



Validation of New Cancer Biomarkers: A Position Statement from the European Group on Tumor Markers

Michael J. Duffy,^{1*} Catharine M. Sturgeon,² György Sölétormos,³ Vivian Barak,⁴ Rafael Molina,⁵
Daniel F. Hayes,⁶ Eleftherios P. Diamandis,⁷ and Patrick M.M. Bossuyt⁸

BACKGROUND: Biomarkers are playing increasingly important roles in the detection and management of patients with cancer. Despite an enormous number of publications on cancer biomarkers, few of these biomarkers are in widespread clinical use.

CONTENT: In this review, we discuss the key steps in advancing a newly discovered cancer candidate biomarker from pilot studies to clinical application. Four main steps are necessary for a biomarker to reach the clinic: analytical validation of the biomarker assay, clinical validation of the biomarker test, demonstration of clinical value from performance of the biomarker test, and regulatory approval. In addition to these 4 steps, all biomarker studies should be reported in a detailed and transparent manner, using previously published checklists and guidelines. Finally, all biomarker studies relating to demonstration of clinical value should be registered before initiation of the study.

SUMMARY: Application of the methodology outlined above should result in a more efficient and effective approach to the development of cancer biomarkers as well as the reporting of cancer biomarker studies. With rigorous application, all stakeholders, and especially patients, would be expected to benefit.

© 2015 American Association for Clinical Chemistry

Biomarkers play an important and sometimes indispensable role in the detection and management of patients

¹ Clinical Research Centre, St Vincent's University Hospital and UCD School of Medicine and Medical Science, Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland; ² Department of Clinical Biochemistry, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom; ³ Department of Clinical Biochemistry, University of Copenhagen North Zealand Hospital, Hillerød, Denmark; ⁴ Oncology Department, Hadassah University Hospital, Jerusalem, Israel; ⁵ Hospital Clinic, University of Barcelona, Barcelona, Spain; ⁶ University of Michigan Comprehensive Cancer Center, Ann Arbor, MI; ⁷ Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; ⁸ Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

* Address correspondence to this author at: UCD Clinical Research Centre, St Vincent's University Hospital, Dublin 4, Ireland. Fax +353-1-269-6018; e-mail michael.j.duffy@ucd.ie.

Received February 24, 2015; accepted April 2, 2015.

Previously published online at DOI: 10.1373/clinchem.2015.239863

© 2015 American Association for Clinical Chemistry

with malignancy. Thus, in asymptomatic individuals, biomarkers may be used to screen for early cancer or premalignant conditions, whereas in symptomatic patients, biomarkers may help in the differential diagnosis of benign and malignant diseases. Following a diagnosis of malignancy, biomarkers may help determine prognosis and identify the most appropriate therapy. In patients who have undergone curative-intent surgery for cancer, biomarkers may be used in follow-up surveillance and for the early detection of possible recurrent disease. For patients receiving systemic treatment, biomarkers may provide a minimally invasive approach for monitoring tumor response [for review, see references (1) and (2)].

Despite the massive number of publications on tumor “biomarkers” in recent years, no new widely used cancer serum biomarker and only a handful of tissue-based biomarkers have entered clinical use in the past 25 years (3). According to Diamandis (3), this is not due to the “lack of biological/biomedical knowledge, powerful technologies or investment of funds.” Rather, the low number of clinically used biomarkers appears largely to be a result of the absence of a clearly defined validation pathway for advancing a newly discovered “biomarker” into the clinic (4).

The aim of this review is thus to discuss the key steps in taking an emerging serum or tissue cancer biomarker from a research laboratory into routine clinical use. The primary focus will be on single-analyte assays, but much of the content will also be relevant to multianalyte or omics-related biomarkers. In addition to focusing on the various steps in validation, we also discuss the registration and reporting of biomarker studies. Although most of the review focuses on cancer biomarkers, the content should also be relevant to the evaluation of disease biomarkers in general. In the review, it is assumed that preliminary results are available, suggesting that the biomarker of interest has clinical potential.

Steps in Cancer Biomarker Validation

SAMPLE COLLECTION AND PROCESSING

An important step in biomarker validation that frequently fails to receive the attention it should is the evaluation of preanalytical factors that may affect the mea-

sured concentration of the biomarker. These factors, which are not unique to cancer biomarkers but apply to all clinically used biomarkers, include both patient/study participant and specimen-related variables (5). Patient-related factors for blood-based assays include age, sex, posture, hydration status, whether fasting or not, and previous or ongoing treatments (e.g., medicines that patients may be taking). In addition, potential confounding factors for serum-based screening/diagnostic biomarkers should be noted. These include benign disease of the liver, kidney, and lung as well as various inflammatory diseases. All of these disorders can result in increased blood concentrations of protein tumor biomarkers.

Specimen-related factors depend on whether blood or tissue is being collected. Factors for consideration when collecting blood include specimen type, i.e., whether it is whole blood, serum, or plasma and if the latter the appropriate anticoagulant. Other factors that may affect the measured concentration include hemolysis (which may lead to spurious increases), centrifugation conditions (time, speed, and temperature) and stability during transport to the laboratory, processing and storage.

Although serum and plasma are frequently used interchangeably, major differences exist in their protein profile. This is because the coagulation during the preparation of serum processes clotting proteins. In contrast, due to the presence of anticoagulants, clotting does not occur when preparing plasma. It is important therefore that either serum or plasma is exclusively used during the validation of a blood biomarker test, unless these fluids have been shown to be interchangeable.

Because of its potential to act as a surrogate for the entire cancer genome, there has been considerable interest recently in measuring circulating tumor DNA (ctDNA)⁹ (6). Owing to its lower background concentration, plasma is generally recommended for this purpose (6). However, the optimum processing steps for ctDNA remain to be established. Similarly, the optimum conditions for the processing of circulating tumor cells (CTC) remain to be investigated. However, recently, special preservative-containing Vacutainer Tubes were successfully used in the conduct of a multiinstitutional trial involving CTC (7).

Because several different factors can affect the stability and thus the measured concentration of a candidate

biomarker, it is important that a pilot study is initially performed to identify the optimum sample collection and storage conditions. This should be performed before clinical validation.

For tissue-based biomarkers, it is important to establish whether fresh or freshly frozen samples are necessary or whether fixed and paraffin-embedded tissue can be used. Although fresh or freshly frozen tissue may be used in research laboratories, tissue-based biomarkers measured for clinical use generally employ formalin-fixed and paraffin-embedded tissue. For the latter, time to fixation as well as the fixation period are among the most critical factors contributing to variability in results, affecting quality of both RNA and protein. Published guidelines for obtaining optimum protein staining include immersion in fixative within 1 h of removal, fixation in 10% neutral buffered formalin for 24 h, dehydration for 1.5–15 h, and embedding in paraffin for 0.5–4.5 h (8). The specific optimum conditions, however, may vary from protein to protein. For extraction of biomarkers from tumor tissue, Peña-Llopis and Brugarolas (9) recently described a method for the simultaneous isolation of high-quality DNA, RNA, and microRNA as well as protein from the same sample (9). In this report, tissue with high tumor cell purity was selected on the basis of the histology of immediately adjacent sections. Other issues relevant to the measurement of tissue-based biomarkers such as intratumor heterogeneity (10) and tumor cell contamination with nonmalignant cells (11) are outside the scope of this report.

Because sample collection and initial processing may be carried out by nonscientific personnel, it is essential that adequate training is provided and detailed standard operating procedures (SOP) used. Any deviation from the SOP for any sample should be noted. Furthermore, for difficult or manual assays, consensus-based guidelines detailing preanalytical, analytical, and reporting issues should be published. Exemplary guidelines are currently available for measuring established biomarkers such as estrogen receptors, progesterone receptors, and human epidermal growth factor receptor 2 (HER2) protein in breast cancer (12, 13), for epidermal growth factor receptor (*EGFR*)¹⁰ mutations and anaplastic lymphoma receptor tyrosine kinase (*ALK*) translocations in non-small cell lung cancer (14), and for *KRAS* mutation testing in colorectal cancer (15). In addition to these guidelines for measuring specific biomarkers, general guidelines, including analyte stability and laboratory

⁹ Nonstandard abbreviations: ctDNA, circulating tumor DNA; CTC, circulating tumor cells; SOP, standard operating procedures; HER2, human epidermal growth factor receptor 2; FDA, US Food and Drug Administration; PRoBE, Prospective Specimen Collection, Retrospective Blinded Evaluation; CA 125, cancer antigen 125; LOE, level of evidence; uPA, urokinase plasminogen activator; PAI-1, plasminogen activator inhibitor-1; LDT, laboratory-developed tests; CE, Conformité Européenne; EU, European Union; PSA, prostate-specific antigen.

¹⁰ Human genes: *EGFR*, epidermal growth factor receptor; *ALK*, anaplastic lymphoma receptor tyrosine kinase; *KRAS*, Kirsten rat sarcoma viral oncogene homolog; *WFDCC2*, WAP four-disulfide core domain 2 (also known as HE4).

Table 1. Parameters used in the analytical validation of biomarkers measured by quantitative assays.^a

Parameter	Definition
Accuracy ^b	How close a result is to the true result.
Precision/imprecision	Closeness of agreement between a series of measurement for the same sample established under specific conditions. Depends on repeatability and reproducibility of assay.
Repeatability	Describes measurements made under the same conditions.
Reproducibility	Describes measurements done under different conditions.
Analytical specificity	Ability of an assay to distinguish the analyte of interest from structurally similar molecules.
Analytical sensitivity	Ability of an assay to detect low quantities of an analyte.
Limit of detection ^c	Lowest amount of analyte that can be reliably distinguished from zero.
Interference (cross-reaction) ^d	Effect of a substance in a sample that alters the correct value of a result.
Carryover ^c	Occurs when a portion of a sample or reaction reagent are unintentionally transferred from one assay reaction into another.
Linearity	The ability of an assay to give concentrations that are directly proportional to the levels of the analyte following sample dilution.
Robustness	Precision of an assay following changes in assay conditions, e.g., variation in ambient temperature, storage condition of reagents.

^a Adapted from Jennings et al. (23).
^b For qualitative assays, accuracy has been defined as the amount agreement between the information in the assay undergoing evaluation and that obtained from the best available method for determining the presence or absence of that analyte (23).
^c Particularly relevant for quantitative assays carried on blood.
^d May be due to chemically related molecules or heterophilic or human anti-mouse antibodies.

QC, for performing analysis of tissue-based molecular biomarkers, have been published by Cree et al. (16).

Analytical Validation of Biomarker Assays

Because multiple assays using different formats may exist for a biomarker, it is necessary to develop and validate each specific assay that will be used in the clinic (17, 18). For serum protein-based biomarker assays, quantitative immunoassays (e.g., automated ELISA) are usually used for this purpose. Tissue-based biomarkers, on the other hand, are most frequently detected by immunohistochemistry at the protein level and by in situ hybridization at the DNA level. Newer assays for nucleic acid-based biomarkers may involve mutation analysis for DNA-based biomarkers (e.g., single genes, panels of genes, whole exome sequencing, or whole genome sequencing) (19) and either reverse transcription-PCR (e.g., Oncotype Dx) (20) or microarray (e.g., MammaPrint) (21) for mRNA-based biomarkers.

Irrespective of the assay format, a new biomarker assay must undergo analytical (technical) validation. Analytical validation involves confirming that the method used for the biomarker measurement is accurate, precise, specific, robust, and stable over time (17, 18, 22–24). Additional criteria for evaluating quantitative methods such as ELISA-type or quantitative PCR assays include

linearity with sample dilution, parallelism, recovery following analyte addition, and functional sensitivity (See Table 1 for definitions of these and related terms).

Consensus-based metrics for defining limits for most of these variables are not available at present. However, according to the National Academy of Clinical Biochemistry (US), clinically used serum-based immunoassays should have interassay CVs of <10% and within-assay variability of <5% (7). Acceptable precision is particularly important at clinical decision concentrations. For immunohistochemistry, reproducibility is generally determined using the κ -statistic or percentage agreement between 2 individual assessors (25). Agreement of $\geq 85\%$ is generally regarded as acceptable (25). Further information on metrological specifications is available in the Stockholm Consensus Statement on quality requirements for laboratory medicine tests (26).

Frequently, the initial assay development and validation occurs in a research laboratory (academic or industrial), where sample numbers are low and urgency of results less important than in a clinical setting (27). In that situation, further validation should take place in the “cut and thrust” of a clinical laboratory, where massive numbers of samples are being processed and rapid result reporting is necessary. This dual site validation also provides evidence for transferability of an assay. Furthermore, for large-volume assays such as serum-based pro-

Table 2. Parameters used in evaluating clinical validity of a biomarker.

Parameter	Definition
Clinical sensitivity	True positive rate, how good is the test in detecting individuals who have the condition of interest.
Clinical specificity	True negative rate, how good is the test in correctly excluding individuals without the condition of interest.
PPV ^{a,b}	Proportion of positive tests that are correct.
NPV ^b	Proportion of negative tests that are correct.
Positive likelihood ratio	How much more likely is a positive test to be found in an individual with the relevant condition than in a person without it.
Negative likelihood ratio	How much more likely is a negative test to be found in an individual without the relevant condition than in a person with it.
AUC	Area under ROC curve. AUC is used to compare different tests, i.e., an AUC value close to 1 indicates good discrimination, whereas an AUC of 0.5 provides no useful diagnostic information.
ROC analysis	A graphical approach for showing accuracy across the entire range of biomarker concentrations.
Hazard ratio	Chance of an event (e.g., disease recurrence, death) occurring in the treatment arm divided by the chance of the event occurring in the control arm, or vice versa.
Relative risk	Ratio of the probability of an event (e.g., disease recurrence, death) occurring in treated group to the probability of the event occurring in the control group.

^a PPV, positive predictive value; NPV, negative predictive value; AUC, area under curve.
^b With PPV and NPV, it is necessary to define the population to which it applies.

tein biomarkers, it is generally necessary to automate the analytical process, a step which is almost always performed by commercial diagnostic companies (27). Following test automation, a further round of validation is necessary, both by the diagnostic company as well as by the clinical laboratory in which the test is being used.

The potential dangers of using unreliable and non-validated immunoassays in biomarker development have been addressed by Diamandis and coworkers (28, 29). Among the suggestions made by these authors to minimize this possibility were to purchase such reagents from companies with a proven quality record, perform local validation, and report any identified problems.

It is widely recommended that analytical validation of an emerging biomarker assay be performed relatively early during its development (30). This is to ensure that it provides the necessary accuracy and robustness necessary for clinical use. For evaluating companion diagnostic biomarkers (biomarkers that are essential for the safe and effective use of a therapeutic product), the US Food and Drug Administration (FDA) has suggested that the development and validation parallel that of its companion drug (31).

Clinical Validation of Biomarker Tests

Clinical validation ensures that the results of the biomarker stratify individuals into different groups, such as

those with or without disease or those likely to have a good outcome vs those at an increased risk of disease recurrence (17, 18, 24). For diagnostic tests, clinical validation is usually reported in terms of diagnostic accuracy. Criteria used for describing diagnostic accuracy of a biomarker test include sensitivity for disease, specificity for disease, positive predictive value, negative predictive value, likelihood ratio, and ROC analysis (see Table 2 for definitions of these and related terms).

Ideally, a new biomarker should aim to fulfill a currently unmet need in cancer detection and/or patient management. Alternatively, if the biomarker test does not address an unmet need, it should provide an advantage over preexisting biomarker(s) for the clinical question being addressed, such as being more accurate and simpler to measure, providing a more rapid result, or being made available at reduced costs (e.g., by reducing frequency of imaging).

Clinical validation is more demanding and time-consuming than analytical validation. It should start with defining the specific context in which the new biomarker is to be used, i.e., which cancer type and whether the biomarker is to be used in screening, aiding diagnosis, determining prognosis, therapy prediction, or patient monitoring. This latter question is important, because it identifies the population of individuals to be used in the clinical validation step, the informative statistics, and the acceptability criteria. Thus samples should be collected

from subjects, similar with respect to disease status in the target group, in which it is intended to use the biomarker (5).

Problems in Clinical Validation

BIAS

One of the most common problems in clinical validation is bias or systematic differences between groups of individuals being compared in diagnostic studies, e.g., patients and controls. According to Ransohoff and Gourlay (32, 33), bias can be defined as “the systematic erroneous association of a characteristic with a group in a way that distorts a comparison with another group.” Thus, the presence of bias may produce positive findings that are unrelated to the clinical reality and furthermore are not reproducible (32, 33). For example, a common problem with diagnostic biomarker studies is that samples are taken from conveniently available archived samples that have been collected from separate populations that are mismatched for age range, sex, race, or other factors that may or may not lead to unintentional bias. Any observed difference in marker concentration between the 2 groups might thus be due to these differences rather than the presence or absence of disease. Another potential source of bias in screening/diagnostic biomarkers is differences in the transport, processing, and storage of samples from control and diseased groups.

For eliminating bias in biomarker clinical validation studies, Pepe et al. (34) have suggested a nested case control study in which samples are collected prospectively before a diagnosis is established and are then evaluated in a blinded fashion, retrospectively [i.e., Prospective Specimen Collection, Retrospective Blinded Evaluation (PRoBE) design]. Only after outcome data are available are random samples from cases and controls selected for the study. This retrospective and random approach minimizes the problem of baseline nonequivalence, because samples are collected without knowledge of disease status or outcome, so both selection of study participants and sample collection should be completely objective (34). Furthermore, systematic biases can be eliminated by requiring similar handling of samples from patients and controls. According to Pepe et al. (34), the PRoBE design can be used in the evaluation of screening, diagnostic and prognostic biomarkers.

The PRoBE design was recently used to compare the sensitivity of 35 biomarkers for the detection of early ovarian cancer (35). In this study, prediagnostic blood samples were collected as part of the PLCO (Prostate–Lung–Colorectal–Ovarian) screening trial. Despite being the earliest biomarker of the 35 investigated, cancer antigen 125 (CA 125) was identified as the most sensitive, outperforming all the others, including WAP four-disulfide core domain 2 (WFDC2; also known as HE4).

In a related study, CA 125 was shown to be at least as good as several multimarker panels previously proposed for ovarian cancer (36). These 2 high-quality studies clearly demonstrate that CA 125 is at present the best single biomarker for the early detection of ovarian cancer.

OVERFITTING

In proteomic and genomic studies, thousands of variables can be measured simultaneously in relatively small numbers of patients. The enormous amount of information generated is frequently used to model the system (the disease) from which the data were generated. These models are often subsequently used to predict various clinical parameters, such as outcome. Many statistical approaches (often based around regression) can be used to model an omics system but they are vulnerable to overfitting, as the number of variables generated via high-throughput omics technologies far exceeds the number of samples examined. This can result in findings or predictions identified in a study population that are not reproduced in populations different from those used to derive the model (37–39). Such model overfitting can be avoided with appropriate internal and external validation studies (37–39).

Internal validation can be conveniently achieved by dividing the study population into 2 independent groups. One of these—the “training set”—is used to train or build the model. The second group—the “validation set”—is then used to test whether the model works equally well in a different group of individuals, independent from those in the training set. Both populations should be similar to the population in which the model is to be used. Alternatively, testing can be carried out iteratively by cross-validation, using different training and test groups. Several approaches are available for performing cross validation, which are discussed in detail in references (37–39).

External validation with a completely different study population is also necessary (37–39). As for internal validation, the samples used should be similar to those for the target population in which the test will be used (17). The more external validations that are performed, especially if at different locations and time points, the more robust and generalizable the model should be. To ensure that accurate and independent outcome data are obtained, the number of patients in the group used for external validation must be large enough for the study to achieve sufficient statistical power (38).

MULTIPLICITIES

In addition to the potential problems posed when investigating large numbers of biomarkers, further statistical problems may occur with other multiples such as disease subset analysis and use of several different endpoints (overall survival, progression-free survival, objective re-

Table 3. Cancer biomarkers that have undergone/or are undergoing validation in level I evidence (LOE I) studies.

Biomarker	Clinical use	Type of validation
FOBT ^a	Screening for colorectal cancer	PRCT
PSA	Screening for prostate cancer ^b	PRCT
CA 125	Screening for ovarian cancer ^b	PRCT
uPA/PAI-1	Determining prognosis in breast cancer	PRCT, pooled analysis
Estrogen receptor	Predicting response to hormone therapy in breast cancer	PRCT, metaanalysis
HER2	Predicting response to anti-HER2 therapy in breast cancer	PRCT
Oncotype DX ^b	Determining prognosis in ER-positive lymph node-negative breast cancer	PRT
MammaPrint ^b	Determining prognosis in lymph node-negative breast cancer	PRT
<i>BRAF</i> mutation	Predicting response to anti- <i>BRAF</i> therapy in melanoma	PRCT
<i>KRAS</i> mutations	Predicting response to anti-EGFR antibodies in colorectal cancer	PRT
<i>EGFR</i> mutations	Predicting response to anti-EGFR kinase inhibitors in non-small cell lung cancer	PRT
CEA	Postoperative surveillance after curative surgery in colorectal cancer	PRT, metaanalyses

^a FOBT, fecal occult blood testing; PRCT, prospective randomized clinical trial; PRT, prospective retrospective trial; ER, estrogen receptor; CEA, carcinoembryonic antigen.
^b PRCT in progress. The fact that a biomarker has undergone validation in a LOE I evidence study does not necessarily qualify that biomarker for clinical use. Thus, CA 125 is not currently recommended for screening asymptomatic women for ovarian cancer although it was evaluated in a prospective randomized trial (which showed that screening with CA 125 and ultrasound failed to reduce mortality from ovarian cancer).

sponse rate, duration of response) (40). According to Berry (40), multiplicities are widespread and indeed can be silent, i.e., unreported or unrecognized. To minimize potential problems with multiplicities, it is essential to have a written protocol prepared in advance that specifies the study aims and methods to be used. Furthermore, all the steps planned to be done and everything that was completed should be reported (40).

Demonstration of Clinical Value

Analytical and clinical validation are not sufficient to recommend use of a tumor biomarker test in standard practice. In addition to these requirements, the test should be shown to have clinical value (utility), meaning that patient outcomes are improved by directing care based on the test results compared to care of the patient without the test. Although a large number of biomarkers have undergone analytical and clinical validation, relatively few have been shown to have clinical value (Table 3). Indeed, historically, most biomarkers entered clinical use with little evidence of clinical value. In the future however, demonstration of clinical value is likely to be necessary and indeed should be mandatory for the adoption of a new biomarker into clinical practice. Furthermore, it should be mandatory for reimbursement.

Clinical value can be demonstrated if high-level evidence exists that measurement of a biomarker or biomarker profile alters patient management in a manner that positively impacts on outcome, compared to out-

come without use of the biomarker (17, 18, 24, 41, 42). Ideally, for demonstrating clinical value, it is necessary to show that the biomarker measurement results in clinical management that increases overall survival, without adversely affecting the patient. Obtaining such evidence requires large studies with numerous patients and is expensive and time-consuming. Alternative outcome measures that have been proposed include increased length of disease-free interval, reduced cost of care [e.g., due to earlier diagnosis, fewer inpatient or outpatient hospital visits, fewer invasive procedures, withdrawal of ineffective treatment, or increased quality of life (less use of toxic therapy)] (18). Widely acceptable outcomes for these end points are presently not available. For example, should the performance of a biomarker test lead to a treatment that extends patient survival by 3 months, 6 months, or longer?

Demonstration of clinical value should be based on obtaining a benefit of sufficient magnitude that is likely to be clinically meaningful and statistically significant. Ideally, this should be obtained using a high level of evidence (LOE), i.e., an LOE I study (43). The gold standard method for obtaining LOE I is testing the biomarker in a prospective randomized trial in which the biomarker evaluation is the primary purpose of the trial. The design of the prospective randomized trial depends on the intended use of the biomarker. Thus, for evaluating a screening biomarker, the target population should be randomized to have or not to have the biomarker of interest measured. For evaluating therapy-predictive bio-

markers, several trial designs have been proposed which have previously been discussed in detail (44–47). Ideally, prognostic biomarkers should be evaluated in patients not receiving systemic adjuvant therapy or, when this is not possible, in patients receiving the standard therapy for that malignancy. This is best carried out prospectively, although a retrospective study of sufficiently high statistical power without bias and using an analytically validated assay may also be acceptable (48).

Although validation in a randomized prospective trial has long been regarded as the gold standard methodology for demonstration of clinical value, such an approach is time-consuming, requires large numbers of study participants, and is expensive. In the event of validation in a prospective randomized trial not being possible, LOE I evidence may be obtained from a prospective-retrospective trial employing archival samples collected from a previously completed prospective study (49). There are, however, some caveats when using such an approach. Thus, with a prospective-retrospective design, investigators must ensure that the following conditions are met (49):

- There is enough suitable archival tissue available from a sufficient number of patients to achieve the necessary statistical power. [It has been suggested that samples from at least two-thirds of the participants in the relevant prospective trial should be available for biomarker testing (49)].
- Study participants included in the biomarker analysis are representative of those participating in the definitive trial.
- The biomarker assay has previously undergone rigorous preanalytical and analytical validation for use in archival tissue.
- The protocol for biomarker evaluation (including sample size and statistical testing) has been finalized before any measurements on archival tissue are made.
- Results obtained with the archival samples are validated with samples from at least one related trial.

Although prospective or retrospective validation studies using samples from randomized trials may be the preferred method for demonstrating clinical utility, these may not always be necessary to advance a new biomarker into clinical use. Thus, according to Lord et al. (50), a randomized trial may not be required if accurate data show that a new biomarker is more clinically specific and/or safer than a previously validated test addressing the same clinical question. However, if the new biomarker has increased clinical sensitivity compared to the existing one, its measurement will result in the diagnosis of an increased number of diseased cases. In this situation, data from previously conducted trials that analyzed only cases detected with the existing biomarker may not be applicable to the additional cases. If so, it may be

necessary to perform a randomized trial to detect possible therapy efficacy in the cases detected with the new biomarker. A randomized trial, however, may not be required if the new biomarker identifies the same disease subset as the existing test or if the response to therapy is similar, irrespective of the disease subtype (50).

Another approach for obtaining LOE I evidence for new biomarkers is to undertake a systematic review of the literature followed by metaanalyses or pooled analyses. Ideally, systematic reviews should include results from both published and unpublished data, individual patient data (when it can be acquired), and studies that used validated assays. Inclusion of data from unpublished studies is particularly desirable because this can eliminate possible publication bias, given that positive studies are more likely to be published than negative studies. According to Egger et al. (51), a meta-/pooled analysis to establish clinical utility should be undertaken only in conjunction with a systematic review. A helpful checklist for evaluating the quality of the individual studies [known as QUADAS (Quality Assessment Studies of Diagnostic Accuracy)] has been published and should be used when performing systematic reviews of test accuracy studies (52).

A good example of a pooled analysis that included both published and unpublished data and results from analytically validated assays was the demonstration of the prognostic value of both urokinase plasminogen activator (uPA) and plasminogen activator inhibitor-1 (PAI-1) in breast cancer (53). The pooled analysis used individual patient data from 11 published and 7 unpublished studies and demonstrated that uPA and PAI-1 were prognostic in both lymph node–negative and lymph node–positive breast cancer patients. The prognostic impact of uPA and PAI-1 in the axillary node–negative patients was subsequently confirmed in a prospective randomized trial (54, 55).

Although evidence of clinical utility should be mandatory before introduction of a biomarker into routine clinical use, such evidence will not necessarily result in its widespread adoption, particularly if measurement of the biomarker is costly, requires nonroutine sample collection, and/or is technically difficult. For example, although uPA and PAI-1 are extensively validated, they are not widely used in clinical practice because fresh or freshly frozen tumor tissue is required. Assays providing similar prognostic value (e.g., Oncotype DX) are performed on the much easier to use paraffin-embedded and formalin-fixed tissue (56, 57) and have been more widely adopted.

Regulatory Approval

Following analytical validation, clinical validation, and demonstration of clinical value, regulatory approval is

required. The process of regulatory approval for new biomarkers varies in different countries. In the US, new biomarkers enter the marketplace for clinical use by either of 2 pathways (58). One of these involves clearance or approval by the FDA and performance of the assay in a CLIA-certified laboratory, while the other is known as the laboratory-developed tests (LDT) pathway.

Biomarker tests are classified as medical devices by the FDA and therefore are subject to the same regulatory procedures as other medical devices. These are classified into 3 groups, according to the intended use and risk to patients (58). Class I devices are of low risk and are usually exempt from review by the FDA before proceeding to market, although the manufacturing company must register the test with the FDA. Class II devices are of moderate risk and are evaluated by the FDA through review of a 510(k) premarket notification. These devices are cleared if submitted evidence confirms that they are equivalent to a legally marketed device previously cleared by the FDA. Class III devices are those that may cause significant risk to patients when used. Approval requires submission of evidence that the device has undergone analytical and clinical validation and that it is safe and effective for patient care. Additional evidence of clinical utility is not currently required, except for companion biomarkers. For these, the FDA now requires their coapproval with the relevant treatment (31).

Although FDA approval/clearance ensures the availability of safe and reliable products, the process of obtaining such approval/clearance can be time-consuming and expensive. This may be a particular problem for academic institutions and small companies. Indeed, obtaining regulatory approval has been regarded as a further hurdle in making new biomarkers available for clinical use.

The second pathway for entry to the US market, LDT, is regulated by the Centers for Medicare and Medicaid Services, under the umbrella of the CLIA 1988 act. LTDs (“home-brew” or “in-house” tests) are usually performed only in the laboratories in which they were developed and validated. Currently, performance of these tests does not require approval by the FDA. However, as with the FDA approval system, LTDs are also performed in CLIA-certified laboratories. Although evidence of analytical and clinical validation is mandatory for performing these tests, demonstration of clinical utility is not. Thus, at present LTDs may be used by doctors for clinical decision-making without evidence of clinical value.

Recently, the FDA announced its intention to regulate LTDs as devices to ensure the reliability and safety of results produced (59). The first LTDs that will have to comply with these new regulations are likely to be the high-risk tests such as companion biomarkers.

These changes were in accord with the previously published recommendations of Hayes et al. (4):

- The FDA should consider reviewing all oncology products, including therapeutics and biomarkers, in one unit.
- Approval for biomarkers should include evidence of both clinical validation and clinical value.
- Approval should require demonstration of clinical value in an LOE I evidence study such as a prospective clinical trial or a prospective–retrospective trial.
- The FDA should reconsider its enforcement discretion for LDT involving tumor biomarker assays and ensure that they are subject to FDA regulatory controls.
- The FDA should recommend that new anticancer drug trials be accompanied with a bank of samples that are collected and stored during the trial. This should be paid for by the relevant sponsoring pharmaceutical company.

In Europe, biomarkers for clinical use require the Conformité Européenne (CE) mark, which represents the manufacturer’s declaration that the product meets the requirements of the applicable European Community directives. CE marking can be obtained either by self-certification by the manufacturer or by submission to the appropriate body in a European Union (EU) country.

Although the EU currently classifies companion biomarkers as *in vitro* devices with low risk, legislation on their approval is currently undergoing review (60). In the future, they are likely to be classified as high individual risk or moderate public health risk. Their use will then necessitate conformity assessment by a notified body (60) and evidence of clinical value will be required.

Postmarketing Evaluation

As previously highlighted by the Test Evaluation Working Group of the European Federation of Clinical Chemistry and Laboratory Medicine, evaluation of a new biomarker should not end with its adoption in clinical practice (18). Thus, it is essential that the analytical performance of a biomarker in a routine laboratory is demonstrated to be as good or better than that observed during its development. This is most effectively achieved through implementation of rigorous internal QC procedures and participation in well-designed external quality assessment or proficiency testing programs, as required of clinical laboratories accredited to national or international standards. An example for which rigorous postmarketing evaluation resulted in improved assay testing was with HER2 for predicting response to trastuzumab (13).

Similarly, the impact of the biomarker test beyond the research and clinical trial settings should be subject to continuing clinical audit (61). For example, with new

Table 4. Guidelines for reporting biomarker studies.

Application	Aims	Guideline	Reference
Biospecimen handling and processing	How to collect, process and store human tissue in a standardized manner	BRISQ ^a	Moore et al. (67)
Diagnostic accuracy	To improve the accuracy and completeness of reporting of studies of diagnostic accuracy and to allow assessment of the potential for bias as well as evaluate its generalizability	STARD	Bossuyt et al. (68, 69)
Prognostic biomarkers ^b	Recommendations for reporting biomarker prognostic studies	REMARK	McShane et al. (70); Altman et al. (71)
Monitoring biomarkers	Recommendations for performing biomarker monitoring studies	MONITOR	Sölétormos et al. (72)
Biomarkers in clinical trials	Describes a risk-management approach for use of biomarkers in clinical trials		Hall et al. (73)
Omics in clinical trials	To establish the readiness of omics-based assays for use in clinical trials		McShane et al. (74); McShane et al. (75)
Immunohistochemistry and in situ hybridization studies	Reporting immunohistochemistry and in situ hybridization	MISFISHIE	Deutsch et al. (76)
Preparing systematic reviews	Evaluating quality of individual studies	QUADAS	Whiting et al. (52); Moher et al. (77)

^a BRISQ, Biospecimen Reporting for Improved Study Quality; STARD, STAndards for the Reporting of Diagnostic accuracy studies; REMARK, REporting recommendations for tumour MARKer prognostic studies; MISFISHIE, Minimum Information Specification For In Situ Hybridization and Immunohistochemistry Experiments; QUADAS, Quality Assessment of Diagnostic Accuracy Studies.

^b Although primarily designed for prognostic biomarkers, these guidelines may also be used for predictive biomarkers.

developments in medical technologies, it may be appropriate to ask the question whether the biomarker still retains its clinical value or if it has been superseded by a new and improved diagnostic procedure. A further issue that requires monitoring relates to whether the biomarker continues to be used in the setting in which it was originally validated. Thus, a biomarker approved for patient monitoring should not be used for screening without appropriate validation for that purpose, as happened with prostate-specific antigen (PSA). Finally, if the new biomarker replaces an older one, a question should be asked if the older test should be withdrawn. An example of a situation where this occurred was the removal of prostatic acid phosphatase as a biomarker following the introduction of PSA.

Registering and Reporting Biomarker Studies

It is widely accepted that the design and reporting of many biomarker studies has been poor, frequently including lack of detail regarding selection of patients and controls and/or sample handling, use of assays that perform poorly, inadequate sample numbers, and/or inappropriate statistical analyses (62). Publication bias resulting from the preferential reporting of positive vs negative results is also problematic. For example, Kyzas et al. (63) found that positive significant results were included in almost all of the published reports on prognostic bio-

markers they reviewed, with one examined database reporting positive findings in 91% of 340 articles and fully negative findings in only 1.5%.

In an attempt to address such shortcomings and improve the quality of biomarker studies, a biomarker registry has recently been introduced with the aim of providing a comprehensive record of completed and ongoing studies relating to cancer biomarkers and also enabling investigators to identify completed but still unpublished studies (64). Importantly, the database also allows identification of biomarker studies with negative results which may not be published (65). This should help to reduce bias in the preparation of meta-/pooled-analysis studies and systematic reviews. It is strongly recommended that investigators involved in LOE 1 evidence biomarkers trials register their study in the above or a related registry.

The quality and transparency of biomarker studies would also be enhanced by improved reporting. Editors of a number of medical journals have appealed for complete descriptions of laboratory methods and sample handling in reports of clinical biomarker studies (66). This information should include the type of specimens analyzed and how they were collected and stored, the analytical instrument and method used, analytical performance during the study (e.g., imprecision, reportable range) and derivation of any reference intervals used. All relevant raw data should also be made publicly available to allow

independent investigators to reanalyze and reinterpret the data (17).

Several other complementary guidelines and checklists have been published in recent years (Table 4). Following these guidelines is strongly encouraged when reporting all levels of biomarker studies, i.e., from preliminary studies to randomized trials. It is highly desirable that authors, editors, and reviewers implement these guidelines urgently.

Conclusion

Although several reports have been published on biomarker validation (17, 18, 27, 34, 43, 44, 78), this report is one of the first to consider all the key steps in the process, from analytical and clinical validation, through demonstration of clinical utility and regulatory approval to reporting (publishing) and registration and finally adoption in clinical practice. The proposals made apply to use of biomarkers at all stages of the patient pathway and are relevant to both serum and tissue biomarkers. From the above, it is clear that the process of developing a clinically useful biomarker is a long and expensive undertaking which requires the multidisciplinary collaboration of academic researchers, hospital clinicians, clinical laboratorians, biostatisticians, regulators, and clinical and scientific staff in diagnostic and pharmaceutical companies. There has previously been considerable expenditure of time, money, and clinical material when attempting to develop new biomarkers that ultimately have had no clinical impact. In the future, applying the methodology outlined in this article should enable much more efficient and effective development and reporting of cancer biomarker studies. With such rigorous applica-

tion, all stakeholders, and especially patients, would be expected to benefit.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors' Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

Employment or Leadership: C.M. Sturgeon, involvement in the Leeds NIHR and DEC projects (UK) which relate to this topic; E.P. Diamandis, *Clinical Chemistry*, AACCC.

Consultant or Advisory Role: D.F. Hayes, Pfizer, OncoImmune LLC, InBiomotion, and Biomarker Strategies.

Stock Ownership: D.F. Hayes, OncoImmune and InBiomotion.

Honoraria: D.F. Hayes, Eli Lilly.

Research Funding: M.J. Duffy, Science Foundation Ireland (Strategic Research Cluster Award, 08/SRC/B1410 to Molecular Therapeutics for Cancer Ireland), the Cancer Clinical Research Trust, and the Irish Cancer Society (BREAST-PREDICT, CCRC13GAL); D.F. Hayes, Novartis, Veridex (Johnson & Johnson), Janssen Diagnostics, and the Breast Cancer Research Foundation and institutional funding from the Fashion Footwear Association of New York/Shoes on Sale and AstraZeneca.

Expert Testimony: None declared.

Patents: D.F. Hayes, Patent numbers 05725638.0-1223-US2005008602, US 8,790,878 B2, US 8,951,484 B2, US provisional patent application original application no. 61/079,642, and US provisional patent application no. 61/593,092.

Other Remuneration: D.F. Hayes, License of patented technology to Janssen Diagnostics by my institution naming me as inventor.

Acknowledgments: M.J. Duffy, C.M. Sturgeon, G. Sölétormos, V. Barak, and R. Molina are members of the EGTM (European Group on Tumor Markers). D.F. Hayes, E.P. Diamandis, and P.M.M. Bossuyt are guest authors.

References

- Duffy MJ. Tumor markers in clinical practice: a review focusing on common solid cancers. *Med Princ Pract* 2013;22:4-11.
- Paoletti C, Hayes DF. Molecular testing in breast cancer. *Annu Rev Med* 2014;65:95-110.
- Diamandis EP. Cancer biomarkers: can we turn recent failures into success? *J Natl Cancer Inst* 2010;102:1462-7.
- Hayes DF, Allen J, Compton C, Gustavsen G, Leonard DG, McCormack R, et al. Breaking a vicious cycle *Sci Transl Med* 2013;5:196cm6.
- Sturgeon CM, Hoffman BR, Chan DW, Ch'ng SL, Hammond E, Hayes DF, et al. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines for use of tumor markers in clinical practice: quality requirements. *Clin Chem* 2008;54:e1-10.
- Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. *Clin Chem* 2015;61:112-23.
- Smerage JB, Barlow WE, Hortobagyi GN, Winer EP, Leyland-Jones B, Sirkalovic G, et al. Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. *J Clin Oncol* 2014;32:3483-9.
- Engel KB, Moore HM. Effects of preanalytical variables on the detection of proteins by immunohistochemistry in formalin-fixed, paraffin-embedded tissue. *Arch Pathol Lab Med* 2011;135:537-43.
- Peña-Llopis S, Brugarolas J. Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications. *Nat Protoc* 2013;8:2240-55.
- McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* 2015;27:15-26.
- Hergenroth M, Kenzelmann M, Gröne HJ. Laser-controlled microdissection of tissues opens a window of new opportunities. *Pathol Res Pract* 2003;199:419-23.
- Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol* 2010;28:2784-95. Erratum in: *J Clin Oncol* 2010;28:3543.
- 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.
- Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *J Mol Diagn* 2013;15:415-53.
- García-Alfonso P, García-Foncillas J, Salazar R, Pérez-Segura P, García-Carbonero R, Musulén-Palet E, et al. Updated guidelines for biomarker testing in colorectal carcinoma: a national consensus of the Spanish Society of Pathology and the Spanish Society of Medical Oncology. *Clin Transl Oncol* 2015;17:264-73.
- Cree IA, Deans Z, Ligtenberg MJ, Normanno N, Edsjö A, Rouleau E, et al. Guidance for laboratories performing molecular pathology for cancer patients. *J Clin Pathol*

- 2014;67:923–31.
17. Micheel C, Nass SJ, Omenn GS, Institute of Medicine. Evolution of translational omics: lessons learned and the path forward. Micheel C, Nass SJ, Omenn GS, eds. Washington (DC): National Academies Press; 2012. Available from http://www.nap.edu/catalog.php?record_id=13297.
 18. Horvath AR, Lord SJ, St John A, Sandberg S, Cobbaert CM, Lorenz S, et al. From biomarkers to medical tests: the changing landscape of test evaluation. *Clin Chim Acta* 2014;427:49–57.
 19. Biesecker LG, Green RC. Diagnostic and clinical genome and exome sequencing. *N Engl J Med* 2014; 370:2418–25.
 20. Markopoulos C. Overview of the use of Oncotype DX® as an additional treatment decision tool in early breast cancer. *Expert Rev Anticancer Ther* 2013;13:179–94.
 21. Slodkowska EA, Ross JS. MammaPrint 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Expert Rev Mol Diagn* 2009;9:417–22.
 22. Khleif SN, Doroshow JH, Hait WN; AACR-FDA-NCI Cancer Biomarkers Collaborative. AACR-FDA-NCI Cancer Biomarkers Collaborative consensus report: advancing the use of biomarkers in cancer drug development. *Clin Cancer Res* 2010;452:16:3299–318.
 23. Jennings L, Van Deerlin VM, Gulley ML; College of American Pathologists Molecular Pathology Resource Committee. Recommended principles and practices for validating clinical molecular pathology tests. *Arch Pathol Lab Med* 2009;133:743–55.
 24. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. *Genet Med* 2009;11:3–14.
 25. Williams PM, Lively TG, Jessup JM, Conley BA. Bridging the gap: moving predictive and prognostic assays from research to clinical use. *Clin Cancer Res* 2012;18: 1531–9.
 26. Fraser CG, Kallner A, Kenny D, Petersen PH. Introduction: strategies to set global quality specifications in laboratory medicine. *Scand J Clin Lab Invest* 1999;59:477–8.
 27. Sturgeon C, Hill R, Hortin GL, Thompson D. Taking a new biomarker into routine use—a perspective from the routine clinical biochemistry laboratory. *Proteomics Clin Appl* 2010;4:892–903.
 28. Prassas I, Diamandis EP. Translational researchers beware! Unreliable commercial immunoassays (ELISAs) can jeopardize your research. *Clin Chem Lab Med* 2014;52:765–6.
 29. Prassas I, Brinc D, Farkona S, Leung F, Dimitromanolakis A, Chrysoja CC, et al. False biomarker discovery due to reactivity of a commercial ELISA for CUZD1 with cancer antigen CA125. *Clin Chem* 2014;60:381–8.
 30. Taube SE, Clark GM, Dancy JE, McShane LM, Sigman CC, Gutman SI. A perspective on challenges and issues in biomarker development and drug and biomarker codevelopment. *J Natl Cancer Inst* 2009;101:1453–63.
 31. Mansfield EA. FDA perspective on companion diagnostics: an evolving paradigm. *Clin Cancer Res* 2014;20:1453–7.
 32. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. *Nat Rev Cancer* 2005;5:1: 42–9.
 33. Ransohoff DF, Gourlay ML. Sources of bias in specimens for research about molecular markers for cancer. *J Clin Oncol* 2010;28:698–704.
 34. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 2001; 93:1054–61.
 35. Cramer DW, Bast RC Jr, Berg CD, Diamandis EP, Godwin AK, Hartge P, et al. Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer Prev Res (Phila)* 2011;4:365–74.
 36. Zhu CS, Pinsky PF, Cramer DW, Ransohoff DF, Hartge P, Pfeiffer RM, et al. A framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer. *Cancer Prev Res (Phila)* 2011;4: 375–83.
 37. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 2003; 95:14–8.
 38. Simon R. Development and validation of therapeutically relevant multi-gene biomarker classifiers. *J Natl Cancer Inst* 2005;97:866–7.
 39. Tinker AV, Boussiotas A, Bowtell DD. The challenges of gene expression microarrays for the study of human cancer. *Cancer Cell* 2006;9:333–9.
 40. Berry D. Multiplicities in cancer research: ubiquitous and necessary evils. *J Natl Cancer Inst* 2012;104: 1124–32.
 41. Febbo PG, Ladanyi M, Aldape KD, De Marzo AM, Hammond ME, Hayes DF, et al. NCCN Task Force report: evaluating the clinical utility of tumor markers in oncology. *J Natl Compr Canc Netw* 2011;9(Suppl 5):S1–32.
 42. Parkinson DR, McCormack RT, Keating SM, Gutman SI, Hamilton SR, Mansfield EA, et al. Evidence of clinical utility: an unmet need in molecular diagnostics for patients with cancer. *Clin Cancer Res* 2014;20:1428–44.
 43. Hayes DF, Bast RC, Desch CE, Fritsche H Jr, Kemeny NE, Jessup JM, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst* 1996;88:1456–66.
 44. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol* 2005;23:2020–7.
 45. Kaplan R, Maughan T, Crook A, Fisher D, Wilson R, Brown L, Parmar M. Evaluating many treatments and biomarkers in oncology: a new design. *J Clin Oncol* 2013;31:4562–8.
 46. Tajik P, Zwinderman AH, Mol BW, Bossuyt PM. Trial designs for personalizing cancer care: a systematic review and classification. *Clin Cancer Res* 2013;19:4578–88.
 47. Sargent DJ, Korn EL. Decade in review-clinical trials: shifting paradigms in cancer clinical trial design. *Nat Rev Clin Oncol* 2014;11:625–6.
 48. Duffy MJ, Napieralski R, Martens JW, Span PN, Spyrtos F, Sweep FC, et al. Methylated genes as new cancer biomarkers. *Eur J Cancer* 2009;45:335–46.
 49. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 2009;101:1446–52.
 50. Lord SJ, Irwig L, Bossuyt PM. Using the principles of randomized controlled trial design to guide test evaluation. *Med Decis Making* 2009;29:E1–12.
 51. Egger M, Smith GD, Sterne JA. Uses and abuses of meta-analysis *Clin Med* 2001;1:478–84.
 52. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
 53. Look MP, van Putten WLJ, Duffy MJ, Harbeck N, Christensen IJ, Thomssen C, et al. Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients. *J Natl Cancer Inst* 2002;94:116–28.
 54. Janicke F, Precht A, Thomssen C, Harbeck N, Meisner C, Untch M, et al. Randomized adjuvant chemotherapy trial in high-risk node-negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1. *J Natl Cancer Inst* 2001;93:913–92.
 55. Harbeck N, Schmitt M, Meisner C, Friedel C, Untch M, Schmidt M, et al. Ten-year analysis of the prospective multicentre Chemo-NO trial validates American Society of Clinical Oncology (ASCO)-recommended biomarkers uPA and PAI-1 for therapy decision making in node-negative breast cancer patients. *Eur J Cancer* 2013;49: 1825–35.
 56. Carlson JJ, Roth JA. The impact of the Oncotype Dx breast cancer assay in clinical practice: a systematic review and meta-analysis. *Breast Cancer Res Treat* 2013; 141:13–22.
 57. Drukker CA, Bueno-de-Mesquita JM, Retèl VP, van Harten WH, van Tinteren H, Wesseling J, et al. A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. *Int J Cancer* 2013;133:929–36.
 58. Füzéry AK, Levin J, Chan MM, Chan DW. Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. *Clin Proteomics* 2013;10:13.
 59. US Food and Drug Administration. Draft guidance for industry, Food and Drug Administration staff, and clinical laboratories: framework for regulatory oversight of laboratory developed tests (LDTs). <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM416685.pdf> (Accessed November 2014).
 60. Pignatti F, Ehmam F, Hemmings R, Jonsson B, Nuebling M, Papaluca-Amati M, et al. Cancer drug development and the evolving regulatory framework for companion diagnostics in the European Union. *Clin Cancer Res* 2014;20:1458–68.
 61. Sturgeon CM, Duffy MJ, Walker G. The National Institute for Health and Clinical Excellence (NICE) guidelines for early detection of ovarian cancer: the pivotal role of the clinical laboratory. *Ann Clin Biochem* 2011; 48:295–9.
 62. Duffy MJ, Crown J. Precision treatment for cancer: role of prognostic and predictive markers. *Crit Rev Clin Lab Sci* 2014;51:30–45.
 63. Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Almost all articles on cancer prognostic markers report statistically significant results. *Eur J Cancer* 2007;43:2559–79.
 64. Andre F, McShane LM, Michiels S, Ransohoff DF, Altman DG, Reis-Filho JS, et al. Biomarker studies: a call for a comprehensive biomarker study registry. *Nat Rev Clin Oncol* 2011;8:171–6.
 65. Rifai N, Bossuyt PM, Ioannidis JP, Bray KR, McShane LM, Golub RM, Hooft L. Registering diagnostic and prognostic trials of tests: is it the right thing to do? *Clin Chem* 2014;60:1146–52.
 66. Rifai N, Annesley TM, Berg JP, Brugnara C, Delvin E, Lamb EJ, Ness PM, et al. An appeal to medical journal editors: the need for a full description of laboratory methods and specimen handling in clinical study reports. *Clin Chem* 2012;58:483–5.
 67. Moore HM, Kelly A, McShane LM, Vaught J. Biospecimen reporting for improved study quality (BRISQ). *Clin Chim Acta* 2012;413:1305.
 68. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD

- initiative. Standards for Reporting of Diagnostic Accuracy. *Clin Chem* 2003;49:1-6.
69. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Inwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003;49:7-18.
70. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM; Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 2005;97:1180-4.
71. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Med* 2012;10:51.
72. Sölétormos G, Duffy MJ, Hayes DF, Sturgeon CM, Barak V, Bossuyt PM, et al. Design of tumor biomarker-monitoring trials: a proposal by the European Group on Tumor Markers. *Clin Chem* 2013;59:52-9.
73. Hall JA, Salgado R, Lively T, Sweep F, Schuh A. A risk-management approach for effective integration of biomarkers in clinical trials: perspectives of an NCI, NCRI, and EORTC working group. *Lancet Oncol* 2014;15:e184-93.
74. McShane LM, Cavenagh MM, Lively TG, Eberhard DA, Bigbee WL, Williams PM, et al. Criteria for the use of omics-based predictors in clinical trials. *Nature* 2013;502:317-20.
75. McShane LM, Cavenagh MM, Lively TG, Eberhard DA, Bigbee WL, Williams PM, et al. Criteria for the use of omics-based predictors in clinical trials: explanation and elaboration. *BMC Med*. 2013 Oct 17;11:220.
76. Deutsch EW, Ball CA, Berman JJ, Bova GS, Brazma A, Bumgarner RE, et al. Minimum information specification for in situ hybridization and immunohistochemistry experiments (MISFISHIE). *Nat Biotechnol* 2008;26:305-12.
77. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264-9.
78. McShane LM, Hayes DF. Publication of tumor marker research results: the necessity for complete and transparent reporting. *J Clin Oncol* 2012;30:4223-32.