Example Wording for Protocol

In making the decision to recommend termination of the study, the Data Safety Monitoring Board shall be guided by a formal stopping rule based on the primary endpoint of 1 year mortality from all causes. The test statistic shall be the normal approximation to the difference in binomial proportions between the treatment and control groups. Additional secondary analyses which adjust for covariates will be performed using logistic regression.

The clinical trial may be stopped early either for reasons of demonstrated efficacy (the DrugX treatment arm has significantly lower 1 year mortality than the placebo arm) or for reasons of futility (the 1 year mortality on the DrugX treatment arm is not sufficiently lower than that on the placebo arm to warrant continuation of the trial).

The formal stopping boundaries will be determined by a symmetric one-sided design (Emerson and Fleming, *Biometrics*, 1989), a family also included in the unified family of group sequential stopping rules (Kittelson and Emerson, *Biometrics*, 1999). In the notation of the latter paper, the stopping rule will be based on a one-sided group sequential design testing a lower alternative hypothesis at a level of significance $\alpha = .025$ with $\beta = .975$, an upper (futility) stopping boundary relationship specified by $P_d = 0.8$ (a boundary relationship intermediate to the O'Brien-Fleming (1979) type boundary and the Pocock (1977) type boundary), and a lower (efficacy) stopping boundary relationship specified by $P_a = 0.8$ (a boundary intermediate between O'Brien-Fleming and Pocock boundaries). It is envisioned that four equally spaced formal interim analyses will be performed during the monitoring of the study. Under such a monitoring schedule and assuming a baseline 1 year mortality rate of 24% on the placebo arm, a sample size of 2,522 patients (1,261 patients on each of the placebo and treatment arms) will provide 80% power to detect a 20% relative reduction in 1 year mortality (i.e., an absolute difference in 1 year mortality rates of .048). The following table provides a more detailed description of the power provided by such a sample size for a range of baseline 1 year mortality rates.

Table 1: Alternatives for which a sample size of 2,522 subjects provides the specified power as a function of placebo arm 1 year mortality rates.

	18% Placebo Mortality			24% Placebo Mortality			30% Placebo Mortality		
Power	DrugX Mort	Abs Diff	Rel Diff	DrugX Mort	Abs Diff	Rel Diff	DrugX Mort	Abs Diff	Rel Diff
50%	.1501	-.0299	-17%	.2064	-.0336	-14%	.2635	-.0365	-12%
80%	.1374	-.0426	-24%	.1920	-.0480	-20%	.2480	-.0520	-17%
90%	.1307	-.0493	-27%	.1845	-.0555	-23%	.2398	-.0602	-20%
95%	.1251	-.0549	-30%	.1782	-.0618	-26%	.2330	-.0670	-22%
97.5%	.1202	-.0598	-33%	.1727	-.0673	-28%	.2271	-.0729	-24%

Under the planned schedule of four equally spaced analyses and assuming a baseline placebo 1 year mortality rate of 24%, Table 2 presents the stopping boundaries at each analysis for the specified stopping rule expressed as the absolute difference in 1 year mortality rates (DrugX – Placebo). Also presented are the Z statistics and fixed sample lower one-sided P values which correspond to those stopping boundaries.

Table 2: Stopping boundaries for a level .025 one-sided symmetric design with $P_d = P_a = 0.8$, four equally spaced analyses, a maximal sample size of 2522 subjects, and a 1 year mortality rate of approximately 21.6% on both treatment arms combined.

		Efficacy (lower) stopping boundary			Futility (upper) stopping boundary		
Analysis	Sample Size	Abs Diff (%)	Z statistic	Fixed P (lower)	Abs Diff	Z statistic	Fixed P (lower)
1	630	-10.20	-3.117	.0009	3.47	1.061	.8556
2	1,261	-5.86	-2.532	.0057	-0.87	-0.376	.3533
3	1,892	-4.24	-2.242	.0125	-2.49	-1.320	.0934
4	2,522	-3.36	-2.057	.0199	-3.36	-2.057	.0199

Thus, according to the above table, if the mortality rate on the combined treatment arms is 21.6%, an absolute difference in mortality of –10.20% or less (e.g., 26.7% 1 year mortality on the placebo arm and 16.5% 1 year mortality on the DrugX arm) when 630 subjects have been accrued to the study (315 subjects on each arm), the stopping rule would suggest that the study be terminated early with a decision that treatment with DrugX results in a statistically significant improvement in 1 year mortality. On the other hand, if at that first analysis there were an absolute difference in mortality of 3.47% or more (e.g., 19.86% 1 year mortality on the placebo arm and 23.53% 1 year mortality on the DrugX arm), the stopping rule would suggest that the study be terminated early with a decision that it was futile to continue the trial because there was not sufficient evidence that any beneficial effect of DrugX was clinically important.

As with the power properties of the study, the absolute difference in mortality rates which correspond to the stopping boundaries is affected by the baseline 1 year mortality rate on the placebo arm as well as any effect of DrugX on 1 year mortality rates. Due to the need to estimate the baseline mortality rates when performing the statistical test, the expression of the stopping boundaries as an absolute difference in mortality rates thus depends on the overall mortality rates on both arms combined. Table 3 presents these correspondences for the specified stopping rule used with a maximal sample size of 2,522 subjects for selected 1 year mortality rates for the combined arms. (Note that the stopping boundary expressed as a Z statistic or a lower fixed sample P value statistic is unaffected by the observed mortality rates, and thus the values given in Table 2 for those statistics pertain to the cases presented in Table 3 as well.)

Table 3: Stopping boundaries expressed as an absolute difference in 1 year mortality rates (DrugX – Placebo) for a level .025 one-sided symmetric design with $P_d = P_a = 0.8$, four equally spaced analyses, and a maximal sample size of 2522 subjects for selected observed mortality rates on both treatment arms combined.

		15.9% Combined Mortality		27.4% Combined Mortality	
Analysis	Sample Size	Efficacy boundary	Futility boundary	Efficacy boundary	Futility boundary
1	630	-9.06%	3.08%	-11.05%	3.76%
2	1,261	-5.20%	-0.77%	-6.35%	-0.94%
3	1,892	-3.76%	-2.21%	-4.59%	-2.70%
4	2,522	-2.99%	-2.99%	-3.65%	-3.65%

Table 4 presents for the setting presented in Table 2 (i.e., a combined mortality rate of 21.6%) the statistical inference that would be reported if the study were to result in

observed treatment effects corresponding to the stopping boundaries. The estimates, P values, and confidence intervals reported in Table 4 have been adjusted for the stopping rule. (Note that the fixed sample P value presented in Table 2 is not appropriate for statistical inference. Instead it is presented to facilitate the use of standard statistical software when computing the test statistic: Lower one-sided P values calculated by standard statistical software (i.e., as would be appropriate for nonsequential fixed sample studies) can be compared to the critical values for that statistic as presented in Table 2 in order to obtain a level .025 sequential hypothesis test.)

Table 4: Statistical inference regarding the effect of DrugX on 1 year mortality (measured as the absolute difference in mortality rates between the DrugX and placebo arms) which would be reported if observed results corresponded exactly to the stopping boundaries for a level .025 one-sided symmetric design with $P_d = P_a = 0.8$, four equally spaced analyses, a maximal sample size of 2522 subjects, and a 1 year mortality rate of approximately 21.6% on both treatment arms combined as presented in Table 2.

		Efficacy (lower) stopping boundary			Futility (upper) stopping boundary		
Analysis	Sample Size	Adjusted estimate	Exact 95% conf intvl	Adjusted P value	Adjusted estimate	Exact 95% conf intvl	Adjusted P value
1	630	-9.5%	(-14.3%, -3.7%)	.001	2.7%	(-3.0%, 7.6%)	.833
2	1,261	-5.3%	(-9.2%, -1.2%)	.007	-1.4%	(-5.6%, 2.5%)	.242
3	1,892	-4.0%	(-7.2%, -0.3%)	.017	-2.7%	(-6.4%, 0.5%)	.048
4	2,522	-3.4%	(-6.7%, 0%)	.025	-3.4%	(-6.7%, 0%)	.025

As noted above, the exact stopping boundaries that are appropriate for the group sequential design will depend upon the exact schedule of interim analyses and the best estimate of the variability of the test statistic as computed from the observed 1 year mortality rates. The intended schedule of interim analyses is four equally spaced analyses subject to minor variations due to the availability of the DSMB. An exception to this schedule may be implemented if during the course of this study the results of an independent study of DrugX become available. In that case, an interim analysis will be planned to allow for an analysis just after the results of that study are made public. In any case, the number and timing of interim analyses of the data for this trial will not be determined by the interim results of this study for the primary endpoint.

Modifications of the stopping rule to account for changes in the schedule of interim analyses and estimates of baseline mortality rates will be made by using the parametric form of the stopping rule as specified above, with constraints imposed for analyses previously performed. Boundaries will be constrained on the scale of the maximum likelihood estimate of the treatment effect, with the current best estimate of the test statistic's variance used at each analysis (*S+SeqTrial User's Manual*, Data Analysis Products Division, MathSoft, Seattle WA, 2000). The one-sided type I error will be maintained at .025, and the maximal sample size will be constrained at 2,522 subjects.

At each formal interim analysis, the DSMB will use the stopping rule computed in the above manner as a guideline in evaluating the trial results with respect to 1 year mortality from all causes. In making a recommendation to terminate the study, the DSMB will of course also consider information on safety endpoints, as well as consistency of outcomes for secondary endpoints and consistency of outcomes within important subgroups as described in the protocol.