and the probability of a high-risk ODX is 47% with UTMC nomogram. The estimated ME scores in the same case are 21.5 (ME1), 21.7 (ME2), and 20.6 (ME3). Using the MDA (ME results between 18 and 25 and mitosis score of 1), this case will result in an actual ODX score of 25 or less with over 95% certainty. In such cases, one could forgo ODX testing using MDA, but this decision cannot be taken based on UTMC Nomogram results. Another model that is used frequently in the research setting is the immunohistochemical score 4 (IHC4) model.[51] It is similar to ME3, but there are subtle differences. ME3 was formulated using a database of cases sent for clinical ODX testing (and, therefore representative of the routine clinical

TABLE 1. Pathologic Complete Response Rate to Neoadjuvant Chemotherapy Based on ME3 Scores Compared With Multigene Assay Scores and Other Models

Test	Number (N)	pCR rates	Reference
BCI	N = 94 (only ER + /HER2-neg cases)	Low: 0% (0/50) Intermediate: 10% (3/31) High: 8% (1/13)	Mathieu MC et al Ann Oncol, 2012[43]
BluePrint	N = 403 (both ER+ and ER-negative)	Luminal A: 2% (1/44) Luminal B: 7% (10/145) HER2: 53% (39/74) Basal-like: 35% (49/140)	Whitworth et al Ann Surg Oncol. 2014[47]
EndoPredict	N = 553	Low: 7% (19/283) High: 17% (45/270)	Bertucci F et al Cancer Lett. 2014[41]
ODX	N = 60	Low/Int. risk: 0% (0/36) High risk: 17% (4/24)	Yardley et al BCRT, 2015[48]
Prosigna	N = 180 (all subtypes)	Luminal A: 9% (5/54) Luminal B: 20% (21/105) HER2 enriched: 14% (1/7) Basal-like: 50% (7/14)	Prat A et al Clin Cancer Res, 2016[45]
IHC4	N = 88	Lowest quartile: 0% Highest quartile: 30%	Sheri A et al BCRT, 2017[46]
ODX	N = 989	Low: 2.2% (5/227) Intermediate: 1.6% (7/450) High: 9.6% (30/312)	Pease AM et al Ann Surg Oncol, 2019[44]
EarlyR	N = 659	Low: 5-10% (20-40/400) Intermediate: 10% (7/69) High: 24% (47/190)	Buechler SA et al[42] Breast, 2019
ME3	N = 191	<18: 0% (0/37) 18 to <31: 4% (5/99) 31 of higher: 36% (20/55)	Farrugia DJ et al Mod Pathol. 2017[39]
ME3	N = 166	<18: 0% (0/64) 18 to 25: 0% (0/45) 25 to <31: 14% (3/21) 31 of higher: 39% (14/36)	Bhargava R et al[40] Mod Pathol. 2021

BCI indicates Breast cancer index®; ER, Estrogen receptor; ME3, Magee Equation 3; ODX, Oncotype DX®; pCR, Pathologic complete response.

practice). In contrast, IHC4 was developed using IHC stain scores for ER, PR, HER2, and Ki-67 on tissue blocks from patients enrolled in the Arimidex, Tamoxifen Alone, or in Combination (ATAC) trial. ME3 uses H-scores for ER and PR, but IHC4 uses H-score for ER and the percentage of positive cells for PR. In addition, the Ki-67 proliferation index needs to be modified in the IHC4 score to be used in the equation. IHC4 has been used in research studies and is comparable with ODX in providing information beyond routine clinical pathologic factors. ME3 has been primarily used to provide an estimate of actual ODX score in routine clinical practice and is increasingly used in the neoadjuvant setting to appropriately select ER + /HER2-negative cancer patients who can benefit from chemotherapy. Moreover, the ME calculator is easily accessible for anyone to use (https://path.upmc.edu/onlineTools/mageeequations.html).

CLINICAL IMPLEMENTATION AND CHALLENGES

MEs are simple to use and freely accessible, but individual institutions should validate their results before implementing them into routine practice. This is no different than validating a new antibody before it is made available for clinical use. We believe that for more widespread acceptance, prospective validation studies would be useful, and so would the increasing awareness among clinicians.

Challenges for Consistent ME Results

MEs are dependent on accurate grading and quality IHC results. Therefore, it is paramount that breast cancer grading is performed by pathologists with experience and interest in breast pathology. IHC staining and reporting of breast cancer receptor results may not sound too complicated, but there is potential to falter at many steps of the way. For an appropriate reporting of any IHC assay, laboratories should be careful with pre-analytical, analytical, and post-analytic factors, as deviation from the standard procedure can alter the results to a significant degree. Fortunately, the guidelines set forth by the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) provide a detailed framework on how testing should be performed.[52–56] Of particular importance is maintaining a cold ischemic time of less than 1 hour. Fortunately, most receptor studies are performed on core biopsy specimens that are often transferred immediately to 10% neutral buffered formalin containers. The time within formalin is less critical, but ASCO/CAP recommends tissue exposure to formalin for 6 to 72 hours. Rapid processing and exposure to formalin for less than 6 hours are not recommended. Any prolonged exposure time should be validated. No particular scoring system is endorsed by ASCO/CAP for ER and PR, but the guidelines suggest reporting the results semiquantitatively. For purposes of MEs, H-scores should be used. An H-score can be converted to an Allred score, but an Allred score cannot be reliably converted to an H-score. We have shown good inter-observer concordance for calculating H-scores among breast pathologists.[57] Moreover, the H-scores for ER and PR have been shown to have a linear correlation with ODX quantitative gene expression levels by reverse transcription polymerase chain reaction assay.[58,59] If steps are taken to assure accurate grading and reliable semiquantitation of the receptor results, then MEs can provide consistent results that can be used in routine practice to make chemotherapy decisions. In a prospective randomized trial, Robertson et al queried medical oncologists if the integration of MEs into routine clinical practice affects whether oncologists order ODX testing.[30] They found that in a practice where Ki-67 is routinely available, the addition of MEs into routine clinical practice was not associated with a reduction in ODX use, but the availability of both Ki-67 and MEs did, however, increase physician comfort with systemic therapy decisions. This is an important issue that is misunderstood by many clinicians. Ki-67 is no doubt an important biomarker; however, it should not be used alone (even if accurately measured) in making chemotherapy decisions. This is because, in ER+ tumors, increasing Ki-67 index predicts chemotherapy benefit when it is accompanied by a concurrent reduction in tumor hormone receptor content. This is what Magee Equations measure and therefore are a better predictor of response to chemotherapy than Ki-67 alone. There are no guidelines from ASCO/CAP regarding Ki-67 staining, but the pre-analytical and analytical variables important for hormone receptors are also important for Ki-67 testing. The issue of variability in Ki-67 results has been comprehensively addressed by the International Working Group (IWG).[60–62] IWG has also determined that the best agreement between pathologists occurs by counting 4 random fields of 100 cells each.[61] Although doable, it is certainly time-consuming. We have always taken a pragmatic approach in reporting the Ki-67 proliferation index. We first estimate the Ki-67 proliferation index. If the estimate falls below 10 or above 50, then the estimate stands as the final Ki-67 proliferation index. If the estimate is between 10 and 50, then 50 to 100 cells are counted in a representative area based on the pathologist's discretion to arrive at the proliferation index. This approach seems to have worked in all our prior correlative studies and also in the neoadjuvant setting, where the Ki-67 proliferation index has been used within ME3 to predict chemotherapy benefit. Another important development that has taken place which will further improve ME calculations among pathologists is the availability of QuPath software from IWG for anyone to use for free. The software may require a few tweaks before implementing in clinical practice but could be a useful tool for pathologists who do not perform quantitative scoring regularly. The software can also calculate the H-score in the field of view. However, field selection for counting will remain the responsibility of the pathologist.

Which ME is the Best?

All equations provide a reasonable estimate of ODX RS and generally provide similar scores, except that ME1 is slightly more influenced by tumor size and ME2 is not

dependent on the Ki-67 proliferation index. However, we examined each equation's ability to safely forgo ODX RS testing when used alongside MDA. It is to be noted that MDA requires all 3 equations, but in the MDA validation study, we also analyzed the data using individual equations and the average MEs score.[38] Using individual equations in the decision algorithm slightly increased the percentage of cases classified as DNS, but the accuracy also decreased slightly, particularly impacting the ability to predict "DNS, expect high-risk" category. Although results for each of the equations are comparable, the use of all equations (rather than just 1) slightly increases the accuracy of results and shall increase the user's confidence to safely forgo ODX testing.

Are MEs Prognostic or Predictive?

The molecular assays that are currently used in breast cancer management were initially devised for prognostic use, but are now routinely used for making therapy decisions in ER+ breast cancers. Low-risk patients are spared chemotherapy, while high-risk patients are offered chemotherapy with the hope of reducing the risk of recurrence and improving survival. This approach appears to have merit, but the principle doesn't always hold as there are many patients with a high risk of recurrence whose risk cannot be always minimized with chemotherapy (large tumor size, multiple positive nodes, but low tumor grade with strong hormone receptor expression and low proliferation). With routine usage, it appears that although these tests have both prognostic and predictive value, they are not interchangeable. Some are driven more strongly by estrogen-related features and others by proliferation.[63] Depending on the composition of the multigene panel, some tests are more predictive, and others are more prognostic. Therefore, some multigene assays are now coupled with clinical parameters (such as using EndoPredict clinical score or EPclin over just the EP score) to provide a more accurate assessment of prognosis. However, in doing so, the predictive ability of the test may be compromised. This applies to MEs as well. In our experience, we have noted that response to chemotherapy (predictive value) is best determined by increased proliferation with a concomitant reduction in tumor hormone receptor content which justifies using ME score as a continuous variable. However, if a patient presents with a large tumor, multiple foci, or lymph node positivity, then these tumor features should also be taken into account for estimating prognosis. Although ME scores can still be used to estimate the benefit of chemotherapy, the prognostic estimation will require one to consider all tumor features rather than just the ME scores or the results of molecular assays.

SUMMARY

MEs are a set of multivariable models that are used to estimate ODX RS scores. When combined with the mitotic activity score, these can be used algorithmically (MDA) to safely forgo ODX testing. The use of MDA in routine practice has the potential to save $300,000 per 100 test requests. MEs, particularly ME3 can be used to select ER + /HER2-negative breast cancer patients who will benefit the most from neoadjuvant chemotherapy. MEs can easily be calculated based on the pathology report using a free online web calculator or phone app. Having the ME scores available to the surgeons or oncologists at the time of patient appointment can enhance a patient's experience by eliminating the long wait time for a clinical decision.

ACKNOWLEDGMENTS

The authors would like to thank all pathologists at UPMC Magee-Womens Hospital who have always meticulously reported semiquantitative immunohistochemical scores without which Magee Equations would not have been possible. They also thank all breast cancer patients who came to seek treatment at UPMC Magee-Womens Hospital over the years.

REFERENCES

1. Eifel P, Axelson JA, Costa J, et al. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst*. 2001;93:979–989.
2. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991;19:403–410.
3. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999;286:531–537.
4. Perou CM, Jeffrey SS, van de Rijn M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A*. 1999;96:9212–9217.
5. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747–752.
6. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98:10869–10874.
7. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003;100:8418–8423.
8. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351:2817–2826.
9. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347:1999–2009.
10. Filipits M, Rudas M, Jakesz R, 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;17:6012–6020.
11. Jerevall PL, Ma XJ, Li H, et al. Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. *Br J Cancer*. 2011;104:1762–1769.
12. Ma XJ, Hilsenbeck SG, Wang W, et al. The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer. *J Clin Oncol*. 2006;24:4611–4619.
13. Ma XJ, Salunga R, Dahiya S, et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. *Clin Cancer Res*. 2008;14:2601–2608.
14. Wallden B, Storhoff J, Nielsen T, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics*. 2015;8:54.
15. Stein RC, Dunn JA, Bartlett JM, et al. OPTIMA prelim: a randomised feasibility study of personalised care in the treatment of women with early breast cancer. *Health Technol Assess*. 2016;20:xxiii–xxix. 1-201.

16. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol*. 2006;24:3726–3734.
17. 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;379:111–121.
18. Flanagan MB, Dabbs DJ, Brufsky AM, et al. Histopathologic variables predict Oncotype DX recurrence score. *Mod Pathol*. 2008;21:1255–1261.
19. Klein ME, Dabbs DJ, Shuai Y, et al. Prediction of the Oncotype DX recurrence score: use of pathology-generated equations derived by linear regression analysis. *Mod Pathol*. 2013;26:658–664.
20. Acs G, Esposito NN, Kiluk J, et al. A mitotically active, cellular tumor stroma and/or inflammatory cells associated with tumor cells may contribute to intermediate or high Oncotype DX Recurrence Scores in low-grade invasive breast carcinomas. *Mod Pathol*. 2012;25: 556–566.
21. Dabbs DJ, Klein ME, Mohsin SK, et al. High false-negative rate of HER2 quantitative reverse transcription polymerase chain reaction of the Oncotype DX test: an independent quality assurance study. *J Clin Oncol*. 2011;29:4279–4285.
22. Chen YY, Tseng LM, Yang CF, et al. Adjust cut-off values of immunohistochemistry models to predict risk of distant recurrence in invasive breast carcinoma patients. *J Chin Med Assoc*. 2016;79:649–655.
23. de Lima MAG, Clemons M, Van Katwyk S, et al. Cost analysis of using Magee scores as a surrogate of Oncotype DX for adjuvant treatment decisions in women with early breast cancer. *J Eval Clin Pract*. 2020;26:889–892.
24. Glasgow A, Sechrist H, Bomeisl P, et al. Correlation between modified Magee equation-2 and Oncotype-Dx recurrence scores using both traditional and TAILORx cutoffs and the clinical application of the Magee Decision Algorithm: a single institutional review. *Breast Cancer*. 2021;28:321–328.
25. Harowicz MR, Robinson TJ, Dinan MA, et al. Algorithms for prediction of the Oncotype DX recurrence score using clinicopathologic data: a review and comparison using an independent dataset. *Breast Cancer Res Treat*. 2017;162:1–10.
26. Hou Y, Moosavi HS, Wei L, et al. Magee Equation Recurrence Score Is Associated With Distal Metastatic Risk in Male Breast Carcinomas: Experience From Two Institutions. *Am J Clin Pathol*. 2018;150:491–498.
27. Hou Y, Tozbikian G, Zynger DL, et al. Using the modified magee equation to identify patients unlikely to benefit from the 21-gene recurrence score assay (Oncotype DX Assay). *Am J Clin Pathol*. 2017;147:541–548.
28. Hou Y, Zynger DL, Li X, et al. Comparison of Oncotype DX With modified magee equation recurrence scores in low-grade invasive carcinoma of breast. *Am J Clin Pathol*. 2017;148:167–172.
29. Remoue A, Conan-Charlet V, Deiana L, et al. Breast cancer tumor heterogeneity has only little impact on the estimation of the Oncotype DX(R) recurrence score using Magee Equations and Magee Decision Algorithm. *Hum Pathol*. 2021;108:51–59.
30. Robertson SJ, Ibrahim MFK, Stober C, et al. Does integration of Magee equations into routine clinical practice affect whether oncologists order the Oncotype DX test? A prospective randomized trial. *J Eval Clin Pract*. 2019;25:196–204.
31. Robertson SJ, Pond GR, Hilton J, et al. Selecting patients for Oncotype DX testing using standard clinicopathologic information. *Clin Breast Cancer*. 2020;20:61–67.
32. Saigosoom N, Sa-Nguanraksa D, OC E, et al. The evaluation of Magee Equation 2 in predicting response and outcome in hormone receptor-positive and HER2-negative breast cancer patients receiving neoadjuvant chemotherapy. *Cancer Manag Res*. 2020;12:2491–2499.
33. Sughayer M, Alaaraj R, Alsughayer A. Applying new Magee equations for predicting the Oncotype Dx recurrence score. *Breast Cancer*. 2018;25:597–604.
34. Turner BM, Gimenez-Sanders MA, Soukiazian A, et al. Risk stratification of ER-positive breast cancer patients: A multi-institutional validation and outcome study of the Rochester Modified Magee algorithm (RoMMa) and prediction of an Oncotype DX((R)) recurrence score <26. *Cancer Med*. 2019;8:4176–4188.
35. Turner BM, Skinner KA, Tang P, et al. Use of modified Magee equations and histologic criteria to predict the Oncotype DX recurrence score. *Mod Pathol*. 2015;28:921–931.
36. Dabbs DJ, Clark BZ, Serdy K, et al. Pathologist's health-care value in the triage of Oncotype DX((R)) testing: a value-based pathology study of tumour biology with outcomes. *Histopathology*. 2018;73: 692–700.
37. Bhargava R, Clark BZ, Dabbs DJ. Breast cancers with Magee Equation score of less than 18, or 18-25 and mitosis score of 1, do not require Oncotype DX Testing: A value study. *Am J Clin Pathol*. 2019;151: 316–323.
38. Bhargava R, Clark BZ, Carter GJ, et al. The healthcare value of the Magee Decision Algorithm: use of Magee Equations and mitosis score to safely forgo molecular testing in breast cancer. *Mod Pathol*. 2020;33:1563–1570.
39. Farrugia DJ, Landmann A, Zhu L, et al. Magee Equation 3 predicts pathologic response to neoadjuvant systemic chemotherapy in estrogen receptor positive, HER2 negative/equivocal breast tumors. *Mod Pathol*. 2017;30:1078–1085.
40. Bhargava R, Esposito NN, O'Connor SM, et al. Magee Equations and response to neoadjuvant chemotherapy in ER+/HER2-negative breast cancer: a multi-institutional study. *Mod Pathol*. 2021;34: 77–84.
41. Bertucci F, Finetti P, Viens P, et al. EndoPredict predicts for the response to neoadjuvant chemotherapy in ER-positive, HER2-negative breast cancer. *Cancer Lett*. 2014;355:70–75.
42. Buechler SA, Gokmen-Polar Y, Badve SS. EarlyR signature predicts response to neoadjuvant chemotherapy in breast cancer. *Breast*. 2019;43: 74–80.
43. Mathieu MC, Mazouni C, Kesty NC, et al. Breast Cancer Index predicts pathological complete response and eligibility for breast conserving surgery in breast cancer patients treated with neoadjuvant chemotherapy. *Ann Oncol*. 2012;23:2046–2052.
44. Pease AM, Riba LA, Gruner RA, et al. Oncotype DX((R)) recurrence score as a predictor of response to neoadjuvant chemotherapy. *Ann Surg Oncol*. 2019;26:366–371.
45. Prat A, Galvan P, Jimenez B, et al. Prediction of response to neoadjuvant chemotherapy using core needle biopsy samples with the prosigna assay. *Clin Cancer Res*. 2016;22:560–566.
46. Sheri A, Smith IE, Hills M, et al. Relationship between IHC4 score and response to neo-adjuvant chemotherapy in estrogen receptor-positive breast cancer. *Breast Cancer Res Treat*. 2017;164:395–400.
47. Whitworth P, Stork-Sloots L, de Snoo FA, et al. Chemosensitivity predicted by BluePrint 80-gene functional subtype and MammaPrint in the Prospective Neoadjuvant Breast Registry Symphony Trial (NBRST). *Ann Surg Oncol*. 2014;21:3261–3267.
48. Yardley DA, Peacock NW, Shastry M, et al. A phase II trial of ixabepilone and cyclophosphamide as neoadjuvant therapy for patients with HER2-negative breast cancer: correlation of pathologic complete response with the 21-gene recurrence score. *Breast Cancer Res Treat*. 2015;154:299–308.
49. Orucevic A, Bell JL, King M, et al. Nomogram update based on TAILORx clinical trial results - Oncotype DX breast cancer recurrence score can be predicted using clinicopathologic data. *Breast*. 2019;46:116–125.
50. Orucevic A, Bell JL, McNabb AP, et al. Oncotype DX breast cancer recurrence score can be predicted with a novel nomogram using clinicopathologic data. *Breast Cancer Res Treat*. 2017;163:51–61.
51. Cuzick J, Dowsett M, Pineda S, 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;29:4273–4278.
52. Allison KH, Hammond MEH, Dowsett M, et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. *J Clin Oncol*. 2020;38:1346–1366.
53. Hammond ME, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol*. 2010;28: 2784–2795.

54. Wolff AC, Hammond ME, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol*. 2013;31:3997–4013.
55. Wolff AC, Hammond ME, Schwartz JN, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol*. 2007;25:118–145.
56. Wolff AC, Hammond MEH, Allison KH, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline focused update. *J Clin Oncol*. 2018;36:2105–2122.
57. Cohen DA, Dabbs DJ, Cooper KL, et al. Interobserver agreement among pathologists for semiquantitative hormone receptor scoring in breast carcinoma. *Am J Clin Pathol*. 2012;138:796–802.
58. Jeon T, Kim A, Kim C. Automated immunohistochemical assessment ability to evaluate estrogen and progesterone receptor status compared with quantitative reverse transcription-polymerase chain reaction in breast carcinoma patients. *J Pathol Transl Med*. 2021;55:33–42.
59. Kraus JA, Dabbs DJ, Beriwal S, et al. Semi-quantitative immunohistochemical assay versus oncotype DX(R) qRT-PCR assay for estrogen and progesterone receptors: an independent quality assurance study. *Mod Pathol*. 2012;25:869–876.
60. Dowsett M, Nielsen TO, A'Hern R, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst*. 2011;103:1656–1664.
61. Leung SCY, Nielsen TO, Zabaglo LA, et al. Analytical validation of a standardised scoring protocol for Ki67 immunohistochemistry on breast cancer excision whole sections: an international multicentre collaboration. *Histopathology*. 2019;75:225–235.
62. Polley MY, Leung SC, McShane LM, et al. An international Ki67 reproducibility study. *J Natl Cancer Inst*. 2013;105:1897–1906.
63. Buus R, Sestak I, Kronenwett R, et al. Molecular drivers of Oncotype DX, Prosigna, endopredict, and the breast cancer index: A TransATAC Study. *J Clin Oncol*. 2021;39:126–135.