Personalized medicine relies on personalized diagnoses; molecularly targeted therapies require molecular diagnostic assays. There is an urgent need for validated diagnostic assays aiding selection of optimal chemotherapy, whether conventional cytotoxic or targeted biological therapies, without which personalized medicine cannot advance. In this edition of the Journal, Mulligan et al. (1) describe the development and validation of a novel approach to select patients who benefit from anthracycline/cyclophosphamide chemotherapy. The signature was developed from the Fanconi Anemia DNA damage repair phenotype. This is both novel and interesting, with clear functional and biological rationales, rare in biomarker development. But can the diagnostic test developed be truly described as validated, as intimated by the article? This is critical for two key reasons: 1) patients and clinicians need evidence that novel diagnostic tests are validated (ie, they are accurate, informative, and add value to clinical decision making), and 2) industry, health-care providers, and payers need to know which novel diagnostic assays provide information that improves health-care provision. These concepts are different from, but not in tension with, the concept of a valid scientific study or research validating a scientific hypothesis. It is not our intention to challenge the authors use of the term “validated” but rather to demonstrate, using this study as an example, the tension that exists between scientific and clinical validation, which differ, not in stringency of evidence required, but in the questions addressed. We believe it critical to recognize the importance of appropriate validation of novel diagnostic tests and to support studies that accelerate their development.

Validation of diagnostic tests for clinical implementation is, some might argue, more demanding than validation of novel therapeutics. There are multiple guidelines relating to this process, including the Evaluation of Genomic Applications in Practice and Prevention (EGAPP), Biospecimen Reporting for Improved Study Quality (BRISQ), Reporting Recommendations for Tumour Marker Prognostic Studies (REMARK), and so on. (2–5). The US Food and Drug Administration and other national regulators draw from guidelines to develop processes for evaluation of diagnostic assays, providing much needed transparency to this challenge. Although space precludes a comprehensive review of all relevant guidelines, we will evaluate the assay proposed by Mulligan et al. (1) against three key stages of assay validation: technical validity, clinical validity, and clinical utility.

Technical validation of diagnostic assays requires evidence the assay is “fit for purpose,” a broad term that involves technical validation steps encompassing accuracy, reproducibility, sensitivity, and specificity (5). Mulligan et al. (1) assessed reproducibility but used a less-rigorous criterion, supporting but not fully validating their assay. They state, without detailing supporting evidence, “total standard deviation estimates were … acceptable as they are <5% of the … reportable range” (page 24 of Supplementary Data). Although this suggests the assay has potential for technical validation, data on assay reproducibility, particularly with respect to diagnosis of patients with “borderline” results is not presented, making it difficult to be confident that the assay robustly stratifies patients, either at central or peripheral hospital laboratories. Experience with single markers [eg, HER2 (6,7)] suggests these are critical components of technical validation, and in our opinion this is as important for multiparameter tests as it is for “simple” single marker assays. Accuracy is more challenging to demonstrate, particularly when a broad phenotypic assay is applied, but evidence shows this challenge can be addressed with appropriately designed studies (8–10). EGAPP (5) guidelines state that “convincing” evidence for analytical validity requires “studies that provide confident estimates of analytic sensitivity and specificity” using representative samples, particularly addressing challenging cases. Such data may be available, and, if so, should be published to demonstrate technical validity of the DNA-damage response deficiency (DDRD) assay.

The authors present several analyses supporting utility of the DDRD signature to predict outcome after chemotherapy in both neoadjuvant and adjuvant settings. They show convincingly that the DDRD signature is associated with either pathological complete response or relapse after chemotherapy but not in patients who did not receive chemotherapy. They used publically available and institutional cohorts, including three neoadjuvant cohorts (n = 51, 66, and 86 patients, respectively) treated with fluorouracil, epirubicin, and cyclophosphamide (FEC)/fluorouracil, doxorubicin, and cyclophosphamide (FAC), and four adjuvant cohorts, of which three were untreated and one (n = 191 patients) was treated “historically” with FAC. The data support a role for the DDRD signature in prediction of residual risk after chemotherapy, but do they clinically validate the formalin-fixed paraffin-embedded DDRD assay for anthracycline/cyclophosphamide chemotherapy response as claimed? We argue that, although these data are interesting and part of a pathway leading to clinical validation, they represent an incomplete clinical validation for the following reasons:

1. Although the DDRD signature has been tested in silico in multiple cohorts, it is clear that the DDRD assay itself was not tested in six of seven of the cohorts, and it is unclear if the final, retrospective cohort of 191 case patients were tested using the specific formalin-fixed paraffin-embedded DDRD assay.

2. All patients tested, who received chemotherapy, received FEC or FAC, making it impossible to discriminate between DDRD as a predictor for response to 5-fluorouracil, anthracyclines or cyclophosphamide.
The primary hypothesis, that DDRD is predictive of response to DNA damaging chemotherapy, has not been tested in these experiments by comparison with, for example, taxane-based chemotherapy. Even comparisons with nonchemotherapy-based treatments are across different patient cohorts using different expression arrays. The authors also have not presented any preclinical evidence that disruption of the DDRD pathway alters response to DNA-damaging agents. Although their scientific hypothesis that the DDRD pathway selects for response to DNA-damaging agents is sound, they present neither a clinical nor preclinical evidence to support the DDRD signatures' ability to aid selection of different therapeutics for patient management.

The final statement above crystallizes the dilemma facing those seeking to validate novel diagnostic assays for clinical use. It is imperative that novel diagnostic assays address an unmet clinical need. If the DDRD assay were validated, which clinical need would it address? Given the potential for the assay to determine response to DNA-damaging agents, the question becomes, does this assay aid decision making when selecting taxane- vs anthracycline- or cyclophosphamide-based treatments? The lack of evidence, either preclinical or clinical, relating to response of patients treated with taxane-based chemotherapy becomes a critical gap in the validation. Validation of the assay in either multiple retrospective (4) or biomarker-stratified, prospective (11,12) trials is a critical component of future validation.

So have we come to bury Mulligan et al. (1) or to praise them? Mulligan et al. (1) have initiated a program of research, which, in our view, may lead to development of a robust and clinically valid diagnostic test. Although claims for validation may be premature, there is at least a strong scientific rationale, often lacking in opportunistic biomarker approaches, as to why the DDRD signature might, in the future, aid treatment selection. Further research will prove or disprove this claim, and we view this study as important and as laying a strong foundation for further studies focused on technical and clinical validation of this assay. Claims to have validated such a test as clinically useful are, however, premature, as careful evaluation has shown and as the authors themselves recognize. Were they wrong to make such claims? We believe that, collectively, editors, reviewers, and authors should focus on clear use of terminology relating to diagnostic tests. This is a critical area for the future of personalized medicine, which requires clear evidence of clinical validity before such claims can be made. Above all, editors, supported by reviewers, are those most able to develop such clarity through reference to available guidelines.

References


Funding

We thank the Government of Ontario for funding, which is provided through the Ontario Ministry of Research and Innovation.

Note

The authors have no conflicts of interest to disclose. The study sponsor had no role in the writing of the editorial or the decision to submit it for publication.

Affiliations of authors: Transformative Pathology, Ontario Institute for Cancer Research, Toronto, ON, Canada (JMSB, MS); Edinburgh Cancer Research UK Centre, University of Edinburgh, Edinburgh, UK (JMSB); Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada (JMSB); Institute of Health Policy Management and Evaluation, University of Edinburgh, Toronto, ON, Canada (KIP); Sunnybrook Odette Cancer Centre, Toronto, ON, Canada (KIP).