

available at www.sciencedirect.com
www.elsevier.com/locate/molonc

Review

Biomarker validation and testing



Daniel F. Hayes*

University of Michigan Comprehensive Cancer Center, 6312 CCC, 1500 E. Medical Center Drive, Ann Arbor, MI 48109-5942, USA

ARTICLE INFO

Article history:

Received 9 October 2014

Accepted 9 October 2014

Available online 18 October 2014

Keywords:

Tumor biomarkers

Validation

Level of evidence

ABSTRACT

A tumor biomarker is a molecular or process-based change that reflects the status of an underlying malignancy. A tumor biomarker may be identified and measured by one or more assays, or tests, for the biomarker. Increasingly, tumor biomarker tests are being used to drive patient management, either by identifying patients who do not require any, or any further, treatment, or by identifying patients whose tumors are so unlikely to respond to a given type of treatment that it will cause more harm than good. A tumor biomarker assay should only be used to guide management if it has analytical validity, meaning that it is accurate, reproducible, and reliable, and if it has been shown to have clinical utility. The latter implies that high levels of evidence are available that demonstrate that application of the tumor biomarker test for a given use context results in better outcomes, or similar outcomes with less cost, than if the assay were not applied. Use contexts include risk categorization, screening, differential diagnosis, prognosis, prediction of therapeutic activity or monitoring disease course. Very few tumor biomarker tests have passed these high bars for routine clinical application. However, if tumor biomarker tests are going to be used to drive patient care, than an understanding, and careful assessment, of these concepts are essential, since “A Bad Tumor Biomarker Test Is as Bad as a Bad Drug.”

© 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

The term “personalized medicine” has recently gained widespread acceptance among both the medical and lay communities. Fundamentally, “personalized medicine” implies getting the right therapy to the right patient at the right time, dose, and schedule. Of course since the beginning of medicine, physicians have tried to determine the correct diagnosis and match appropriate therapy to the patient at hand with the best evidence available (Schilsky, 2009). However,

over the last five to ten years, the tools to aid clinicians in their quest to personalize medicine have become increasingly sophisticated, and perhaps no more so than in the field of oncology. The revolution in molecular biology over the last three decades has provided a much better understanding of the aberrant pathways that drive the malignant process. The pharmaceutical industry has exploited this better understanding of tumor biology to develop therapeutic agents that are targeted to these aberrant pathways. Finally, immunologic and molecular genetic technologies that were unthinkable as

* Tel.: +1 734 615 6725; fax: +1 734 647 9271.

E-mail address: hayesdf@umich.edu

<http://dx.doi.org/10.1016/j.molonc.2014.10.004>

1574-7891/© 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

recently as a decade ago have permitted the generation of diagnostic approaches that illuminate the specific changes in cancer versus normal cells.

In spite of these advances, there seems to be more hype than reality. Very few molecular diagnostic tests have gained recommendation by major guidelines bodies, and only a few tumor biomarker tests have proven successful in the marketplace (Hayes et al., 2013). Further, some tumor biomarker assays are commercially available without documented evidence that they improve patient care, and yet are being ordered and used by many clinicians. What has led to this relative state of chaos? The remainder of this review will be dedicated to the theme that “A Bad Tumor Marker Test Is as Bad as a Bad Drug (Hayes et al., 2013),” detailing the current state of affairs and knowledge about what is needed to take a tumor biomarker test from a good idea to clinical reality.

2. What is a tumor biomarker test?

It is important to understand the distinction between a tumor biomarker and a test for it (Institute of Medicine, 2012). A tumor biomarker is an indication that a normal tissue is likely to or has become malignant, and/or it provides an indication of how a malignancy will behave, either naturally or in the context of therapy. A tumor biomarker might be a molecular change, such as in a nucleic acid, protein, or metabolite. It might also be a process change, such as an alteration in tissue appearance. Further, the presence of a benign process within malignant tissue might also be considered a tumor biomarker, such as neovascularization, that in itself is not malignant but may provide an indication of the expected biology of the cancer. Tumor biomarkers may be detected and/or monitored in tissue, blood, or relevant secretions, such as urine, stool, sputum, or breast nipple aspirates.

A tumor biomarker test is used to identify or measure the perturbations reflected by the tumor biomarker. There may be one or more assays or tests that provide some indication of the status of the tumor biomarker. These may measure the same thing, or they may measure very different perturbations in the biomarker. The *erbB2* gene, which encodes for the HER2 protein, provides a good example of this issue. There are at least 3 commercially available assays for *in situ* hybridization to determine amplification of the gene, several assays, mostly based on immunohistochemistry, that quantify relative expression of the HER2 protein in cancer tissue, and others that quantify relative expression of the HER2 message (Wolff et al., 2013a,b). Recently, mutations in *erbB2* that activate the protein without over-expression have been reported. Each of these may give related indication of HER2 activity, but they are all very different and may or may not provide useful similar clinical information.

3. How is a tumor biomarker test used in the clinic?

To develop and validate a tumor biomarker test, several critical issues must be addressed. First, and foremost, one must establish the intended use or context (Table 1). These include risk categorization, screening, diagnosis, prognosis, prediction of

Table 1 – Intended use contexts for tumor biomarker tests.

- Risk categorization
- Screening for new cancer
- Differential diagnosis
 - Cancer vs. benign
 - Epithelial vs. hematopoietic vs. mesenchymal
 - Organ of origin
- Prognosis
 - Early stage
 - Metastatic
- Prediction of therapy activity
 - Early stage
 - Metastatic
- Monitoring disease status
 - Early stage
 - Metastatic

therapeutic response, and monitoring (Henry and Hayes, 2006). A tumor biomarker test might be used to place an unaffected individual into one or more categories of risk, in which he/she might take preventive or screening strategies that would otherwise be unacceptable. Perhaps the best examples of this use is the presence or absence of a germline Y chromosome. Men do not generally undergo screening or prevention for breast cancer, while women do not need to be concerned about their risk of prostate cancer. A second use context is screening for the presence of a new cancer. Few if any tumor biomarker tests have been successfully developed for this role, although as an example, use of human papilloma virus assays have been incorporated into standard of care for screening for cervical cancers. Diagnosis, or more accurately differential diagnosis, is an important issue in pathology. Tumor biomarker tests, principally immunologically-based, are used on occasion to distinguish benign from malignant tissues, and more frequently to determine that an undifferentiated cancer is epithelial versus hematopoietic or mesenchymal.

The most commonly used tumor biomarker assays are used to predict the future behavior of an established cancer. The term “prognostic factor” refers to a tumor biomarker test that infers a high or low risk of a cancer-related event assuming the patient receives no more therapy than he/she has already received, if any. The most widely accepted prognostic factors in cancer are the size of the primary tumor, the presence or absence of regional lymph nodes or distant metastases. These have been codified into the now classic “TNM” staging system maintained by the Joint Commission on Cancer (AJCC, 2010).

In contrast, predictive factors, also designated response modifier elements, are used to estimate the relative likelihood that a cancer will respond to a class of, or even individual, therapeutic agents. Perhaps the oldest and most widely used example of a predictive tumor biomarker is the estrogen receptor (ER), which may be measured in many ways using different assays. Regardless, patients with ER negative breast cancers do not benefit from endocrine (anti-estrogen) therapy,

while nearly one-half of those with ER positive breast cancers do (Hammond et al., 2010a,b).

Finally, serial tumor biomarker tests may be used to monitor the course of therapy, or even follow-up, to determine if a patient should remain on a given management plan or, perhaps, should have his/her strategy altered due to apparent progression. For example, there are assays for several circulating proteins, such as carcinoembryonic antigen, CA19-9, CA125, prostate specific antigen, and MUC-1 protein, that are commonly used to monitor patients with colorectal, pancreatic, ovarian, prostate, and breast cancers, respectively.

4. When should a tumor marker test be used to guide clinical care?

Although not absolutely, in general application of a new therapeutic strategy, especially a new drug, requires the presence of high levels of evidence that the drug is safe and effective. Consensus definitions of these two terms are reasonably accepted in the field, although one might argue over the degree of toxicities that a patient will tolerate or the exact clinical endpoint that is considered “meaningful.” The regulatory framework in most developed countries for therapeutics is consistent and well-understood by all involved. For example, introduction of a new anti-cancer drug into the clinical arena usually requires evidence from prospective randomized clinical trials that at least event-free, if not overall, survival is improved with statistical significance at a reasonably low cost of toxicity.

In contrast, the regulatory environment for review and commercial use of tumor biomarker tests is much less clear (Hayes et al., 2013). Clearance or approval of a tumor biomarker test by the Food and Drug Administration (FDA) does not necessarily mean that it improves patient outcomes

Table 2 – Important definitions for tumor biomarker semantics.

Analytical validity

- Does the tumor biomarker test accurately and reliably measure the analyte of interest in the appropriate patient specimen?

Clinical validity

- Does the tumor biomarker test accurately and reliably identify a clinically or biologically defined disorder, or separate one population into two or more groups with distinct clinical or biological outcomes or differences?

Clinical utility

- Are there high levels of evidence that use of the tumor biomarker test to guide clinical decisions result in improved measurable clinical outcomes compared with those if the biomarker test results were not applied?

Modified from (Teutsch et al., 2009).

or should be used. Further, because of FDA enforcement discretion, laboratory developed tests (LDTs) can be generated and used to direct clinical care without FDA approval, as long as they are performed within a laboratory that follows good laboratory practices according to the Clinical Laboratory Improvement Amendments (CLIA) Act of 1988. Use of a tumor biomarker to guide treatment decisions within a clinical trial has always required application of an investigational device exemption (IDE) to the FDA. Recently, the FDA has announced that it will carefully review the enforcement discretion decision regarding use of an LDT to care for patients in routine care, but it is expected that discussion about this decision will evolve over the next several years.

Taken together, these circumstances have led to a “vicious cycle” in which tumor marker tests are generally felt to have less value than therapeutics for cancer management (Figure 1) (Hayes et al., 2013). As a result, clinical decisions to use a tumor biomarker test, recommendation by guidelines

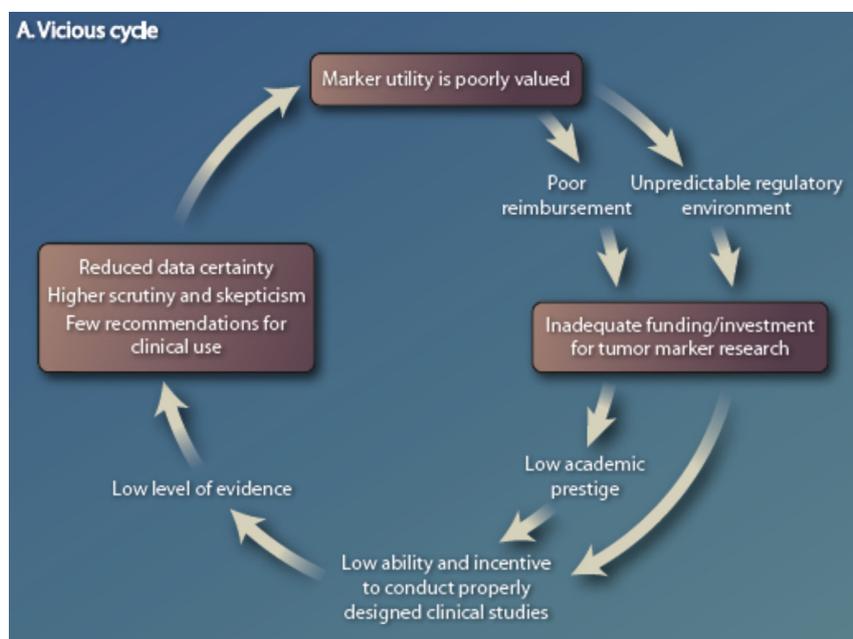


Figure 1 – The vicious cycle of tumor biomarker research. From (Hayes et al., 2013) with permission.

Table 3 – Elements of tumor marker studies that constitute levels of evidence determination.

Category	A	B	C	D
Trial design	Prospective	Prospective using archived samples	Prospective/observational	Retrospective/observational
Clinical trial	PRCT designed to address tumor marker	Prospective trial not designed to address tumor marker, but design accommodates tumor marker utility. Accommodation of predictive marker requires PRCT	Prospective observational registry, treatment and follow up not dictated	No prospective aspect to study
Patients and patient data	Prospectively enrolled, treated, and followed in PRCT	Prospectively enrolled, treated, and followed in clinical trial and, especially if a predictive utility is considered, a PRCT addressing the treatment of interest	Prospectively enrolled in registry, but treatment and follow up standard of care	No prospective stipulation of treatment or follow up; patient data collected by retrospective chart review
Specimen collection, processing, and archival	Specimens collected, processed and assayed for specific marker in real time	Specimens collected, processed, and archived prospectively using generic SOPs. Assayed after trial completion	Specimens collected, processed, and archived prospectively using generic SOPs. Assayed after trial completion	Specimens collected, processed and archived with no prospective SOPs
Statistical Design and analysis	Study powered to address tumor marker question	Study powered to address therapeutic question; underpowered to address tumor marker question	Study not prospectively powered at all. Retrospective study design confounded by selection of specimens for study	Study not prospectively powered at all. Retrospective study design confounded by selection of specimens for study
Validation	Result unlikely to be play of chance Although preferred, validation not required	Result more likely to be play of chance than A, but less likely than C Requires one or more validation studies	Result very likely to be play of chance Requires subsequent validation studies	Result very likely to be play of chance Requires subsequent validation

From (Simon et al., 2009) with permission.

a PCRT = prospective randomized controlled trial; SOPs = standard operating practices.

bodies to assist in these decisions, and reimbursement decisions from third party payers for their use, have been relatively arbitrary, with insufficient data to support or refute the relative value of the tests. If indeed clinicians are to truly provide personalized oncology, this vicious cycle needs to be broken. Therefore, over the last two decades, several experts have attempted to develop structured recommendations for criteria that might be used in a manner analogous to those applied to decisions regarding new therapeutics.

Teutsch et al., representing the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative of the United States Centers for Disease Control, suggested three important semantic definitions for tumor biomarker research: Analytical validity, clinical validity, and clinical utility (Table 2) (Teutsch et al., 2009). Analytical validity implies that the test for the tumor biomarker is accurate and reliable in the type of specimen to which it will be applied. Clinical validity refers to evidence that the tumor biomarker test divides a single population into two or more distinct groups, based on biology or clinical outcomes, with statistical significance. In

contrast, clinical utility requires that high levels of evidence demonstrate that use of the tumor biomarker test improves clinical outcomes or that clinical outcomes are identical with less cost or toxicity.

Table 4 – Revised determination of levels of evidence using elements of tumor marker studies.

Level of evidence	Category from Table 1	Validation studies available
I	A	None required
I	B	One or more with consistent results
II	B	None or inconsistent results
II	C	2 or more with consistent results
III	C	None or 1 with consistent results or inconsistent results
IV–V	D	NA

From (Simon et al., 2009) with permission.

a NA = Not applicable, since LOE IV and V studies will never be satisfactory for determination of medical utility.

Efforts to organize tumor biomarker test results into these categories, and to grade the levels of evidence that would help determine if a marker has clinical utility, have been proposed (Hayes et al., 1996; Simon et al., 2009). The ideal level of evidence should come from a prospective trial in which the tumor biomarker clinical utility for a specific use is the main objective. Indeed, several trial designs have been proposed to accomplish such a task (Freidlin et al., 2010; Sargent et al., 2005).

Prospective trials are time consuming and costly. Therefore, there are precious few examples of prospective trials in which the primary objective is to determine clinical utility of

a specific tumor biomarker test. However, one obvious difference between therapeutics and diagnostics is that the latter can be assessed using archived specimens that have been collected and stored for future use. In this regard, Simon, Paik and Hayes suggested that high levels of evidence can be ascertained by performing “prospective retrospective” analyses of a tumor biomarker test using archived specimens (Simon et al., 2009). However, these criteria are quite rigorous, necessitating use of specimens collected from patients who participated in prospective trials that addressed the specific use intended for the tumor biomarker test. They ranked the

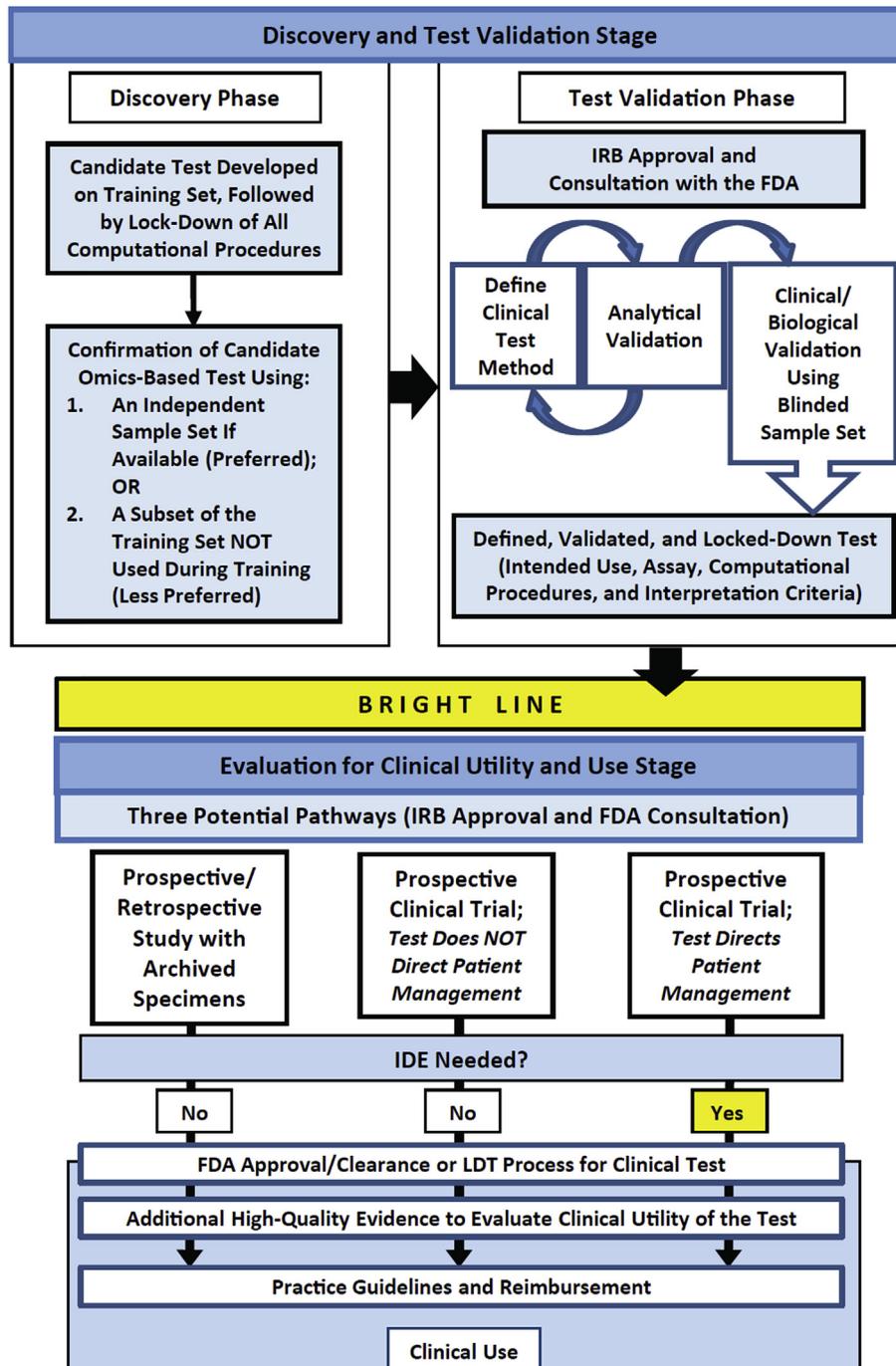


Figure 2 – A roadmap for tumor biomarker test development (adapted from Omics – based test development (Institute of Medicine, 2012)) with permission. LDT = laboratory developed test. See.

types of studies in a hierarchal fashion, to determine if a specific tumor biomarker test has clinical utility for a specific clinical use (Simon et al., 2009). Table 3 shows this hierarchy with the requirements that must be met to fit each category, while Table 4 provides the required elements to reach level 1 evidence for clinical utility.

In 2012, a committee convened by the Institute of Medicine (IOM) distilled these concepts into a roadmap to guide investigators from a good idea to generating clinical utility for a tumor biomarker (Figure 2) (Institute of Medicine, 2012). The process outlined in Figure 2 represents a process from new biomarker discovery to development of a “locked down” test for that biomarker that has high analytical validity and has some evidence of clinical validity, as defined by EGAPP. Preferably, but not absolutely, at least in the United States, this step should be performed within a CLIA-approved laboratory, which ensures that good laboratory practices are followed. When the investigator is satisfied that the tumor biomarker test has sufficiently high analytical and clinical validity, it should be taken across the “bright line” to determine if it has clinical utility for a given use context using one of the strategies discussed in the preceding paragraph. If a prospective clinical trial is pursued to test for clinical utility, the FDA should be consulted to determine if an IDE is, or is not, required (See IOM report (Institute of Medicine, 2012)).

5. How should the results of tumor biomarker test investigations be reported?

A critical component of determining the relative level of evidence to support analytical validity and clinical utility of a tumor biomarker test is the quality of the reporting of the studies that are used to evaluate the marker assay. Several efforts have been proposed to standardize reporting of tumor biomarker test studies, analogous to those required for therapeutic investigations (McShane and Hayes, 2012; Moher et al., 2001). These have included suggested descriptions of the pre-analytical factors that could substantially influence reproducibility of the assay; the so-called Biospecimen reporting for improved study quality (BRISQ) criteria (Moore et al., 2012). Furthermore, the Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK), which have been recently updated, are now required by a many oncology journals for manuscript submissions describing tumor biomarker results (Altman et al., 2012). In an effort to bring even more transparency to the field, a registry for prospective or prospective retrospective tumor biomarker studies has been established, so that investigators can document that their studies were truly prospectively planned (Andre et al., 2011). It is hoped that eventual participation in this or similar registries will decrease the now-rampant problem of publication bias for tumor biomarker studies, much in the way that the required registration in clinicaltrials.gov has done for therapeutic trials (McShane and Hayes, 2012).

6. Summary

As we enter the era of truly personalized medicine in oncology, it is critical that we continue to apply the scientific

method to the consideration of what tumor biomarker tests to use to guide patient management. Maintaining this level of rigor may become even more difficult, yet will remain even more important, as the fields of genomic-based therapies continue to evolve. Applying diagnostic or therapeutic strategies because they make sense, or because they are appealing, is a seductive but dangerous approach. It is essential that any tumor biomarker test used to guide treatment management have both analytical validity and clinical utility. As we move into this brave new world, we must keep in mind that “A Bad Tumor Biomarker Test Is as Bad as a Bad Drug.”

REFERENCES

- AJCC, 2010. Breast. In: Edge, S.B., Byrd, D.R., Compton, C., Fritz, A.G., Greene, F.L., Trotti, A. (Eds.), *AJCC Staging Manual, Seventh ed.* Springer, New York, Dordrecht, Heidelberg, London.
- Altman, D.G., McShane, L.M., Sauerbrei, W., Taube, S.E., 2012. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *Plos Med.* 9, e1001216.
- Andre, F., McShane, L.M., Michiels, S., Ransohoff, D.F., Altman, D.G., Reis-Filho, J.S., Hayes, D.F., Pusztai, L., 2011. Biomarker studies: a call for a comprehensive biomarker study registry. *Nat. Rev. Clin. Oncol.* 8, 171–176.
- Freidlin, B., McShane, L.M., Korn, E.L., 2010. Randomized clinical trials with biomarkers: design issues. *J. Natl. Cancer Inst.* 102, 152–160.
- Hammond, M.E., Hayes, D.F., Dowsett, M., Allred, D.C., Hagerty, K.L., Badve, S., Fitzgibbons, P.L., Francis, G., Goldstein, N.S., Hayes, M., Hicks, D.G., Lester, S., Love, R., Mangu, P.B., McShane, L., Miller, K., Osborne, C.K., Paik, S., Perlmutter, J., Rhodes, A., Sasano, H., Schwartz, J.N., Sweep, F.C., Taube, S., Torlakovic, E.E., Valenstein, P., Viale, G., Visscher, D., Wheeler, T., Williams, R.B., Wittliff, J.L., Wolff, A.C., 2010a. 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.* 28, 2784–2795.
- Hammond, M.E., Hayes, D.F., Dowsett, M., Allred, D.C., Hagerty, K.L., Badve, S., Fitzgibbons, P.L., Francis, G., Goldstein, N.S., Hayes, M., Hicks, D.G., Lester, S., Love, R., Mangu, P.B., McShane, L., Miller, K., Osborne, C.K., Paik, S., Perlmutter, J., Rhodes, A., Sasano, H., Schwartz, J.N., Sweep, F.C., Taube, S., Torlakovic, E.E., Valenstein, P., Viale, G., Visscher, D., Wheeler, T., Williams, R.B., Wittliff, J.L., Wolff, A.C., 2010b. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch. Pathol. Lab. Med.* 134, e48–e72.
- Hayes, D.F., Allen, J., Compton, C., Gustavsen, G., Leonard, D.G., McCormack, R., Newcomer, L., Pothier, K., Ransohoff, D., Schilsky, R.L., Sigal, E., Taube, S.E., Tunis, S.R., 2013. Breaking a vicious cycle. *Sci. Transl. Med.* 5, 196cm196.
- Hayes, D.F., Bast, R.C., Desch, C.E., Fritsche Jr., H., Kemeny, N.E., Jessup, J.M., Locker, G.Y., Macdonald, J.S., Mennel, R.G., Norton, L., Ravdin, P., Taube, S., Winn, R.J., 1996. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J. Natl. Cancer Inst.* 88, 1456–1466.

- Henry, N.L., Hayes, D.F., 2006. Uses and abuses of tumor markers in the diagnosis, monitoring, and treatment of primary and metastatic breast cancer. *Oncologist* 11, 541–552.
- Institute of Medicine, 2012. Evolution of Translational Omics: Lessons Learned and the Path Forward. The National Academies Press, Washington, D.C.
- McShane, L.M., Hayes, D.F., 2012. Publication of tumor marker research results: the necessity for complete and transparent reporting. *J. Clin. Oncol.* 30, 4223–4232.
- Moher, D., Schulz, K.F., Altman, D.G., 2001. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann. Intern. Med.* 134, 657–662.
- Moore, H.M., Kelly, A., McShane, L.M., Vaught, J., 2012. Biospecimen reporting for improved study quality (BRISQ). *Clin. Chim. Acta* 413, 1305.
- Sargent, D.J., Conley, B.A., Allegra, C., Collette, L., 2005. Clinical trial designs for predictive marker validation in cancer treatment trials. *J. Clin. Oncol.* 23, 2020–2027.
- Schilsky, R.L., 2009. Personalizing cancer care: American Society of Clinical Oncology presidential address 2009. *J. Clin. Oncol.* 27, 3725–3730.
- Simon, R.M., Paik, S., Hayes, D.F., 2009. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J. Natl. Cancer Inst.* 101, 1446–1452.
- Teutsch, S.M., Bradley, L.A., Palomaki, G.E., Haddow, J.E., Piper, M., Calonge, N., Dotson, W.D., Douglas, M.P., Berg, A.O., 2009. The evaluation of genomic applications in practice and prevention (EGAPP) initiative: methods of the EGAPP working group. *Genet. Med.* 11, 3–14.
- Wolff, A.C., Hammond, M.E., Hicks, D.G., Dowsett, M., McShane, L.M., Allison, K.H., Allred, D.C., Bartlett, J.M., Bilous, M., Fitzgibbons, P., Hanna, W., Jenkins, R.B., Mangu, P.B., Paik, S., Perez, E.A., Press, M.F., Spears, P.A., Vance, G.H., Viale, G., Hayes, D.F., 2013a. 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. *Arch. Pathol. Lab. Med.*
- Wolff, A.C., Hammond, M.E., Hicks, D.G., Dowsett, M., McShane, L.M., Allison, K.H., Allred, D.C., Bartlett, J.M., Bilous, M., Fitzgibbons, P., Hanna, W., Jenkins, R.B., Mangu, P.B., Paik, S., Perez, E.A., Press, M.F., Spears, P.A., Vance, G.H., Viale, G., Hayes, D.F., 2013b. 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, 3997–4013.