Randomized decision designs when the number of available subjects precludes a more standard study design

Lynette Smith and James R. Anderson
University of Nebraska College of Public Health

Background: In a standard clinical research study comparing two treatments, the size of the study is determined such that, should a specific clinically important difference in outcome exist, the study would have some fixed power (testing at some fixed level of statistical significance) to detect that difference. So, for instance, for a time-to-event outcome, if we wished to be 80% confident (testing at the 5% level of statistical significance [2-sided]) to detect a one-third reduction in the risk of treatment failure, the sample size and follow-up would have to be sufficiently large to have observed 191 failures.

However, there are clinical situations in which the application of standard clinical trials design doesn't work. This work was motivated by an interest in comparing two treatments for recurrent Wilm's tumor. Wilm's tumor is a pediatric kidney cancer, with an incidence of about 7.6 per million population under the age of 15 (1) (http://seer.cancer.gov/publications/childhood/renal.pdf) with the highest incidence seen in the age group 0-4. More than ninety percent of patients with Wilm's tumor survive 5 years and deaths after 5 years are rare. It is estimated that within the network of hospitals and medical centers affiliated with the Children's Oncology Group, an NCI-funded cooperative group studying the biology and treatment of childhood cancers (2), about 30 patients are diagnosed with Wilm's tumor in first recurrence and about 20 per year are likely to enroll on a treatment study for recurrent Wilm's tumor. About 40% of patients with Wilm's tumor in first relapse are expected to survive without experiencing a subsequent relapse long-term with standard treatment (3).

Few clinical trials enroll subjects for more than 5 years, since after that period; the questions asked in the study may no longer be of great interest. Thus, if a study comparing treatments for recurrent Wilm's tumor was proposed, it would likely enroll no more than 20 patients per year for 5 years or 100 patients total.

As an alternative to the standard clinical trial design, Don Berry, in an unpublished paper, suggests an alternative to the standard clinical trial design could be the use of decision analysis, with the goal of "delivering good medicine to patients who have or will have the disease or condition".

An alternative study design, assuming exponential failure

Suppose a limited number (N) of patients is available for study and the traditional randomized trial design comparing 'standard' to 'new' is not feasible. We propose a study design which seeks to maximize the number of study subjects receiving the better of the two treatments.

Consider a study that enrolls subjects at a uniform rate over time for a fixed enrollment period (e.g., 20 subjects per year for 5 years). Assume that $n_1/2$ patients are randomized to each of 2 treatments and then time-to-event is compared using the log-rank test. All remaining ($N-n_1$) patients are then assigned treatment 2 (treatment 1) based on whether the Cox model test statistic (or, equivalently, the normalized log-rank test) is negative (positive).

We then wish to maximize the number of study subjects receiving the better of the two treatments through the choice of n_1.

Now for equal treatment assignment, after the observation of d_i events, the Cox model test statistic, $S = \sqrt{d_i}(\hat{\beta}/2)$, is approximately normally distributed with mean $\sqrt{d_i}\,(\beta^*/2)$ and variance 1, where β^* is the true value for β, the logarithm of the relative risk, and the difference between the null variance and the variance at β^* is ignored.

The number of patients expected to be enrolled on regimen 2, R_1, can be written as

$$E(R_1) = 0.5\,(n_1) + (N - n_1) * \text{Prob}\{ S < 0\} = \text{Prob}\,\{ N(0,1) < -[\sqrt{d_i}\,(\beta^*/2)]\,\}.$$

It is straightforward to calculate the expected number of failures observed after the enrollment of n patients, when patients enter uniformly over time and the failure rate is exponential(λ). From this, the value of $E(R_1)$ can be calculated for various choices of N, n_1, and λ.

Table 1 displays the optimal values for n_1, the expected total number of patients assigned to regimen 2 for the optimal n_1, and expected total number of failures observed at the optimal n_1 under the null hypothesis, for various values for λ, assuming a total sample size of 100.

Table 1
Optimal values for n_1, expected total number of patients assigned to regimen 2 for the optimal n_1, and the expected number of failures observed at the optimal n_1 under the null hypothesis, for various values for λ (N=100)

	Optimal value for n_1			Expected value for patients assigned to regimen 2			Expected number of failures at optimal n_1 at null		
λ value	RR=0.5	RR=0.67	RR=0.75	RR=0.5	RR=0.67	RR=0.75	RR=0.5	RR=0.67	RR=0.75
0.25	40	44	46	70.7	63.4	59.1	8.5	10.2	11.0
0.50	35	41	43	74.8	66.8	62.6	11.7	15.4	16.7
0.75	32	39	41	77.1	68.7	64.2	13.4	18.6	20.2
1.00	30	37	39	78.5	70.0	65.2	14.6	20.3	22.0

Realistic true reductions in the risk of failure are likely to be in the range of 0.67-0.75, and thus it appears that if the total sample size is 100, the optimal n_1 is near 40, at a time when about 15-20 of the total expected 100 failures have been observed. Thus, the design which maximized the number of subjects receiving the better treatment was to randomize about 40 patients (having observed about 20% of the total expected information) and then assign the remaining 60 patients to the regimen with the better observed outcome.

The increase in the number of patients receiving the more effective regimen under this design, can be substantial (Table 1). For true reductions in the risk of failure in the range of 0.67-0.75, the ratio of the number of patients treated on the more effective regimen to that of patients treated on the less effective regimen is in the range of 1.5 to 2.3.

Applying the alternative decision design when some patients are expected to survival event-free long-term

The event-free survival for patients with Wilm's tumor in first relapse can be reasonably estimated by a "cure" model, where 40% of patients can be expected to survive long-term, with the event-free survival for the remaining 60% following an exponential distribution with parameter λ=0.75.

We hypothesized that, in this settings, the n_1 corresponding to observing about 15-20% of the expected information would be optimal in terms of maximizing the sample size on regimen 2. To study this, we performed a simulation study, simulating 1000 replicates of a study enrolling an average of 20 subjects per year over 5 years, where the event-free survival for regimen 1 was modeled as: $S(t) = 0.4 + (0.6)*\exp(-0.75t)$ and the cure rate for regimen 2 was set at values from 0.4 to 0.6.

We simulated studies where the decision to assign the remaining patients to one of the regimens was made at the end of 1, 2, 3 or 4 years of enrollment (corresponding to expected n_1 values of 20, 40, 60 and 80 patients).

The simulation results indicated that the number of study subjects receiving the better treatment was maximized at an n_1 value near 40 (corresponding to observing a mean of about 11 events (corresponding to 18% of the total expected number of events [E=60]).

Table 2 displays the expected total number of patients assigned to regimen 2 when n_1 = 40 for various choices of the cure rate for regimen 2.

Table 2
Simulation expected total number of patients assigned to regimen 1 when n_1 = 40 for various choices of the cure rate for regimen 1 (cure rate for regimen 2 = 40%) (1000

Cure rate for regimen 2	Approximate relative risk	Expected sample size for regimen 1
40%	1.00	49.5
45%	0.87	54.4
50%	0.75	58.0
55%	0.65	63.7
60%	0.55	66.1

For a true improvement in the cure rate from 40% to 50% (relative risk ~0.75), the ratio of the number of patients treated on the more effective regimen to that of patients treated on the less effective regimen is 1.38; for a cure rate for regimen 1 of 60%(relative risk ~0.75, this ratio is 2.0.

While the optimal decision time for the cure model (~20% of the expected information) appears to be similar to that for the exponential model, the gain for a fixed sample size is less, because the total information available for the cure model is less.

Discussion

In settings where the available patient pool makes standard Phase III clinical studies impossible, we propose to use a study model that, after the randomization of n_1 patients, forces a decision to treat all subsequent patients on one of the treatments. We show that, for a sample size of 100 and exponential(λ) failure times, the optimal n_1 is about 40 for values of λ and relative risk differences which might be seen is a typical

Phase III cancer clinical trial. A sample size of 40, assuming uniform enrollment of patients over time, corresponds to about 15-20% of the total expected information.

We also showed through simulation that when some proportion of patients can be expected to survive event-free long-term and the event-free survival can be described using a cure model, the optimal for a sample size of 100 is again about 40, corresponding to about 20% of the total expected information.

We propose an "all-or-nothing" decision after the enrollment of n_1 patients. However, one might consider the assignment of additional patients that was based on some measure of the difference observed in event-free survival at the time that n_1 patients had been enrolled. One possibility would be to compute the estimate of the relative risk and assign the N-n_1 patients so that ($1/RR_{est}$) patients are assigned to regimen 2 for each patient assigned to regimen 1, where RR_{est} is the estimate of the event rate ratio, regimen 2 compared to regimen 1. Alternatively, one could compute CP(regimen 2), the conditional probability that the outcome at the end of the trial would favor regimen 2 if the remaining data followed the trend observed in the first n_1 patients (4) and assign VP(regimen 2) of the remaining the N-n_1 patients to regimen 2.

References

1. Leslie Bernstein, Martha Linet, Malcolm A. Smith, Andrew F. Olshan. Renal Tumors. In: Ries LAG, Smith MA, Gurney JG, Linet M, Tamra T, Young JL, Bunin GR (eds). *Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975-1995*, National Cancer Institute, SEER Program. NIH Pub. No. 99-4649. Bethesda, MD, 1999.
2. O'Leary M, Krailo M, Anderson JR, Reaman GH; Children's Oncology Group. Progress in childhood cancer: 50 years of research collaboration, a report from the Children's Oncology Group. Semin Oncol. 2008 35:484-93.
3. Malogolowkin M, Cotton CA, Green DM, Breslow NE, Perlman E, Miser J, Ritchey ML, Thomas PR, Grundy PE, D'Angio GJ, Beckwith JB, Shamberger RC, Haase GM, Donaldson M, Weetman R, Coppes MJ, Shearer P, Coccia P, Kletzel M, Macklis R, Tomlinson G, Huff V, Newbury R, Weeks D; National Wilms Tumor Study Group. Treatment of Wilms tumor relapsing after initial treatment with vincristine, actinomycin D, and doxorubicin. A report from the National Wilms Tumor Study Group. Pediatr Blood Cancer. 2008 50:236-41.
4. Lan KKG and Wittes J. The B-value: A tool for monitoring data. Biometrics 1988 44:579-85.