Biostatistics Primer

What a Clinician Ought to Know—Prognostic and Predictive Factors

Lorinda Simms, MSc, P Stat, Helen Barraclough, MSc,† and Ramaswamy Govindan, MD‡§*

Several prognostic factors in oncology have been established over the years, such as performance status, tumor size, and disease stage. The identification of prognostic and predictive factors is becoming increasingly important in medical research, particularly as scientific discoveries have led to better understanding of diseases and genetics, resulting in tailored therapy. Advances in drug discovery and better understanding of the mechanism of action, may also identify factors that may be prognostic and/or predictive. Prognostic or predictive factors may include patient characteristics such as age, ethnicity, sex, or smoking status, disease characteristics such as disease stage or nodal status, and molecular markers such as HER2 amplification and K ras mutation.

It can be challenging to distinguish whether a factor is prognostic or predictive, based on what is reported in the literature. This article is intended to help the reader assess whether a factor is prognostic and/or predictive.

Key Words: Prognostic, Predictive, Treatment interaction, Statistics.

(*J Thorac Oncol.* 2013;8: 808-813)

In the context of randomized clinical trials (RCTs), a large number of baseline factors are collected, and these factors have the potential to be prognostic and/or predictive. Any baseline factor may be prognostic and/or predictive; some examples of such factors are age, sex, race, and disease stage. Only baseline values of factors are assessed to be prognostic and/or predictive in RCTs because these are not affected by the treatment administered.[1,2]

The baseline values for factors that do change over time, such as performance status and prostate antigen level, are within the scope of this article; however, the postbaseline time-dependent changes and impact on outcomes are outside the scope of this article. The prognostic and predictive factor concepts discussed in this article are focused on RCTs. However, these concepts can be applied similarly to observational research and beyond.

Why Do Clinicians Need to Know About Prognostic and Predictive Factors?

Prognostic and predictive factors enable clinicians to make informed decisions about when to initiate, stop, or change therapy for a patient (prognostic factors), or what specific therapy to choose for an individual patient (predictive factors). Whereas randomization is important in clinical trial design to ensure balance between treatment arms, incorporation of prognostic factors into the randomization process (such as through stratification or minimization) increases balance within those important prognostic factors that may influence outcomes.[3] Predictive factors are important in tailoring treatment to the appropriate patients. Uses of prognostic and predictive factors are summarized in Table 1.

What are Prognostic and Predictive Factors, and How Are They Different?

In the simplest case, a prognostic factor is a variable that is assessed before starting any treatment; based on the value (i.e., level) of this factor, the clinician can expect that a patient may have a better or worse clinical outcome (such as survival or response), regardless of what treatment the patient receives.[4] For example, it is well recognized in oncology that patients with good baseline performance status do better than patients with worse performance status, or women often do better than men, regardless of the therapy being given. In this case, the relationship between the level of a factor (e.g., good performance status) with a patient's outcome (e.g., survival) does not depend on the specific treatment given.

A factor is predictive when the treatment effect (such as response rate [RR]), within the levels of the factor (e.g., men and women), varies depending on the treatment given.[5] That is, the relationship between the patient's level of a factor (e.g., women) with a patient's outcome (e.g., RR) depends on the specific treatment given.

A factor may be either prognostic and/or predictive for a particular endpoint (e.g., survival) and may not have any relationship at all with other endpoints such as

*Statistics and Information Sciences, Eli Lilly and Company, Toronto, Ontario, Canada; †Asia Pacific Statistical Sciences, Eli Lilly and Company, Sydney, Australia; and ‡Department of Medicine, Division of Oncology, Washington University School of Medicine, St. Louis, Missouri; and §Alvin J. Siteman Cancer Center at Washington University School of Medicine, St. Louis, Missouri.

Disclosure: Lorinda Simms is employed by and holds stock in Eli Lilly Canada, Inc. Helen Barraclough is employed by Eli Lilly Australia.

Address for correspondence: Ramaswamy Govindan, MD, Division of Medical Oncology, Washington University School of Medicine, 660 S. Euclid, Box 8056, St. Louis, MO 63110. E-mail: rgovinda@im.wustl.edu

ISSN: 1556-0864/13/0806-0808

TABLE 1. Uses for Prognostic and Predictive Factors

Prognostic Factor	Predictive Factor
Contributes to understanding of disease process, such as when to initiate, stop, or change treatment.	Helps a physician determines whether a certain treatment is a good option for a specific patient, based on the individual patient's characteristics.
Is used in clinical trial inclusion/exclusion criteria to ensure homogeneity of the study population.	Can also be used in clinical trial inclusion/exclusion criteria to further assess the treatment effect, specifically in patients most likely to have increased benefit and/or decreased risk and to further quantify the benefit/risk.
Is used in clinical trial randomization process and/or statistical analyses to ensure balance between treatment arms, and to perform subgroup or adjusted analyses (if there is an imbalance in prognostic factors between the treatment arms).	Is used in clinical trial statistical analyses to evaluate or quantify a treatment-by-factor interaction.

progression-free survival or RR. In addition, predictive factors are often only predictive for a particular therapy or class of therapies. Furthermore, a factor may be both prognostic and predictive at the same time. Identification of an important predictive factor can change the standard of care, such that in the future, this factor can then become a prognostic factor. That is, type(s) of patients who are likely to do poorly with a new treatment would receive other treatments; the remaining patient type(s) who show benefit from this new treatment may then also show a prognostic effect. The differences between prognostic and predictive factors are summarized in Table 2.

How to Determine Whether a Factor is Prognostic or Predictive?

Assessing whether a baseline factor is prognostic or predictive can be done either prospectively (i.e., hypothesis testing by prespecifying this analysis in the protocol and/or statistical analysis plan) or retrospectively (i.e., hypothesis generating using exploratory analyses). The validity of prespecified analyses is obtained predominantly when the results are reproduced and consistent across studies, with an underlying biologic plausibility.[1,2] Increased confidence in such analysis is greatly improved in large studies that are well-designed and minimize error and biases through randomization, blinding, adequate follow-up, and prespecified analyses.[2]

Analyses to assess whether a factor is prognostic and/or predictive typically use regression methods. The type of regression method depends on the type of endpoint. For example, typically Cox regression will be used for *time-to-event* endpoints such as overall survival or progression-free survival, logistic regression will be used for *binary* endpoints such as responder (Yes/No), and linear regression will be used for *continuous* endpoints such as change in tumor size. Flowchart 1 illustrates a common approach to inferential models used to assess whether a baseline factor is prognostic and/or predictive; inferential models are prespecified and use methods to control for multiple testing. Flowchart 2 illustrates a common approach to exploratory model building to assess prognostic and/or predictive baseline factors.

Hypothetical Examples Illustrating Different Scenarios

The following examples discuss hypothetical results to illustrate different combinations of prognostic and/or predictive factors. In each example, Tx A is the control, and Tx B is the experimental treatment.

Example 1—Sex: Neither prognostic nor predictive

In this example both women and men have a 30% RR on Tx A and a 60% RR on Tx B (Fig. 1A). Hence, the size and magnitude of the treatment effect (i.e., improvement of 30% on Tx B) was the same in both sexes. In addition, the RR for both treatments was similar (in this case, the same) for both sexes.

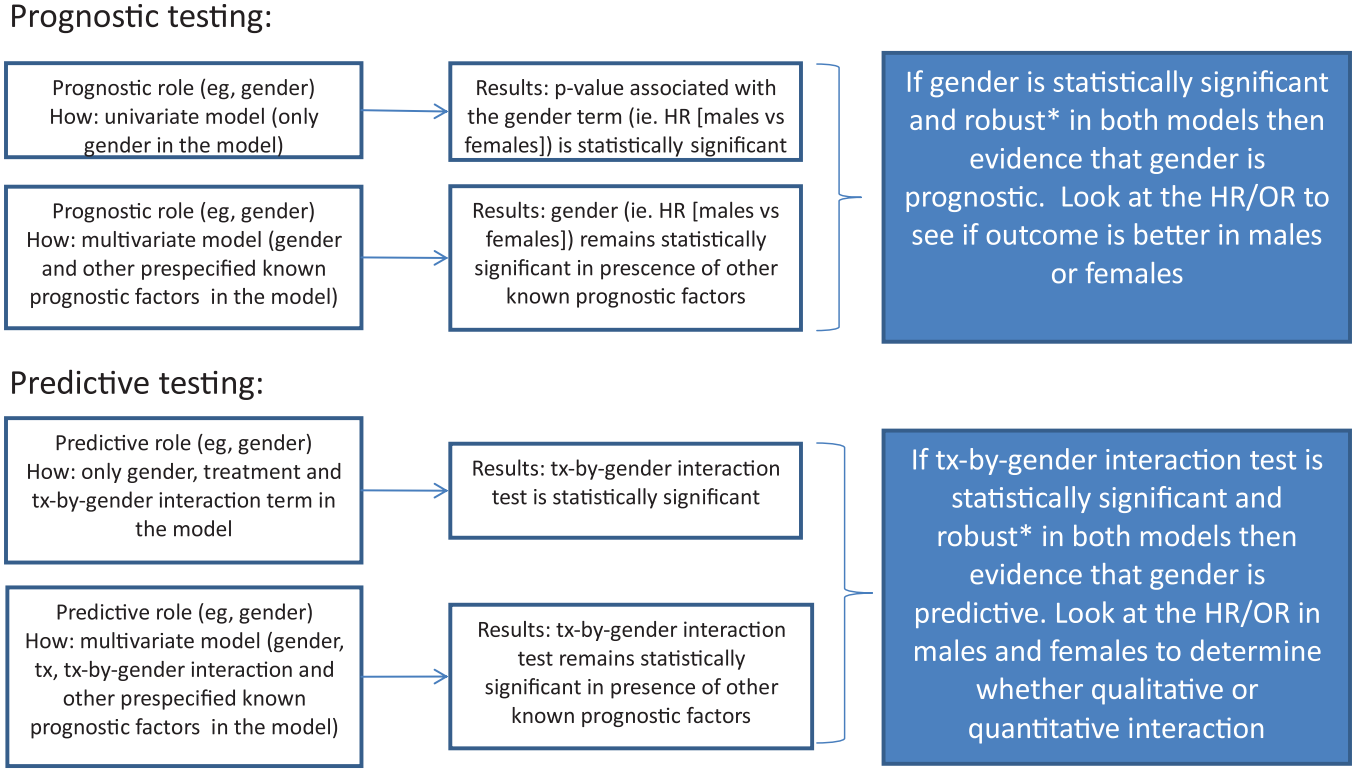
In a regression model evaluating the prognostic effect of sex, the *p* value for the sex effect was 0.75, supporting that sex is not prognostic.

Example 2—Sex: Prognostic but not predictive

In a hypothetical clinical trial, women on Tx A had an RR of 30%, whereas those on Tx B had an RR of 60%, a difference of 30% (Fig. 1B). In contrast, men on Tx A had an RR

TABLE 2. Differences between Prognostic and Predictive Factors

Prognostic Factor	Predictive Factor
Is independent of treatment.	Is dependent on a specific treatment.
The magnitude and direction of the treatment effect are the same across subsets of patients, but the response is consistently higher in one subset than another across the treatment arms (e.g., women have higher response than men on Tx A and Tx B).	The magnitude and/or the direction of the effect changes from one subset to another (e.g., smaller effect of Tx A compared with Tx B in women than in men).
Results are replicated across multiple studies of multiple different treatments within the same (or similar) disease state.	Results are replicated across several studies, containing the same experimental treatment (however, potentially having different control arms) within the same (or similar) disease state.



*Robustness refers to general consistency and similarity of point estimates; the point estimates do not need to be identical, but should be in the same direction, and similar magnitude. For any notable differences observed, possible reasons should be explored further.
For example, a HR of 0.85 (95% CI: 0.75-0.95) for a univariate model, and a multivariate model with a HR of 0.80 (95% CI: 0.72, 0.88) are sufficiently similar in magnitude and direction, and considered to be robust.
However, a HR of 0.85 (95% CI: 0.75-0.95) for a univariate model, but a multivariate model with a HR of 0.50 (95% CI: 0.35, 0.65), would not be considered to be robust, since the magnitude has substantially changed, even though the direction has not. This example suggests that there may be imbalances in baseline factors that may be impacting the analyses.

FLOWCHART 1. Prespecified analyses: Inferential regression models.

of 20%, whereas those on Tx B had an RR of 50%, also a difference of 30%. Thus, both women and men had a treatment difference in response of 30%, but men had lower responses than women (20% versus 30% on Tx A and 50% versus 60% on Tx B). This lower response for men, compared with that for women, suggests that sex is prognostic.

Furthermore, when statistical modelling was used, the *p* value for the sex effect was 0.04, confirming the prognostic effect of sex. However, when a treatment-by-sex interaction test was performed, the *p* value was 0.65, supporting that sex is not predictive.

Example 3—Sex: Not prognostic, but predictive (quantitative interaction)

In several hypothetical trials, there has not been any indication that women have a better or worse response than men, suggesting sex has no prognostic effect (Fig. 1C). In this hypothetical clinical trial, women had a 30% RR on Tx A compared with 60% for Tx B, a difference of 30%. In contrast, men had a 30% RR on Tx A versus 45% on Tx B, a difference of only 15%. In this example, the magnitude of the treatment effect (difference in RR) differs, suggesting a quantitative interaction.

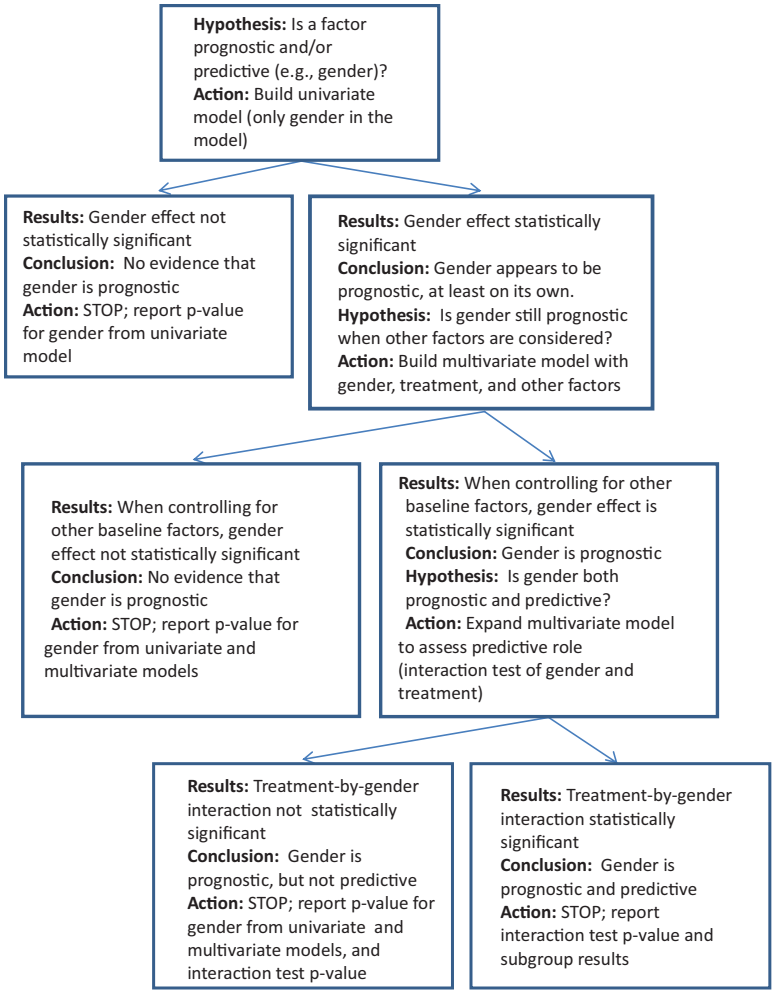
Formal statistical evaluation using an interaction test between treatment and sex resulted in a *p* value of 0.025, supporting sex to be predictive but not prognostic because the RR on the control arm (Tx A) was the same for both sexes.

Example 4—Sex: Not prognostic but predictive (qualitative interaction)

In the same hypothetical example, women again had a 30% improved RR on Tx B compared with Tx A (60% versus 30%) (Fig. 1D). However in men, the RR was worse on Tx B than on Tx A by 20%. Hence the direction of the treatment effect changed between the sexes, suggesting a qualitative interaction. A statistical model evaluating these data showed evidence of a predictive effect, with an interaction *p* value of 0.001.

Example 5—Estrogen receptor status: Prognostic (positively) and predictive (positively)

In breast cancer, patients with estrogen receptor (ER) positive status, tend to live longer than those with ER negative status (Fig. 1E). In a hypothetical trial, patients with ER



Note: this is a common approach to model building, other approaches may be used. One approach may be to assess for predictive factors only, rather than both prognostic and predictive.

FLOWCHART 2. Model building to determine prognostic and predictive factors.

positive disease had a 5-year survival rate of 70%, whereas those with ER negative disease had a 5-year survival rate of 50%; statistical analysis confirmed the prognostic value of ER status with a *p* value of 0.02 (not shown in Fig. 1E). Similar results have been shown across other trials, with other treatments.

In addition, in this hypothetical trial, patients who were ER positive and received Tx B had a 5-year survival rate of 80%, compared with 60% for those who received Tx A. In contrast, patients who were ER negative and received Tx B had a 5-year survival rate of 50% versus 40% for those who received Tx A. The interaction test of treatment and ER status produced a *p* value of 0.01, also supporting the predictive value of ER status.

These hypothetical results show that ER positivity is both prognostic and predictive, because ER-positive patients had a better prognosis than those who are ER negative. In addition, ER-positive patients also had a larger treatment effect with Tx B than those treated with Tx A, compared with patients who are ER negative and were treated with either agent.

Example 6—HER2 receptor status: Prognostic (negatively) and predictive (positively)

This example (Fig. 1F) continues from example 5. The HER2 receptor status in breast cancer has been shown to be associated with poor prognosis. In various hypothetical trials, the illustrated results show that HER2 is both prognostic and predictive, because patients who are HER2 positive had a worse prognosis than those who are HER2 negative on the control arm, Tx A (i.e., negatively prognostic). In addition, patients who are HER2 positive also had a smaller treatment effect with Tx B than those treated with Tx A, compared with patients who are HER2 negative (i.e., positively predictive; a quantitative interaction is shown, with a positive benefit for both HER2-negative and -positive patients—although, a larger magnitude for HER2-negative, and a small magnitude for HER2-positive patients).

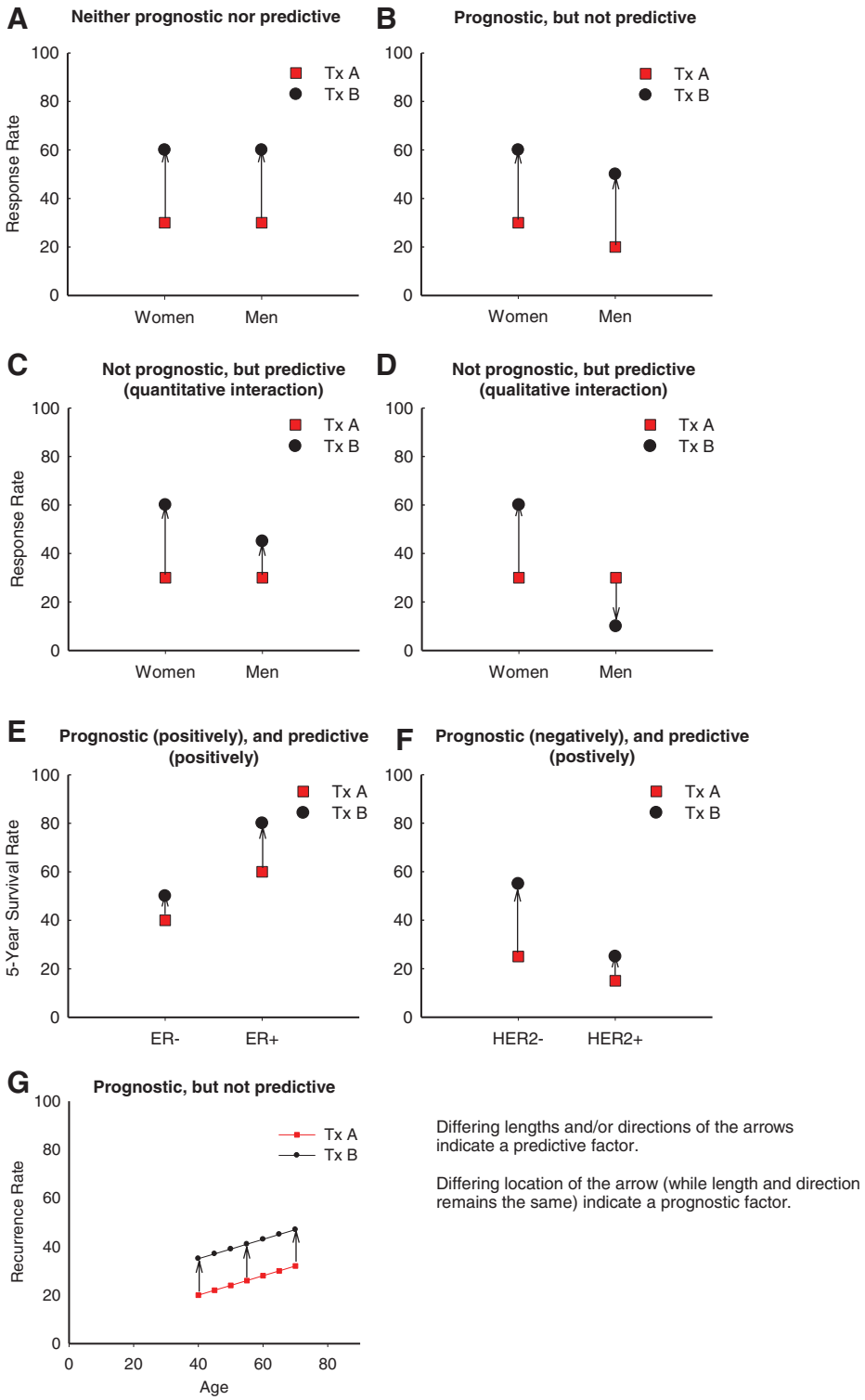


FIGURE 1. Hypothetical examples of prognostic and predictive factors.

Example 7—Age (continuous): Prognostic but not predictive

This example (Fig. 1G) follows on from the previous examples. However, in this case, the factor of interest is continuous, as opposed to being binary.

In locally advanced cancers, increasing age has been shown to increase the risk of cancer recurrence. In the hypothetical data shown here, for every 5-year increase in age (starting at the age of 40 years) for patients on Tx A, there was a 2% increase in the recurrence rate; hence, at 40 years of age,

the risk of recurrence was 20%, whereas at 70 years, it was 32%. For patients receiving TxB, the risk of recurrence for a 40-year old was 35%, whereas the risk of recurrence at age 70 was 47%. The treatment difference in the risk of recurrence was 15%, at both 40 and 70 years of age. When statistical modeling was used, the *p* value for age was 0.025, confirming the prognostic effect of age. However, when a treatment-by-age interaction test was performed, the *p* value was 0.45, indicating that age is not predictive.

SUMMARY

Identification of prognostic factors in a given disease setting is important. These prognostic factors enable appropriate planning of further clinical research and determine which type of patients would be appropriate for a given trial. Understanding of prognostic factors enables the clinician to evaluate other research and determine its applicability to his or her patients, through an assessment of how similar the patients included in that research are to those in the clinician's practice.

Predictive factors are critically important in being able to tailor therapy to a given patient. If a patient has a certain profile, and it is known that a certain therapy works better for patients with that particular profile, then clinicians are able to ensure that the patients get the best therapy to prolong their survival and/or maintain quality of life.

Clinicians need clear descriptions and analyses from published literature to determine whether a factor is prognostic or predictive. This can be achieved by giving the complete regression analysis in a table, including all the factors and results (point estimates; 95% confidence intervals; and *p* values) from that regression model.[6] Interaction tests need to be clearly described as such to ensure accuracy of interpretation (e.g., either by using the actual term *interaction test* or by adequately describing the variable—e.g., *treatment-by-sex interaction*).

One limitation of interactions tests is that they often have low statistical power, hence, a significant interaction *p* value may not be observed, even if the factor is truly predictive.[7] Of note, studies are usually designed and powered for testing a treatment effect on the primary endpoint in the overall study population, not for subgroup analyses and/or an interaction test.[7–9] Multiple significance testing can also make it seem that there is a subgroup effect, even when there is no real effect (i.e., observed by chance alone); appropriate testing using gatekeeping, and/or *p* value adjustment should be considered.[1,7,10,11] Ideally, studies should be adequately designed and/or powered for subgroup analyses and/or tests of interaction.[1]

The strongest evidence in support of prognostic and predictive factors is replication and consistency of findings across several large trials, and/or in confirmatory studies, with the addition of biologic rationale to support the findings. Subgroup results (either prognostic or predictive) are generally considered hypothesis generating until they can be reproduced across several large studies (or in a meta-analysis), with an appropriate underlying biologic plausibility.[1,2,7,9,10] Further evaluation of prognostic and predictive factors in clinical trials is urgently needed to advance our understanding of diseases and enable further tailoring of therapy to obtain the best patient outcomes.

ACKNOWLEDGMENTS

The authors thank Richard Walton, Jonathan Denne, and Yanping Wang for their constructive review and suggestions, which improved this article.

REFERENCES

1. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991;266:93–98.
2. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992;116:78–84.
3. Bulpitt CJ. Subgroup analysis. *Lancet* 1988;2:31–34.
4. National Cancer Institute Dictionary of Cancer Terms. Definition of prognostic factor. Available at: http://www.cancer.gov/dictionary?cdrid=44246. Accessed November 16, 2012.
5. National Cancer Institute Dictionary of Cancer Terms. Definition of predictive factor. Available at: http://www.cancer.gov/dictionary?CdrID=44245. Accessed November 16, 2012.
6. Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? *BMJ* 2001;322:226–231.
7. International Conference on Harmonization (ICH) Topic E9. Statistical principles for clinical trials, 1998.
8. Barraclough H, Govindan R. Biostatistics primer: what a clinician ought to know: subgroup analyses. *J Thorac Oncol* 2010;5:741–746.
9. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000;355:1064–1069.
10. Pocock SJ, Hughes MD. Estimation issues in clinical trials and overviews. *Stat Med* 1990;9:657–671.
11. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987;317:426–432.

