The University of Texas
M. D. Anderson Cancer Center

2008-0133: A Bayesian randomized trial of Image-Guided Adaptive Conformal Photon vs Proton Therapy, with Concurrent Chemotherapy, for Locally Advanced Non-Small Cell Lung Carcinoma: Treatment Related Pneumonitis and Locoregional Recurrence

Version 17

Acceoted on 2/19/2013

13.2 Sample Size Calculation and Data Analysis Plan

13.2.1 Hypothesis

The primary endpoint of the study is time to treatment failure, defined as the interval from the time of randomization to the development of treatment-related pneumonitis (TRP) or local failure, whichever occurs first. We are interested in studying whether IGAPT can reduce the TRP rate compared with IGAXT. Although local disease control rates are thought to be similar for IGAPT and IGAXT, this assumption needs to be prospectively tested, and hence we chose a combined endpoint of TRP and local failure as the primary endpoint.

Our primary outcome is time to treatment failure. Treatment failure is defined as TRP or local failure. These 2 types of failure are equally important. Based on the preliminary data, it is assumed that the time to treatment failure follows a log-normal distribution, with 6-month and 12-month treatment failure rates for the IGAXT arm of 30% and 40%, respectively (27). We also assume that under the null hypothesis (H0), failure in the IGAPT arm is the same as failure in the IGAXT arm, and under the alternative hypothesis (H1), IGAPT can reduce the corresponding failure rates to 20% and 25%, respectively.

Eligible patients will be assigned to receive either IGAXT or IGAPT based on the Bayesian adaptive randomization method. Adaptive randomization will allocate more patients to the treatment that yields more favorable outcomes based on the interim observed data. This design is appealing ethically because if a difference is found in treatment efficacy, more patients in the trial can be treated with the more effective treatment. If no difference is found in treatment efficacy, patients will be assigned with equal probability to either treatment, as is the case in the conventional randomized trials with equal allocation ratio. With a maximum of 150 eligible patients, we will have 81% power to detect this difference with a one-sided type I error rate of 10% or less (see subsequent sections and Table 4 [Operating Characteristics] for details). Under H1, there is a 48% chance of early stopping with about 58% of the patients randomized into the IGAPT group in a median sample size of 128. The sample size calculation assumes an accrual rate of 7 patients per month with 32 months of accrual plus an additional 24 months of follow-up. The total sample size is up to 250 patients, to account for patients being ineligible or lost to follow-up.

Patients randomized to IGAPT may be denied insurance reimbursement for this treatment, and therefore, they may elect not to be treated with IGAPT. These patients will be transferred to Group 4, as described in section 7.5, and they will be treated with the highest dose of IGAXT possible (74 CGE, 66 CGE).

13.2.2 Model specification for Bayesian adaptive randomization design using time-to-treatment-failure as the endpoint.

The time to treatment failure is assumed to follow a log-normal distribution. A Bayesian lognormal regression model with adaptive randomization will be used

Specifically, we will conclude that IGAPT is better if, given the current data, the probability of time to treatment failure for IGAPT being longer than for IGAXT is greater than P_L. The proposed design allows continuous monitoring. The trial can be stopped early should the interim results indicate a high probability of one treatment being better than the other. Specifically, we set $P_L = 0.70$ for interim monitoring. If the trial continues to the end without being stopped early, we set $P_L = 0.59$ in the final analysis. The probability calculation is confined to within the 24-month period, which corresponds to the clinical implication of the relevance and importance of recurrence or TRP during this period.

Patients will be randomized to the achievable radiation dose [74, 66] as described in section 7.5. We assume that approximately 1/3 of patients will be randomized to each dose level. However, we do not expect any differences between doses with respect to the treatment effects on TRP or local failure. The calculations of randomization probabilities and decision rules described above will include all data across dose levels.

13.2.4 Operating characteristics

The operating characteristics from 1,000 simulation runs are summarized in Tables 4a, 4b and 4c. According to the specifications given in the above two sections, a total of 150 eligible patients are uniformly enrolled during the 24-month accrual period. An event of either grade $\geq$ 3 TRP or local failure is defined as a treatment failure. The first 20 patients are equally randomized into one of the two groups. After enrolling the first 20 patients and after one event is observed in each group, patients are adaptively randomized based on the posterior 1-year failure rate estimates of the two groups. Failure or losses to follow-up for randomized patients are updated as they occur, i.e., the observed failure or loss to follow-up is updated as it happens. In addition, when new patients enter the study, the censoring time is updated for all patients who are still on study but have not yet developed an events. The early stopping rule applies after a total of 20 events is observed in the entire sample. The decision rules for declaring that one treatment is better than the other are given in the previous section.

Under H**0** (Scenario 0 in Table 4a), where IGAPT is assumed to produce the same 1-year treatment failure rate (40%) as IGAXT, there is a 4.9% chance of early stopping because IGAPT is better than IGAXT and a 5.3% chance of early stopping because IGAXT is better. The 2.5th, 50th, and 97.5th percentiles for the proportion of patients randomized into the IGAPT group are 37%, 50%, and 65%, respectively. The median sample sizes are 75 and 74 for the IGAPT and IGAXT groups, respectively. There is an 8.7% chance to claim IGAPT as being better and an 8.1% chance to claim IGAXT as being better at the end of the study. The average trial length is 3.70 years.

Under H**1 (**Scenario 1 in Table 4a), where IGAPT is assumed to produce lower 1-year treatment failure rates than IGAXT (25% versus 40%), there is a 48% chance of early stopping because IGAPT is better than IGAXT, but there is virtually no chance of early stopping because IGAXT is better. The 2.5th, 50th, and 97.5th percentiles for the proportion of patients randomized into the IGAPT group are 44%, 58%, and 73%, respectively. The median sample sizes are 74 for the IGAPT group and 54 for the IGAXT group. There is an 81% chance to claim IGAPT as being better and virtually no chance to claim IGAXT as being better at the end of the study. The average trial length is 2.70 years.

Additional simulations summarized in Table 4a show that we have more than a 90% chance to claim IGAPT as being better and virtually no chance to claim IGAXT as being better at the end of the study for a situation where the treatment failure rate decreases with dose for both IGAXT and IGAPT (Scenario 2). We also have more than an 80% chance to claim IGAPT as being better and virtually no chance to claim IGAXT as being better at the end of the study for a situation where the difference in treatment failure rates between IGAXT and IGAPT decreases with dose (Scenario 3).

Simulations summarized in Table 4b show that with 10% of patients randomized to IGAPT and then denied insurance coverage for their treatment, we still have at least 80% power for Scenarios 1-3 and about a 10% significance level under Scenario 0. With 10% of patients randomized to IGAPT and then denied insurance coverage for their treatment the trial will need to enroll a few more patients on average, it will last a little longer, and the proportion of patients treated on IGAPT will be a 2%-3% less than if there was no insurance denial.

Simulations summarized in Table 4c show that with 25% of patients randomized to IGAPT and then denied insurance coverage for their treatment, we still have at least 80% power for Scenarios 1 and 2, about 80% power for Scenario 3, and about a 10% significance level under Scenario 0. With 25% of patients randomized to IGAPT and then denied insurance coverage for their treatment the trial will need to enroll a few more patients on average, it will last a little longer, and the proportion of patients treated on IGAPT will be a 7%-8% less than if there was no insurance denial.

In both cases of 10% or 25% insurance denial, the trial will have only modest (<10%) increases in the total sample size and study duration.

Table 4. Operating characteristics (n=150 eligible patients). Assumes 90% of patients at 74 CGE and 10% of patients at 66 CGE. Also assumes 10% of patients assigned to IGAPT dropped due to insurance denial.

		Scenario 0 (H_0)		
True 1-Year Treatment Failure Rates for Doses 74, 66 CGE:		IGAXT 0.40, 0.40	IGAPT 0.40, 0.40	Total Sample Size (including insurance denial)
	Percentile			
Distribution of number of patients after adaptive randomization	2.5%	33	27	74
	50.0%	77	70	156
	97.5%	96	90	163
Distribution of the proportion of patients randomized to IGAPT	2.5%		0.33	
	50.0%		0.47	
	97.5%		0.63	
Probability of early stopping		0.050	0.059	
Probability of final selection		0.084	0.104	
Average trial length (years)		3.77		
		Scenario 1 (H_1)		
True 1-Year Treatment Failure Rates for Doses 74, 66 CGE:		IGAXT 0.40, 0.40	IGAPT 0.25, 0.25	Total Sample Size (including insurance denial)
	Percentile			
Distribution of number of patients after adaptive randomization	2.5%	26	38	76
	50.0%	58	71	153
	97.5%	86	95	163
Distribution of the proportion of patients randomized to IGAPT	2.5%		0.41	
	50.0%		0.56	
	97.5%		0.71	
Probability of early stopping		0.0	0.473	
Probability of final selection		0.0	0.811	
Average trial length (years)		2.80		
		Scenario 2		
True 1-Year Treatment Failure Rates for Doses 74, 66 CGE:		IGAXT 0.40, 0.30	IGAPT 0.25, 0.15	Total Sample Size (including insurance denial)
	Percentile			
Distribution of number of patients after adaptive randomization	2.5%	27	41	81
	50.0%	58	69	153
	97.5%	90	94	163
Distribution of the proportion of patients randomized to IGAPT	2.5%		0.39	
	50.0%		0.55	
	97.5%		0.71	
Probability of early stopping		0.0	0.493	
Probability of final selection		0.0	0.815	
Average trial length (years)		2.76		
		Scenario 3		
True 1-Year Treatment Failure Rates for Doses 74, 66,CGE:		IGAXT 0.40, 0.40	IGAPT 0.20, 0.25	Total Sample Size (including insurance denial)
	Percentile			
Distribution of number of patients after adaptive randomization	2.5%	26	40	81
	50.0%	46	68	117
	97.5%	82	96	162
Distribution of the proportion of patients randomized to IGAPT	2.5%		0.43	
	50.0%		0.58	
	97.5%		0.73	
Probability of early stopping		0.0	0.692	
Probability of final selection		0.0	0.952	
Average trial length (years)		2.26		

13.2.5. Statistical analyses and additional statistical considerations

Standard statistical methods including descriptive statistics and exploratory data analysis will be applied for checking data quality; identifying outliers, patterns, or associations; and providing summaries of the data distribution. Patients' medical-demographic information at baseline will be tabulated by treatment group to assess comparability between the IGAXT and IGAPT groups. Student's *t* tests or Wilcoxon rank sum tests will be used to compare continuous variables between the two different groups. Chi-square tests or Fisher's exact tests will be applied to assess the association between two categorical variables. Time-to-event outcomes, including overall survival, progression-free survival, time to radiation esophagitis, time to treatment-related pneumonitis (TRP), time to local failure, as well as time to treatment failure, will be estimated by using the Kaplan-Meier method. Log-rank tests will be used to test the difference in time-to-event distributions between treatment groups. Cox proportional hazards models will be used for multi-covariate time-to-event analysis to test the treatment effect adjusted by other important covariates such as dose and inflammatory cytokines. In addition, the association of relevant pharmacogenetic endpoints and gene polymorphisms with the time to the development of TRP or local failure will also be analyzed by the Cox model and other exploratory tools. Toxicity data will be summarized using frequency tables. Associations between the types and severity of toxicity and treatment will be evaluated as well. We will also perform subset analyses by dose level. Patients who are not randomized (Group 3 and Group 4) will be analyzed using descriptive statistics separately from those who are randomized. Patients in Group 3 and Group 4 will be analyzed separately. Exploratory analyses may also be performed by combining Group 3 and Group 4.