Reliable Evaluation of Prognostic & Predictive Genomic Tests

Richard Simon, D.Sc.
Chief, Biometric Research Branch
National Cancer Institute
http://brb.nci.nih.gov

Different Kinds of Biomarkers

- **Prognostic biomarkers**
  - **Measured before treatment to indicate long-term outcome for patients untreated or receiving standard treatment**
  - **May reflect both disease aggressiveness and effect of standard treatment**
- **Predictive biomarkers**
  - **Measured before treatment to identify who will benefit from a particular treatment**

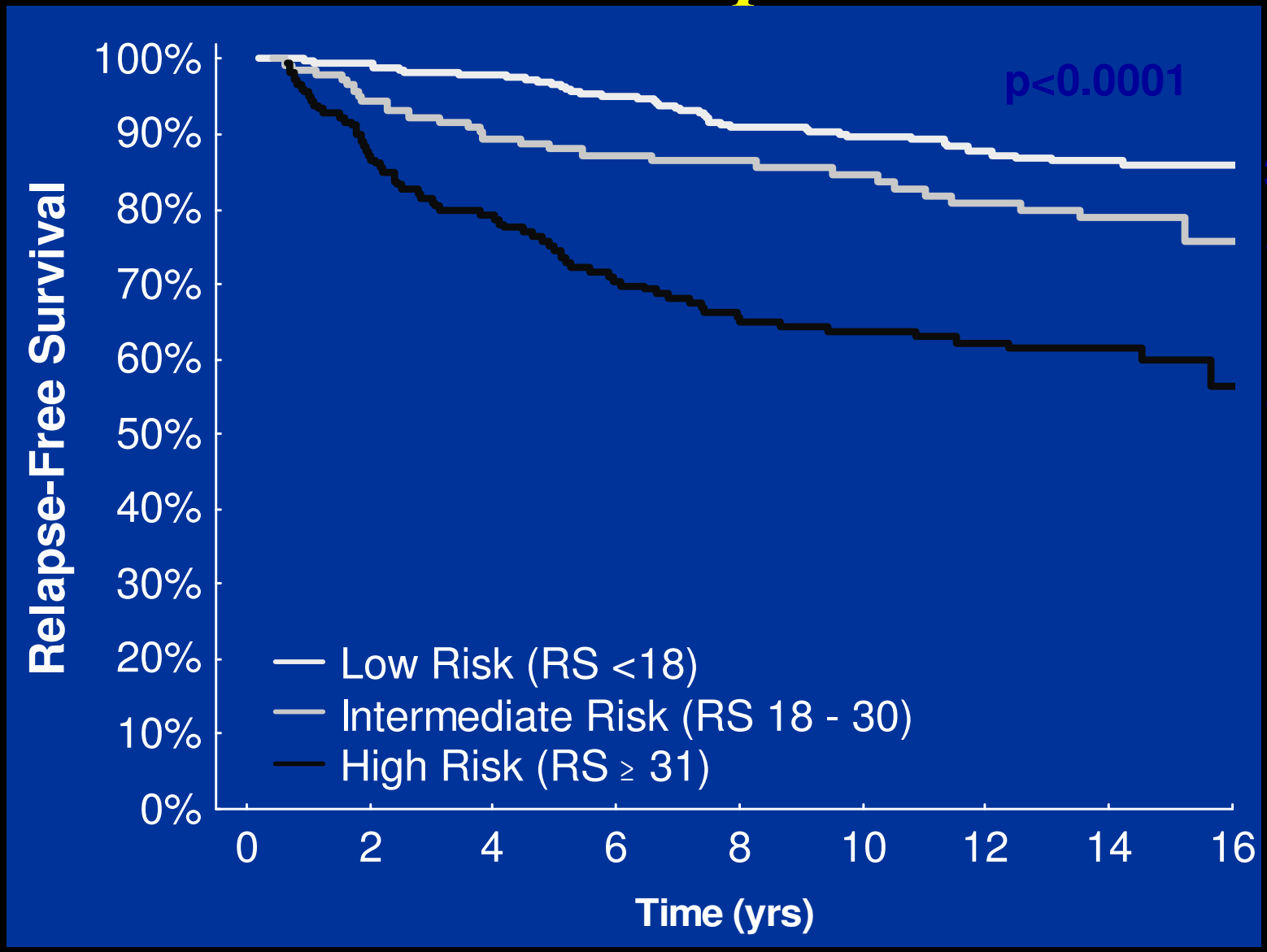
Prognostic & Predictive Biomarkers

- **Many cancer treatments benefit only a minority of patients to whom they are administered**
- **Being able to predict which patients are likely to benefit could**
  - **save patients from unnecessary toxicity, and enhance their chance of receiving a drug that helps them**
  - **Help control medical costs**
  - **Improve the success rate of clinical drug development**

Prognostic Biomarkers in Node Negative Breast Cancer

- **To identify patients who are likely to be cured by surgery/radiotherapy and hormonal therapy and therefore are unlikely to benefit from adjuvant chemotherapy**
  - **Oncotype Dx recurrence score based on expression of 21 genes measured by RT-PCR on FFPE diagnostic biopsy**

B-14 Results—Relapse-Free Survival



Paik et al, SABCS 2003

Key Features of OncotypeDx Development

- **Identification of important therapeutic decision context**
- **Prognostic marker development was based on patients with node negative ER positive breast cancer receiving tamoxifen as only systemic treatment**
  - **Use of patients in previously conducted NSABP clinical trials**
- **Staged development and validation**
  - **Separation of data used for test development from data used for test validation**
- **Development of robust assay with rigorous analytical validation**
  - **21 gene RTPCR assay for FFPE tissue**
  - **Quality assurance by single reference laboratory operation**

Prognostic Factors in Oncology

- **Most prognostic factors are not used because they are not therapeutically relevant**
- **Most prognostic factor studies are not conducted with a medical indication clearly in mind**
  - **They use a convenience sample of patients for whom tissue is available.**
  - **Generally the patients are too heterogeneous to support therapeutically relevant conclusions**

TAILORx Clinical Trial for Prosepctive Evaluation of Oncotype Dx

- **Prospectively register patients with breast cancer**
  - **Node negative, HR positive, HER2 negative, age < 75, standard eligibility for chemorx**
  - **All patients receive hormonal therapy**
  - **900 sites participating**
- **Perform Oncotype Dx assay**
- **If OncotypeDx RS < 11**
  - **Withhold chemotherapy**
  - **Sized to evaluate whether 10-year DFS is > 95% vs < 93.5%**
- **If RS 11-25**
  - **Randomize to +- chemotherapy**
  - **Sized to detect 3% reduction in 5-year DFS from baseline of 90% with chemo**
- **If RS > 25**
  - **Administer chemotherapy**

Predictive Biomarkers

- **Predictive markers to identify patients whose tumors are likely (or unlikely) to benefit from specific drugs.**
- **Particularly important for molecularly targeted drugs**
- **Usually single gene/protein**
  - **HER2 for anti-Her2 rx in breast cancer**
  - **KRAS for anti-EGFR antibodies in colorectal cancer**

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 OCTOBER 23, 2008 VOL. 359 NO. 17

K-ras Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer

Christos S. Karapetis, M.D., Shirin Khambata-Ford, Ph.D., Derek J. Jonker, M.D., Chris J. O'Callaghan, Ph.D., Dongsheng Tu, Ph.D., Niall C. Tebbutt, Ph.D., R. John Simes, M.D., Haji Chalchal, M.D., Jeremy D. Shapiro, M.D., Sonia Robitaille, M.Sc., Timothy J. Price, M.D., Lois Shepherd, M.D.C.M., Heather-Jane Au, M.D., Christiane Langer, M.D., Malcolm J. Moore, M.D., and John R. Zalcberg, M.D., Ph.D.*

ABSTRACT

BACKGROUND

Treatment with cetuximab, a monoclonal antibody directed against the epidermal growth factor receptor, improves overall and progression-free survival and preserves the quality of life in patients with colorectal cancer that has not responded to chemotherapy. The mutation status of the *K-ras* gene in the tumor may affect the response to cetuximab and have treatment-independent prognostic value.

METHODS

We analyzed tumor samples, obtained from 394 of 572 patients (68.9%) with colorectal cancer who were randomly assigned to receive cetuximab plus best supportive care or best supportive care alone, to look for activating mutations in exon 2 of the *K-ras* gene. We assessed whether the mutation status of the *K-ras* gene was associated with survival in the cetuximab and supportive-care groups.

RESULTS

Of the tumors evaluated for *K-ras* mutations, 42.3% had at least one mutation in exon 2 of the gene. The effectiveness of cetuximab was significantly associated with *K-ras* mutation status ($P=0.01$ and $P<0.001$ for the interaction of *K-ras* mutation status with overall survival and progression-free survival, respectively). In patients with wild-type *K-ras* tumors, treatment with cetuximab as compared with supportive care alone significantly improved overall survival (median, 9.5 vs. 4.8 months; hazard ratio for death, 0.55; 95% confidence interval [CI], 0.41 to 0.74; $P<0.001$) and progression-free survival (median, 3.7 months vs. 1.9 months; hazard ratio for progression or death, 0.40; 95% CI, 0.30 to 0.54; $P<0.001$). Among patients with mutated *K-ras* tumors, there was no significant difference between those who were treated with cetuximab and those who received supportive care alone with respect to overall survival (hazard ratio, 0.98; $P=0.89$) or progression-free survival (hazard ratio, 0.99; $P=0.96$). In the group of patients receiving best supportive care alone, the mutation status of the *K-ras* gene was not significantly associated with overall survival (hazard ratio for death, 1.01; $P=0.97$).

CONCLUSIONS

Patients with a colorectal tumor bearing mutated *K-ras* did not benefit from cetuximab, whereas patients with a tumor bearing wild-type *K-ras* did benefit from cetuximab. The mutation status of the *K-ras* gene had no influence on survival among patients treated with best supportive care alone. (ClinicalTrials.gov number, NCT00079066.)

From Flinders Medical Centre and Flinders University, Adelaide, Australia (C.S.K.); Bristol-Myers Squibb Research and Development, Princeton, NJ (S.K.-F.); Ottawa Hospital Research Institute, University of Ottawa, Ottawa (D.J.J.); National Cancer Institute of Canada Clinical Trials Group, Kingston, ON (C.J.O., D.T., S.R., L.S.); Austin Health, Melbourne, Australia (N.C.T.); National Health and Medical Research Council Clinical Trials Centre, University of Sydney, Sydney (R.J.S.); Allan Blair Cancer Centre, Regina, SK, Canada (H.C.); Cabrini Hospital and Alfred Hospital, Melbourne, Australia (J.D.S.); Queen Elizabeth Hospital and University of Adelaide, Adelaide, Australia (T.J.P.); Cross Cancer Institute, Edmonton, AB, Canada (H.-J.A.); Bristol-Myers Squibb, Wallingford, CT (C.L.); Princess Margaret Hospital, Toronto (M.J.M.); and Peter MacCallum Cancer Centre and University of Melbourne, Melbourne, Australia (J.R.Z.). Address reprint requests to Dr. Karapetis at the Department of Medical Oncology, Flinders Medical Centre, Flinders Dr., Bedford Park, SA 5042, Australia, or at c.karapetis@flinders.edu.au.

*Other participants in the CO.17 trial from the National Cancer Institute of Canada Clinical Trials Group and the Australasian Gastro-Intestinal Trials Group are listed in the Supplementary Appendix, available with the full text of this article at www.nejm.org.

N Engl J Med 2008;359:1757-65.

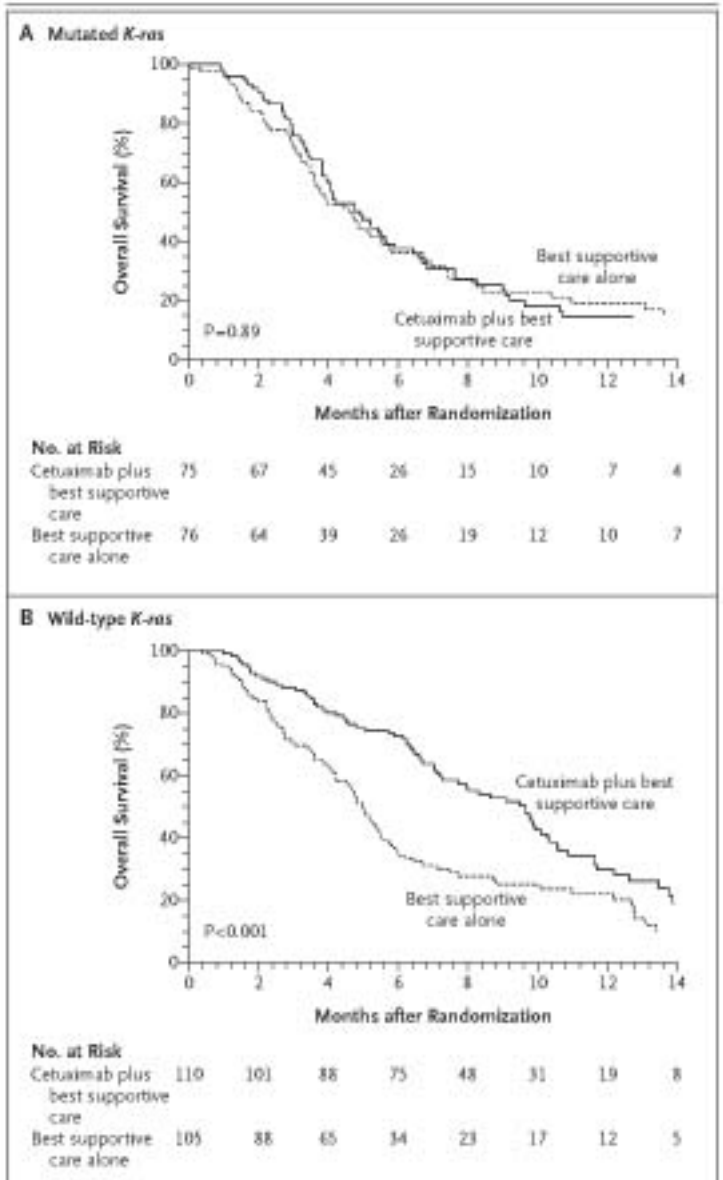


Figure 1. **Kaplan–Meier Curves for Overall Survival According to Treatment.**

Panel A shows results for patients with mutated *K-ras* tumors, and Panel B for patients with wild-type *K-ras* tumors. Cetuximab as compared with best supportive care alone was associated with improved overall survival among patients with wild-type *K-ras* tumors but not among those with mutated *K-ras* tumors. The difference in treatment effect according to mutation status was significant (test for interaction, P=0.01).

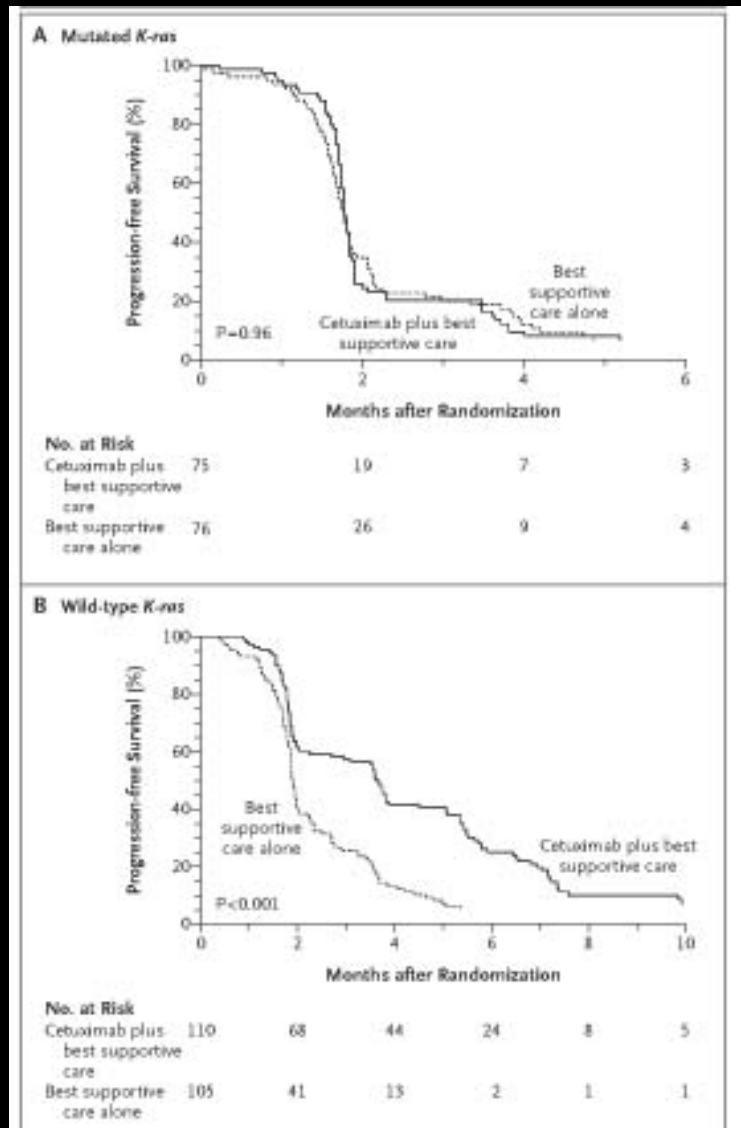


Figure 2. **Kaplan–Meier Curves for Progression-free Survival According to Treatment.**

Panel A shows results for patients with mutated *K-ras* tumors, and Panel B for patients with wild-type *K-ras* tumors. Cetuximab as compared with best supportive care alone was associated with improved progression-free survival among patients with wild-type *K-ras* tumors but not among those with mutated *K-ras* tumors. The difference in treatment effect according to mutation status was significant (test for interaction, P<0.001).

Home
Meetings
Abstracts & Virtual Meeting
Practice Resources
Education & Training
News
Press Center
ASCO News & Forum
• Feature Articles
Podcasts
Legislative & Regulatory
Quality Care & Guidelines

About ASCO
ASCO Bookstore
Careers in Oncology
Downloads & Technology
Foundation Grants & Awards
Membership
Research Policy
State Affiliates

Information for Patients CancerNet
The Journal of Clinical Oncology
The Journal of Oncology Practice
ASCO News & Forum
ASCO Cancer Foundation
The ASCO Press Center

ASCO Releases its First Provisional Clinical Opinion (PCO)

Patients with metastatic colorectal cancer who are candidates for anti-EFGR therapy should have their tumors tested for *KRAS* gene mutations, according to ASCO's first Provisional Clinical Opinion (PCO).

If a patient has a mutated form of the *KRAS* gene, the Society recommends *against* the use of anti-EFGR antibody therapy, based on recent studies indicating this treatment is only effective in patients with the normal (wild-type) form of the *KRAS* gene. It is estimated that 40% of patients with colon cancer have the *KRAS* mutation.

"Personalized medicine is the next frontier in cancer care," said Richard L. Schilsky, MD, ASCO President. "Using *KRAS* testing to guide colorectal cancer treatment is a prime example of where cancer care is heading."

"Basing cancer treatment on the unique genetic characteristics of the tumor or the individual with cancer will improve patient outcomes and help avoid unnecessary costs and side effects for patients who are unlikely to benefit," Dr. Schilsky added.

PCOs are intended to offer timely preliminary clinical direction to oncologists following the publication or presentation of potentially practice-changing data from major studies. ASCO's PCO on *KRAS* gene testing was given prior to the January 15-17, 2009 Gastrointestinal Cancers Symposium in San Francisco, California. The Symposium was co-sponsored by ASCO, the American Gastroenterological Association (AGA), the American Society for Radiation Oncology (ASTRO), and the Society of Surgical Oncology (SSO).

Among the 500 presentations was an important economic and scientific study that discussed the possibility of more than half a billion dollars in savings for the United States healthcare system. The study showed that routine testing for *KRAS* gene mutations in patients with metastatic colorectal cancer could save the U.S. health system up to $604 million per year by identifying who would benefit from the drug cetuximab.

Information on the PCO is currently available on ASCO.org, and the entire report will be published in the February, 1 2009 issue of the *Journal of Clinical Oncology* (JCO).

Validation = Fitness for Intended Use

Types of Validation

- **Analytical validation**
  - **Accuracy in measurement of analyte**
  - **Robustness and reproducibility**
- **Clinical validation**
  - **Correlation with clinical state or outcome**
- **Clinical utility**
  - **Actionable**
  - **Use results in patient benefit**

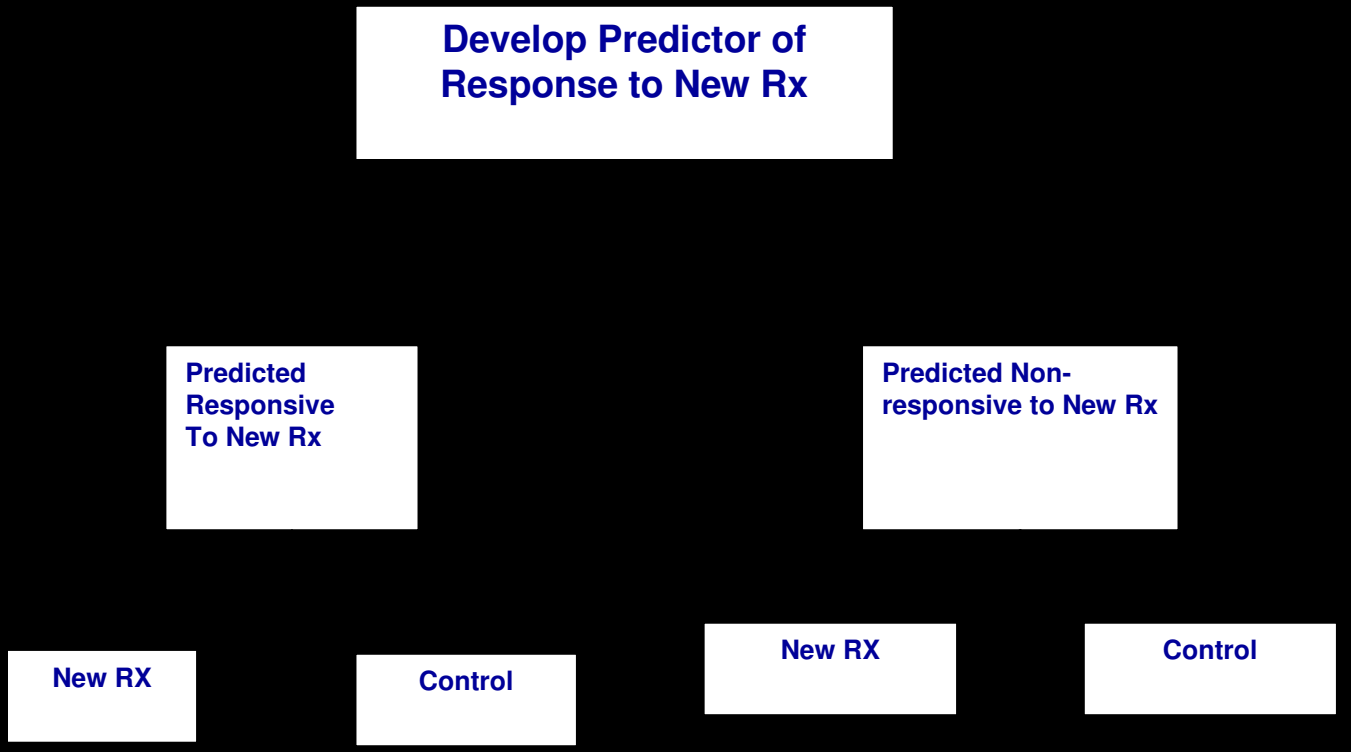
Clinical Utility

- **Benefits patient by improving treatment decisions**
- **Depends on context of use of the biomarker**
  - **Treatment options and practice guidelines**
  - **Other prognostic factors**

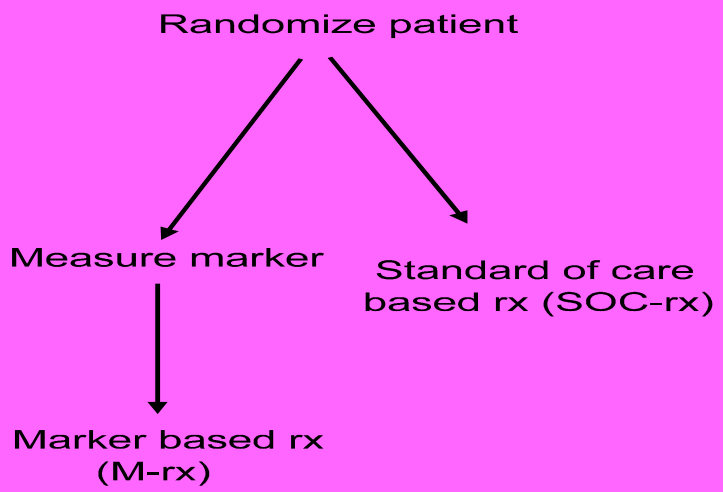
Optimal Designs for Evaluating the Clinical Utility of a Prognostic Biomarker in Breast Cancer

- **Prospective trial to identify such patients and**
  - **withhold chemotherapy**
    - **TAILORx**
  - **or randomize to chemorx vs withhold chemorx**
    - **MINDACT**
- **Prospective-retrospective analysis**
  - **Prospective plan for analysis of archived specimens from previous clinical trial in which patients did not receive chemotherapy**
  - **NSABP B14 evaluation of OncotypeDx**

Optimal Design for Evaluating Predictive Biomarker



Marker Strategy Design



The marker strategy design is also generally very inefficient in terms of the number of patients required for randomization. Sample size requirements for randomized clinical trials are often proportional to the reciprocal of the square of the size of the treatment effect to be detected with a specified statistical power. For the marker strategy design, only the overall treatment effect between the two randomized groups can be evaluated, and the size of that effect is generally quite small because many patients will receive the same treatment regardless of the group to which they are randomized. If the analysis is to demonstrate that withholding a standard therapy for test-negative patients is not inferior, then sample size problems are compounded, and even with a huge sample size, the results are unlikely to be convincing.

Prospective-Retrospective Study

COMMENTARY

Use of Archived Specimens in Evaluation of Prognostic and Predictive Biomarkers

Richard M. Simon, Soonmyung Paik, Daniel F. Hayes

The development of tumor biomarkers ready for clinical use is complex. We propose a refined system for biomarker study design, conduct, analysis, and evaluation that incorporates a hierarchal level of evidence scale for tumor marker studies, including those using archived specimens. Although fully prospective randomized clinical trials to evaluate the medical utility of a prognostic or predictive biomarker are the gold standard, such trials are costly, so we discuss more efficient indirect "prospective–retrospective" designs using archived specimens. In particular, we propose new guidelines that stipulate that 1) adequate amounts of archived tissue must be available from enough patients from a prospective trial (which for predictive factors should generally be a randomized design) for analyses to have adequate statistical power and for the patients included in the evaluation to be clearly representative of the patients in the trial; 2) the test should be analytically and preanalytically validated for use with archived tissue; 3) the plan for biomarker evaluation should be completely specified in writing before the performance of biomarker assays on archived tissue and should be focused on evaluation of a single completely defined classifier; and 4) the results from archived specimens should be validated using specimens from one or more similar, but separate, studies.

J Natl Cancer Inst 2009;101:1–7

Many biomarker studies are conducted with convenience samples of specimens, which just happen to be available and are assayed for the marker, with no prospectively determined subject eligibility, power calculations, marker cut-point specification, or analytical plans. Such studies are very likely to result in highly biased conclusions and truly deserve to be pejoratively labeled as "retrospective." However, if a "retrospective" study is designed to use archived specimens from a previously conducted prospective trial, and if certain conditions are prospectively delineated in a written protocol before the marker study is performed, we argue that it might be considered a "prospective–retrospective" study.

Such a study should carry considerably more weight toward determination of clinical utility of the marker than a simple study of convenience, in which specimens and an assay happen to be available. Having multiple studies of different candidate biomarkers based on archived tissues from the same prospective trial would, however, present a greater opportunity for false-positive conclusions than a single fully prospective trial focused on a specific biomarker. Consequently, independent confirmation of findings for specific biomarkers in multiple prospective–retrospective studies is important (see below).

2) Analytical issues. For a tumor marker study to be sufficient to change clinical practice, the test itself should be ready for clinical practice. For studies to change clinical practice, the investigator should carefully and prospectively plan to use reagents, conditions, and cut points that have been previously determined to be accurate and reproducible. These considerations include fixed reagent supply sources, concentrations, and incubation times among many other possible variables. In addition, the investigator should have demonstrated with statistical confidence the analytical concordance of results between archived specimens and clinical samples for that specific assay. Examples of these concerns include whether the sample was prepared for analysis in a tissue microarray or as a whole section, and whether and how it was subjected to antigen retrieval.

Clinical Study Design

As noted in the first required condition, the investigator should have a clear idea of the specific intended use for the assay. In general, this will be as a prognostic factor to decide if any further treatment is necessary or as a predictive factor to determine whether a particular type of therapy is likely to be effective. To establish medical utility of a prognostic marker, a randomized trial is sometimes not necessary. For example, a prospective single-arm trial in which chemotherapy is withheld from patients at a low risk of recurrence is used in the portion of the TAILORx clinical trial designed to validate the very favorable prognostic outcomes in the low recurrence score population. Assuming that preanalytical factors are well controlled and match current practice activities and that the clinical data are collected in a fashion typical of a clinical trial, archived tissue from a sufficiently large population of untreated patients may be adequate to permit accurate estimates of recurrence based on tumor marker subgroups for determination of clinical utility of the marker.

Suggested Revision of LOEs

In the original American Society of Clinical Oncology LOE scale, "retrospective studies" were determined to be LOE II or worse (3). We now propose an updated revision of the LOE scale, in which more precise definitions are provided for the types of studies that might be used to analyze the clinical utility of a biomarker and in which retrospective studies using archived specimens might reach level I evidence. The LOE for the medical utility of a biomarker relates to key factors involving patients, specimens, assays, and statistical analysis plans (Tables 1 and 2).

Use of Archived Tissues To Determine Clinical Utility of Tumor Markers

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. Focused analysis plan for marker question developed prior to doing assays	Study not prospectively powered at all. Retrospective study design confounded by selection of specimens for study. Focused analysis plan for marker question developed prior to doing assays	Study not prospectively powered at all. Retrospective study design confounded by selection of specimens for study. No focused analysis plan for marker question developed prior to doing assays
Validation	Result unlikely to be play of chance Although preferred, validation not required	Result more likely to be play of chance that 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

Simon R.M., Paik S, Hayes DF. J Natl Cancer Inst. 101:1446-52, 2009

Use of Archived Tissues To Determine Clinical Utility of Tumor Markers

Category	A
Trial Design	Prospective
Clinical trial	PCT/PRCT designed to address tumor marker
Patients and patient data	Prospectively enrolled, treated, and followed in PCT/PRCT
Specimen collection, processing, and archival	Specimens collected, processed and assayed for specific marker in real time
Statistical Design and analysis	Study powered to address tumor marker question.
Validation	Result unlikely to be play of chance Although preferred, validation not required

Simon R.M., Paik S, Hayes DF. J Natl Cancer Inst. 101:1446-52, 2009

Use of Archived Tissues To Determine Clinical Utility of Tumor Markers

Category	B
Trial Design	Prospective using archived samples
Clinical trial	Prospective trial not designed to address tumor marker, but design accommodates tumor marker utility. Accommodation of predictive marker requires PCT/PRCT
Patients and patient data	Prospectively enrolled, treated, and followed in clinical trial and, especially if a predictive utility is considered, a PRCT addressing the treatment of interest
Specimen collection, processing, and archival	Specimens collected, processed, and archived prospectively using generic SOPs. Assayed after trial completion.
Statistical Design and analysis	Study powered to address therapeutic question; underpowered to address tumor marker question. Focused analysis plan for marker question developed prior to doing assays
Validation	Result more likely to be play of chance that A, but less likely than C. Requires one or more validation studies

Simon R.M., Paik S, Hayes DF. J Natl Cancer Inst. 101:1446-52, 2009

Use of Archived Tissues To Determine Clinical Utility of Tumor Markers

Category	C
Trial Design	Prospective /observational
Clinical trial	Prospective observational registry, treatment and follow up not dictated
Patients and patient data	Prospectively enrolled in registry, but treatment and follow up standard of care
Specimen collection, processing, and archival	Specimens collected, processed, and archived prospectively using generic SOPs. Assayed after trial completion.
Statistical Design and analysis	Study not prospectively powered at all. Retrospective study design confounded by selection of specimens for study. Focused analysis plan for marker question developed prior to doing assays
Validation	Result very likely to be play of chance. Requires subsequent validation studies

Simon R.M., Paik S, Hayes DF. J Natl Cancer Inst. 101:1446-52, 2009

Use of Archived Tissues To Determine Clinical Utility of Tumor Markers

Category	D
Trial Design	Retrospective/observational
Clinical trial	No prospective aspect to study
Patients and patient data	No prospective stipulation of treatment or follow up; patient data collected by retrospective chart review
Specimen collection, processing, and archival	Specimens collected, processed and archived with no prospective SOPs
Statistical Design and analysis	Study not prospectively powered at all. Retrospective study design confounded by selection of specimens for study. No focused analysis plan for marker question developed prior to doing assays
Validation	Result very likely to be play of chance. Requires subsequent validation

Simon R.M., Paik S, Hayes DF. J Natl Cancer Inst. 101:1446-52, 2009

Revised LOI Scale: Use of Archived Tissues

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

Simon R.M., Paik S, Hayes DF. J Natl Cancer Inst. 101:1446-52, 2009

Archived tissue specimens from high-quality datasets can therefore be of great importance for establishing the medical utility of a prognostic or predictive biomarker. We argue that it is appropriate to use archived tissue specimens from large prospective clinical trials to do so. For such an evaluation to be more useful than just for generating hypotheses, however, several conditions must be satisfied:

1) Archived tissue, adequate for a successful assay, must be available on a sufficiently large number of patients from the pivotal trials to permit appropriately powered analyses and to ensure

that the patients included in the biomarker evaluation are clearly representative of the patients in the pivotal trials. Although no minimal requirement can be stated as universally applicable, we would suggest that samples from at least two-thirds of the patients be available for analysis.

2) Substantial data on analytical validity of the test must exist that ensure that results obtained from the archived specimens will closely resemble those that would have been obtained from analysis of specimens collected in real time. Assays should be conducted blinded to the clinical data.

3) The analysis plan for the biomarker evaluation must be completely developed before the performance of the biomarker assays. Both the analysis plan for the biomarker study and the design of the trial(s) whose samples were selected for analysis should be appropriate for the evaluation of a companion diagnostic had it been undertaken at the outset. The analysis should be focused on a single, completely defined, diagnostic classifier. For multigene classifiers, the mathematical form of combining the individual components, weights, and cut points should be specified beforehand. In general, the analysis should not be exploratory, and practices that might lead to a false-positive conclusion should be avoided.

4) The results must be validated in at least one or more similarly designed studies using the same assay techniques.

Conclusions

- **Claims of medical utility for prognostic and predictive biomarkers based on analysis of archived tissues can have either a high or low level of evidence depending on several key factors.**
- **These factors include the analytical validation of the assay, the nature of the study from which the specimens were archived, the number and condition of the specimens, and the development prior to assaying tissue of a focused written plan for analysis of a completely specified biomarker classifier.**
- **Studies using archived tissues from prospective clinical trials, when conducted under ideal conditions and independently confirmed can provide the highest level of evidence.**
- **Traditional analyses of prognostic or predictive factors, using non analytically validated assays on a convenience sample of tissues and conducted in an exploratory and unfocused manner provide a very low level of evidence for clinical utility.**