Clinical Trials

New challenges for 21st century clinical trials

Richard Simon

The online version of this article can be found at:

Additional services and information for *Clinical Trials* can be found at:

Email Alerts: http://ctj.sagepub.com/cgi/alerts

Subscriptions: http://ctj.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 17 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
http://ctj.sagepub.com/cgi/content/refs/4/2/167

New challenges for 21st century clinical trials

Richard Simon

The genomics and biotechnology revolutions sweeping biology will influence clinical trials in several important ways. First, they will help elucidate the molecular basis of diseases and thereby facilitate the development of more effective treatments. It will become increasingly clear that many of the entities currently treated in clinical trials are distinct at a molecular level and unlikely to be responsive to the same treatments. This is already clear for breast cancer and lung cancer. For example, breast tumors that do not express estrogen receptors are not responsive to treatments that block estrogen stimulation such as tamoxifen or aromatase inhibitors. Patients with breast cancers that are driven by amplified Her-2 gene experience a dramatic halving of the hazard of recurrence or death from treatment with antibodies that block the Her-2 receptor [1,19].

In cancer, progress has been limited because most treatments used have been non-specific DNA poisons that treat the symptom of tumor proliferation rather than the molecular basis of the proliferation [2]. Developing drugs that treat the underlying molecular basis of the diseases will require further elucidation of the oncogenic mutations that cause the diseases. Statisticians who learn sufficient genomics and biology have important roles in this discovery process [3]. The drugs that target these key mutations can be expected to be effective against subsets of our current heterogeneous diagnostic categories. If we continue to treat broad patient populations with the new generations of drugs, we may fail to recognize effective drugs because the overall effects will be very diluted. The current approach of treating broad populations of patients is based on an assumption that qualitative treatment-by-subset interactions are unlikely. However, increasing knowledge of tumor biology indicates that such qualitative interactions are highly likely. A positive drug effect for a subset and a zero drug effect for the complementary set of patients represents a qualitative interaction [4].

The new generation of drugs will need to be developed in conjunction with diagnostics using reproducible assays of pre-treatment specimens that identify the subset of patients likely to benefit from the drug [5]. Simon and Maitournam [6–8] have evaluated the relative efficiency of targeted randomized clinical trials compared with standard broad eligibility clinical trials. They compared the designs in terms of the required numbers of randomized and screened patients. The relative efficiencies depend on the sensitivity and specificity of the diagnostic assay and the prevalence of patients who are likely to benefit from the new treatment. For most reasonable assays, if the prevalence of sensitive patients is low, the number of patients randomized in the targeted design is much less than the number required for traditional broad clinical trials. The number of patients required to screen for the targeted design may be either less than or greater than the number of randomized patients needed for traditional designs, depending on assay performance characteristics. Simon and Zhao have developed and made available web-based software for performing those calculations for both binary response endpoints and for survival endpoints (http://linus.nci.nih.gov/~brb/samplesize).

Freidlin and Simon [9] have investigated a new adaptive randomized clinical trial design in which a genomic classifier of the patients likely to benefit from the new treatment relative to the control treatment is identified. Simultaneously, the trial is used in a statistically valid manner to test hypotheses about the treatment effect in that subset. In the adaptive design, the treatment effect for all randomized patients is tested using a 0.04 threshold for statistical significance. If the overall test is not significant, then the sensitive subset is identified using the first half of the patients and the treatment effect within that subset for patients accrued during the second half of the trial is evaluated using a threshold of significance of 0.01. Several variants of this design are possible, including extending accrual for the adaptively identified subset to achieve a larger

Biometric Research Branch, National Cancer Institute, 9000 Rockville Pike, Bethesda, MD 20892-7434, USA
Author for correspondence: Richard Simon, Biometic Research Branch, National Cancer Institute, 9000 Rockville Pike, Bethesda, MD 20892-7434, USA. E-mail: rsimon@nih.gov

sample size of the 'enriched' population. This type of analysis plan for allocating the type I error between an overall test of treatment effect for all randomized patients and for a test for a pre-specified subset is also discussed by Simon and Wang [10].

Statisticians have taught clinicians to distrust subset analysis, particularly if the treatment effect was not significant for the overall population of randomized patients. This conventional wisdom was certainly good advice for the post-hoc data dredging type of subset analysis. But it is not good advice for prospectively planned subset analysis using an analysis strategy that preserves the experimentwise type I error. The conventional wisdom not to trust a subset analysis unless the overall analysis is significant is often not biologically sound when the pre-specified subset of patients likely to be sensitive to the new treatment is based on a diagnostic linked to the molecular target of the new treatment. The conventional wisdom is also not appropriate, even when the pre-specified subset is based on an empirical multi-gene characterization of the sensitive patients. The criteria developed by the U.S. Food and Drug Administration for 'validated biomarkers' should not be applied to predictive classifiers of the subset of patients likely to benefit from a new drug [11]. Those criteria require that the biology of the disease is sufficiently well understood that there is compelling evidence that biomarker level reflects disease activity. They were developed for biomarkers proposed as surrogates for clinical benefit. The concept of 'validity' of a biomarker has meaning only in the sense of being 'fit for purpose', however, and the purpose of biological measurements used for treatment selection is completely different from the purpose of a biological measurement used as a surrogate of patient benefit. Unfortunately, there is currently extensive confusion about the use of the term 'biomarker' and in the concept of validation. This confusion threatens to enmesh the development and utilization of predictive classifiers in the regulatory complexities of surrogate endpoints. The proper validation of a surrogate for clinical benefit for a class of drugs requires a series of randomized clinical trials in which the candidate surrogate and clinical benefit are measured, with the demonstration that treatment differences with regard to the two measures are concordant [12–14]. This is very demanding to accomplish, and generally it is more expedient to perform the phase III evaluation of a new treatment using a direct measure of clinical benefit as endpoint.

There is a concern about approval of drugs for a defined subset of patients in the event that the treatment might subsequently be used more broadly by practitioners. This concern has in the past been the basis for broad eligibility trials. Demonstrating effectiveness of a targeted cancer drug in a subset of patients that is considered to represent a diagnostically distinct sub-category of disease should not be taken as evidence that the drug would be effective for other patients. Hence, the labeling indication for use of such a drug should be restricted to the defined subset on which it was tested and found effective. However, failure to test the drug in other diagnostic subsets of the disease should not necessarily be grounds for failing to make the drug available in the subset for which it was shown to be effective. We should treat molecular diagnoses, not traditional symptomatic disease categories [5]. In cases in which the diagnostic test is linked to the intended molecular target of the drug, it may not be ethically justifiable to treat 'classifier-negative' patients with the drug. However, this may require the U.S. Food and Drug Administration to re-consider its interpretation of regulations for licensing of diagnostics. If it can be demonstrated that the classifier can be measured reproducibly and that it defines a set of patients for whom a new drug is effective, then it is difficult to argue that the drug should not be approved because physicians may use the drug in classifier-negative patients, or because a sponsor has failed to test the drug in classifier-negative patients. The situation is somewhat different from the licensing of a diagnostic for use in guiding treatment decisions for a regimen already in broad use [15].

In using a predictive classifier to target a clinical trial or an analysis plan, it is essential that the data used to develop the predictive classifier be distinct from the data used to test the new treatment in the subset determined by the classifier. The process of developing the classifier may be subjective, incorporating biological knowledge, assay measurement considerations as well as considering a variety of algorithmic feature selection and classifier-type comparisons. It is not appropriate to standardize or regulate this process. However, evaluation of a new treatment in a subset determined by a classifier should not be exploratory or subjective. It should generally involve a randomized clinical trial of the new treatment versus control for the classifier-positive patients, or a randomized trial in which entry is not restricted to classifier-positive patients, but in which a specific analysis plan involving the classifier is defined and the experimentwise error is preserved. It is not sufficient to merely state that the trial will be 'stratified' by the predictive classifier. A specific analysis plan should be specified in the protocol. The field must move away from a passive mode of using phase III trials for repeatedly developing new classifiers, tweaking existing classifiers and re-evaluating the components of classifiers. Phase III clinical trials should move toward a

prospective mode of evaluating treatment effect in a subset determined by a completely specified previously developed classifier, or in the classifier-positive and-negative patients using a pre-defined analysis plan with experimentwise type I error preserved [16–18].

For disease settings where qualitative interactions are likely, dismissing a prospectively specified subset analysis unless the overall effect is significant is not sound either biologically or statistically. It is true that broadly active drugs that do not require expensive companion diagnostics are preferable. But this ideal may not be attainable in many diseases. In cancer, the cost of the diagnostic is likely to be relatively small compared with the cumulative cost of the targeted drug. Statisticians must strive to design clinical trials that are biologically sound, rather than out of concern that the drug might be misused. Insisting on evaluation of drugs for broad populations without examining the effects in pre-defined subsets may perpetuate small treatment effects, marginal benefit to adverse event ratios, inconsistency of results across trials and the need for large trials. This type of direction is unsustainable, either economically or intellectually, in 21st century science. It would deny statisticians the ability to claim that the randomized clinical trial represents good science and a foundation for evidence-based medicine. It would also deprive our societies of the immense potential of the genomic and biotechnology revolutions to develop predictive medicines that are highly effective for diseases defined at the molecular level.

The genomic and biotechnology revolutions present us with powerful tools for improving the health of patients. Randomized clinical trials should continue to play an essential role in the evaluation of new drugs. However, statisticians face important challenges in moving from an inference posture of 'retrospective correlation' to one that brings about reliable predictive medicine. The creation of effective predictive medicine based upon patient genetics and disease genomics is an achievable goal and offers many benefits to patients and society. Meeting this challenge will require statisticians to develop new statistical designs, new analysis methods, new conventional wisdom and new levels of knowledge of genomics, biotechnology and disease biology. This will also require changes to the education of biostatisticians, and new partnerships among academia, industry and government.

References

1. **Joensuu H, Kellokumpu-Lehtinen PL, Bono P** *et al.* Adjuvant docetaxel or vinorelebine with or without trastuzimab for breast cancer. *N Engl J Med* 2006; **354**: 809–20.
2. **Simon R.** Targets for treatment success. *Nature Clin Pract Oncol* 2006; **3**: 1.
3. **Simon R.** Bioinformatics in cancer therapeutics-hype or hope? *Nature Clinic Pract Oncol* 2005; **2**: 223.
4. **Gail M, Simon R.** Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* 1985; **41**: 361.
5. **Papadopoulos N, Kinzler KW, Vogelstein B.** The role of companion diagnostics in the development and use of mutation-targeted cancer therapies. *Nature Biotechnol* 2006; **24**: 985–95.
6. **Maitournam A, Simon R.** On the efficiency of targeted clinical trials. *Stat Med* 2005; **24**: 329–39.
7. **Simon R, Maitournam A.** Evaluating the efficiency of targeted designs for randomized clinical trials: Supplement and Correction. *Clin Cancer Res* 2006; **12**: 3229.
8. **Simon R, Maitournam A.** Evaluating the efficiency of targeted designs for randomized clinical trials. *Clin Cancer Res* 2005; **10**: 6759–63.
9. **Freidlin B, Simon R.** Adaptive signature design: An adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin Cancer Res* 2005; **11**: 7872–78.
10. **Simon R, Wang SJ.** Use of genomic signatures in therapeutics development. *Pharmacogen J* 2006; **6**: 166–73.
11. **FDA.** Guidance for industry: Pharmacogenomics data submissions. Rockville MD, Food and Drug Administration, U.S. Department of Health and Human Services.
12. **Burkowski T, Molenberghs G, Buyse M, Geys H, Renard D.** Validation of surrogate endpoints in multiple randomized clinical trials with failure time end points. *J Royal Stat Soc C* 2001; **50**: 405.
13. **Korn EL, Albert PS, McShane LM.** Assessing surrogates as trial endpoints using mixed models. *Stat Med* 2004; **24**:163–82.
14. **Torri V, Simon R, Russek-Cohen E, Midthune D, Freidman M.** Relationship of response and survival in advanced ovarian cancer patients treated with chemotherapy. *J Natl Cancer Inst* 1992; **84**: 407.
15. **Simon R.** When is a genomic classifier ready for prime time? *Nature Clin Pract: Oncol* 2004; **1**: 2–3.
16. **Simon R.** A roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 2005; **23**: 7332–41.
17. **Simon R.** Development and validation of therapeutically relevant multi-gene biomarker classifiers. *J Natl Cancer Inst* 2005; **97**: 866–67.
18. **Simon R.** Development and evaluation of therapeutically relevant predictive classifiers using gene expression profiling. *J Natl Cancer Inst* 2006; **98**: 1169–71.
19. **Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE, Davidson NE, Al E.** Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 2005; **353**: 1673–84.