

# Clinical Trial Designs for Therapeutic Cancer Vaccines

Richard Simon  
National Cancer Institute

Dr. Richard Simon  
Chief, Biometric Research Branch  
National Cancer Institute  
9000 Rockville Pike  
MSC #7434  
Bethesda MD 20892-7434  
Phone (301)496-0975  
Fax (301)402-0560  
[rsimon@nih.gov](mailto:rsimon@nih.gov)

## **1. Introduction**

Many aspects of the usual paradigm for the clinical development of cytotoxic anti-cancer drugs are not appropriate for therapeutic cancer vaccines (1). In this chapter we explore some of these differences and recommend designs and strategies more suited for the development of effective therapeutic vaccines.

## **2. Dose-Escalation Studies**

### **2.1 Safety Studies**

Phase I studies usually involve dose escalation in cohorts of three patients, starting with a very low dose, in an effort to identify a maximally tolerated dose. Phase I studies are generally conducted in patients with advanced metastatic disease who have failed all other available treatments. Tumor vaccines are often based on DNA constructs, viral vectors and cytokines that have been determined as safe in previous clinical trials.

Peptide vaccines generally seem inherently safe as long as the cytokine adjuvants are used in combinations and doses previously demonstrated to be safe. Consequently, no phase I safety study should be required for most therapeutic cancer vaccines.

On the other hand, a novel virus or plasmid used as a recombinant vaccine vector for the first time should be evaluated for safety. A dose escalation design may be appropriate but patients whose immune systems have not been compromised by extensive chemotherapy are the most relevant subjects. If such vectors are proven to be nontoxic at substantial

doses, then subsequent dose-escalation safety trials using the same vectors but with different recombinant inserts may not be required.

## 2.2 Immunogenicity Studies

Feasibility issues limit the maximum doses of certain 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.

For cancer vaccines, it is not always the case that more is better. 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 (2,3). Thus, for subsequent trials using similar peptides, an intermediate fixed dose of 1 mg was chosen for vaccination, bypassing repetitive phase I studies.

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.

Korn et al. (4) defined a sequential procedure for finding a biologically active dose, although not necessarily the minimal active dose. During an initial accelerated phase one patient per dose level is treated until a biological response is seen. After the first response is seen, cohorts of 3-6 patients are treated per dose level. With 0-1 biological responses among the 3 patients at a dose level, escalate to the next level for the next cohort of patients. With 2 or 3 responses out of the 3 patients, expand the cohort to a total of 6 patients. With 5 or 6 biological responses out of the 6 patients, declare that dose level to be the biologically active level and terminate the trial. With fewer than 5 biological responses out of the 6 patients, a new cohort of 3 patients is accrued at the next higher dose level, etc. Korn et al. (4) describe some of the statistical properties of this sequential design.

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. Consider, for example, planning a study of two dose levels to 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 (5). 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.

### **3. Phase II Studies**

The general 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.

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 (6,7).

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.(8) 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, clinical endpoints and/or immunological endpoints are commonly used. Clinical endpoints include tumor shrinkage, reduction in tumor marker levels or delay in time to tumor progression.

### 3.1 Single-Arm Trials Using Clinical or Immunologic Response Rate

If tumor shrinkage is the endpoint, then phase II designs used for cytotoxics can be employed (9,10,11). 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  $p_0$  (10). 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 ( $\alpha$ ) is the probability of declaring the regimen active when the true response probability is  $p_0$ . The false negative rate ( $\beta$ ) 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 using tumor regression as endpoint.

An optimum two stage design can also be used with a binary immunologic response endpoint. In such a case, however, the values of  $p_0$  and  $p_1$  will generally be much higher than for a tumor regression endpoint. Optimum two-stage designs for any values of  $p_0$ ,  $p_1$ ,  $\alpha$ , and  $\beta$  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  $p_1-p_0$ .

A variety of alternative single-arm designs for evaluating binary endpoints have been published. For example, Garnsey-Ensign developed three stage designs (11), and Thall and Simon (12) developed continuous monitoring Bayesian designs. The essential characteristics of this class of designs is that the endpoint is binary and that the objective is to evaluate the response rate of the regimen on it's own, not in comparison to the response rate for some other regimen. If the objective is comparative, then even for a single arm trial, the design and method of analysis should take into account the variability in the estimate of response rate for the external control regimen. Methods such as that of Makuch and Simon (13), Dixon and Simon (14) and Thall and Simon (12,15 ) attempt to take that variability into account.

For therapeutic vaccines, the current situation for most diseases (other than melanoma) is that few if any partial or complete tumor responses have been observed with any regimen but that varying degrees of immunogenicity have been obtained. It is usually very difficult to compare the degree of immunogenicity obtained with different regimens by



different investigators because of differences in assays, variation in procedures and reagents, and differences in patient selection. It is even difficult to compare the degree of immunogenicity obtained by the same investigator in different studies with different regimens because of assay variability. There is also generally some uncertainty in what measures of immunogenicity are most appropriate. There are currently no measures that can be considered true surrogates for clinical response.

It takes fewer patients to determine whether a regimen causes any clinical responses than it does to compare it to another regimen with regard to immunologic response rate. The optimum two-stage design recommended above for evaluating clinical response rate has a first stage of only 9 patients. Consequently, a reasonable phase II development strategy is to design phase II trials using the optimal two-stage design for distinguishing a 5% clinical response rate from a 25% clinical response rate with error rates of 10%. If after accruing the 9 patients in the first stage, no clinical responses are seen, then the trial is terminated. If one or more clinical responses are seen, accrual should continue unless the level of immunologic response is so inadequate that the investigator would like to make some modifications to the regimen. In cases where accrual is terminated after 9 patients because of lack of clinical responses, the immunological activity of the regimen for the 9 patients accrued will provide information for modifications of the vaccine regimen.

### 3.2 Multiple Arm Screening Trials Using Immunological Response Rate

One of the complexities of therapeutic vaccine development is the many options available for attempting to enhance immunological recognition of a specified tumor antigen. In addition to the vector or mode of presentation of the antigen to the immune system, there are alternative adjuvants, preparative regimens, routes and schedules of administration. Because of the difficulty and time required for clinical trials, it is best to optimize vaccines to the extent possible using animal models. Nevertheless, there may be several vaccine candidates available for clinical trial. One approach would be to perform a two-stage 9-24 patient clinical trial on all candidate regimens, stopping at 9 patients unless partial remissions are seen. An alternative strategy is to perform a multi-arm phase II trial to optimize the regimen with regard to immunogenicity before focusing on clinical endpoints. The reason for using a multi-arm randomized phase II design is to ensure comparability of patients on the different regimens and to control for assay variability.

### 3.2.1 Factorial Screening Designs

Two types of randomized phase II trials are potentially relevant for optimizing a vaccine regimen. One method is the use of a phase II factorial design. Suppose that there are  $m$  binary factors that represent dimensions in which a basic vaccine may be modified. For example, one factor might be route of administration and another might be whether a specified adjuvant is administered. Since there are  $m$  binary factors, there are  $2^m$  possible combinations of levels of the factors. For example, let the levels of each factor be denoted 0 or 1. Then with two factors the possible combinations of factors are (0,0),

(0,1), (1,0), and (1,1). The study is conducted by randomizing  $N$  patients into the  $2^m$  treatment groups. If the factors affect immunogenicity independently, then in comparing the two levels of one factor, one can ignore the other factors. Actually, a stratified or model based analysis is more powerful than simple pooling, but the point is that the two levels of each factor are compared based on the assumption that the difference does not depend on the levels of the other factors. Consequently, comparing level 0 of a factor to level 1 of that factor involves comparing average immunogenicity for the  $N/2$  patients with the factor at level 0 to average immunogenicity for the  $N/2$  patients with the factor at level 1. The trial is sized, i.e.  $N$  is selected, for  $m$  independent two-arm comparisons involving  $N/2$  patients per arm, not for one  $2^m$  arm comparison. If there are 3 factors ( $m=3$ ), then there are  $2^3 = 8$  arms to the trial. If  $N=32$ , then 4 patients are randomly assigned to each of the arms, but the comparisons of the levels of each factor involve comparing average immunogenicity for two groups of 16 patients. The value of  $N$  is selected based on the manner in which immunogenicity is measured (e.g. continuous scale or binary), assay and biological variability among patients, and size of difference to be detected. Suppose, for example, that immunogenicity is measured on a continuous scale, and let  $x$  denote the change in immunogenicity from baseline for a patient after treatment. Assume that  $x$  is approximately normally distributed and let  $\sigma$  denote the standard deviation for  $x$  for different patients receiving the same vaccine regimen. Let  $\delta$  denote the size of the difference in mean value of  $x$  we wish to be able to detect in comparing vaccine groups, and let  $\alpha$  and  $\beta$  denote the type 1 and type 2 error rates for the comparison. Then  $N/2$  patients are required in each of the two groups with:

$$N = 4 \left( \frac{z_{\alpha/2} + z_{\beta}}{\delta / \sigma} \right)^2 \quad (1)$$

where  $z_{\alpha/2}$  is the  $100(\alpha/2)$ 'th percentile of the standard normal distribution and  $z_{\beta}$  is the  $100\beta$ 'th percentile. For 5% type 1 error and 80% power, we have  $z_{\alpha/2}=1.96$  and  $z_{\beta}=0.84$ . To detect a difference in means that represents one standard deviation of inter-patient variability in immunogenicity requires  $N=32$  patients randomized. This gives 16 patients in each level of each binary factor. With 3 binary factors, there are 8 treatment groups. Hence randomly assigning 4 patients per treatment group will satisfy this requirement. The required sample size is very dependent on the  $\delta/\sigma$  ratio. Reducing  $\sigma$  by improving the assay reproducibility will increase this ratio for a fixed  $\delta$ . The quantity  $\sigma$  reflects both biological variability and assay variability, and so using a more homogeneous group of patients may also serve to reduce  $\sigma$ .

### 3.2.2 Randomized Selection Design

An alternative approach to optimizing a vaccine regimen is to conduct a randomized phase II trial of the variants and to select the regimen that has the best average immunogenicity in the trial. This type of approach has been described by Simon et al. (16), Strauss and Simon (17) and Yao et al. (18). The analysis does not result in any conclusions of which factors are important to immunogenicity or which regimens are

significantly better than which other regimens, but merely a selection of a regimen which is most promising for further investigation. There are two approaches to establishing sample size per treatment group for such selection designs. One approach, described in Simon et al. (16) is to require that the sample size per treatment be large enough to assure with high probability that if one treatment is superior to all other treatments by a specified amount  $\delta$ , then it will have the largest sample mean and will therefore be selected. With normally distributed measures of immunogenicity, the probability of correct selection depends on the ratio of  $\delta/\sigma$  and on the number of treatment arms. If there is one best arm and the rest are inferior by  $\delta$ , then the probability of correct selection decreases as the number of arms increases. Table 3 shows the sample size required to have a probability of correct selection of 0.90 as a function of  $\delta/\sigma$  and the number of treatment arms. For  $\delta/\sigma$  values of 0.75 or greater, the design requires fewer than 15 patients per arm for up to 8 randomized arms.

Another approach to establishing sample size for randomized selection designs is based on the assumption that the true mean immunogenicity for a regimen can be regarded as a random draw from some hypothetical super-distribution of activity levels (17,18). By studying more regimens in a randomized trial, one has a greater chance of including a very active regimen. If the number of patients available for the trial is fixed, there is a trade-off between the number of arms in the trial and the number of patients per arm. If we assume that the super-distribution is normal with mean  $\mu$  and standard deviation  $v$ , then we can compute the expected mean immunogenicity level for the regimen that has the best sample mean in the randomized K arm selection trial. Table 4 shows the

expected mean immunogenicity levels as a function of the number of randomized arms and the sample size per arm when the total number of patients is fixed at 50. Four cases are shown: 2 arms of 25 patients, 5 arms of 10 patients, 10 arms of 5 patients, and 25 arms of 2 patients. The mean  $\mu$  and standard deviation  $v$  of the super-distribution are set at 0 and 1 respectively, and the table shows results for different values of the standard deviation  $\sigma$  of immunogenicity measurement for patients receiving the same vaccine regimen. It can be seen from the table, that in most cases the best treatment is identified by studying 25 treatment arms, each with only 2 patients. This is not really practical and reflects the unrealistic assumption that an unlimited number of regimens are available and that the activities of these regimens are independent and can be regarded as draws from a normal distribution. Nevertheless, the model does highlight the principle that there is an opportunity cost to studying few regimens thoroughly. The approach provides some justification for screening many regimens with a smaller sample size than is used for non-selection based trials.

### 3.3 Controlled Phase II Trials with Time to Tumor Recurrence or Progression Endpoint

Therapeutic vaccines may be more effective in patients with lower tumor burdens, and may slow progression rather than cause regression of bulk tumor. 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.

Evaluating the effect on a regimen on time to progression of sub-clinical disease is very 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, 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 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 (14) provide formulas for computing the number of patients required in the single arm trial.

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 ( $\alpha$ ) because the results of phase III trials are viewed as definitive and are used as a basis for marketing approval and practice guidelines. In the development of

cancer vaccines, there is a role for what might be called a “phase 2.5” trial (1). 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 marketing 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 an  $\alpha=0.10$ , only about 87 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) (19). 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.



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. The number of events required to have power  $1-\beta$  for detecting a treatment effect of size  $\delta$  with a one-sided statistical significance level of  $\alpha$  is approximately:

$$E = 2 \left( \frac{z_{\alpha} + z_{\beta}}{\ln(\delta)} \right)^2 \quad (2)$$

where  $\delta$  is the ratio of median survivals to be detected if the survival distributions are exponential (19). In the calculation of the previous paragraph,  $\delta=1.67$ ,  $\alpha=0.10$ ,  $\beta=0.20$  and consequently  $E = 35$  events. The number of patients needed to obtain 35 total events depends on the accrual rate, accrual period and follow-up period. 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. If the disease is rapidly progressive and all patients are followed until progression, then only 35 patients need to be randomized to observe 35 events.

Two different vaccine regimens can be evaluated in a randomized controlled phase II trial with time to progression endpoint by utilizing a three arm design. One arm would be the control group that does not receive either vaccine. For separate evaluation of each vaccine group without adjustment of the significance level for the fact that two vaccines are being evaluated, the number of events and patients required increases by one third compared to the two-arm trial. For example, if 44 patients per arm are required for the two-arm trial above, then 44 patients per arm are required for the three arm trial.

Trials using time to progression endpoints can be terminated early if interim results are not promising. One simple strategy is to perform an interim analysis when half of the planned total number of events have been observed. Accrual can be terminated if the number of events in the treatment group is greater than the number of events in the control group at that time. This interim analysis does not effect the type 1 error rate and causes negligible loss in statistical power (20). More sophisticated and efficient interim analysis plans for early termination when results are not promising are also possible (21). For a trial with multiple vaccine arms and one control arm, the interim monitoring can be used to evaluate each vaccine arm and stop accrual to those for which results are not promising.

Randomized phase 2.5 trials may be structured so that all patients first receive tumor reduction with other modalities prior to randomization.

#### **4. Phase III Trials**

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 of therapeutic cancer vaccines do not differ in important respects from phase III trials of conventional treatments; a randomized trial is required in both cases with a medically relevant endpoint and an appropriate control group.

## **5. Summary**

Therapeutic cancer vaccines have characteristics that require a new paradigm for phase I and phase II clinical development. Effective development plans may take advantage of some of the following observations:

Dose ranging safety trials are not appropriate for many cancer vaccines.

Dose ranging trials to establish an optimal biologic dose are often not practical. We have presented an efficient design of Korn et al. (4) to identify an immunogenic dose.

Vaccine efficacy can be efficiently evaluated with tumor response as endpoint utilizing a two stage design with only 9 patients in the first stage. If no partial or complete responses are observed in the initial 9 patients, accrual to the trial is terminated.

Optimization of vaccine delivery by comparing results of single arm phase II studies using immunological response as endpoint is problematic because of assay variation and potential non-comparability of patients in different studies.

Randomized screening studies can be used to efficiently optimize vaccine immunogenicity. Efficiency in use of patients depends on having assay variation and inter-patient variability small relative to the difference in immunogenicity to be detected.

Phase II studies using time to progression as endpoint are most interpretable if they employ randomized designs with a no-vaccine control group. Such designs may use an inflated type 1 error rate, and need not be prohibitively large if patients with rapidly progressive disease are studied. Interim monitoring plans may effectively limit the size of the trials by terminating accrual early when results are not consistent with the targeted improvement.

### **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 (1) on which much of this chapter is based.

## References

1. 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.
2. 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.
3. 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.
4. Korn EL, Rubinstein LV, Hunsberger SA, et al.: Clinical trial designs for cytostatic agents and agents directed at novel molecular targets. In Buolamwini JK and Adjei AA (eds.) *Novel Anticancer Drug Protocols (Methods in Molecular Medicine Series)*, Humana Press, 2003 (In Press).
5. 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.

6. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *Journal of Clinical Oncology* 1983;1:710.
7. 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.
8. 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.
9. Fleming TR. One sample multiple testing procedure for phase II clinical trials. *Biometrics* 1982;38:143.
10. Simon, R. Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials* 1989;10:1-10.
11. 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.
12. Thall PF, Simon R, Estey E: A new statistical strategy for monitoring safety and efficacy in single-arm clinical trials. *Journal of Clinical Oncology* 14:296-303, 1996.

13. Makuch, RW, and Simon, RM.: Sample size considerations for non-randomized comparative studies. *J. Chron. Dis.* 33: 175-181, 1980.
14. Dixon DO, Simon R: Sample size considerations for studies comparing survival curves using historical controls. *Journal of Clinical Epidemiology* 1988;41:1209-1213.
15. Thall, PF, and Simon, R. Incorporating historical control data in planning phase II clinical trials. *Stat. in Med.* 9:215-228, 1990.
16. Simon, R, Wittes, RE, and Ellenberg, SS.: Randomized phase II clinical trials. *Cancer Treatment Rep.* 69: 1375-1381, 1985.
17. Strauss N and Simon R. Investigating a sequence of randomized phase II trials to discover promising treatments. *Statistics in Medicine* 14:1479-1489,1995.
18. Yao TJ, Begg CB, Livingston PO: Optimal sample size for a series of pilot trials of new agents. *Biometrics* 52:992-1001, 1996.
19. 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.

20. Thall, PF, Simon, R, Ellenberg, SS, and Shrager, R. Optimal two-stage designs for clinical trials with binary response. *Stat. in Med.* 7: 571-579, 1988.

21. Lan, G, Simon, R, and Halperin, M.: Stochastically curtailed tests in Long-term clinical trials. *Communications in Statistics - Sequential Analysis* 1: 207-219, 1982.



### 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**

**Number of Patients Per Arm for Randomized Selection Design**

Number of treatment arms	Patients Per Arm		
	$\delta/\sigma=0.5$	$\delta/\sigma=0.75$	$\delta/\sigma=1.0$
2	13	6	4
3	21	9	6
4	24	11	6
5	27	13	7
6	30	14	8
7	31	14	8
8	35	15	9

**Table 3**

**Expected Immunogenicity of Selected Regimen for Randomized Selection Designs That Utilize 50 Patients**  
Assumes  $\mu=0$  and  $v=1$

Number of regimens	Patients per regimen	Expected Immunogenicity of Selected Regimen		
		$\sigma=2$	$\sigma=1$	$\sigma=0.75$
2	25	0.52	0.55	0.56
5	10	0.99	1.10	1.13
10	5	1.15	1.40	1.44
25	2	1.14	1.60	1.73

**Table 4**