Sunil Badve · Yesim Gökmen-Polar *Editors*

Molecular Pathology of Breast Cancer

Springer

Molecular Pathology of Breast Cancer

Sunil Badve · Yesim Gökmen-Polar
Editors

Molecular Pathology of Breast Cancer

Springer

Editors
Sunil Badve
Department of Pathology and Laboratory Medicine
Indiana University School of Medicine
Indianapolis, IN
USA

Yesim Gökmen-Polar
Department of Pathology and Laboratory Medicine
Indiana University School of Medicine
Indianapolis, IN
USA

ISBN 978-3-319-41759-2 ISBN 978-3-319-41761-5 (eBook)
DOI 10.1007/978-3-319-41761-5

Library of Congress Control Number: 2016950893

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom

—Clifford Stoll
Lawrence Berkeley National Laboratory, USA

The last few years has seen the deluge of data regarding the alterations in breast cancers. Recent advances in technology also permit analysis of single cells for these alterations. However, clinicians and scientists faced with an onslaught of this data from the scientific and lay press are finding it difficult to distinguish data from information. The major question that arises is—how does it affect the lives of my patients? My research?

Molecular Pathology of Breast Cancer seeks to provide an overview of the recent advances in breast cancer and bring together the techniques, data, and knowledge to provide some understanding and wisdom. We believe that this work will represent a new and important resource for clinicians and scientists, by serving as a "ready reckoner." The chapters, written by experts in the field, provide valuable information to those already involved in and familiar with the complexities of breast cancer. In order to introduce the territory to the novices, the chapters, while being detailed, have been kept short and the discussions brief. The hope is to make the topics "meaningful" but less intimidating for the audience.

It is clear that advances in molecular biology have provided exhaustive data regarding breast cancer. However, it is necessary to separate the wheat from the chaff. We are extremely grateful to the cadre of authors, who have graciously donated their time and energy to make this hard work possible. In a series of chapters within the book, these experts have presented most recent research and highlighted the direction for future research.

Indianapolis, IN, USA

Sunil Badve
Yesim Gökmen-Polar

Contents

Translation of Biomarkers into Clinical Practice

1

Lisa Meier McShane, Tracy G. Lively and Hala R. Makhlouf

Biomarkers have long played a key role in the clinical management of breast cancer. Their use continues to expand beyond the classic biomarkers such as hormone receptors (ER and PR) for guiding use of endocrine therapy and HER2 status for guiding use of HER2-targeting agents. In recognition of the critical role that biomarkers play in drug development and in patient care, the U.S. Food and Drug Administration (FDA) and the U.S. National Institutes of Health (NIH) have recently partnered to develop a standardized glossary of terminology related to biomarkers and clinical outcomes. In that glossary, it is stated that a biomarker is "a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions. Molecular, histologic, radiographic, or physiologic characteristics are types of "biomarkers" (FDA-NIH Biomarker Working Group 2016). Biomarkers can be used individually, or in combination as a "signature", at multiple points along a patient's clinical trajectory to guide clinical care decisions.

The American Society of Clinical Oncology (ASCO) recently issued guidelines for clinical use of biomarkers (beyond T, N and M staging) to aid in decisions on systemic therapy for women with metastatic breast cancer (Van Poznak et al. 2015) and to aid in decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer (Harris et al. 2016). For the metastatic setting, no biomarkers were fully endorsed by the guideline committee except for estrogen receptor (ER), progesterone receptor (PR), and HER2 (Human Epidermal Growth Factor Receptor 2) status in combination with clinical evaluation, patient preferences, and judgment; CEA, CA 15-3, and CA 27.29 were regarded by the committee as appropriate for use adjunctive to decisions regarding therapy but not for use in isolation (Van Poznak et al. 2015). The committee that examined biomarkers for the early-stage invasive breast cancer setting found sufficient evidence to recommend clinical use of OncotypeDX®, EndoPredict®, Prosigna™, Breast Cancer Index℠ and uPA/PAI-1 in specific subgroups of breast cancer, in addition to the well-established estrogen and progesterone receptor (ER/PR) and HER2 biomarkers (Harris et al. 2016). Although the number of biomarkers recommended for clinical use has increased in

L.M. McShane (✉) · T.G. Lively · H.R. Makhlouf
Division of Cancer Treatment and Diagnosis, National Cancer Institute, 9609 Medical Center Drive, Bethesda, MD 20892, USA
e-mail: lisa.mcshane@nih.gov

H.R. Makhlouf
Department of Pathology, Faculty of Medicine, Ain Shams University, 38 Al Abbassia, Greater Cairo, Cairo 11566, Egypt

S. Badve and Y. Gökmen-Polar (eds.), *Molecular Pathology of Breast Cancer*, DOI 10.1007/978-3-319-41761-5_1

the last few years, there is still a large gap between that number and the number of biomarker studies published, perhaps reflecting a lack of appreciation of requirements for translation of a biomarker into clinical practice or other challenges inherent in that process.

There are several challenges in translation of biomarker research results to a clinical test that is useful for making patient treatment decisions. Variation in assay methods used to measure the biomarker (or signature) across potentially many studies comprising the evidence base can make it difficult to interpret the literature and determine the specific assay methods that are optimal. Additionally, pre-analytical factors, which refer to the conditions under which biospecimens are collected, processed or stored prior to analysis, can sometimes have a profound impact on the ability to measure a biomarker reliably or even to measure it at all (Moore et al. 2011). Heterogeneity due to pre-analytic and analytic factors may be further compounded by differences in clinical populations or treatment settings studied. All of the pre-analytical, analytical, and clinical issues must be confronted when developing a biomarker test or evaluating its usefulness for clinical care. Multi-disciplinary expertise is needed to determine which biomarkers are the most informative and reliable for making specific clinical decisions, and to develop the most promising biomarkers into clinical-grade tests.

Biomarker tests need to be rigorously evaluated to establish their readiness for clinical use. Pathologists and clinicians must understand how to appropriately select, apply, and interpret clinical tests, be able to judge if a test has been appropriately validated, and have an appreciation of the potential risks and benefits associated with use of a given test. These requirements apply regardless of whether pathologists or laboratorians develop a version of a biomarker test for use in their laboratory or provide advice concerning use of biomarker tests performed by outside laboratories. Understanding the general process by which biomarker tests are developed and validated is critical in making an informed judgement about the clinical readiness of any particular biomarker-based test.

1.1 From Biomarker to Biomarker Test

Clinical use of a biomarker requires a reliable method to measure it. The constellation of elements that enable measurement comprise the *biomarker test*, which is defined as "an assessment system comprising three essential components: (1) materials for measurement; (2) an assay for obtaining the measurement; and (3) method and/or criteria for interpreting those measurements" (FDA-NIH Biomarker Working Group 2016). For biomarker signatures, the test would also include a procedure for combining measurements of multiple biomarkers, such as output from omics assays which include those based in the disciplines of "genomics, transcriptomics, proteomics, metabolomics, and epigenomics" (Micheel et al. 2012). The result of combining the measurements is typically a risk score developed from a statistical model or a categorization output by an algorithm that classifies each case into one of multiple possible categories based on the pattern detected in the biomarker measurements. Such models or algorithms will be referred to here as *multivariable biomarker predictors* or in the case of high throughput omics technologies, *omics predictors*.

Many biomarkers used in treatment decisions for breast cancer have undergone an evolution in methods for measurement. For example, clinical measurement methods for estrogen receptor (ER) have evolved from ligand binding assays performed on tumor cytosols which produced continuous measurements in units of fmol/mg (typically with ≥ 3 or 20 called positive) to immunohistochemical assays that could be performed on formalin-fixed tumor tissue and which produced semi-quantitative measurements (Hammond et al. 2010). For some biomarkers, acceptable measurement methods have been well established; whereas, for other biomarkers a variety of measurement methods exist, often with little understanding of the degree of concordance that might be expected among results obtained by different methods. The clinical impact of discordance in biomarker measurements due to assay methodology may vary depending on the

density of biomarker values in the patient population that are near key clinical decision points and the degree of discordance between assays near those points. Committees convened jointly by the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) developed best practice guidelines for testing HER2 status (Wolff et al. 2007, 2013) and the hormone receptors ER and PR (Hammond et al. 2010) which specify acceptable pre-analytic and analytic conditions and procedures in order to promote consistency and reliability of testing.

In contrast, there are examples of biomarkers used with some regularity for the care of patients with breast cancer that continue to be measured by a variety of different approaches with insufficient attention paid to the impact of the different measurement methodologies. Stuart-Harris et al. (2008) reviewed literature on the nuclear proliferation marker Ki67 in breast cancer. They reported that among the 43 studies reporting use of an immunohistochemical assay for assessment of Ki67 in early breast cancer, 7 different antibodies for IHC, single or in combination had been used; among those studies, 19 different cutpoints, ranging from 0 to 30 %, had been used for determination of high expression. Further, recent reports by Polley et al. (2013a, b) demonstrated a concerning lack of concordance due to scoring approach alone when eight different laboratories across the world experienced in Ki67 immunohistochemistry evaluated a common set of stained breast cancer tissues represented on a tissue microarray slide. Pathmanathan et al. (2014) demonstrated how varying the Ki67 cutpoint in increments of 5 % could substantially alter the accuracy with which Ki67 assessments could predict survival following breast cancer diagnosis. Together with the lack of concordance found by Polley et al. (2013a, b), this sensitivity of prognostic ability to cutpoints suggests that the clinical value of Ki67 assessments will likely vary across laboratories performing the testing. These examples illustrate the need for greater attention to the specific methods used to measure biomarkers and better understanding of the impact of pre-analytic and analytic heterogeneity on clinical performance of biomarkers.

Details of both pre-analytic conditions and analytical methods should be provided when investigators publish reports of biomarker studies and these should also be provided by laboratories offering biomarker tests. Checklists have been developed to provide guidance on what information is important to report in publications involving use of biospecimens (Moore et al. 2011) and biomarkers used in prognostic (McShane et al. 2005; Altman et al. 2012) and diagnostic (Bossuyt et al. 2003a, b, 2015) studies. Further useful information can be found on the EQUATOR website (EQUATOR Network 2016) which provides a wealth of checklists and guidance for reporting a wide variety of health research studies. Laboratories that offer biomarker tests for clinical use should clearly state any important pre-analytical requirements for biospecimens, provide information about the particular testing procedures they use, and provide clear instructions about how test results should be interpreted. Such details should also be provided in any clinical study protocol that involves investigational use of a biomarker test. These steps would help to make biomarker test development more efficient and ensure that clinical biomarker tests were properly used and their results interpreted appropriately.

1.2 Clinical Uses for Biomarker Tests

Evaluation of clinical performance of a biomarker test must start with a clear statement of the intended use of the test in clinical decision making. Uses most relevant to therapy decisions include forecasting prognosis, therapy selection, or monitoring for disease recurrence or progression; these will be the focus in this chapter. Intended use must also consider the clinical context, including disease stage and treatments received, or other clinical or pathologic factors that define subgroups of patients whose disease is managed differently in routine practice. A major reason for failure of many biomarkers to be

translated to a test used in clinical practice is that correlations between biomarker values and outcomes observed in exploratory studies in heterogeneous patient populations often do not translate to information that is meaningful or useful in clinical management.

Prognostic biomarkers are "used to identify likelihood of a clinical event, disease recurrence or progression" (FDA-NIH Biomarker Working Group 2016). Presence of malignant cells in lymph nodes of patients who undergo surgical resection of their tumor predicts higher likelihood of developing recurrent disease. Patients with breast cancer who carry certain germline BRCA1 or BRCA2 mutations are at higher risk of developing a second primary breast cancer or ovarian cancer (Brekelmans et al. 2007; Bergfeldt et al. 2002). In oncology, the term prognosis has generally been used in the clinical context of patients receiving either no therapy (beyond primary surgery) or a uniform standard therapy that all patients are likely to receive. If the term prognosis is used in other settings, for example in the context of a targeted therapy, or in a setting where patients could have received any of several therapies, then there is risk that prognostic and predictive effects (see definition of predictive biomarker below) can become confused. Indeed, potential targets for new therapies are often first discovered as prognostic biomarkers. If a new therapeutic can inhibit a biomarker with negative prognostic effect, that new therapy might improve clinical outcome. This was the situation with HER2 overexpression which was first identified as a negative prognostic factor; HER2-targeted therapies were subsequently developed that substantially improved survival for patients in both the metastatic and adjuvant settings. The importance of distinguishing these terms is discussed further in the section **Evaluation of clinical utility**.

Predictive biomarkers are "used to identify individuals who are more likely than similar patients without the biomarker to experience a favorable or unfavorable effect from a specific intervention or exposure" (FDA-NIH Biomarker Working Group 2016). The term "predictive" has been used somewhat variably in oncology so it is important to be clear about the context. One way in which the term predictive biomarker has been used is in the setting of selecting between two different treatments (one of which could be no further treatment (beyond surgery, possibly with radiation) as in an adjuvant setting for early stage breast cancer), usually with focus on a time-to-event endpoint (e.g., overall survival, recurrence-free survival, disease-free survival, progression-free survival). In this context alternate terms for predictive biomarker are *treatment-effect modifier*, *treatment-guiding* or *treatment-selection biomarker*. The term *treatment-selection biomarker* will be used here; it means that the effect of a particular treatment relative to some other treatment varies depending on the value of the biomarker. The biomarker could predict benefit, lack of benefit, or even harm from a particular treatment. In the simplest setting of a binary biomarker, one could say that a positive biomarker result defines a population that benefits from treatment A relative to B (e.g., longer survival when a patient receives treatment A compared to treatment B) but the biomarker negative group either does equally well under A and B or does better under B than A. A classic example of a treatment-selection predictive biomarker in breast cancer is hormone receptor status to guide use of endocrine therapy. Patients whose tumors are negative for hormone receptors are unlikely to benefit from endocrine therapy (with or without concomitant chemotherapy), whereas the group of patients whose tumors are positive will have an overall reduced rate of recurrence and longer survival if they receive endocrine therapy.

Biomarkers are increasingly used to enrich or select the patient population for clinical trials of targeted anti-cancer agents. This is an approach used for development of new therapeutics, but it has implications for the eventual regulatory approval of the new therapeutic and its approved indications for clinical use. Varied terminology has been used to refer to such biomarkers, including predictive biomarkers, selection biomarkers, or enrichment biomarkers. This type of biomarker will be referred to here as an *enrichment-predictive biomarker*. The key distinction between an enrichment-predictive biomarker

and a treatment-selection biomarker is that for an enrichment-predictive biomarker there is no or very little clinical evaluation of the new drug in the "biomarker negative" subgroup. This drug development path might have been chosen because there was little or no biological rationale for why the new drug should work in the biomarker negative group or because in pre-clinical studies drug effects were observed only in models (e.g., cell lines, animal models, xenografts) that were positive for the biomarker. For example, if the drug is a monoclonal antibody one might not expect it to work for patients whose tumors do not express the target antigen. However, such assumptions are sometimes too simplistic and might not account for off-target effects of the drug or might be based on cutpoints for defining positivity that are not optimal. When varied assays are used to assess an enrichment-predictive biomarker it is important to consider whether any particular assay being used identifies a patient population similar to the one identified by the enrichment biomarker actually used in the pivotal clinical trials of the therapeutic agent. Further elaboration with an example is discussed in the section **Evaluation of clinical utility**.

In settings where chemotherapy is given as the first treatment with or without subsequent surgery (e.g., as neoadjuvant therapy or for metastatic disease), biomarkers which can predict tumor response (or possibly prolonged progression-free survival or stable disease) may be of interest. Such biomarkers are often called predictive biomarkers, but they are indicated for a purpose slightly different than the predictive biomarkers just described for therapy selection. They will be denoted *response-predictive biomarkers* here. Rather than comparing between treatments, response predictive biomarkers may be used to indicate likelihood of drug activity—either tumor objective response (complete or partial response) or prolonged stable disease or time to progression. Importantly, drug activity as assessed by tumor response does not necessarily translate to a clinical benefit in terms of prolonged overall or disease-free survival.

Monitoring biomarkers are "measured serially for assessing status of a disease or medical condition or for evidence of exposure to (or effect of) an environmental agent or medical product. Monitoring biomarkers may also be used to indicate toxicity or assess safety, or to provide evidence of exposure, including exposures to medical products" (FDA-NIH Biomarker Working Group 2016). In oncology, blood-based biomarkers and image-based biomarkers are widely used for monitoring patients following initial therapy to detect signs of persistent, recurrent or progressive disease. CT scans to assess tumor burden (which can be considered a "biomarker") are used routinely to monitor for progression in advanced disease. Serum biomarkers such as CEA, CA 15-3, and CA 27.29 have also been widely used for monitoring in metastatic disease (Van Poznak et al. 2015), and more recently circulating tumor cells or cell-free DNA have been investigated for their potential usefulness in monitoring [e.g., circulating tumor cell evaluation in the randomized trial S0500 (NCT00382018; Smerage et al. 2014)].

1.3 Principles in Determination of Fitness of a Biomarker Test for an Intended Clinical Use

Evaluation of the suitability of a biomarker test for a particular clinical use requires a series of studies to address analytical validity, clinical validity, and clinical utility. The nature of these studies will depend on the type of assay methodology used for measurement and the intended use of the biomarker test. Some biomarker tests used in clinical care for breast cancer, for example, CELLSEARCH® Circulating Tumor Cell Kit, Prosigna™, and MammaPrint® have been reviewed and cleared by the FDA (U.S. FDA 2006, 2013, 2015). Others such as standard immunohistochemical tests including ER, PR, and HER2 are performed routinely in essentially all laboratories which analyze breast tumor specimens; they may be performed using a commercial assay kit or using a test developed in

the laboratory offering it. Other tests such as OncotypeDX® (Genomic Health 2016) are performed at a central commercial laboratory. Biomarker tests performed in CLIA-certified laboratories may have never been reviewed by the FDA; however, CLIA-certified laboratories are required to validate their assays and perform quality monitoring, and many participate in proficiency testing and education programs such as those offered by the College of American Pathologists (College of American Pathologists 2016). Regardless of the level of FDA or other external review that a biomarker test has undergone, it is important that the appropriate evaluations have been performed by some qualified party to ensure that the test can be used safely and its results can be relied upon to have a particular clinical interpretation.

1.3.1 Analytical validity

Analytical validity refers to "establishing that the performance characteristics of a test, tool, or instrument are acceptable in terms of its sensitivity, specificity, accuracy, precision, and other relevant performance characteristics using a specified technical protocol (which may include specimen collection, handling and storage procedures)" (FDA-NIH Biomarker Working Group 2016). Analytical validity pertains to a test's technical performance but says nothing about its clinical usefulness. Design of analytic validation studies will depend on the specific type of assay under evaluation, but there are several helpful references providing general guidance (Jennings et al. 2009; Linnet and Boyd 2012; Pennello 2013; Becker 2015). A particularly good reference for analytic validation of immunohistochemical assays is CLSI document I/LA28-A2 (CLSI 2010). Some researchers have published analytical validation studies that they conducted and these may also serve as useful guides; examples of published analytical validation studies for tests used for breast cancer include those for CELLSEARCH® (Allard et al. 2004) and several for omics tests including Prosigna™ gene expression ROR score (Nielsen et al. 2014) and the OncotypeDX® Risk Score (Cronin et al. 2007). Publication of skillfully executed analytical validation studies should be encouraged to disseminate best practices.

1.3.2 Clinical validity

Clinical validity refers to "establishing that the test, tool, or instrument acceptably identifies, measures, or predicts the concept of interest where "concept" refers to a "clinical, biological, physical, or functional state, or experience" (FDA-NIH Biomarker Working Group 2016). Clinical validity is established by showing that the biomarker test results are related to the concept of interest in the relevant clinical setting, typically by demonstrating a statistically significant association and quantifying its strength in an appropriately designed study. For example, if a biomarker test is intended to predict disease-free survival, one might demonstrate that patient biomarker values measured at diagnosis are statistically significantly associated with disease-free survival time using Cox proportional hazards regression (Cox 1972) or other type of survival analysis, as appropriate. To demonstrate clinical validity of a biomarker test for monitoring for recurrence following treatment in the adjuvant setting one might, for example, demonstrate that the biomarker value measured at one year after the end of therapy is associated with likelihood of disease recurrence within the following year using an approach such as a landmark analysis (Anderson et al. 1983). For a response-predictive biomarker test that reports a continuous biomarker value, one could show that the biomarker value associates with likelihood of tumor response, for example by showing that the area under the receiver operating characteristic curve is significantly greater than the chance value of 0.5 (Hanley and McNeil 1982; Zou et al. 2007). Although these examples illustrate how associations could be estimated and tested, more is needed to establish that it is beneficial to use a biomarker test to guide clinical care. This concept of benefit from use of a test relates to the notion of clinical utility, which is discussed in depth in the next section.

1.3.3 Clinical Utility

Clinical utility for a biomarker test refers to a conclusion that use of the test "will lead to a net improvement in health outcome or provide useful information about diagnosis, treatment, management, or prevention of a disease. Clinical utility includes the range of possible benefits or risks to individuals and populations" (FDA-NIH Biomarker Working Group 2016). Assessment of clinical utility for a biomarker test is predicated on the test's analytical and clinical validity already having been established.

A laboratory or clinician may wish to evaluate the evidence for clinical utility of a biomarker test that is offered by another laboratory, or they may wish to evaluate clinical utility for a test that they newly developed. A laboratory might also wish to offer its own version of a biomarker test which has already been developed and confirmed to have clinical utility as performed by another laboratory. For the last situation in which a laboratory's intent is to transport the biomarker test to an in-house test, it is important for the laboratory to confirm that the test, as that specific laboratory performs it, delivers results highly concordant with those of the test as it was performed in prior studies that established the test's analytical and clinical validity and confirmed its clinical utility; if test results are not highly concordant, it is incumbent upon the laboratory to demonstrate that the test as performed in-house maintains its clinical performance and utility. For all of these situations a thorough understanding of acceptable approaches for establishing clinical utility is necessary.

The approach to demonstrating clinical utility of a biomarker test will depend on the intended clinical use. The three clinical uses as prognostic, predictive, and monitoring tests are elaborated on here. The first step in evaluation of clinical utility is a clear statement of the intended use; this includes careful definition of the patient population to which the test will be applied and the clinical decision that the test will inform. Too often biomarker studies are carried out using convenience sets of specimens with more attention paid to discovering statistically significant correlations than to what clinical decision the biomarker might help to inform. Such studies of convenience rarely lead to clinically helpful or viable biomarker tests (McShane and Polley 2013; Simon et al. 2009). Investigators aiming to develop biomarker tests for clinical care should focus on the intended clinical use as early as possible in the development process to ensure that the clinical studies forming the evidence base are performed in the relevant patient population and clinical context.

1.4 Evaluation of Clinical Utility

1.4.1 Prognostic Biomarker Utility

To establish that a prognostic biomarker test has clinical utility one should be able to demonstrate that it can identify patients for whom different prognoses, as forecast by the test, would lead to different clinical management decisions and that those decisions lead to a net benefit for the patient. Additionally, the information provided by the test should either add to existing routinely used prognostic indicators or the test should provide information comparable to existing indicators and be more reliable, convenient, or less invasive or expensive.

The 2016 *ASCO Clinical Practice Guideline for Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women with Early-Stage Invasive Breast Cancer* cites evidence of clinical utility for prognosis for OncotypeDX®, EndoPredict®, Prosigna™, Breast Cancer Index℠ and uPA/PAI-1 for women with ER/PgR positive/HER2-negative (node-negative) breast cancer (Harris et al. 2016). In the indicated group of patients each of these tests was able to identify a subgroup with sufficiently good outcome in the absence of chemotherapy (e.g., low risk of disease recurrence) that chemotherapy would not be recommended. The ASCO guidelines committee did not find sufficient evidence for clinical utility for prognosis for any biomarkers in node-positive or HER2-positive disease.

Reasons that the ASCO biomarkers guidelines for early stage breast cancer did not recommend all prognostic biomarkers assessed or any prognostic

markers outside of the setting of ER/PgR positive/HER2-negative node-negative breast cancer were the lack of sufficient data in the other subgroups or presentation of results only from patient cohorts heterogeneous with respect to standard prognostic variables and or treatments. As an example of why it can be important to study a group of patients who are relatively homogeneous with respect to standard prognostic variables, one can compare the prognostic ability of the

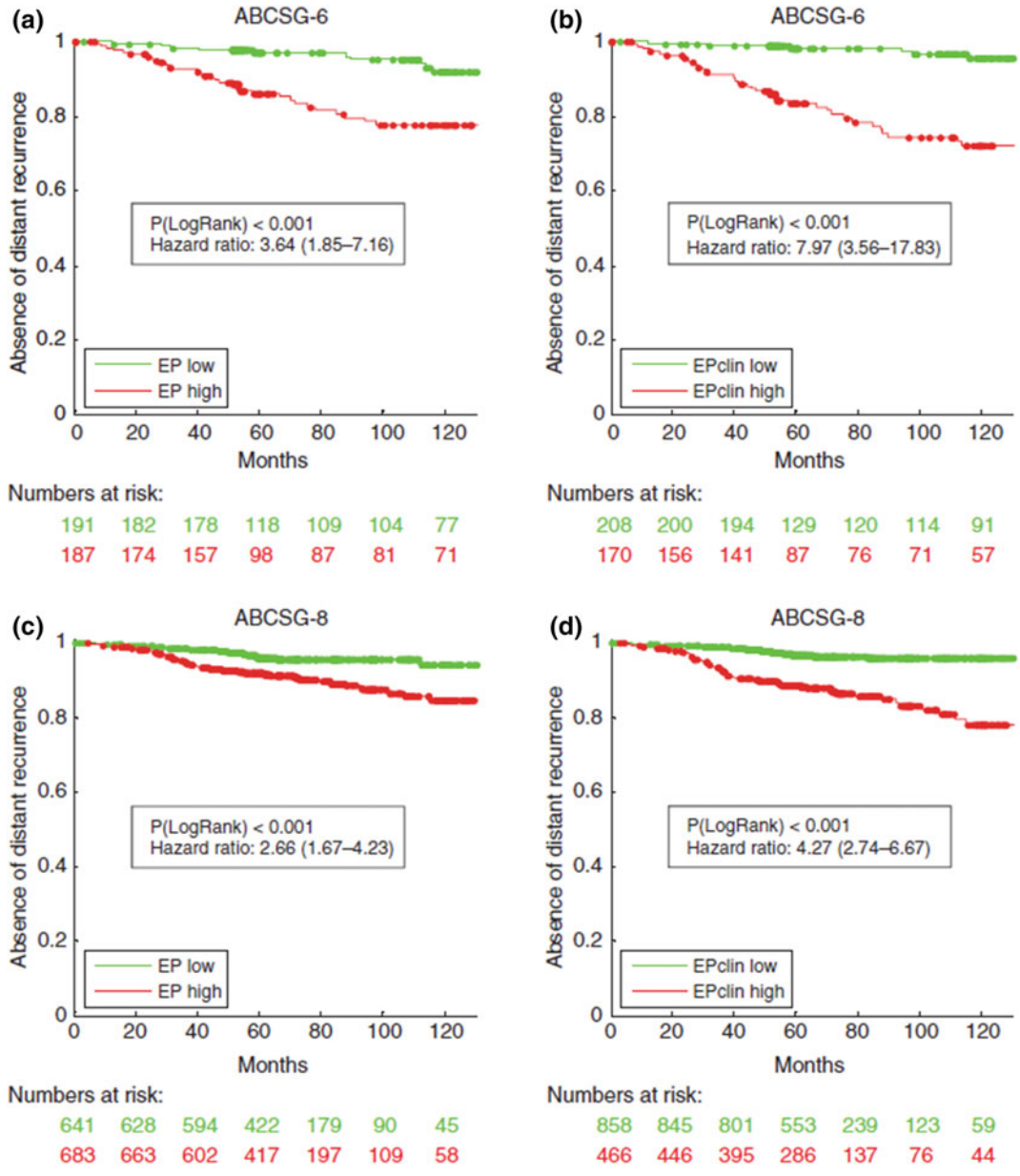

Fig. 1.1 Kaplan-Meier plots of distant recurrence by EP and EPclin risk groups. Distant recurrence according to EP risk groups (**a** and **c**) and EPclin risk groups (**b** and **d**) in patients from the 2 validation cohorts (ABCSG-6 *top*; ABCSG-8 *bottom*). Cutoff points for EP were prespecified at 5 (3.3 for EPclin) in the training set. *Numbers in parentheses* indicate the 95 % CI of the HR (Reprinted from Fig. 2 in Filipits et al. 2011)

dichotomized EndoPredict® EP score within subgroups defined by standard prognostic variables (Filipits et al. 2011). Significant differences in distant recurrence rate between EP low-risk and EP high-risk patients were observed in validation sets from the ABCSG-6 (Fig. 1.1a) and ABCSG-8 (Fig. 1.1c) trials. At 10 years, the distant recurrence rates for patients with EP low and EP high were 8 % (3–13 %) and 22 % (15–29 %) in ABCSG-6 ($P < 0.001$) and 6 % (2–9 %) and 15 % (11–20 %) in ABCSG-8 ($P < 0.001$), respectively. The subgroup defined as low risk by dichotomized EP score in both trials demonstrated 10-year distant recurrence rate less than 10 %. Similar results are shown for the EPclin score which incorporates additional prognostic variables nodal status and tumor size, although the separation between the survival curves appears wider (Figs. 1.1b, d). If the analyses are segregated by nodal status, then in the combined trial cohorts only the low risk group within the node-negative patients, and not the low risk group within the node-positive patients, achieves a distant recurrence rate less than 10 % (Fig. 1.2).

For the MammaPrint® test (Agendia, Inc. Irvine CA), the ASCO guidelines committee could not establish clinical utility due to ambiguity regarding the patient population and treatment setting in which it could be confidently used. The studies of the 70-gene prognosis signature (which was commercially developed into MammaPrint®) included patients with mixed prognostic variables such as positive and negative nodal status and both hormone receptor positive and negative tumors. Patients with hormone receptor positive tumors did not uniformly receive endocrine therapy, and some patients received chemotherapy while others did not (Harris et al. 2016). This heterogeneity among the studied patients made it impossible to determine whether the risk groups identified by the 70-gene prognosis signature were useful independently of standard prognostic variables or were indicating patients most likely to benefit or not from endocrine therapy or from chemotherapy. These examples illustrate that clinical context is critically important for determination of the clinical utility of a prognostic biomarker and studies should be designed with clinical context in mind.

Although it is customary for biomarkers to be categorized into two or more risk groups for clinical decision making, it is important to understand that any type of categorization of a continuous risk score results in a loss of information. For continuous prognostic risk scores it is usually possible to display the risk of the event (e.g., recurrence) at some fixed timepoint as a function of the risk score value, and these risk scores may include standard prognostic variables.

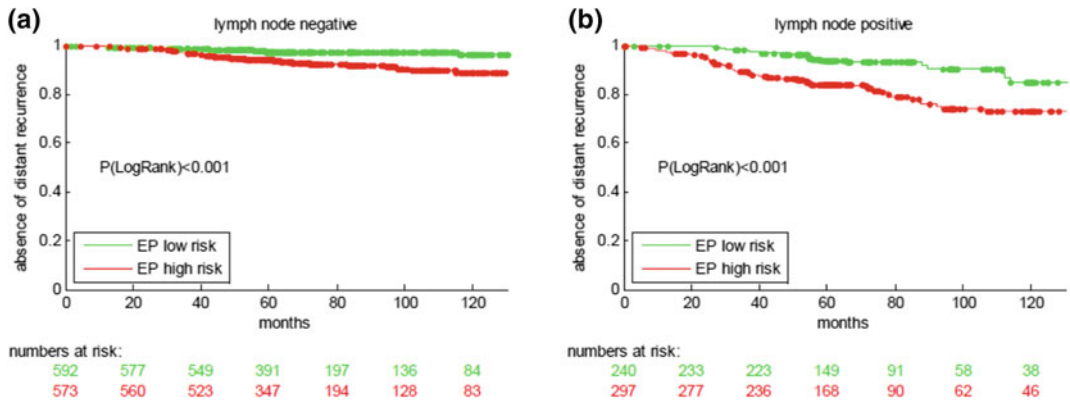

Fig. 1.2 Kaplan-Meier plots of distant recurrence by EP risk groups. Distant recurrence according to EP risk groups separately by nodal status (lymph node negative *left*; lymph node positive *right*) for combined ABCSG-6 and ABCSG-8 validation cohorts. Cutoff points for EP were prespecified at 5 in the training set. Ten-year distant recurrence-free survival is less than 90 % in the low risk lymph node positive group (Extracted from Fig. 9S in Filipits et al. 2011 online supplement)

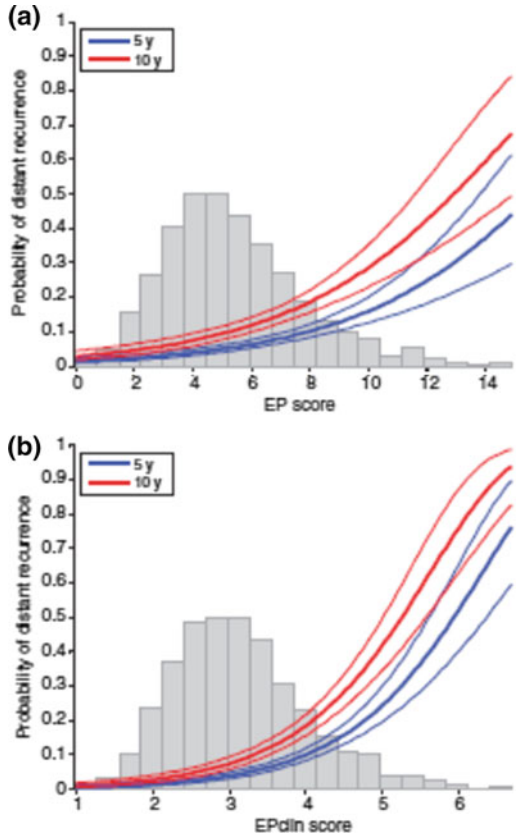

Fig. 1.3 Estimated probability of distant recurrence as continuous functions of the EP risk score (**a**) and the EPclin risk score (**b**). The continuous relation between the respective score and the probability of developing a distant recurrence within the first 5 and 10 years after surgery is described by an independent model for each score generated from all ABCSG-6 and ABCSG-8 data (n = 1702). The thin curves indicate the 95 % CI. The gray histogram in the background shows the distribution of scores for the patients (Reprinted from Fig. 1 in Filipits et al. 2011)

The Kaplan-Meier plots shown in Fig. 1.3a, b depict the prognostic ability of the EP and EPclin continuous scores in the combined ABCSG-6 and ABCSG-8 trial cohorts. These plots allow one to predict distant recurrence risk within 5 years and 10 years as a function of a risk score. The gray-shaded histograms also provide a visualization of the distribution of risk scores in the study population. Variation in absolute risk within each of the low and high risk scores is evident in these figures. This variation is not captured if the risks are reported only in aggregate for each of the low and high risk groups.

1.4.2 Predictive Biomarker Utility

The goal in demonstrating clinical utility for a predictive biomarker is to establish that the biomarker will guide a decision to select a particular treatment over a certain other treatment (the second treatment potentially being no further treatment) and that the selected treatment is associated with benefit for the patient. Criteria for establishing clinical utility vary somewhat for the three types of predictive biomarkers (treatment selection, enrichment-predictive, and response-predictive). Here we highlight some basic design considerations, and references are provided for readers interested in more extensive discussions. An excellent book length treatment of trial designs for predictive medicine is the book edited by Matsui et al. (2015).

1.4.2.1 Considerations for Treatment-Selection or Enrichment-Predictive Biomarker Utility

To establish clinical utility for a treatment-selection or enrichment-predictive biomarker, data from a trial in which there is a randomization between the treatments of interest is generally needed. Either the trial must be conducted prospectively or there must be an adequate number of specimens available from an appropriate completed randomized trial. Note that while it may be tempting to claim that a biomarker is predictive for benefit from a particular therapy when it is associated with more favorable outcome for those patients, such an effect may only be reflecting a prognostic effect that would be present independent of treatment (Polley et al. 2013a, b). Three basic phase III trial designs, or combinations or variations of these designs, are typically used to demonstrate clinical utility for treatment-selection or enrichment-predictive biomarkers: (1) the enrichment design, (2) the stratified design, and (3) the strategy design (Sargent et al. 2005; Freidlin et al. 2010). As will be discussed next, all of these designs require randomization, but they differ in other respects such as patient selection, treatment allocation, and the conclusions they support.

The enrichment design measures the biomarker on each patient at study entry and then randomizes only those patients whose tumors are positive between the experimental therapy (which is hypothesized to be better for patients who are biomarker positive) and some alternative standard therapy; this design can establish definitive evidence for clinical utility of the experimental therapy in the population selected by the enrichment-predictive biomarker if the experimental therapy is demonstrated to be superior to the standard therapy in that group. No information is provided by this design regarding which treatment is better for biomarker-negative patients; it is assumed that existing evidence suggests that biomarker-negative patients are not likely to benefit from the experimental therapy and thus they are not randomized. An enrichment design does not require that the biomarker used for enrichment perfectly identifies the group of patients who benefit from the experimental therapy. The biomarker only needs to be "good enough" so that the treatment effect is sufficiently amplified to be detected statistically in the enriched patient group. Even if imperfect, biomarker enrichment will have implications for the labeling of the new therapy, if the experimental agent is successful in trials leading to approval.

The drug development path for trastuzumab in breast cancer is an example for which the pivotal trials used biomarker enrichment. The metastatic trials enrolled only patients whose tumors were positive by a clinical-trial grade immunohistochemical assay for HER2, and in the adjuvant setting the pivotal trials enrolled patients whose tumors were positive for HER2 by either immunohistochemical (IHC; protein) or in situ hybridization (ISH; gene amplification) assays (Wolff et al. 2007). Due to apparent benefit of trastuzumab in patients whose tumors were negative on central testing but positive on a local assay used for study entry in the pivotal adjuvant studies, a new adjuvant trial, NSABP B-47 (NCT01275677), is underway to determine whether there is benefit of trastuzumab for patients whose tumors are HER2-Low. HER2-Low is defined in the B-47 trial as follows: 1 + by IHC; or 2 + by IHC and ratio of HER2 to chromosome enumeration probe 17 (CEP17) must be <2.0 or, if a ratio-based test was not performed, the HER2 gene copy number must be <4 per nucleus. Patients whose tumors are negative by both IHC and ISH are not eligible for the B-47 trial. This example illustrates the difficulties in the initial identification and subsequent refinement of an enrichment-predictive biomarker.

If it is desired to establish that a biomarker has clinical utility for treatment selection, then the stratified design is the most efficient design to use in most situations. The stratified design randomizes all patients between treatment A (thought to be better for biomarker positive patients) and treatment B (usually some standard therapy used irrespective of biomarker status) with stratification of the randomization by biomarker status to ensure balance of the biomarker values across treatment arms. To show clinical utility of the biomarker for identifying the population of patients who will have an overall better outcome with treatment A compared to treatment B, one must demonstrate that in the biomarker positive subgroup (in the simplest case of a binary biomarker) outcome is superior with treatment A, whereas, in the biomarker negative subgroup treatment B is ether the same or better than A. This design is the most informative in that it clearly distinguishes which treatment has greatest overall benefit in each biomarker subgroup.

A variant of the stratified design is what is sometimes referred to as an *all-comers design*. For this design only the analysis, and not the randomization, is stratified by the biomarker. The biomarker analysis may occur at the same time as the primary trial analysis or many years later using archived specimens. If carried out with appropriate rigor, such retrospective analyses of specimens from all-comers trials (a type of prospective-retrospective study) can provide a high level of evidence for clinical utility (Simon et al. 2009). Risks in using the all-comer design are that the biomarker measurements might not be available on some portion of the patients who are randomized (reducing the statistical power for the analyses) or the group of patients for whom biomarker measurements are available are

Table 1.1 Comparison of biomarker-driven clinical trial designs

Feature	Biomarker-stratified design	Enrichment design	Biomarker-strategy design, with biomarker assessment in the control arm	Biomarker-strategy design, without biomarker assessment in the control arm
Questions design can answer	What is the best treatment in each biomarker-defined subgroup? What is the best treatment in the overall study population? Is the biomarker-directed treatment strategy better than the control in the overall study population? (indirect assessment) Is the biomarker prognostic? Predictive?	What is the best treatment in the biomarker-positive patients?	Is the biomarker-directed treatment strategy better than the control treatment in the overall study population? (direct assessment) What is the best treatment in the biomarker-positive subgroup? Is the biomarker prognostic?	Is the biomarker-directed treatment strategy better than the control treatment in the overall study population? (direct assessment) What is the best treatment in the biomarker-positive subgroup? (indirect assessment) Is the biomarker prognostic? (indirect assessment)
Questions design cannot answer		What is the best treatment in the biomarker-negative subgroup? Is the biomarker prognostic? Predictive?	What is the best treatment in the biomarker-negative subgroup? Is the biomarker predictive?	What is the best treatment in the biomarker-negative subgroup? Is the biomarker predictive?
Advantages	Provides efficient assessment of relative treatment efficacy in each biomarker-defined subgroup and in the whole group	If the assumption that the biomarker reliably identifies the group likely to benefit from the experimental therapy is true, then the design provides an efficient test of efficacy of the experimental treatment in that subgroup, particularly if the biomarker positivity rate is low	Can be used for evaluation of complex biomarker-guided treatment strategies with a large number of treatment options or biomarker categories	Biomarker assessment is limited to the biomarker-directed arm (resource consideration) No issues associated with withholding the biomarker status from the control-arm patients Compliance not influenced by patient knowledge of the biomarker status in the control arm Can be used for evaluation of complex biomarker-guided treatment strategies with a large number of treatment options or biomarker categories
Disadvantages	The design is not feasible for evaluation of biomarker strategies with a large number of treatment options	If the experimental therapy is beneficial in a subgroup but the biomarker does not correctly identify this subgroup, a promising therapy may be missed A positive trial does not prove the utility of the biomarker because the relative treatment efficacy may be the same in the unevaluated biomarker-negative patients	A positive trial does not prove the utility of the biomarker because the experimental treatment may be better than the control treatment for all patients regardless of biomarker status Inefficiency	A positive trial does not prove the utility of the biomarker because the experimental treatment may be better than the control treatment for all patients regardless of biomarker status Inefficiency

Reprinted from Table 1 in Freidlin et al. (2010)

non-representative of the full patient group in such a way that the relationship between biomarker and treatment effect is distorted. In many situations these potential biases will not be a major problem, particularly if specimen collection is mandatory for trial eligibility.

The strategy design is another design which is sometimes used to establish that a biomarker has

clinical utility for treatment selection, but it has some limitations. This design can be viewed as a test of the combination of the biomarker test and associated treatment assignment algorithm; patients are randomized to have biomarker testing or not. Patients randomized to the arm with biomarker testing receive the treatment designated by a pre-defined algorithm based on biomarker value (e.g., experimental targeted therapy for biomarker-positive patients and standard therapy for biomarker-negative patients in the simplest case of a binary biomarker). Patients assigned to the no-testing arm receive a standard treatment. Use of the strategy design is usually discouraged because it is statistically inefficient (because biomarker-negative patients receive the same treatment on both arms) and does not allow for separation of biomarker and treatment effects; however, it may be the only viable option in situations where a biomarker takes many possible values or the treatment assignment algorithm is complex.

Summaries of biomarker-driven clinical trial designs and questions they are able to address are given by Table 4.1 in Micheel et al. (2012). Freidlin et al. (2010) discuss advantages and disadvantages for these designs (Table 1.1) and interim monitoring considerations as well as providing many examples of actual trials that used these designs or hybrids of them (Freidlin et al. 2010). Additionally, care must be taken in the statistical design of the stratified (or all comers) design to consider sequence of testing within biomarker subgroups and appropriate type I error control (Freidlin and Korn 2014). Further statistical details are beyond the scope of this discussion.

1.4.2.2 Considerations for Response-Predictive Biomarker Utility

Evaluation of clinical utility for a response-predictive biomarker requires consideration of both long and short term endpoints due to the uncertainties in the association between a near term response endpoint and a long term event-free survival (EFS) endpoint which may include overall survival as well as recurrence, progression, or other events. In a neoadjuvant setting, the ability to achieve a tumor response might offer the advantage of allowing change in surgical management from mastectomy to lumpectomy, resulting in less morbidity and a more favorable cosmetic outcome for a patient. However, it must also be considered whether the reduced surgery could lead to less favorable long term event free survival (EFS) or whether a delay in surgery due to administration of pre-operative therapy could have a detrimental effect on long term EFS, especially if the pre-operative therapy is at best modestly effective. In an advanced disease setting where surgery is not an option and where it is believed that a therapeutic agent will have long term EFS benefit only if it demonstrates activity in the form of tumor shrinkage, a biomarker would have clinical utility as a response-predictive biomarker if it can be established to reliably predict when a tumor will not respond. The clinical utility of such a biomarker would lie in its ability to identify futile treatments, sparing the patient toxicity and potentially allowing selection of an effective treatment more quickly.

To demonstrate clinical utility of a response-predictive biomarker in either the neoadjuvant or metastatic setting, generally a randomized trial would be needed comparing use of the biomarker to not using it; or, in rare instances it might be possible to rely on extensive historical data to establish that acting on the response-predictive biomarker leads to a net benefit to patients through some combination of positive effects on short and long term endpoints. Another challenge is that a biomarker could predict response for two different treatments but provide no information about which treatment would lead to better survival; higher response rate does not necessarily translate to better survival outcome. For example, meta-analyses of neoadjuvant clinical trials in breast cancer that collected both pathologic complete response and event-free survival outcomes were unable to demonstrate that a certain magnitude of difference in pathologic complete response rates reliably translates to a particular magnitude of difference in event-free survival (Cortazar et al. 2014; Berruti et al. 2014; Korn et al. 2016). All of these examples highlight the need for clinical evaluation of

Table 1.2 Comparison of clinical trial phases I–IV of therapeutic trials and tumor biomarker-monitoring trials

Phases of clinical trials, clinical validity				
Type of trial	Phase I	Phase II	Phase III	Phase IV
Therapeutic oncology trial	Explores toxicity and optimal dosage and/or schedule of a new therapy or a new use of an old therapy	Estimates whether the new therapy shows evidence of antineoplastic activity. Usually conducted for a specific disease condition	Compares, through randomization, the new therapy that showed promising results in phase II trials with the current standard of care	Evaluates the benefits, side effects, risks, and optimal use of the therapy over an extended period through long-term surveillance of patients
Tumor biomarker-monitoring trial	Explores the kinetics of the biomarker and the correlation between a change in tumor burden and a change in serial biomarker concentrations	Estimates the monitoring performance of serial biomarker measurements to identify, exclude, and predict a change in tumor burden	Compares, through randomization, whether early biomarker-guided intervention produces a clinical change that improves patient outcomes	Evaluates the change in long-term outcome in terms of overall survival and adverse effects after the biomarker-guided intervention has been introduced into routine use

Reprinted from Table 1 in Söletormos et al. (2013)

response-predictive biomarkers to consider the impact of their use on both short and long term endpoints.

1.4.3 Monitoring Biomarker Utility

Biomarkers are frequently used to monitor disease status during therapy for signs of treatment response, toxicity, resistance or disease progression, or after therapy to detect signs of recurrence or progression. In order to establish clinical utility of such biomarkers it must be shown that clinically significant changes can be detected above the background noise and that detecting those signals leads to a benefit that can be realized by changing therapeutic management. Demonstration of an association between a monitoring biomarker and a clinical outcome may be sufficient to establish clinical validity but it is insufficient to establish clinical utility. Generally it must be shown that the monitoring biomarker test can detect the change in disease status with sufficient lead time before the appearance of clinical signs and that with that lead time there are clinical decisions or actions which can be taken to improve outcome for the patient. Examples of clinical management changes when biomarker monitoring occurs during therapy include a switch to a new therapeutic agent or to a different treatment modality, a change in dose or schedule of the current therapeutic agent, or possibly terminating treatment completely. After completion of therapy, biomarker monitoring may be used to detect recurrent or progressive disease to allow for decisions regarding resumption of therapy. Monitoring biomarkers could also be used to guide decisions regarding initiation of therapy in an active monitoring situation for some in situ breast cancers. Demonstration of clinical utility for a monitoring biomarker typically requires a randomized trial in which patient clinical outcomes resulting from a strategy which acts on the biomarker is compared to that from a strategy independent of the biomarker. The *European Group on Tumor Markers* outlined a process for the rigorous evaluation of tumor biomarker-monitoring trials (Söletormos et al. 2013) as summarized in Table 1.2.

The S0500 trial is an example of a biomarker-monitoring trial in which the role of circulating tumors cells (CTCs) in managing chemotherapy for women receiving first line chemotherapy for metastatic breast cancer (newly metastatic disease or progressive metastatic disease while on hormonal therapy) was assessed (Smerage et al. 2014). Patients were first grouped according to CTC level at baseline. Arm A comprised those patients who did not have increased CTCs at baseline and who were recommended to remain on initial therapy until progression. Those patients who initially had increased CTCs but experienced a decrease in CTCs after 21 days of therapy were recommended to remain on initial therapy (arm B). Patients with persistently increased CTCs after 21 days of therapy were randomly assigned to continue initial therapy (arm C1) or change to an alternative chemotherapy (arm C2). This trial design permitted several questions to be addressed about the role of CTCs in the monitoring setting. A comparison of arm A to arms B + C1 addresses the prognostic ability of baseline CTCs in the context of unchanging standard therapy. A comparison of arms C1 and C2 addresses whether patients with persistently elevated CTCs after 21 days of therapy benefit from a change in chemotherapy; this treatment comparison constitutes an enrichment trial (enrichment for patients with persistently elevated CTCs after 21 days of therapy) embedded in the larger trial. The S0500 study confirmed that baseline CTCs were prognostic but was unable to demonstrate that a switch of cytotoxic chemotherapy was beneficial for those patients with persistently elevated CTCs. A question that is not addressed is whether patients who did not have elevated CTCs at day 21 would have benefitted from a change in chemotherapy. Nonetheless, it is unlikely that there would be interest in addressing that question given the null trial results for those patients with persistently elevated CTCs.

1.5 Regulatory Considerations

In the United States, CLIA regulations require that laboratories performing biomarker tests and returning the results to a patient or the patient's physician must follow good laboratory practices (CMS 2016), but there are not specific CLIA requirements for clinical validation or documentation of comparability of test results between different laboratories. The FDA has longstanding regulatory processes for approval or clearance of biomarker tests which are marketed as devices, but there has been confusion regarding what types of tests meet the definition of a laboratory developed test (LDT) not requiring FDA review, versus fall under the regulatory system for medical devices. Consequently there is the potential for gaps in the evidence supporting biomarker tests offered by some laboratories, particularly if those laboratories do not participate in other quality assurance programs such as those offered through the College of American Pathologists (College of American Pathologists 2016).

Historically FDA has defined an LDT as "an IVD [in vitro diagnostic] that is intended for clinical use and designed, manufactured and used within a single laboratory" (U.S. FDA 2014b). This definition would not cover, for example, a laboratory test developed by a commercial or health system central laboratory and offered through multiple laboratories within its network; such tests are technically subject to FDA review because, strictly speaking, they are not LDTs although rarely have they been reviewed by FDA. FDA's recent draft guidance on a proposed new regulatory approach for LDTs signals its intent to consider increased regulation of both IVDs meeting the traditional LDT definition as well as an expanded definition that would include IVDs that are offered by a CLIA-certified laboratory as an "LDT" (and have not undergone any FDA review for clearance or approval) even

though they might not meet the strict historical definition of an LDT (U.S. FDA 2014b).

CLIA requirements, which apply any time a test result is retuned to a patient or the patient's physician, must be adhered to regardless of whether the biomarker test is being performed for investigational purposes in the context of a clinical trial or is being used for routine clinical care. Researchers conducting clinical trials in which biomarker tests will be used must also be aware that such use might require an Investigational Device Exemption (IDE) (U.S. FDA 2014a). Applications for IDEs undergo review for evaluation of the potential risks associated with use of the test weighed against possible benefits with particular emphasis on analytical performance of the biomarker tests.

It is important for laboratories and clinical investigators to remain current in their understanding of, and compliance with, regulatory requirements. A large percentage of biomarker tests currently in use for guiding clinical care decisions have received little external review. Whether through increased regulatory oversight, or wider adoption of best practices for development and evaluation of biomarker tests, it is critical to ensure the safety and efficacy of biomarker tests used in clinical decision making.

1.6 Discussion

Biomarker-based tests are increasingly being used in oncology and are integral to the implementation of precision medicine. Best practices for the development and evaluation of these tests need to be followed, just as rigorous processes are required for the development of new therapeutics. Of paramount importance to this evaluation process is careful consideration of intended use, which includes the clinical setting, patient and specimen characteristics, and the decisions that are to be informed by use of the biomarker test. The goal of this chapter was to outline the principles of analytical and clinical validation and considerations for assessment of clinical utility to promote enhanced understanding of the translational process and wider adoption of best practices. Adherence to these best practices will increase the chances that biomarker tests will perform reliably on real-world clinical materials and that the results can be relied upon to have particular clinical interpretations leading to clinical decisions that benefit patients.

References

Allard WJ, Matera J, Miller MC et al (2004) Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. Clin Cancer Res 10:6897–6904

Altman DG, McShane LM, Sauerbrei W et al (2012) Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. BMC Med 10:51. http://www.biomedcentral.com/1741-7015/10/51

Anderson JR, Cain KC, Gelber RD (1983) Analysis of survival by tumor response. J Clin Oncol 1(11):710–719

Becker R Jr (2015) Analytical validation of in vitro diagnostic tests. In: Matsui S, Buyse M, Simon R (eds) Design and analysis of clinical trials for predictive medicine. Chapman and Hall/CRC, Boca Raton, pp 33–49

Bergfeldt K, Rydh B, Granath F et al (2002) Risk of ovarian cancer in breast-cancer patients with a family history of breast or ovarian cancer: a population-based cohort study. Lancet 360:891–894

Berruti A, Amoroso V, Gallo F et al (2014) Pathologic complete response as a potential surrogate for the clinical outcome in patients with breast cancer after neoadjuvant therapy: a meta-regression of 29 randomized prospective studies. J Clin Oncol 32:3883–3891

Bossuyt PM, Reitsma JB, Bruns DE et al (2003a) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for reporting of diagnostic accuracy. Clin Chem 49(1):1–6

Bossuyt PM, Reitsma JB, Bruns DE et al (2003b) The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 49(1):7–18

Bossuyt PM, Reitsma JB, Bruns DE et al (2015) An updated list of essential items for reporting diagnostic accuracy studies. Clin Chem 61(12):1446–1452. doi:10.1373/clinchem.2015.246280

Brekelmans CTM, Tilanus-Linthorst MMA, Seynaeve C et al (2007) Tumour characteristics, survival and prognostic factors of hereditary breast cancer from BRCA2-, BRCA1- and non-BRCA1/2 families as compared to sporadic breast cancer cases. Eur J Cancer 43:867–876

Centers for Medicare & Medicaid Services (CMS) (2016) Clinical laboratory improvement amendments (CLIA).

http://www.cms.gov/CLIA/05_CLIA_Brochures.asp. Accessed 29 Feb 2016

Clinical Laboratory Standards Institute (CLSI) (2010) I/LA28-A2: quality assurance for design control and implementation of immunohistochemistry assays, approved guideline, 2nd edn. Clinical Laboratory Standards Institute, Wayne, PA

College of American Pathologists (2016) 2016 surveys and anatomic pathology education programs. http://www.cap.org/ShowProperty?nodePath=/UCMCon/Contribution%20Folders/WebContent/pdf/2016-surveys-catalog.pdf. Accessed 29 Feb 2016

Cortazar P, Zhang L, Untch M (2014) Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet 384:164–172

Cox DR (1972) Regression models and life-tables. J R Stat Soc B 34(2):187–220

Cronin M, Sangli C, Liu M-L et al (2007) Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor–positive breast cancer. Clin Chem 53(6):1084–1091

EQUATOR Network (2016) Enhancing the quality and transparency of health research. http://www.equator-network.org/reporting-guidelines/. Accessed 29 Feb 2016

FDA-NIH Biomarker Working Group (2016) Glossary. In: BEST (Biomarkers, EndpointS, and other Tools) resource. U.S. Food and Drug Administration (FDA) and National Institutes of Health (NIH). Available via Internet. http://www.ncbi.nlm.nih.gov/books/NBK338448/. Accessed 29 Feb 2016

Filipits M, Rudas M, Jakesz R et al (2011) 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 17:6012–6020

Freidlin B, Korn EL (2014) Biomarker enrichment strategies: matching trial design to biomarker credentials. Nat Rev Clin Oncol 11:81–90

Freidlin B, McShane LM, Korn EL (2010) Randomized clinical trials with biomarkers: design issues. J Natl Cancer Inst 102(3):152–160

Genomic Health (2016) Oncotype DX breast cancer assay. http://www.oncotypedx.com/. Accessed 29 Feb 2016

Hammond MEH, Allred DC, Dowsett M et al (2010) American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen/progesterone receptors in breast cancer. J Clin Oncol 28(16):2784–2795

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36

Harris LN, Ismaila N, McShane LM et al (2016) 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 34 (10):1134–1150

Jennings L, Van Deerlin VM, Gulley ML et al (2009) Recommended principles and practices for validating clinical molecular pathology tests. Arch Pathol Lab Med 133:743–755

Korn EL, Sachs MC, McShane LM (2016) Statistical controversies in clinical research: assessing pathologic complete response as a trial-level surrogate end point for early-stage breast cancer. Ann Oncol 27:10–15

Linnet K, Boyd JC (2012) Selection and analytical evaluation of methods with statistical techniques. In: Burtis CA, Ashwood ER, Bruns DE (eds) Tietz textbook of clinical chemistry and molecular diagnostics, 5th edn. Elsevier Saunders, St Louis, pp 7–47

Matsui S, Buyse M, Simon R (eds) (2015) Design and analysis of clinical trials for predictive medicine. Chapman and Hall/CRC, Boca Raton

McShane LM, Polley M-YC (2013) Development of omics-based clinical tests: the challenge of achieving statistical robustness and clinical utility. Clin Trials 10:653–665

McShane LM, Altman DG, Sauerbrei W et al (2005) Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 97 (16):1180–1184

Micheel CM, Nass S, Omenn GS (eds) (2012) Evolution of translational omics: lessons learned and the path forward. The National Academies Press, Institute of Medicine, Washington DC

Moore HM, Kelly AB, Jewell SD et al (2011) Biospecimen reporting for improved study quality (BRISQ). Cancer Cytopathol 119:92–101

Nielsen T, Wallden B, Schaper C et al (2014) Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. BMC Cancer 14:177

Pathmanathan N, Balleine RL, Jayasinghe UW et al (2014) The prognostic value of Ki67 in systemically untreated patients with node-negative breast cancer. J Clin Pathol 67:222–228. doi:10.1136/jclinpath-2013-201793

Pennello GA (2013) Analytical and clinical evaluation of biomarkers assays: when are biomarkers ready for prime time? Clin Trials 10:666–676

Polley M-YC, Freidlin B, Korn EL et al (2013a) Statistical and practical considerations for clinical evaluation of predictive biomarkers. J Natl Cancer Inst 105(22):1677–1683

Polley M-YC, Leung SCY, McShane LM et al (2013b) An international Ki67 reproducibility study. J Natl Cancer Inst 105(24):1897–1906

Sargent DJ, Conley BA, Allegra C et al (2005) Clinical trial designs for predictive marker validation in cancer treatment trials. J Clin Oncol 23:2020–2027

Simon RM, Paik S, Hayes DF (2009) Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 101:1446–1452

Smerage JB, Barlow WE, Hortobagyi GN et al (2014) Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. J Clin Oncol 32:3483–3489

Söletormos G, Duffy MJ, Hayes DF et al (2013) Design of tumor biomarker-monitoring trials: a proposal by the European group on tumor markers. Clin Chem 59 (1):52–59

Stuart-Harris R, Caldas C, Pinder SE et al (2008) Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients. Breast 17:323–334

U.S. Food and Drug Administration (FDA) (2006) 510(k) Summary for CELLSEARCH® circulating tumor cell kit. http://www.accessdata.fda.gov/cdrh_docs/pdf6/k062013.pdf. Accessed 29 Feb 2016

U.S. Food and Drug Administration (FDA) (2013) 510(k) Summary for Prosigna™ breast cancer prognostic gene signature assay. http://www.accessdata.fda.gov/cdrh_docs/pdf13/k130010.pdf. Accessed 29 Feb 2016

U.S. Food and Drug Administration (FDA) (2014a) FDA decisions for investigational device exemption clinical investigations: guidance for sponsors, clinical investigators, institutional review boards, and Food and Drug Administration staff. U.S. Department of Health and Human Services, Food and Drug Administration, Rockville MD. Available via Internet. http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm279107.pdf. Accessed 29 Feb 2016

U.S. Food and Drug Administration (FDA) (2014b) Framework for regulatory oversight of laboratory developed tests (LDTs): draft guidance for industry, Food and Drug Administration staff, and clinical laboratories. U.S. Department of Health and Human Services, Food and Drug Administration, Rockville MD. Available via Internet. http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm416685.pdf. Accessed 29 Feb 2016

U.S. Food and Drug Administration (FDA) (2015) MammaPrint® FFPE 510(k) substantial equivalence determination decision memorandum: assay and instrument combination template. http://www.accessdata.fda.gov/cdrh_docs/reviews/k141142.pdf. Accessed 29 Feb 2016

Van Poznak C, Somerfield MR, Bast RC et al (2015) Use of biomarkers to guide decisions on systemic therapy for women with metastatic breast cancer: American Society of Clinical Oncology clinical practice guideline. J Clin Oncol 33(24):2695–2704

Wolff AC, Hammond ME, Schwartz JN et al (2007) 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 25(1):118–145

Wolff AC, Hammond ME, Hicks DG et al (2013) 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 31(31):3997–4013

Zou KH, O'Malley AJ, Mauri L (2007) Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation 115:654–657