European Journal of Cancer 175 (2022) 187–192

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Current Perspective

Selecting patients with HER2-low breast cancer: Getting out of the tangle

Ximena Baez-Navarro [a,*], Roberto Salgado [b,c], Carsten Denkert [e], Jochen K. Lennerz [h], Frédérique Penault-Llorca [f], Giuseppe Viale [g], John M.S. Bartlett [d], Carolien H.M. van Deurzen [a]

[a] *Department of Pathology, Erasmus MC, Rotterdam, the Netherlands*
[b] *Department of Pathology, GZA-NZA Hospitals, Antwerp, Belgium*
[c] *Division of Research, Peter MacCallum Cancer Centre, Melbourne, Australia*
[d] *Cancer Research UK Edinburgh Centre, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK*
[e] *Institute of Pathology, Philipps-University Marburg and University Hospital Marburg (UKGM), Baldingerstr. 1, 35043 Marburg, Germany*
[f] *Centre Jean Perrin, Université Clermont Auvergne, INSERM, U1240 Imagerie Moléculaire et Stratégies Théranostiques, F-63000 Clermont Ferrand, France*
[g] *Department of Pathology, European Institute of Oncology IRCCS, University of Milan, Milan, Italy*
[h] *Department of Pathology, Center for Integrated Diagnostics, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA*

Received 15 July 2022; received in revised form 10 August 2022; accepted 21 August 2022

KEYWORDS
HER2-low;
Breast cancer;
Immuno histochemistry;
Biomarker;
Trastuzumab-deruxtecan

Abstract The promising effect of antibody–drug conjugates on breast cancer with low expression of HER2 (HER2-low) raises many questions regarding the optimal selection of patients for this treatment. A key question is whether HER2 immunohistochemistry, an assay optimised to detect HER2 amplification, is reliable enough to assess HER2 protein levels to select patients with HER2-low breast cancer in daily pathology practices worldwide. Moreover, whether this assessment can be performed with sufficient reproducibility between pathologists in daily practices is debatable. Herein, we address the historical track record of the CAP-ASCO HER2 Guidelines, the reported limited reproducibility by pathologists of HER2 immunohistochemistry in the non-amplified cases, and the performance variation of different antibodies. Based on this summary, we propose solutions to improve the robustness to enable reliable identification of patients with HER2-low breast cancer.

* *Corresponding author*: Erasmus MC Cancer Institute, Dept. of Pathology, PO Box 2040, 3000 CA, Rotterdam, the Netherlands.
E-mail address: x.baeznavarro@erasmusmc.nl (X. Baez-Navarro).

https://doi.org/10.1016/j.ejca.2022.08.022

1. Historical and current HER2 testing

The implementation of targeted therapies against the human epidermal growth factor receptor 2 (HER2) has considerably improved the outcome of patients with HER2-positive breast cancer (BC) [1]. These HER2-targeted therapies are effective in patients with HER2 amplification or overexpression, while there is no benefit for patients without this alteration. The American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) guidelines were developed in 2007 and have been reviewed and updated in 2013 and 2018, aiming to standardise the detection of HER2-positive BC patients that could benefit from HER2-blockade [2–4].

Novel treatment options have been developed, including antibody–drug conjugates (ADCs). Some of these drugs target cancer cells with low levels of HER2 (HER2-low), such as Trastuzumab-deruxtecan (T-DXd), which is already recommended for HER2-low metastatic BC in NCCN guidelines [5,6]. Several clinical trials use T-DXd in patients with HER2-low metastatic BC that progressed during endocrine therapy and/or that have previously been treated with chemotherapy [7,8]. The DESTINY-Breast(DB)-04 study shows a superior progression-free survival and overall survival in patients with HER2-low BC treated with T-DXd compared to standard chemotherapy [7]. The DB-06 study, focused on comparing the use of T-DXd versus investigator's choice chemotherapy in HER2-low hormone receptor positive metastatic BC, is ongoing [9]. In this study, patients with very low levels of HER2 expression, the so-called ultralow category, are also included but those with no detectable HER2 protein expression remain excluded.

2. ASCO/CAP guidelines for HER2 testing over the years

The ASCO/CAP guidelines for HER2 testing were published in 2007, 2013 and 2018 respectively [2–4]. Regarding HER2-low, the most important adaptation is on the changing definition of HER2 0 and 1+ scores by immunohistochemistry (IHC). According to the 2007 ASCO/CAP guideline, IHC 0 was defined as an absence of membrane staining and the 1+ score was defined as a weak, incomplete membrane staining in any proportion of tumour cells or weak complete staining in <10% of cells [3]. In the next ASCO/CAP guidelines, published in 2013 and 2018, these cut-off points were modified. The definition of IHC 0 was adapted to either no staining or membrane staining that is incomplete and is faint/barely perceptible within <10% of the invasive tumour cells [2,4]. The HER2-ultralow group, as defined in the DB-06 study (IHC >0, <1), is also included in this category, which would be classified as IHC 1+ according to the 2007 guidelines, but as IHC 0 according to the 2013 and 2018 version [9]. Currently, these historical changes have major consequences.

3. Can assays used for identification of HER2-overexpressing cancers be used to select patients for ADC targeted approaches?

Xu et al. showed a broad range of HER2 expression by mRNA between tumours that have no detectable HER2 by IHC [10]. mRNA levels for tumours across 'HER2-low' cancers were overlapping. That is some tumours with no detectable IHC expression of HER2 have mRNA levels equivalent to, if not higher than, some tumours with IHC1+ or even 2+ cancers. These findings are consistent with that IHC assays currently in use for detecting cancers eligible for HER2 targeted (non-ADC) therapies are calibrated to detect levels of HER2 expression 100–10,000 fold above normal breast, and the levels of HER2 required for ADC response are at, or below, the lower limit of detection for these assays [11].

4. Current challenges in the assessment of HER2-low BC (Fig. 1)

There is a 'physiological' expression of HER2 on all BC cells and in the normal breast, but as discussed above current IHC assays are not calibrated to stain this physiological level of expression in normal cells [11]. A critical issue is the fundamental lack of understanding as to what 'HER2-low', let alone the 'HER2-ultralow' cancers actually represent. Tarantino and colleagues demonstrated an enrichment of ER-low tumours among HER2-low tumours, suggesting that the prognostic analysis of HER2-low patients may be confounded by the adverse effect of low ER-expression [12]. In absence of a definition of HER2-low disease, we are confronted with a pathological conundrum: if we cannot define it, how can we measure it?

It is clear that the IHC assay currently used for HER2 detection is incorrectly calibrated to be used in this setting of HER2-low. It would be like using a weighbridge calibrated for weighing elephants when you are trying to accurately weigh mice. Additionally, the influence of pre-analytical factors on expression-levels in the low-HER2 end, like the variations in several metastatic sites and optimal fixation time of the breast specimens is still unknown [13].

Some previous studies have evaluated the interobserver agreement on HER2 scoring. Schettini et al. reported a series of 100 cases evaluated by 5 expert breast pathologists according to the 2018 guidelines [14]. The overall agreement was good (kappa = 0.79), however, there were 35 discordant cases, and the highest number of discordances were observed between IHC 0 versus 1+ (15 cases). Recently, Fernandez and colleagues reported data from the College of Pathologists, where there was only 26% concordance between 0 and 1+ [15].

Furthermore, Scott et al. reported that the Ventana 4B5 antibody clone on the Ventana autostainer

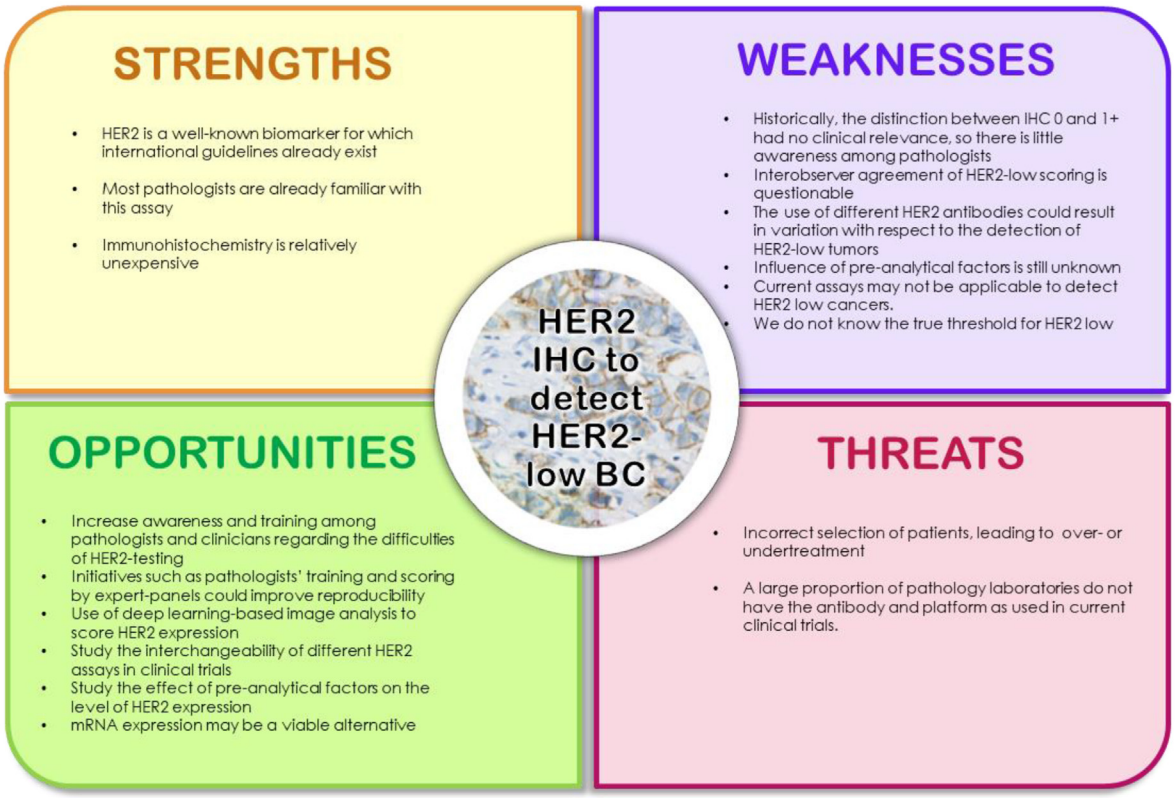

Fig. 1. Strength, Weakness, Opportunity and Threat (SWOT) analysis regarding HER2 immunohistochemistry as an assay to detect HER2-low breast cancer.

identified a higher proportion of HER2-low cases than the HercepTest (27.4% versus 9.2%) [16]. Notably, the DB-06 study uses the Ventana HER2 4B5 assay on the Ventana autostainer.

5. How can we get out of the tangle (Table 1)?

Do we have the right biomarker, but the wrong assay? As with many clinical challenges, problems arise when the complexity of the issue is poorly understood.

The critical questions

a) Is target expression level actually the right metric for us to consider the selection of patients for treatment with ADCs? If the answer is yes, do we know what level of expression is the correct threshold?
b) What is the appropriate assay technology to select cases with sufficient target expression for ADC therapy?
c) How do we ensure that assays developed are calibrated for accurate assessment of target expression?

To date, there are limited data which address point (a) above, which perhaps explains the ongoing controversy around points (b) and (c). More broadly, clinical trials assess the safety and efficacy of an intervention (e.g. a drug) and biomarker-stratified trials simultaneously serve to validate a biomarker, based on a biomarker–treatment interaction effect. Clinical trials therefore do not necessarily serve to validate the assay of the vendor used in the trial as the sole determinant of treatment response. Similarly, the concept of centralised testing used in many trials, does not adequately reflect testing in the daily practice of pathologists and oncologists.

Increasing the awareness among pathologists and oncologists on the importance of an adequate detection of HER2 scores 0, 1+ and 2+ is a start. There may be three main elements attributing to the observed discordances between pathologists and assays, namely: 1. We don't have a clear definition of the correct expression level to target for ADC response; 2. The analytical validity of the assay as defined by the limit of detection, sensitivity, specificity, etc. and these may not cover the thresholds needed; and 3. the relevance of asking the pathologist to distinguish HER2 scores of '0' and '1+', based on staining in 10% of tumour cells when seeking to identify 'HER2-low' cases. It seems plausible that a HER2-score of '0', when there's no staining at all, might be easier to assess by pathologists than a staining that has between 0% and 10% stained tumour cells.

Is a change of the ASCO/CAP Guideline back to the 2007 criteria foreseeable as part of the solution? The DB-trials did not include patients without HER2 expression (DB-06 enrols ultralow, but excludes non-expressors), so the optimal lower threshold for HER2 expression remains unclear. In the DAISY trial, patients

Table 1
Proposed Solutions to improve the HER2-low assay narrative.

Proposed Solutions to improve the HER2-low assay narrative
Increase awareness among pathologists on the importance of appropriate assessment of HER2 scores 0, 1+ and 2+ and potential issues surrounding HER2-testing
Industry should perform in concert with all stakeholders bridging concordance studies using DESTINY-trial samples, with other HER2-assays, and using standardised controls before a drug is definitively approved.
Bridging concordance studies between assays and interpretation of the results hereof should be done in an impartial manner, and results should not be interpreted merely and only in support of the companion diagnostic if the data suggests otherwise.
Concordance studies between pathologists should be based on a sufficient number of participating pathologists, with appropriate statistical guidance using documents such as for example the FDA-guidance 'Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic tests'.
Industry should support, in concert with all stakeholders, relabelling or revising approved HER2-companion diagnostics if there is evidence that the labelling may lead to uncertainty in the identification of patients for treatments.
ASCO/CAP may reconsider the 2007 definition of HER2 score 0, namely without any expression at all, contrary to the 2013 and 2018 definition of HER2 0 that necessitates the evaluation of 10% of tumor cells for a HER-score of 0.
Industry should support, in concert with all stakeholders, relabelling or revising of the HER2 companion diagnostics if equivalent clinical validity has been demonstrated with other HER2-related biomarkers, providing access to clinical trial tissues to validate other assays.
Pathways for regulatory acceptance of other HER2-assays that are equivalent to the companion diagnostic, but less expensive and easier to implement in daily practice, should be developed by governments and regulatory agencies ideally before a drug is labelled together with a companion diagnostic.
Early engagement by all stakeholders in External Quality Control Schemes on HER2-assays to allow rapid development of guidelines and quality standards is essential, preferably before an assay is approved by the regulatory agencies.
Clinical practice guidelines developed by professional organisations like ASCO, ESMO, etc., should endorse not just a companion diagnostic assay used in the trial, but any rigorously technically validated equivalent laboratory assays that can define substantially the same population as the companion diagnostic.
Regulators should require data confirming the analytical validity of the companion diagnostic in the distributed setting in which it would be applied, at a level of rigour comparable to that required to show efficacy of the drug in question. This includes making public all details of the analytical validity data by the central laboratory on the assay used in the trial.

with IHC 0 according to current ASCO/CAP guidelines were included (n = 40) [17]. In this study, 30% of patients with IHC 0 tumours (<10% expression) responded to T-DXd [17]. This raises the question whether every patient could potentially benefit from this therapy, albeit to a different extend, which in turn puts the clinical relevance of differentiating IHC 0 versus IHC 1+ into perspective. Re-analysing the IHC 0 cohort in the DAISY trial using a redefined HER2 0 score with no expression at all, would be interesting, although the group of non-expressors is likely too small to draw firm conclusions. Clearly, more information about this subset is needed.

The Ventana HER2 4B5-assay should not be considered as the only performant 'companion diagnostic' for accurate HER2-low assessment by regulatory and reimbursement agencies, insurance companies and by the scientific, clinical and pathology community, simply because it was the assay used in the DB-trials [18]. The DB-trials have validated HER2-low as the biomarker and has not necessarily analytically validated the assay of the vendor. As for PDL1, the question on interchangeability of different assays arises. The different assays may not be technically fully interchangeable as the assays may be different, but nowadays, using appropriate reference materials, we know that most, but not all PDL1-assays have a high degree of correlation. So, even if technically speaking the assays are not interchangeable, they may be interchangeably used in daily practice. This validation should be developed in a partnership between academia and industry. This concept also applies to HER2, and to all biomarkers. For example, for ER-assessment in BC, there's not one assay for treatment of a patient with exemestane, another one for treatment with tamoxifen, another one for fulvestrant. As absurd as it sounds, the HER2 situation is quite similar. In daily practice one HER2-assay is used for all different HER2-drugs. Conceptually, this reasoning would have impact on the workflow in daily practice since many pathology laboratories do not have the Ventana Platform, which was used in the DB-trials. Under current value-based healthcare paradigms, neither the acquisition of another staining system, nor outsourcing with delays in care seem justifiable. What are pathology laboratories supposed to do? In this context, the IMPassion130-study has assayed different PDL1-assays, and found that all predicted benefit, albeit with slightly different hazard ratios eliciting debate in the academic community on how to interpret these findings [19]. This emphasises again the importance of a correct use and interpretation of statistical methods to inform on the performance of assays. Importantly, different hazard ratios of benefit to treatment nor metrics as median overall survival for different assays inform on the performance of the assay. Therefore, it is highly recommended to perform a similar analysis in the DB-04 trial, using different HER2-assays, as used in the daily practice, and present this data to the regulatory authorities, next to the assay used in the trial. Currently, different

HER2-assays from different vendors from trials of different companies are being used in different tumour types, and not all use the same criteria to define HER2 low. The consequences of these practices for the clinical and pathology practices around the globe should be mitigated by the scientific community, and this includes partnering with the industry and the regulatory authorities as it is increasingly evident that the same mistakes as for the PDL1-saga are being repeated with the HER2-low paradigm.

Regarding concordance-assessment, there is currently no standard for approach for concordance assessment between pathologists. Most often Kappa-values are used. However, Kappa-values are rarely informative on the clinical impact of the observed (dis)concordance [20]. For example, the interrater reliability of auscultation of breath sounds among physical therapists is poor [21], yet nobody will claim that we should stop listening to the lungs of our patients. We recommend using publicly available FDA-Guidance for the evaluation of diagnostic tests [22].

Nevertheless, IHC interpretation remains semi-quantitative and observer-dependent, making it susceptible to human error. The use of deep learning-based image analysis removes some, but not all of these limitations. Moutafi et al. and Gustavson et al. have implemented deep learning–based image analysis model to generate a novel continuous score to quantify HER2 [23,24]. But also here, caution and much more efforts are needed to develop and standardise deep-learning approaches that can be extrapolated to different hospital settings, since there is still no HER2-deep learning assay available that has demonstrated robustness if implemented in different hospital settings.

6. Conclusion

Novel ADCs are a promising therapeutic option for many BC patients. However, there are numerous issues that need to be addressed urgently to optimise the assessment of low HER2 expression to reliably select those patients that could benefit from these drugs. Here we propose to the scientific community, industry as well as the regulatory authorities a set of tools to mitigate the described issues (Table 1). To avoid further chaos akin to the situation with PDL1 we need to learn from our previous experience, partner with industry and the regulatory authorities to optimise identification of patients as accurate and reasonably possible for treatment with powerful drugs such as T-DXd.

Author contributions

XBN, RS and CvD performed the conceptualisation of the article, wrote the manuscript, and designed the tables and figures. RS, CD, JKL, FPL, GV, JMSB contributed with relevant and updated literature, and provided an expert opinion on the topic. All the authors revised and approved of this manuscript.

Conflict of interest statement

CvD received funding from AstraZeneca. The funder had no role in the design or writing of this manuscript.

RS reports non-financial support from Merck and Bristol Myers Squibb (BMS), research support from Merck, Puma Biotechnology and Roche, and personal fees from Roche, BMS and Exact Sciences for advisory boards.

JB reports consultancies from Insight Genetics, Inc. BioNTech AG Biotheranostics, Inc. Pfizer, Rna Diagnostics Inc., oncoXchange/MedcomXchange Communications Inc. Herbert Smith French Solicitors and OncoCyte Corporation. He is involved in a scientific advisory board of MedcomXchange Communications Inc. He received honoraria from NanoString Technologies, Inc. Oncology Education, Biotheranostics, Inc, MedcomXchange Communications Inc. and research funding from Thermo Fisher Scientific Genoptix, Agendia, NanoString Technologies, Inc. Stratifyer GmbH Biotheranostics, Inc. Travel, accommodation expenses were received from Biotheranostics, Inc., NanoString Technologies, Inc, and the Breast Cancer Society of Canada. He reports the following patents: 1) CIN4 predicts benefit from anthracycline, National Phase Application, (Canada, Jan. 11, 2017), 2) Systems, Devices and Methods for Constructing and Using a Biomarker, National Phase Application, 15/328,108 (United States, Jan. 23, 2017); 15824751.0 (Europe, Jan. 12, 2017); (Canada, Jan. 12, 2017), 3) Targeting the Histone Pathway to Detect and Overcome Anthracycline Resistance (IP Title), National Phase Application, PCT/CA2016/000247, Patent Application #: 3000858 (Country of filing: Canada – Patent Application Date: Apr. 4, 2018) and 4) Immune Gene Signature Predicts Anthracycline Benefit, PCT (international application), PCT/CA2016/000305, Filing date: Dec. 7, 2016.

FPL reports personal fees from Astrazeneca, BMS, Daiichy-Sankyo, MSD, Roche, Seagen, Diaceutics, Veracyte, research support from Astrazeneca, Daiichy Sankyo, Roche, MSD and BMS.

GV reports personal fees from AstraZeneca, BMS, Daiichy-Sankyo, MSD, Roche, Seagen, Pfizer, and research support from AstraZeneca and Roche.

All remaining authors have declared no conflicts of interest.

Acknowledgements/funding

RS is supported by the Breast Cancer Research Foundation (BCRF, grant nr. 17-194).

References

[1] Early Breast Cancer Trialists' Collaborative group. Trastuzumab for early-stage, HER2-positive breast cancer: a meta-analysis of 13 864 women in seven randomised trials. Lancet Oncol 2021;22:1139–50.

[2] Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, 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. Arch Pathol Lab Med 2018;142:1364–82.

[3] Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, 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;25:118–45.

[4] Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, 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.

[5] Zhang H, Katerji H, Turner BM, Hicks DG. HER2-Low breast cancers. Am J Clin Pathol 2022;157:328–36.

[6] NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines).. Breast cancer, vol. 2022. National Comprehensive Cancer Network; 2022.

[7] Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, et al. Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer. N Engl J Med 2022;387(1):9–20.

[8] Modi S, Park H, Murthy RK, Iwata H, Tamura K, Tsurutani J, et al. Antitumor activity and safety of trastuzumab deruxtecan in patients with HER2-low-expressing advanced breast cancer: results from a Phase Ib study. J Clin Oncol 2020;38:1887–96.

[9] Study of trastuzumab deruxtecan (T-DXd) vs investigator's choice chemotherapy in HER2-low, hormone receptor positiveMetastatic breast cancer (DB-06), vol. 2022; 2020. ClinicalTrials.gov.

[10] Xu K, Bayani J, Mallon E, Pond GR, Piper T, Hasenburg A, et al. Discordance between immunohistochemistry and ERBB2 mRNA to determine HER2 low status for breast cancer. J Mol Diagn 2022;24(7):775–83.

[11] Reeves JR, Going JJ, Smith G, Cooke TG, Ozanne BW, Stanton PD. Quantitative radioimmunohistochemical measurements of p185(erbB-2) in frozen tissue sections. J Histochem Cytochem 1996;44:1251–9.

[12] Tarantino P, Jin Q, Tayob N, Jeselsohn RM, Schnitt SJ, Vincuilla J, et al. Prognostic and biologic significance of ERBB2-low expression in early-stage breast cancer. JAMA Oncol 2022;8(8):1085–227.

[13] Schrijver WA, van der Groep P, Hoefnagel LD, Ter Hoeve ND, Peeters T, Moelans CB, et al. Influence of decalcification procedures on immunohistochemistry and molecular pathology in breast cancer. Mod Pathol 2016;29:1460–70.

[14] Schettini F, Chic N, Braso-Maristany F, Pare L, Pascual T, Conte B, et al. Clinical, pathological, and PAM50 gene expression features of HER2-low breast cancer. NPJ Breast Cancer 2021;7:1.

[15] Fernandez AI, Liu M, Bellizzi A, Brock J, Fadare O, Hanley K, et al. Examination of low ERBB2 protein expression in breast cancer tissue. JAMA Oncol 2022;8:1–4.

[16] Scott M, Vandenberghe ME, Scorer P, Boothman AM, Barker C. Prevalence of HER2 low in breast cancer subtypes using the VENTANA anti-HER2/neu (4B5) assay. J Clin Oncol 2021;39.1021–1021.

[17] Mosele F, Lusque A, Dieras V, Deluche E, Ducoulombier A, Pistilli B, et al. Unraveling the mechanism of action and resistance to Trastuzumab deruxtecan (T-DXd): biomarker analyses from patients from DAISY trial. ESMO Breast Cancer Annual Congress. 2022.

[18] Salgado R, Bellizzi AM, Rimm D, Bartlett JMS, Nielsen T, Holger M, et al. How current assay approval policies are leading to unintended imprecision medicine. Lancet Oncol 2020;21:1399–401.

[19] Gavrielatou N, Shafi S, Gaule P, Rimm DL. PD-L1 expression scoring: non-interchangeable, non-interpretable, neither, or Both. J Natl Cancer Inst; 2021.

[20] Delgado R, Tibau XA. Why Cohen's Kappa should be avoided as performance measure in classification. PLoS One 2019;14:e0222916.

[21] Brooks D, Thomas J. Interrater reliability of auscultation of breath sounds among physical therapists. Phys Ther 1995;75:1082–8.

[22] Administration, U.S.F.a.D. Statistical Guidance on reporting results from studies evaluating diagnostic tests; draft guidance for industry and FDA reviewers draft guidance-not for implementation. 2003.

[23] Gustavson, M., Haneder, S., Brieu, N., Spitzmueller, A., Kapil, A., Schneider, K. et al. Novel approach to HER2 quantification: digital pathology coupled with AI-based image and data analysis delivers objective and quantitative

[24] Moutafi M, Robbins CJ, Yaghoobi V, Fernandez AI, Martinez-Morilla S, Xirou V, et al. Quantitative measurement of HER2 expression to subclassify ERBB2 unamplified breast cancer. Lab Investig 2022.