

The Development of Novel Agents for the Treatment of Colorectal Cancer

A Critical Review of Current Practice and Some Suggestions for the Future

John L. Marshall, MD, and Edmund A. Gehan, PhD

Dr. Marshall is Chief of the Division of Hematology and Oncology at Georgetown University's Lombardi Comprehensive Cancer Center, where Dr. Gehan is Professor Emeritus of Biostatistics in the Department of Biostatistics, Bioinformatics, and Biomathematics.

Address correspondence to:
John L. Marshall, MD
Division of Hematology and Oncology, Lombardi Comprehensive Cancer Center; 3800 Reservoir Road NW; Washington, DC 20057; Phone: 202-444-7064; Fax: 202-444-1229; E-mail: marshallj@georgetown.edu.

Abstract: The pathways to approval of new therapeutic agents in the United States and globally rely on the performance of well-designed clinical trials demonstrating both safety and efficacy. We recognize that the last decade has seen some great successes in improving outcomes for patients with cancer and, specifically, for patients with colorectal cancer. The development of novel agents active in colon cancer has led to improved survival and cure rates. However, because of the number of agents now available and the established practice patterns, it is becoming increasingly difficult to test new agents in cancer, particularly in colorectal cancer. The focus of this article is to review the current clinical trial designs with a critical eye and propose novel approaches to bringing new agents into the armamentarium of agents effective against colorectal cancer. Our current standards for drug development are increasingly problematic, and it is imperative that we develop new expectations and supporting standards in cancer drug development.

The problems we face in clinical research today are numerous. First, patient participation in oncology clinical research is an embarrassing 3% and decreasing. Second, the regulatory burden on investigators makes the performance of clinical research increasingly difficult, expensive, and time-consuming. Third, there is little incentive for either patients or physicians to participate in clinical research. Fourth, for reasons not entirely clear, current standards of care tend to be accepted even when they are far from optimal. Fifth, we increasingly rely on very large randomized trials designed to detect very small differences between treatment groups. Although we have made small gains in cancer therapy in the last decade, we must recognize the shortcomings of our traditional approaches. Large trials have become prohibitively expensive. We generally fail to identify the patients who actually benefit from specific treatments, resulting in overtreatment of patients and unnecessary exposure of a high proportion of patients to toxic treatments.

Our successes have been, in fact, relatively small. Although we have boasted about our success in colorectal cancer for the last 5 years, the reality is that we have increased median survival by approximately

Keywords

Colorectal cancer, clinical trial design, targeted therapy

12–15 months in the metastatic setting and have improved progression-free survival (PFS) at 3 years in the adjuvant setting by approximately 5%. Although this is better than no advance at all, colon cancer remains a very deadly disease and our treatments are still not very effective.

The Time is Right for a Change

Our improved understanding of cancer biology and recognition of the potential for individualized therapeutic decisions makes now the perfect time to reconsider our clinical research infrastructure and clinical trial designs. It is imperative we modify our methodology in order to achieve significantly greater and more rapid benefits for patients. Unfortunately, our current practice fails to stimulate these developments because physicians and industry benefit financially from the treatment of all patients, supporting continued broad, empiric treatment decisions. We must, therefore, shift our strategy to individualized treatments so that the subsequent reduction in the cost of drug development and increase in efficacy will have positive consequences in the world of cancer medicine. With the high price of empiric therapy, it would not be a surprise if payers begin to force this issue upon us.

The Current Standards of Clinical Research

To begin, we will briefly review traditional cancer drug trial design. Phase I clinical trials are typically performed in patients with a broad spectrum of refractory advanced cancers. No clinical activity is being sought or is required to select agents to continue to phase II. Biologic endpoints are an afterthought, with the primary focus being safety and pharmacokinetics. Next, phase II single-agent or drug combination trials are performed using response rate as a primary trigger to justify carrying treatments further to phase III trial designs. Our “bar” in phase II has been very low, with agents that demonstrate only minor clinical activity being viewed as “positive” and advanced to phase III with little more biologic support. In phase III, typically, the current standard of care is compared to an investigational arm, often the standard of care plus the new agent. The trials are generally very large in size and are designed to detect very small differences. The investment of patients and money is not in proportion to the clinical gains being sought. Viewed objectively, our “successes” represent very small actual clinical gains.

As an example, there are currently two phase III studies in colorectal cancer testing essentially the same hypothesis—whether bevacizumab (Avastin, Genentech) added to FOLFOX (5-fluorouracil [5-FU], leucovorin, and oxaliplatin) regimen improves 3-year disease-free survival (DFS) by 5%. Each of these studies involves

thousands of patients and has an expected duration of at least 5 years before a final analysis. The ECOG 5202 study will accrue 625 patients per year for 5.5 years (a total of 3,610 patients to allow for some dropouts) so that 1,375 high-risk stage II patients can be randomized to FOLFOX treatment (control group) or bevacizumab plus FOLFOX (experimental treatment) to determine with high statistical power (88%) whether the 3-year DFS is improved from an expected 80% for FOLFOX treatment to 85% for bevacizumab plus FOLFOX. Over 2,200 low-risk stage II patients “will be followed under observation” and not test any hypothesis. A final analysis is projected at 8.5 years after the start of the study. This will be in 2014, by which time, 320,000 stage II colon cancer patients will have been diagnosed in the United States alone. What should we be telling our patients in the meantime?

The NSABP C-08 study will accrue 2,632 stage II and III patients in 2.5 years and randomize them to the same treatments as in ECOG 5202 to determine with high statistical power (90%) whether the 3-year DFS is improved from an estimated 76.5% for the control arm to 81.8% for the arm with bevacizumab. There will be interim analyses for efficacy and safety, but the expected time for final analysis is 5 years for DFS and 7 years for survival. Both of these phase III studies involve an enormous allocation of resources in time, efforts for patients and investigators, and money to determine whether there is a modest improvement in DFS for patients receiving bevacizumab.

Using similar trials, we have claimed great success. In fact, we have increased median survival by only 1 year at the cost of toxicity, impairment in quality of life, and significant financial burden on patients and healthcare institutions. A recent evaluation has suggested that the cost for the treatment of colorectal cancer has surpassed our normal standards for life-prolonging procedures.¹ In addition, we are continuing to treat patients empirically in a trial-and-error fashion, exposing them to cytotoxic agents while knowing that a significant proportion will not benefit from that treatment. Currently, we are not successfully accruing patients to such studies and the enthusiasm for adding ever more agents to our current regimens is diminishing.

In refractory metastatic colorectal cancer, clearly, traditional trial designs are still appropriate. Here, response rates and PFS are valid trial endpoints, which are obtained relatively rapidly in moderate-sized studies. This is particularly true of the highly refractory patient; however such patients are unlikely to demonstrate a high response rate. The concept of using phase II randomized designs in this patient population is increasingly exciting and should become a standard as we move forward in the refractory

setting. But we should raise our expectations and seek higher response rates, a better understanding of target expression, and more refined combinations of agents. This effort will require substantial changes in our drug development culture, including collaborations between pharmaceutical companies, increased acceptance by patients and payers, and changes in regulatory standards.

New Windows to Test New Agents

Our “success” in colon cancer has generated two novel indications where new agents could be tested, either as single agents or in combination, requiring fewer patients and therefore quicker evaluations of the agents. These “windows” are ideal for smaller trials with higher expectations. The first is in patients who have responded to initial chemotherapy for colorectal cancer but do not have a curative option such as surgical resection. The current standard here is changing. Traditionally, patients continued to receive chemotherapy until their disease progressed. The neurotoxicity seen with repeated doses of oxaliplatin, however, has forced us to develop novel strategies for these patients. One approach would be to “lighten up” on the chemotherapy and use some form of maintenance therapy. Alternatively, two recent studies suggest that chemotherapy-free intervals have no negative impact on patient survival.^{2,3} These two studies open a new window in which to test agents: using patients who have responded to initial treatment with standard chemotherapy, stop the first-line agents, and test the investigational agent(s) using either PFS or subsequent response as endpoints.

An example of such a trial design is a phase II study in metastatic colorectal cancer with the primary objective of evaluating a novel therapy versus placebo in patients who have had clinical benefit with initial treatment (complete or partial response or stable disease). Patients are stratified by type of response and randomized to receive the therapy or placebo to determine whether the PFS for patients receiving the therapy is significantly greater than that for patients receiving the control treatment. The primary endpoint is the percentage of patients having PFS times of 12 months or longer and it is assumed that PFS time is exponentially distributed with a median PFS time in the control group of 4 months. It has been further assumed that patients can be recruited to be studied over a period of 18 months and followed for an additional 12 months, so that the projected time for the final analysis is 30 months after the start of the randomized study. As this is a phase II study, the false-positive error (alpha) has been set at 10% (one-sided) and the statistical power at 90%. The sample size for such a study depends on the projected improvement in median PFS time in the group receiving the new therapy. Table 1 shows the required total sample

Table 1. Total Number of Randomized Patients Required for Phase II Study of Novel Therapy Versus Control, Comparing Percentages of Patients Progression-Free and Surviving 12 Months or More After Randomization

Randomization (experimental: control)	Number of patients in PFS*			
	At 5 months	At 6 months	At 7 months	At 8 months
1:1	830	250	132	88
2:1	963	291	153	102

Statistical characteristics of trial: statistical power of 90% and false positive (alpha) error of 10% (one-sided).

*Median PFS time in Control Group= 4 months.

sizes when the median PFS time in the experimental group is 5, 6, 7, or 8 months.

For a projected median PFS time in the experimental group of 8 months, the total sample size needed for a 1:1 randomization would be 88 patients (44 patients per group) and 102 patients for a 2:1 randomization (68 patients in experimental group, 34 patients in control group). The latter type of randomization, though requiring more patients, would yield more precise characterization of the PFS experience in the experimental treatment group while maintaining the statistical characteristics of the trial. If the search for new therapies is directed toward those projected to provide substantial benefit to patients, the sample sizes in the table demonstrate that efficient studies can be conducted without requiring sample sizes in the multiple thousands. New therapies confirmed via this approach could be further evaluated in extended phase II studies in which all patients receive the experimental treatment, possibly leading to faster regulatory approvals. Clearly, greater benefit will be recognized more favorably at a regulatory level. As can be seen in Table 1, agents can be tested to determine their single-agent effect (even against a placebo) using as few as 88 patients in frontline metastatic colorectal cancer. Of course, the total number of patients entering the trial will be larger, because only the patients with clinical benefit (complete or partial response or stable disease) will be eligible for the maintenance portion of the study.

The second new “window” is in patients with metastatic colon cancer whose metastatic lesions have been resected and who are now without evidence of disease. Because of differences in the number and location of metastases as well as the techniques used to remove/ablate lesions, this growing patient population is clinically highly variable. Despite the variability, this group is recognized as being at high risk for recurrent disease, making it ideal

for testing new agents in a high-risk adjuvant (or neo-adjuvant) setting. Given that we do not understand the role of postoperative chemotherapy in patients who have had preoperative chemotherapy, this is another window where agents could be tested. A series of phase II studies could be conducted using PFS time as a primary endpoint and the objective of identifying promising new agents for phase III studies of adjuvant therapies in stages II and III. Effective new agents in the phase II studies could be added to the current standard therapy (FOLFOX) and tested against standard therapy alone in randomized phase III studies. For example, a phase II study of a new agent could be designed to determine whether the percentage of metastatic patients with PFS at 6 months after complete resection of lesions increases from 40% (null hypothesis) to 60% (alternative hypothesis). An optimal design of the Simon type⁴ would require a maximum of 46 patients using false positive and false negative error rates both equal to 10%. The study could stop after accrual of only 18 patients if the percentage of patients progression-free at 6 months were 39% (7/18) or less. Conducting new agent studies in a series of phase II studies of this type would utilize a moderate number of patients and permit the identification of promising new agents quickly.

In patients with metastatic colorectal cancer, there are multiple endpoints that could signal a promising new therapy: high response rate, high percentage of responding patients who have a long period of PFS, and/or a high percentage of patients for whom a molecular target is affected. Conversely, another endpoint could be the rate of adverse outcomes, and an unacceptably high rate would indicate that the agent is unsafe for patients. Clinically, it is important to define multiple endpoints of efficacy and toxicity, so that successful outcomes would predict whether new agents or combinations are likely to be effective in the adjuvant setting. Fortunately, biostatisticians have considered methodology for monitoring phase II trials with multiple endpoints—usually efficacy and toxicity, though extensions could be made to multiple efficacy and toxicity endpoints. Thall and colleagues have offered a general Bayesian strategy for monitoring multiple outcomes in single-arm phase II trials.^{5,6} Each patient's outcome is characterized by a multinomial variable that defines a specific combination of events (efficacy and toxicity) that occur during the trial. The events could occur at differing time points during the study. Bayesian criteria for the termination of studies with too-low rates of efficacy or too-high toxicity outcomes are defined. The stopping rule for each endpoint is determined in order to achieve a specified level of improvement in the efficacy endpoint for the experimental treatment relative to a benchmark rate assumed for standard therapy and to control the rate of adverse outcomes. Phase II trials con-

ducted using a strategy of this type would utilize patients efficiently and allow multiple endpoints to be evaluated for a new therapy. The sample size would not be fixed with this strategy, though a maximum could be chosen. The actual sample size would depend on the definition of the multiple stopping rules and the outcomes in the trial.

Using Molecular Targets in Adjuvant Therapy

Our current standard of care is to offer patients with stage III disease 6 months of adjuvant oxaliplatin/5-FU-based chemotherapy. In stage II disease, our current recommendations are to have a discussion with patients about the relative risks and benefits and offer patients 5-FU/oxaliplatin-based chemotherapy. In stage III disease we are essentially treating a large group of patients knowing that a relatively small group will benefit. Specifically, of 100 patients treated for stage III disease, only approximately 25 are expected to benefit from the chemotherapy. Therefore approximately 75 of every 100 patients are being exposed to chemotherapy with no benefit. The numbers in stage II disease are even more striking, with only approximately 3–6% of all patients treated with postoperative adjuvant chemotherapy benefiting. This overtreatment cannot be taken lightly. Not only are the toxicities and costs significant, but small fractions of patients die from the adjuvant therapy itself.

Currently, to test a new agent in the adjuvant setting, huge clinical trials must be performed, randomizing patients between the standard of care and the new treatment, and a very long period of time is required before the data can be validated. Trial size is of course inversely related to the desired improvement in outcome. Our acceptance of large trials designed to detect small differences is a reflection of our complacency with the current system. Our most urgent need is change in the design of adjuvant clinical trials. Ideally we should identify patients who are at high risk for recurrent disease and treat only these patients. Molecularly based tests are being developed that we hope will enable us to identify the patients who are at high risk for recurrent disease. If we could reliably identify this patient population, the size of clinical trials, as well as the time-frame for follow-up, would be dramatically reduced.

To demonstrate this, we propose a trial design utilizing a genetic test that enables us to identify the high-risk patient. Genomic technologies such as DNA microarray expression profiling are providing biomarker signatures that may predict which patients are most likely to respond to a new treatment. Drugs targeted to molecular outcomes are only likely to be effective in patients whose tumors express the target. Simon and Maitournam refer to targeted clinical trial designs as those that are tailored for patients pre-

Table 2. Number of Patients in Experimental and Control Groups, and Statistical Power of Targeted and Untargeted Studies, Required for Varying Percentages of Patients Who Express the Target

Patients Expressing Target (%)	Targeted study			Untargeted Study		
	Experimental	Control	Power (%)	Experimental	Control	Power (%)
100	50	50	92.2	50	50	92.2
80	40	40	86.3	50	50	79.2
60	30	30	76.7	50	50	58.4
40	20	20	61.5	50	50	34.8
20	10	10	39.2	50	50	15.7

Test is of difference in response rates between experimental and control groups at statistical significance level of 5% (one-sided test).

dicted to respond to therapy.⁷ They studied the efficiency of targeted designs compared with traditional phase III randomized designs with broader eligibility criteria. For binary outcomes (eg, response rate, percentage of patients progression-free at a particular time), they considered an experimental versus a control therapy and compared the number of randomized patients required for the untargeted design (with broad entry criteria) to the number of screened patients needed for the targeted design (with eligibility limited to patients passing the screen). For the targeted design, they assumed that a reliable screening assay was available that could predict whether the patient was more likely to respond to experimental than control treatment. Under their assumptions, a targeted design will always require fewer patients than an untargeted design when fewer than 50% of patients are expected to benefit from the experimental regimen. Though the number of randomized patients for the targeted design may be less than that for the untargeted design, the number of patients needed to be screened for the targeted design should be considered. The targeted design may not be worthwhile if a very large number of patients needs to be screened to determine those patients likely to benefit and if there would also be restricted applicability of the experimental treatment.

The savings in number of patients for targeted designs can be dramatic, as illustrated by a study of trastuzumab in metastatic breast cancer. Trastuzumab (Herceptin, Genentech) is a monoclonal antibody against the HER2 receptor, which is overexpressed in 25–30% of breast cancers. A targeted, randomized phase III study of standard chemotherapy with or without trastuzumab in 469 patients whose tumors overexpressed HER2 demonstrated highly significant results in 1-year survival rates (78% vs 67%).⁸ If it is assumed that the antibody is ineffective in assay-negative patients, then the ratio of the number of

patients needed for the untargeted design relative to the targeted design is approximately 16:1, mainly because the improvement in 1-year survival would be only 2.4% in the untargeted design.

The advantages of targeted clinical trial designs over untargeted designs depend on having a reliable assay that demonstrates which patients express the target, the percentage of patients expressing the target, and the relative response rates in patients expressing versus not expressing the target. Suppose we assume that 100 patients are screened and there are varying percentages of patients that express the target, as shown in Table 2. Suppose further that the response rate to experimental treatment is 60% in patients expressing the target and only 30% in patients not expressing the target (the same as the response rate to control treatment). Table 2 compares the statistical power of targeted versus untargeted studies in this circumstance.

The sample size is always 50 patients per group in the untargeted study, whereas the targeted study will be conducted in a smaller number of patients expressing the target. The statistical power is less for the untargeted study except when the percent of patients expressing the target is 100%. The difference in power favoring the targeted study increases as the percentage of patients expressing the target decreases. As an example, when 60% of patients express the target, the statistical power is 76.7% for the targeted study and only 58.4% for the untargeted study. Though the sample size is larger in the untargeted study, the expected response rate (48%) for the experimental treatment is only 18% larger than that for the control treatment (30%). In the targeted study, the expected difference in response rates between the experimental (60%) and control (30%) treatment would be 30%.

When a reliable assay is not available for determining the patients who express the molecular target, Friedlin and

Simon propose an adaptive clinical trial design for developing and prospectively testing a gene expression signature that identifies sensitive patients.⁹ Though their design is described in the context of DNA microarray expression profiling, it can be easily adapted to using single nucleotide polymorphism gene typing or proteomic profiling. The proposed design has two stages, a first stage in which a classifier identifying sensitive patients is developed and a second stage in which the treatment effect is evaluated in patients identified as sensitive in the first stage. The clinical trial is designed to accrue a total of N patients, n_1 in stage 1 and n_2 in stage 2 ($n_2=N-n_1$). A key characteristic of the design is the development of a classifier based on logistic regression analysis that predicts whether a patient is more likely to benefit from the experimental (E) treatment than the control (C) treatment. The classifier does not restrict entry of patients in stage 2, but is utilized prospectively to identify the subset of patients likely to be sensitive to the targeted therapy.

The final analysis is carried out by comparing E to C in all N patients (to detect a possible overall treatment effect) and also in the subset of sensitive patients accrued in stage 2 (to detect a possible targeted treatment effect). The authors recommend that the overall statistical significance level for the clinical trial be .05, using .04 for the significance level of the statistical test comparing E to C in all N patients and using .01 for the statistical test comparing E to C in the subset of sensitive patients. If E is substantially more effective than C in the sensitive patients, the study should still have good statistical power.

In simulation studies based on a trial of 400 patients ($n_1=n_2=200$), Friedlin and Simon demonstrate that when only 10% of the eligible patient population have a sensitive signature, the adaptive trial approach would have much higher statistical power than the traditional approach, mainly because of the cases where the E versus C comparison was statistically significant in the sensitive subset of patients from stage 2. As the percentage of sensitive patients increases to 25% or more, the advantages of the adaptive design over the traditional design in statistical power diminish. In the situation where the experimental treatment effect is present for all patients, the test of E versus C in all patients will still have good statistical power because the significance level is .04. Hence, the proposed design has the advantage of permitting a patient classifier to be developed in stage 1 and good statistical power for detecting an experimental treatment effect when the proportion of sensitive patients is low to moderate, without compromising the ability to detect an overall treatment effect when the percentage of sensitive patients is high.

A challenge in colorectal cancer is to find a reliable assay for predicting which patients are likely to benefit from a new therapy and incorporating the assay into targeted designs of future adjuvant therapy trials. It is hoped that the comparison of a traditional clinical trial and its cost and duration compared to a novel targeted approach would be an incentive for us to invest heavily in identifying a valid gene or molecular biology test that would identify the high-risk patients and expose only that group to chemotherapy.

Conclusion

The time is right to redirect our clinical research efforts away from traditional trial designs to those that incorporate molecular targeting. More importantly, we must shift away from the acceptance of small gains as meaningful and focus on larger, more meaningful gains. If we do, more agents will make it to the market faster at a lower cost. Fewer patients will be exposed empirically to agents that will not work for them. Most importantly, outcomes for patients with cancer will improve. We must break down the perceived and/or real barriers as we move forward in the quest. We must meet the urgent needs of our patients, and their families who are looking to us for answers. Once we all do that, the pathway to the future is much clearer and much more urgent.

References

1. Yu-Ning Wong Y, Meropol NJ, Sargent D, Goldberg R, Beck JR. Direct cost survival analysis of treatment of metastatic colorectal cancer. *J Clin Oncol*. 2006;25(18S pt 1): Abstract 3515.
2. Labianca R, Floriani I, Cortesi E, et al. Alternating versus continuous FOLFIRI in advanced colorectal cancer: a randomized GISCAD trial. *J Clin Oncol*. 2006;25(18S pt 1): Abstract 3505.
3. Maindrault-Göebel F, Lledo G, Chibaudel B, et al. OPTIMO2, a large randomized phase II study of maintenance therapy or chemotherapy-free intervals after FOLFOX in patients with metastatic colorectal cancer: a GERCOR study. *J Clin Oncol*. 2006;25(18S pt 1): Abstract 3504.
4. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials*. 1989;10:1-10.
5. Thall PF, Simon R, Estey E. Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Stat Med*. 1995;14:357-379.
6. Thall PF, Simon R. New statistical strategy for monitoring safety and efficacy in single-arm clinical trials. *J Clin Oncol*. 1996;14:296-303.
7. Simon R, Maitournam A. Evaluating the efficiency of targeted designs for randomized clinical trials. *Clinical Cancer Res*. 2004;10:6759-6763.
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;344:783-792.
9. Friedlin 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;11:7872-7878.