

Clinical Trial Designs for Therapeutic Vaccine Studies

Richard Simon, D.Sc.

Biometric Research Branch

National Cancer Institute

Richard Simon, D.Sc.
Chief, Biometric Research Branch
National Cancer Institute
6130 Executive Blvd, Room 8134
Rockville MD 20852-7434
phone (301) 496-0975
fax (301) 402-0560
e-mail rsimon@nih.gov

Cancer vaccines are different in many ways from cytotoxic drugs. Consequently some of the paradigms for the early clinical development of cytotoxics are not applicable to the development of therapeutic vaccines. In contrast, many of the principles of phase III clinical trials are applicable to cancer vaccines. In this chapter we will describe design strategies for the efficient early clinical development of cancer vaccines.

1.0 Phase I Clinical Trials

The major objective of phase I trials of cytotoxics is determining a dose of the drug or regimen that is safe for use in subsequent clinical studies. For cytotoxics it is assumed that higher doses have greater anti-tumor effects and so phase I trials are usually designed to estimate the maximal tolerated dose.

Tumor vaccines are often based on DNA constructs, viral vectors and cytokines that have been determined as safe from previous clinical trials. Peptide vaccines generally seem inherently safe so long as the cytokine adjuvants are used in combinations and doses previously demonstrated to be safe. For example, peptide vaccines based on nonmutated melanoma antigens such as MART-1/Melan A and gp100 were initially evaluated in a phase I trial at the National Cancer Institute at doses ranging from 0.1 to 10 mg. However, no toxicity was encountered even at the highest dose (1,2).

On the other hand, initial clinical experience with a novel virus or plasmid as a recombinant vaccine vector should be conducted in a Phase I setting. However, if such vectors are proven to be nontoxic even at substantial doses, then subsequent trials using the same vectors but with different recombinant inserts may not require dose escalation studies.

For cancer vaccines, it is not always the case that more is better. For example, in studies of peptide vaccines based on non-mutated melanoma antigens, in vitro analysis did not reveal any correlation between peptide dose and the generation of specific T cell reactivity from the peripheral blood lymphocytes of vaccinated patients (1,2). Thus, for subsequent trials using similar peptides, an intermediate fixed dose of 1 mg was chosen for vaccination, bypassing repetitive phase I studies.

Feasibility issues limit the maximum doses of certain recombinant proteins, viruses or whole tumor cell vaccines that can be produced for administration to patients. In many cases, the dose selected will be based on pre-clinical findings or on practical considerations.

Dose ranging to find the minimal active dose may be feasible but the 3-6 patients per dose level used in conventional toxicity trials may not be adequate. Those small sample sizes are only sufficient to exclude high toxicity rates. Suppose that an assay is used in a

binary manner to define immunogenic response. Table 1 shows the probability of no immunogenic responses in n patients as a function of the true immunogenic response probability. If one wants a dose at which the immunogenic response probability is at least 30% say, then if you observe no immunogenic responses in 7 patients it would be appropriate to escalate to the next dose level. Simon's optimal two-stage phase II designs to be described below, or other phase II designs can be used more generally to distinguish a response probability of less than p_0 (e.g. 0.05) from one greater than p_1 (e.g. 0.30). But unless p_1 is much greater than p_0 , the required number of patients will much larger than for cytotoxic phase 1 trials (see Table 2).

Trying to determine whether there is a dose-response relationship involves comparing immunological responses for different dose levels. Such trials, if designed properly, require larger sample sizes. For example suppose one wishes to plan a study of two dose levels and test whether there is a relationship between dose and immunologic response. If the immunologic response probabilities at the two dose levels are 50% and 90%, then 20 patients treated at each dose level are required for a one-sided statistical significance level of 0.10 and a statistical power of 0.90 (3). Larger sample sizes are required to detect smaller differences. Using more than two dose levels allows one to treat somewhat fewer patients at each dose level, but the total number of patients required to detect a dose-response relationship will actually be much larger than if only two dose levels are tested. This is because the two most extreme dose groups are the most informative for detecting a dose-response relationship.

Trying to characterize the shape of the dose-activity relationship or finding an optimum biologic dose is an even more ambitious objective that is rarely practical in a phase I tumor vaccine study.

Phase I trials of cytotoxics are generally conducted in end-stage patients for whom all established therapy has been exhausted. End-stage patients without intact immune systems may have very little likelihood of benefit or toxicity from a tumor vaccine, however. In some cases the potential toxicity of the regimen may be based on immune stimulation and will not be seen in anergic patients. Hence, such patients contribute little information about potential toxicities for individuals with intact immune systems.

2.0 Phase II Clinical Trials

The objectives of the phase II vaccine trial are similar to those of the phase II cytotoxic trial. The primary objective is to determine whether the regimen has biologic activity that is likely to translate into patient benefit. The second objective is to optimize the regimen so that the biologic activity is likely to translate into patient benefit in phase III trials.

With cytotoxics, the generally accepted endpoint for phase II trials is objective tumor response; that is, tumor shrinkage by at least 50%. Tumor shrinkage is not a direct measure of patient benefit, although it sometimes is predictive of benefit. The most commonly accepted direct measures of patient benefit are survival, disease free survival

and symptomatic relief. Therapeutic effect on these endpoints cannot be reliably established outside of a phase III trial with an appropriate control group not receiving the experimental therapy. Investigators sometimes like to infer that a regimen prolongs survival because the responders live longer than the non-responders but this analysis has long been known to be invalid (4,5).

Tumor shrinkage is generally used as the endpoint for phase II trials of cytotoxics for two reasons. First, because response represents biological activity that can be attributed to the therapy (i.e. tumors rarely shrink spontaneously by 50%). Secondly, if the degree, duration and abundance of responses are sufficient, then it is plausible to hope that tumor response may translate into patient benefit. There are many cytotoxic regimens which were active in phase II trials but which subsequently had no identifiable effect on survival in phase III trials. Torri et al.(6) performed a meta-analysis of randomized trials to quantify the relationship between improvement in response rate and improvement in survival outcome for advanced ovarian cancer studies. They found that a very substantial improvement in response rate was necessary to have any identifiable effect on survival.

For phase II tumor vaccine studies, the investigator has the choice of using clinical endpoints or immunological endpoints. Clinical endpoints include tumor shrinkage, reduction in tumor marker levels or delay in time to tumor progression. Tumor shrinkage is always a welcome occurrence in a phase II tumor vaccine trial. The question is whether a regimen should be precluded from a phase III evaluation of early stage patients if it

does not shrink metastases in more advanced cases. Opinion is divided on this important point.

2.1 Tumor shrinkage endpoint

If tumor shrinkage is the endpoint, then phase II designs used for cytotoxics can be employed (7,8,9). Simon's "optimal two-stage" designs are widely used for phase II cytotoxic trials to test whether a regimen has a response rate above a background level p_0 (8). Frequently, $p_0 = 0.05$ is used. With clinical response, this assumes that no more than 5% of the patients will have apparent responses caused by variability in response assessment or spontaneous remissions. The 2-stage design incorporates an early termination point, which allows the investigator to discontinue patient accrual if a desired endpoint has not been achieved in the first stage of the trial.

At the conclusion of the clinical trial, the regimen will be declared active or inactive. Table 2 shows several designs with 10% false positive rate, 10% false negative rate and $p_0 = 0.05$. The false positive rate (α) is the probability of declaring the regimen active when the true response probability is p_0 . The false negative rate (β) is the probability of declaring the regimen inactive when its true response probability is the target response rate p_1 , the level of activity that we wish to be able to detect. In the first stage, N_1 evaluable patients are entered and treated. If no responses are observed, then the trial is terminated and the regimen is declared inactive. Otherwise accrual continues to a total of N evaluable patients. At that point accrual is complete. If the total number of responses is

at least A, then the regimen is declared active. The last column of the table indicates the probability of early termination after the first stage when the true response probability is p_0 . For example, if $p_0 = 5\%$ and the target response rate is 25%, then 9 patients are treated in the first stage of the trial. If no responses are observed, the trial is terminated.

Otherwise, accrual is continued to a total of 24 patients. If at least 3 responses are seen in the 24 patients, the regimen is declared active. The probability of declaring a regimen active when it's true response rate is 5% or less is 10%. The probability of missing the activity of a regimen with a true response rate of 25% is 10%. With a regimen having a response rate of 5%, the probability of stopping after only 9 patients is 63%.

This design with $p_1 = 25\%$ and $p_0 = 5\%$ seems reasonable for many initial vaccine trials but designs based on other parameter values are easily generated using computer program OTSD (optimum two-stage design) available at <http://lib.stat.cmu.edu/designs>. The required number of patients depends strongly on the difference $p_1 - p_0$. These designs are based on a binary measure of response. We will deal with time to progression endpoints in a later section.

2.2 Tumor marker endpoints

Traditional phase II trials in patients with clinically measurable tumors are often not considered appropriate as initial trials of whether tumor vaccines have biological effect.

Vaccine trials are best conducted in patients with intact immune systems and for some diseases this precludes inclusion of patients with gross tumor.

The strategy of vaccination in a minimal disease state and using a sensitive tumor marker or molecular probe to measure reduction or disappearance of sub-clinical tumor mass can be an effective alternative for development of tumor vaccines. This approach was used in obtaining promising initial results for an idiotypic lymphoma vaccine which is now in randomized phase III testing using conventional clinical endpoints (10). Although the relevance of a molecular marker of sub-clinical disease to long-term prognosis may be in question, such a marker can provide a measure of anti-tumor effect that can be measured in patients with minimal residual disease and is thus useful for early vaccine trials

2.3 Time to tumor progression endpoint

Evaluating whether a treatment delays recurrence or progression is particularly important for tumor vaccines. Patients without clinical evidence of disease may have more intact immune systems and be more appropriate candidate for tumor vaccines than patients with more advanced measurable metastatic disease. This will depend, however, on the type of therapy used to put them in the clinically disease free status.

Evaluating the effect on a regimen on time to progression of sub-clinical disease is particularly problematic in a single arm phase II trial. It is easy to devise a definition of

disease stabilization, i.e. lack of recurrence or progression for a specified period of time, but the validity of the definition depends on the existence of data that establish that such stabilization does not occur in the absence of treatment. This is difficult to establish reliably because of the usual difficulties of identifying comparable non-randomized controls and because of special difficulties involved with measuring time to disease progression in a consistent manner for different cohorts of patients. Consequently, we believe that the use of disease stabilization or time to progression as an endpoint in single arm trials should only be considered when data from a specific set of contemporaneous controls from the same institution are available. In such a case, rather than attempting to define disease stabilization in a valid manner as a dichotomous endpoint (e.g. present or absent based on some threshold), it is preferable to compare the time to progression for the patients in the phase II trial to the distribution of time to progression of a specific set of control patients not receiving the vaccine regimen. Dixon and Simon (11) provide formulas for computing the number of patients required in the single arm trial.

Generally, single arm trials of disease stabilization, time to tumor progression or recurrence are not sufficiently reliable. The randomized phase 2.5 design described below as a more satisfactory approach.

Phase III trials are generally randomized comparisons of a new regimen compared to a standard treatment using an endpoint of established medical importance to the patient such as survival or quality of life. Phase III trials are usually planned using a 5% type one

error parameter (α) because the results of phase III trials are viewed as definitive and are used as a basis for marketing approval and practice guidelines. We propose that in the development of cancer vaccines, there is a role for what might be called a “phase 2.5” trial. Such a clinical trial would also be randomized, but may use an endpoint measuring biological anti-tumor activity even though the endpoint might not be established as a valid surrogate for survival or quality of life. The phase 2.5 trial might also be based on an elevated statistical significance level since the objective of the trial would not be for regulatory approval or for establishing general practice guidelines.

To detect a large effect of a treatment in delaying tumor progression in a rapidly progressive disease such as pancreatic cancer or melanoma with visceral metastases does not require many patients in a randomized trial. With exponentially distributed times to progression, a 40% reduction in the hazard of progression corresponds to a 67% increase in median time to progression. In order to have 80% power ($\beta=0.20$) for detecting this size of effect using a traditional $\alpha=0.05$, only about 117 patients are required (assuming accrual rate of about 3 patients per month, median time to progression of 12 months for control group and follow-up time of 24 months after end of accrual) (12). This can be reduced to 87 patients if $\alpha=0.10$. Hence, 44 patients randomized to vaccine and the same number randomized to control, one can conduct a randomized “phase 2.5” trial for evaluating whether the vaccine reduces the hazard of progression by 40%. This design would be a “phase 2.5” design because of the unconventional use of a one-sided $\alpha = 0.10$ significance level and because time to progression might not be established as representing clear patient benefit. The phase 2.5 design is similar to the phase III design

in the respect that it contains a control group for evaluating the experimental regimen and the intent is comparative.

The basis of the efficiency is that the disease is rapidly progressive and a large treatment effect is targeted. If the median time to progression for the control group were 6 months instead of 12, an even smaller sample size would be required. When the disease is not rapidly progressive the efficiency illustrated here decreases. Statistical power for detecting a specified reduction of the hazard of an event is determined by the number of events, not the number of patients. With a slowly progressive disease, it may take many patients to be entered in order to observe a specified number of events unless the follow-up time following the close of accrual is very long. Also, for a rapidly progressive disease, a large reduction in hazard, e.g.40%, is associated with a moderate absolute increase in median time to event; e.g. 6 months increased to 10 months, 12 months increased to 20 months. Consequently, with a rapidly progressive disease there is greater justification for targeting a relatively large treatment effect.

2.4 Immunological endpoints

For some types of cancers, patients with measurable tumors and intact immune systems are available but existing vaccines do not produce clinical tumor responses. In such circumstances, immunological endpoints may guide attempts to optimize the vaccine and its delivery before randomized phase 2.5 trials or randomized phase III trials are

conducted. A phase III evaluation of a regimen may not be warranted, however, until the regimen demonstrates activity based on a clinical endpoint.

When an immunological endpoint is used, the protocol should provide specific information about the variability of the measurement of that endpoint. Three sources of variability should be distinguished: variability among assay results on the same specimen (e.g. lymphocytes, serum, or tumor tissue), variability among specimens from the same patient drawn at different times, and variability among patients. Documenting this variability is essential for interpretation of results of the trial. For example, such data permit one to define a threshold for change in the immunological endpoint that can be regarded as statistically significant. If these data are not available at the outset, it may be possible to develop them during the course of the clinical trial (e.g. from multiple baseline blood drawings for the patients to be vaccinated or from drawings for control patients). It is important to prepare plans for analysis of immunological endpoints in advance and include these plans in the protocol in order to ensure that needed data are available and to reduce subjectivity in the analysis.

Acknowledgement

I am indebted to my colleagues Seth M. Steinberg, Michael Hamilton, Allan Hildesheim, Samir Khleif, Larry W. Kwak, Crystal L. Mackall, Jeffrey Schlom, Suzanne L. Topalian and Jay A. Berzofsky with whom I collaborated in the National Cancer Institute Cancer Vaccine Working Group and who co-authored the paper (13) on which much of this chapter is based.

References

- (1) Salgaller ML, Marincola F, Cormier, JN, and Rosenberg, SA. Immunization against epitopes in the human melanoma antigen gp100 following patient immunization with synthetic peptides. *Cancer Research* 1996;56:4749-4757.
- (2) Cormier JN, Salgaller ML, Prevette T, Barracchini KC, Rivoltini L, Restifo NP, Rosenberg SA, and Marincola FM. Enhancement of cellular immunity in melanoma patients immunized with a peptide form MART-1/Melan A. *The Cancer Journal from Scientific American* 1997;3:37-44.
- (3) Casagrande JT, Pike MC and Smith PG. The power function of the “exact” test for comparing two binomial distributions. *Applied Statistics* 1978;27:176-180.

- (4) Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *Journal of Clinical Oncology* 1983;1:710.
- (5) Simon R and Makuch RW. A nonparametric graphical representation of the relationship between survival and the occurrence of an event: application to responder versus non-responder bias. *Statistics in Medicine* 1984;3:1.
- (6) Torri, V, Simon R, Russek-Cohen, E, Midthune, D, Friedman, M. Relationship of response and survival in advanced ovarian cancer patients treated with chemotherapy, *J. Natl. Cancer Inst.* 1992;84:407-414.
- (7) Fleming TR. One sample multiple testing procedure for phase II clinical trials. *Biometrics* 1982;38:143.
- (8) Simon, R. Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials* 1989;10:1-10.
- (9) Garnsey-Ensign L, Gehan EA, Kamen D and Thall PF. An optimal three-stage design for phase II clinical trials. *Statistics in Medicine* 1994;13:1727.
- (10) Bendandi M, Gocke CD, Kobrin CB, Benko FA, Sternas LA, Pennington R, Watson TM, Reynolds CW, Gause BL, Duffey PL, Jaffe ES, Creekmore SP, Longo DL, Kwak LW. Complete molecular remissions induced by patient

- specific vaccination plus granulocyte-monocyte colony stimulating factor against lymphoma. *Nature Medicine* 1999;5:1171-7.
- (11) Dixon DO, Simon R: Sample size considerations for studies comparing survival curves using historical controls. *Journal of Clinical Epidemiology* 1988;41:1209-1213.
- (12) Rubinstein LV, Gail MH and Santner TJ. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *Journal of Chronic Diseases* 1981;34:469-479.
- (13) Simon RM, Steinberg SM, Hamilton M, Hildesheim A, Khleif S, Kwak LW, Mackall CL, Schlom J, Topalian SL and Berzofsky JA. Clinical trial designs for the early clinical development of therapeutic cancer vaccines. *Journal of Clinical Oncology* 2001;19:1848-1854.

Finding the Minimum Active Dose

Probability of Immunologic Response	Number of Patients Treated at Dose	Probability of No Immunologic Responses
0.20	11	0.09
0.25	9	0.08
0.30	7	0.08
0.40	5	0.08
0.50	4	0.06

Table 1

Optimal Two-Stage Designs (8)

Target Response Rate (p_1)	First Stage Sample Size (N_1)	Maximum Sample Size (N)	Number of Responses Required For Activity (A)	Probability of Early Termination
20%	12	37	4	.54
25%	9	24	3	.63
30%	7	21	3	.70
35%	6	12	2	.74

Table 2