

*Perspective***Evaluating the Efficiency of Targeted Designs for Randomized Clinical Trials****Richard Simon and Aboubakar Maitournam**

Biometric Research Branch, National Cancer Institute, Bethesda, Maryland

ABSTRACT

Purpose: Genomic technologies make it increasingly possible to identify patients most likely to benefit from a molecularly targeted drug. This creates the opportunity to conduct targeted clinical trials with eligibility restricted to patients predicted to be responsive to the drug.

Experimental Design: We evaluated the relative efficiency of a targeted clinical trial design to an untargeted design for a randomized clinical trial comparing a new treatment to a control. Efficiency was evaluated with regard to number of patients required for randomization and number required for screening.

Results: The effectiveness of this design, relative to the more traditional design with broader eligibility, depends on multiple factors, including the proportion of responsive patients, the accuracy of the assay for predicting responsiveness, and the degree to which the mechanism of action of the drug is understood. Explicit formulas were derived for computing the relative efficiency of targeted *versus* untargeted designs.

Conclusions: Targeted clinical trials can dramatically reduce the number of patients required for study in cases where the mechanism of action of the drug is understood and an accurate assay for responsiveness is available.

INTRODUCTION

Many cancer therapeutics benefit only a subset of treated patients. Genomic technologies such as DNA microarray expression profiling are providing biomarkers that facilitate the prediction of which patients are most likely to respond to a given regimen (1, 2). Molecularly targeted drugs are of increasing importance in cancer therapeutics, and such drugs are only expected to be effective for patients whose tumors express the target (3, 4). Thus, clinical trials may be increasingly tailored for patients who are predicted to respond to therapy (5). We call

these targeted designs. As discussed in this article, we studied the efficiency of targeted designs in comparison with traditional randomized designs with broader eligibility criteria. We evaluated efficiency in the context of a binary outcome end point. Although many clinical trials use survival or time-to-progression end points, the binary end point setting is more tractable, and we obtained results that are intuitive and should be useful in understanding the factors that effect efficiency generally. For the untargeted and targeted design, we considered the comparison of a control *versus* experimental treatment with the same number of randomized patients in the two groups.

We compared the two designs with regard to the number of randomized patients required. We also compared the number of randomized patients for the untargeted design to the number of screened patients required for the targeted design. We assume that in the targeted design patients are screened using an assay that indicates whether the patient is likely to benefit from the new treatment. If the control arm is an active treatment, then the screening classifier should provide an indication of whether the patient is more likely to respond to the new regimen than to the control arm. Our efficiency comparisons are based on using the formula of Ury and Fleiss (6) for planning sample size for comparing proportions because of its known accuracy for approximating the tables of Casagrande, Pike, and Smith for the power of Fisher's exact test (7).

MATERIALS AND METHODS

We considered a population of patients consisting of an R+ portion who were predicted to be responsive to the new treatment and a remainder portion R-. The R- strata constituted a proportion γ of the population. Patients were randomized between the control and the experimental groups. p_c denotes the response probability in control group and was assumed to be the same for R- and R+ patients. The response probability in the treatment group was $p_c + \delta_0$ and $p_c + \delta_1$ for the R- and R+ patients, respectively. The response probability p_e for the experimental treatment group in the untargeted design was a weighted average of $p_c + \delta_0$ and $p_c + \delta_1$ with weights γ and $1-\gamma$, respectively.

For the targeted design we added the symbol T. The response probability in the experimental group was $p_e^T = p_c + \delta_1$. We consider the one-sided test of the null hypothesis $p_c = p_e$ against the alternative hypothesis $p_e > p_c$.

Let n and n^T denote the number of patients needed to randomize in the untargeted and targeted design respectively to achieve the same statistical power for testing the null hypothesis. The expressions for n and n^T are indicated in the Appendix. The relative efficiency of the untargeted and the targeted designs can be expressed in the form:

$$n/n^T = \left[\frac{\delta_1}{\gamma\delta_0 + (1-\gamma)\delta_1} \right]^2 f \quad (A)$$

Received 3/11/04; revised 5/14/04; accepted 5/19/04.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article may be found at <http://linus.nci.nih.gov/~brb/TechReport.htm>.

Requests for reprints: Richard Simon, National Cancer Institute, 9000 Rockville Pike, Bethesda MD 20892-7434. Phone: 301-496-0975; Fax: 301-402-0560; E-mail: rsimon@nih.gov.

The factor f , defined in the Appendix, is often close to 1 (see Supplementary Data), and the relative efficiency can be expressed in a simple and intuitive manner as the squared ratio of treatment effect for the targeted design and that for the untargeted design.

We considered cases where the R- patients do not benefit from the new treatment ($\delta_0 = 0$: case 0) and cases where they benefit half as much as the R+ patients ($\delta_0 = \delta_1/2$: case 1). The response probabilities for control group p_c were taken equal to 0.1 or 0.5, and the improvement in response probability for the R+ patients (δ_1) was 0.2 or 0.4. The calculations shown below were done with $\alpha = 0.025$ and power 80% ($\beta = 0.2$).

RESULTS

Number of Randomized Patients. Fig. 1 shows the number of randomized patients required for the untargeted design relative to the targeted design as a function of the proportion of R+ patients. In all of the cases considered, the targeted design required fewer randomized patients than the untargeted design (ratios >1). However, the advantage of the targeted design was much greater for scenarios where the experimental treatment was completely ineffective for the R- patients (case 0) compared with the scenarios where the experimental treatment was assumed to be partially effective for the R- patients (case 1).

The advantage of the targeted design can be seen analytically in the case 0 by taking $\delta_0 = 0$ in equation A. This gives:

$$n/n^T = \left[\frac{1}{(1-\gamma)} \right]^2 f \tag{B}$$

For $\gamma = 0.5$, equation B gives a relative efficiency of $4f$. When p_c is not close to 0, the value of f is close to 1 (see Supplementary Data), and the untargeted design requires about four times

as many randomized patients. When p_c is close to 0, the factor f is <1 , and the relative efficiency is <4 .

Although the advantage of the targeted design was not as great when the experimental treatment was somewhat effective for both groups of patients (case 1), the reduction in required number of patients can still be substantial. For case 1 with $\delta_0 = \delta_1/2$, the relative efficiency formula (A) reduces to:

$$n/n^T = \left[\frac{1}{1-\gamma/2} \right]^2 f \tag{C}$$

For $\gamma = 0.5$, the relative efficiency is $(4/3)^2 f$, which is ~ 1.75 when f is close to 1. Hence, the untargeted design requires $\sim 75\%$ more randomized patients in this scenario.

If we assume that only R+ patients benefit from the new treatment but that selection of patients is determined by an imperfect assay, then the treatment benefit for assay negative patients can be shown to be $\delta_1(1 - NPV)$ where NPV denotes the negative predictive value of the assay; that is, the probability that the true status is R- when the assay is negative (see Supplementary Data). The treatment benefit for assay positive patients is $\delta_1 PPV$ where PPV is the positive predictive value of the assay. NPV and PPV values of at least 0.9 are highly desirable for enabling the efficiency of the targeted design to be achieved. Decreasing NPV is equivalent to increasing δ_0 , whereas decreasing PPV is equivalent to decreasing δ_1 .

Number of Patients Screened for Targeted Design. To randomize n^T patients with the targeted design, a greater number of patients must be screened. If $n^T/(1-\gamma)$ patients are screened, then the expected number of randomized patients is n^T because $1-\gamma$ is the proportion of R+ patients. This is, of course, not required for the untargeted design. Fig. 2 shows the ratio of the number of randomized patients for the untargeted design (n) to the number of screened patients for the targeted design $n^T/(1-\gamma)$.

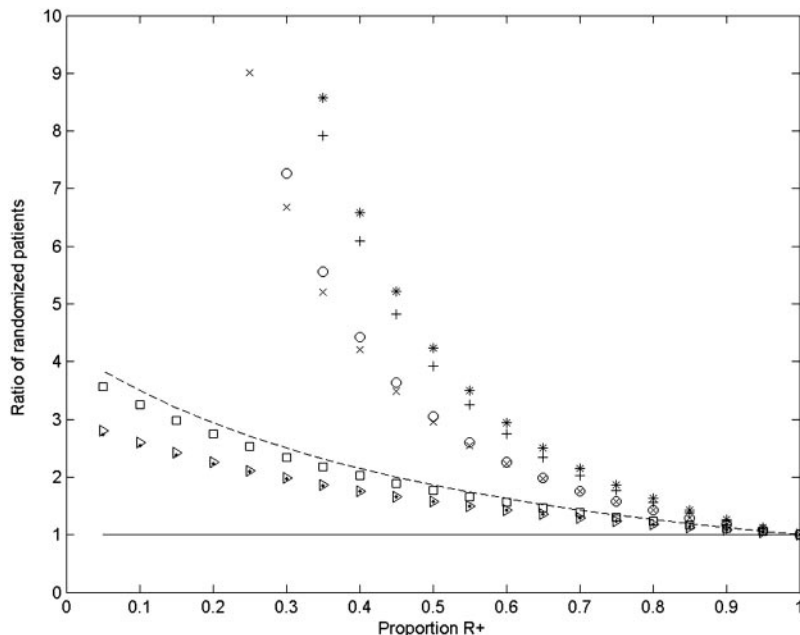


Fig. 1 Number of randomized patients required for untargeted design relative to that for targeted design. The horizontal axis represents the proportion of patients who express the target and are expected to be responsive to the new treatment. Case 0: No treatment effect for R- patients ($\delta_0 = 0$); O: $\delta_1 = 0.2, p_c = 0.1$; X: $\delta_1 = 0.4, p_c = 0.1$; +: $\delta_1 = 0.2, p_c = 0.5$; *: $\delta_1 = 0.4, p_c = 0.5$. Case 1: Treatment effect for R- patients is half as large as that for R+ patients ($\delta_0 = \delta_1/2$); >: $\delta_1 = 0.2, p_c = 0.1$; <: $\delta_1 = 0.4, p_c = 0.1$; square: $\delta_1 = 0.2, p_c = 0.5$; -: $\delta_1 = 0.4, p_c = 0.5$.

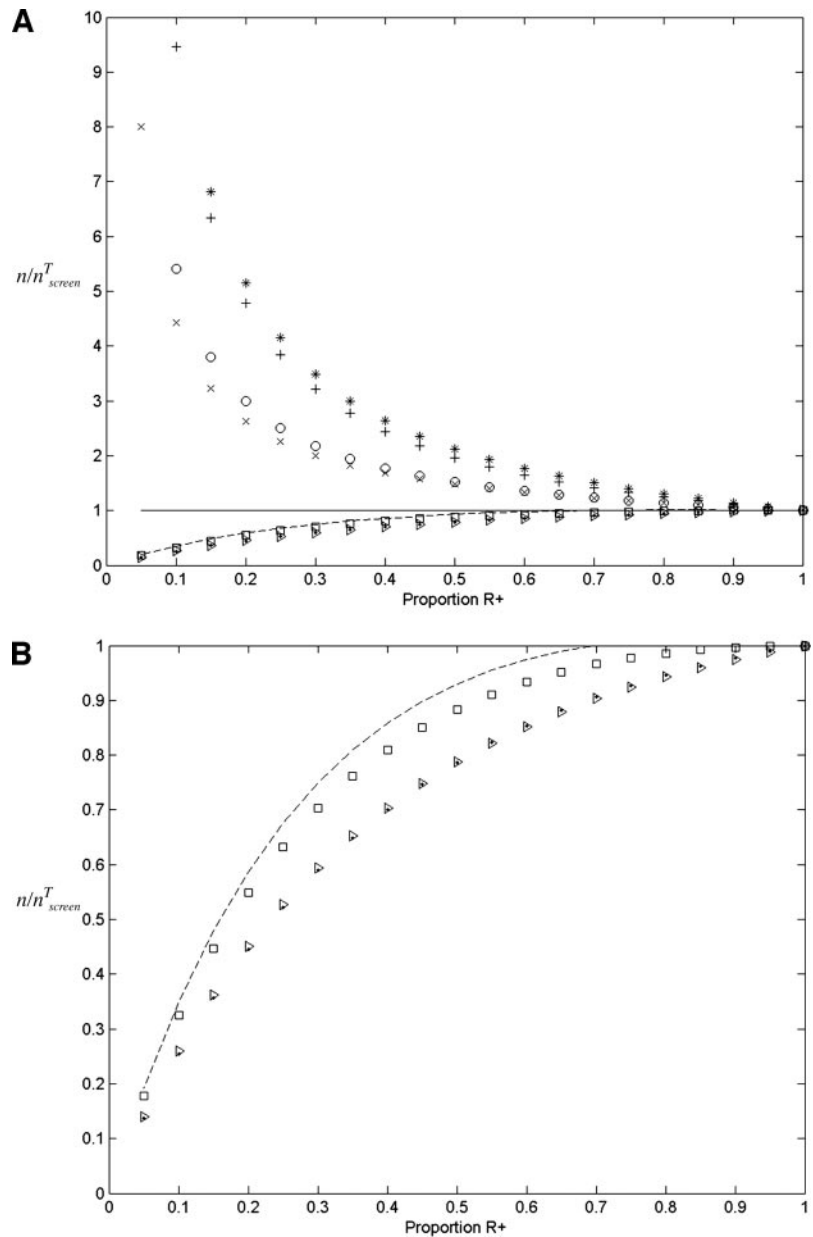


Fig. 2 A, ratio of number of randomized patients for untargeted design to number of screened patients for targeted design. The horizontal axis represents the proportion of patients who express the target and are expected to be responsive to the new treatment. Case 0: No treatment effect for R- patients ($\delta_0 = 0$): \circ : $\delta_1 = 0.2, p_c = 0.1$; \times : $\delta_1 = 0.4, p_c = 0.1$; $+$: $\delta_1 = 0.2, p_c = 0.5$; $*$: $\delta_1 = 0.4, p_c = 0.5$. Case 1: Treatment effect for R- patients is half as large as that for R+ patients ($\delta_0 = \delta_1/2$): \triangleright : $\delta_1 = 0.2, p_c = 0.1$; \cdot : $\delta_1 = 0.4, p_c = 0.1$; \square : $\delta_1 = 0.2, p_c = 0.5$; $-$: $\delta_1 = 0.4, p_c = 0.5$. B, zoom view of A below ratio 1.

Again, there is a clear separation between the scenarios where R- patients do not benefit from the experimental treatment (case 0) and those scenarios where R- patients benefit partially from the experimental treatment (case 1).

In the case 0 scenarios, the targeted design is more efficient than the untargeted design, not just with regard to number of randomized patients required, but even with regard to required number of patients to screen. For case 0, the ratio of randomized to screened patients is

$$n/n_{screen}^T = \left[\frac{1}{(1-\gamma)} \right] f \tag{D}$$

which is the same as equation B except for the absence of the square.

For the scenarios of case 1 where the R- patients benefit partially from the experimental treatment (Fig. 2), the number of patients required for screening with the targeted design is always greater than the number required to randomize for the untargeted design. Analytically the ratio is:

$$n/n_{screen}^T = \left[\frac{1-\gamma}{(1-\gamma/2)^2} \right] f. \tag{E}$$

For $\gamma = 0.5$ and $f = 1$, this ratio equals 16/18.

Examples. Trastuzumab is a monoclonal antibody against the Her2 receptor, which is overexpressed in 25% to 30% of breast cancers (8). A targeted randomized Phase III trial of standard chemotherapy with or without Trastuzumab

was conducted in 469 patients with metastatic breast cancer whose tumors overexpressed Her2 based on immunohistochemical analysis in a central laboratory. The results were highly statistically significant favoring the Trastuzumab arm with regard to several end points including 1-year survival rate (78% versus 67%). If we assume that the antibody is completely ineffective in assay negative patients ($\delta_0 = 0$) then equation B indicates that the ratio of number of randomized patients for untargeted versus targeted trial is ~ 16 to 1 ($1-\gamma = 0.25$, $f = 1$). A targeted trial of the size actually conducted provides 90% power for detecting a 9.6% improvement in the 1-year survival rate above a baseline of 67% with a two-sided 5% statistical significance level ($p_c = 0.67$, $\delta_1 = 0.096$). If the trial were untargeted then the formula for n in the Appendix with $p_c = 0.67$, $\delta_0 = 0$, $\delta_1 = 0.096$, $1-\gamma = 0.25$, and $p_e = p_c + \gamma\delta_0 + (1-\gamma)\delta_1$ indicates that 23,586 randomized patients would be required, because the overall treatment improvement in 1-year survival rate would be only 2.4%. If assay-negative patients also benefit from Trastuzumab, then the sample size for the untargeted trial would be reduced. For example, suppose the assay-negative patients benefit half as much as the assay-positive patients ($\delta_0 = 0.048$ and consequently $p_e = 0.67 + 0.06$). Then the overall treatment benefit for the untargeted trial is a 6% improvement in 1-year survival rate, and the untargeted trial requires only 1,256 total patients. This is still 2.67 times as many randomized patients as required for the targeted trial. Also, although immunohistochemical assays are not precise, it seems unlikely that the assay-negative patients would achieve half the benefit as the assay positive patients.

Gefitinib is a small molecule inhibitor of epidermal growth factor receptor (EGFR) kinase activity. Two untargeted Phase III trials of standard chemotherapy with or without gefitinib in 2,130 chemotherapy naïve patients with advanced non-small cell lung cancer failed to demonstrate any benefit of gefitinib (9, 10). Two reports indicated recently that response to gefitinib alone in patients treated previously can be predicted on the basis of somatic mutations in the tyrosine kinase domain of the *EGFR* gene (11, 12). In the report by Lynch *et al.* (11), 8 of 9 responders had such mutations compared with none of seven nonresponders. Only $\sim 10\%$ of non-small cell lung cancer patients have tumors with such mutations. Conducting an untargeted Phase III trial for a molecularly targeted drug with a $1-\gamma$ value of only 10% is almost a futile proposition, as noted by Dancey and Freidlin (13) if R- patients do not benefit. Even the 2,130 patients actually randomized is insufficient. For example, even if the 1-year survival rate is increased by an enormous 40% over the baseline level of 40% ($\delta_1 = 0.40$, $p_c = 0.40$) for patients whose tumors have the mutation, then the formula of the Appendix with $\delta_0 = 0$ and $1-\gamma = 0.10$ indicates that for 90% power and a 2-sided 5% significance level a nontargeted trial requires 3,248 randomized patients. If the size of the treatment benefit for patients with mutations is a more realistic 20% increase in 1-year survival rate, then the nontargeted design requires 12,806 patients. The formula in the Appendix for n^T indicates that the targeted design requires only 34 and 138 total randomized patients, respectively, in these two conditions. For gefitinib, the assumption $\delta_0 = 0$

seems reasonable based on the available data and on the accuracy of a genotype assay. At the time the Phase III trials were initiated, an accurate assay for predicting gefitinib activity was not available. Overexpression of EGFR did not correlate well with gefitinib response, and the relationship between mutations in the tyrosine kinase domain of the *EGFR* gene and response to gefitinib was not known. The relationship was discovered, however, based on materials collected in the Phase II trials. The Phase II response rates suggested that the proportion of patients responsive to gefitinib was very small, and consequently it may have been beneficial to attempt to characterize Phase II responders based on genotype or gene expression profile data before launching the Phase III trials.

DISCUSSION

On the basis of the modeling assumptions used here, the targeted clinical trial design often requires fewer randomized patients than the untargeted design. The degree of reduction depends heavily on the availability of an assay for identifying all patients who will benefit from the new treatment and the prevalence of such patients. When the new treatment benefits only a subset of patients and those patients can be accurately identified, then the targeted design can require many fewer randomized patients than the untargeted design. Under these conditions, the number of patients required for screening with the targeted design will be less than the number required for randomization with the untargeted design. When $<50\%$ of the patients are predicted to benefit from the experimental regimen, the untargeted design becomes impractically inefficient under the conditions described here.

When the experimental treatment has multiple pathways of effect or when the negative predictive value of the assay is inadequate, then the advantages of the targeted design are more limited. The targeted design may still require fewer randomized patients, and an estimate of its efficiency can be obtained from Fig. 1 or using equation C. Under these conditions, however, the targeted design may require more patients to be screened. The increment in the number of screened patients can be estimated from Fig. 2, A and B, or from equation E. Unless there are substantial savings in the required number of randomized patients, the targeted design may not be deemed worthwhile because of the need to screen patients and the potential restriction in applicability of the experimental treatment. When the experimental treatment is partially effective for the R- patients and the proportion of R+ patients is very low, both targeted and untargeted designs become very expensive. The untargeted design becomes expensive because a huge number of randomized patients is required, because the average treatment effect for all of the randomized patients becomes very small. The targeted design becomes expensive, because a large number of patients must be screened.

Several aspects of the model used here are idealizations. One limitation is our use of a binary outcome end point. Although many cancer clinical trials are based on survival or disease-free survival end points, the use of binary outcome here enables us to avoid technical complexities, which might mask the most essential issues. Our results highlight the potential

value of codevelopment with a therapeutic of an accurate assay for predicting responsive patients and the value of understanding the mechanisms of action of new therapeutics so that its potential realm of effectiveness can be efficiently established. Our evaluation presupposes the existence of Phase II data in unselected populations that enables the development or validation of an assay for identifying patients who are most likely to respond to the treatment. The development of predictively accurate assays is difficult, and limitations in the assay will decrease the potential efficiency gains that can be achieved from a targeted Phase III trial.

APPENDIX

The Ury and Fleiss expression for the sample size of the untargeted design is:

$$n = (n_d/4)[1 + \sqrt{2w + 1}]^2$$

where

$$n_d = [z_\alpha \sqrt{2pq} + z_\beta \sqrt{p_c(1-p_e) + p_c(1-p_c)}]^2 / \bar{\delta}^2,$$

$$\bar{\delta} = \gamma\delta_0 + (1-\gamma)\delta_1,$$

$$p_e = p_c + \bar{\delta},$$

$$\bar{p} = (p_c + p_e)/2,$$

$$\bar{q} = 1 - \bar{p}, \text{ and}$$

$$w = \bar{\delta}^2 / [(z_\alpha + z_\beta)^2 pq].$$

The constants z_α and z_β denote the 100 (1- α) and 100 (1- β) percentiles of the standard normal distribution.

For the targeted design we add the symbol T. The response probability in the experimental group is $p_e^T = p_c + \delta_t$ and the expression for the sample size becomes:

$$n^T = (n_d^T/4)[1 + \sqrt{1 + 2w_T}]^2$$

where,

$$n_d^T = [z_\alpha \sqrt{2\bar{p}\bar{q}_T} + z_\beta \sqrt{p_e^T(1-p_e^T) + p_c(1-p_c)}]^2 / \bar{\delta}_1^2,$$

$$\bar{p}_T = (p_e^T + p_c)/2,$$

$$\bar{q}_T = 1 - \bar{p}_T \text{ and}$$

$$w_T = \bar{\delta}_1^2 / [(z_\alpha + z_\beta)^2 \bar{p}_T \bar{q}_T]$$

The relative efficiency of the two designs with regard to number of randomized patients is therefore given by equation (A) with f defined by

$$f = \left[\frac{z_\alpha \sqrt{2pq} + z_\beta \sqrt{p_c(1-p_e) + p_c(1-p_c)}}{z_\alpha \sqrt{2\bar{p}_T \bar{q}_T} + z_\beta \sqrt{p_e^T(1-p_e^T) + p_c(1-p_c)}} \frac{1 + \sqrt{1 + 2w}}{1 + \sqrt{1 + 2w_T}} \right]^2.$$

REFERENCES

1. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene expression profiling and supervised machine learning. *Nat Med* 2002;8:68–74.
2. Rosenwald A, Wright G, Chan WC, et al. The Use of Molecular Profiling to predict survival after chemotherapy for diffuse large B-cell lymphoma. *N Eng J Med* 2002;346:1937–47.
3. Balis FM. Evolution of anticancer drug discovery and the role of cell-based screening. *J Natl Cancer Inst* 2002;94:78–9.
4. Grunwald V, Hidalgo M. Developing inhibitors of the epidermal growth factor receptor for cancer treatment. *J Natl Cancer Inst* 2003;95:851–67.
5. Fox E, Curt GA, Balis FM. Clinical trial design for target-based therapy. *Oncologist* 2002;7:401–9.
6. Ury HK, Fleiss JL. On approximate sample sizes for comparing two independent proportions with the use of Yates' correction. *Biometrics* 1980;36:347–51.
7. Casagrande JT, Pike MC. An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics* 1978;34:483–6.
8. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer that Overexpresses HER2. *N Engl J Med* 2001;344:783–92.
9. Giaccone G, Herbst RS, Manegold C, et al. Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: a phase III trial-INTACT 1. *J Clin Oncol* 2004;22:777–84.
10. Herbst RS, Giaccone G, Schiller JH, et al. Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: a phase III trial-INTACT2. *J Clin Oncol* 2004;22:785–94.
11. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129–39.
12. Paez GJ, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science (Wash DC)* 2004;1126:1–10.
13. Dancey JE, Freidlin B. Targeting epidermal growth factor receptor—are we missing the mark? *Lancet* 2003;362:62–4.