Template for time-to-event group sequential design

Keaven M. Anderson

January 2, 2014

1 Introduction

This vignette provides a template for time-to-event sample size calculations for fixed designs using `nSurv` and group sequential designs using `gsSurv`. Edit this template (from gsDesign R package source `/inst/doc/gsSurvTemplate.rnw`) so that you can reuse it on a regular basis for sample size calculations for time-to-event study planning. The template only uses the simplest options with a single stratum and exponential failure and dropout rates. The template can be modified to accomodate multiple strata and/or piecewise exponential failure and dropout rates; this was not chosen here since the simplest options are a) often used and b) simplest to learn and apply for beginners. Note that we produce tabular, textual and graphical output; examining the source file to see how this is done will enable you to easily customize to fit your purposes. You will need the knitr R package, which I find simpler to use than Sweave (although you could make minor edits and use Sweave). I have found using knitr and the RStudio development environment to be a good combination. Within this template, we suppress printing of all of the code used to generate results. The file `/inst/doc/gsSurvTemplateInstructions.pdf` is an alternate version of this document that shows code associated with this template.

We apply the Lachin and Foulkes (1986) sample size method and extend it to group sequential design. This method fixes the duration of a study and varies enrollment rates to power a trial. Alternate text when enrollment rate is fixed and enrollment duration is allowed to vary: *We use the Lachin and Foulkes (1986) basic power equation to compute sample size along the lines of Kim and Tsiatis (1990) where enrollment rates are fixed and enrollment duration is allowed to vary to enroll a sufficient sample size to power a study.* The sample size method assumes proportional hazards between treatment groups. Accrual of events of events over time is approximated assuming failure rates are exponential.

2 Introduction

2.1 Fixed design sample size

This section can be deleted if you are only interested in group sequential design. The median time-to-event is assumed to be 12 months in the control group. The trial is designed to demonstrate superiority of experimental treatment over control with an assumed hazard ratio of 0.75 with 80% power and a one-sided Type I error rate of 2.5. The total sample size is 580 and a total of 379 endpoints is required for analysis. Planned recruitment duration is 24 months and the minimum follow-up planned is 12 months. Thus, the total expected study duration is 36 months. Enrollment is assumed to be piecewise constant at rates of 6.9 for months 0 - 1, 10.4 for months 1 - 3, 17.3 for months 3 - 6, 27.7 for months 6 - 24. The assumed dropout rate is 0.1% per month.

3 Group sequential design

For a comparative trial we consider a 2-arm group sequential design with overall survival as the primary endpoint as shown in Table 1 (a second version of the table is in 2. Timing, number of events, sample size, boundaries (Z-values, nominal p-values, approximate hazard ratios) are shown as well as the probability of crossing study boundaries under the null and alternate hypotheses. Bounds are determined by Hwang-Shih-DeCani spending functions with $\gamma = -10$ (α-spending) and $\gamma = 2$ (β-spending). The median time-to-event is assumed to be 12 months in the control group. The trial is designed to demonstrate superiority of experimental treatment over control with an assumed hazard ratio of 0.75. The total sample size is 701 and a total of 459 endpoints is required for the final analysis. Planned recruitment duration is 24 months and the minimum follow-up planned is 12 months. Thus, the total expected study duration is 36 months. Enrollment is assumed to be piecewise constant at rates of 8.4 for months 0 - 1, 12.6 for months 1 - 3, 21 for months 3 - 6, 33.6 for months 6 - 24. The assumed dropout rate is 0.1% per month. There is a single interim analysis planned after 184 events have accrued which is expected after approximately 19.9 months.

Following are plots of the Z-values (Figure 1) and approximate hazard ratios (Figure 2) at the design bounds.

References

Kyungmann Kim and Anastasios˜A. Tsiatis. Study duration for clinical trials with survival response and early stopping rule. *Biometrics*, 46:81–92, 1990.

John˜M. Lachin and Mary˜A. Foulkes. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics*, 42:507–519, 1986.

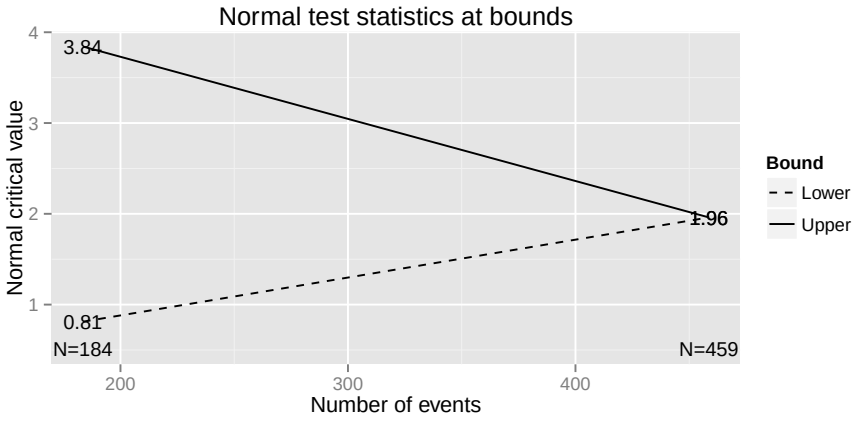

Figure 1: Z-value bound plot

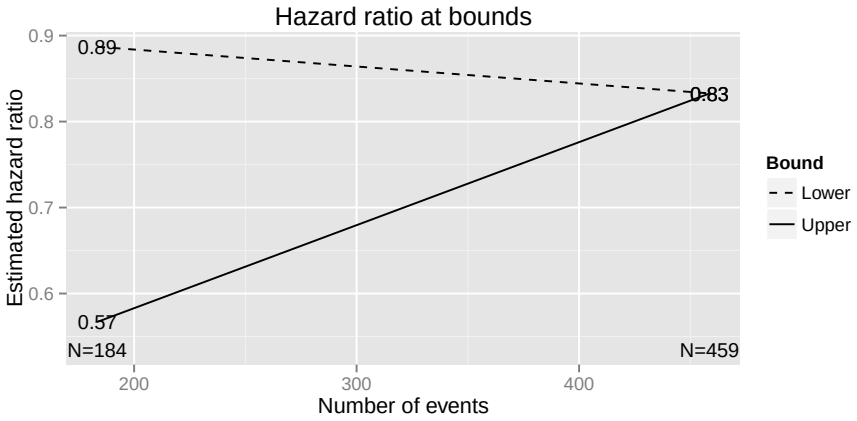

Figure 2: Hazard ratio bound plot

Analysis	Value	Futility	Efficacy
IA 1: 40%	Z-value	0.81	3.84
N: 563	HR	0.89	0.57
Events: 184	p (1-sided)	0.2084	1e-04
19.9 months	P{Cross} if HR=1	0.7916	1e-04
	P{Cross} if HR=0.75	0.1274	0.0293
Final analysis	Z-value	1.96	1.96
N: 701	HR	0.83	0.83
Events: 459	p (1-sided)	0.025	0.025
36 months	P{Cross} if HR=1	0.9801	0.0199
	P{Cross} if HR=0.75	0.2	0.8

P{Cross} is the probability of crossing the given bound (efficacy or futility) at or before the given analysis under the assumed hazard ratio (HR). Design assumes futility bound is discretionary (non-binding), but smaller upper boundary crossing probabilities shown here assume trial stops at first boundary crossing (binding bounds).

Table 1: Overall survival trial design with HR=0.75, 80% power and 2.5% Type 1 error.

Analysis	Value	Efficacy	Futility
IA 1: 40%	Z	3.8427	0.8119
N: 564	p (1-sided)	0.0001	0.2084
Events: 184	HR at bound	0.5670	0.8870
Month: 20	P(Cross) if HR=1	0.0001	0.7916
	P(Cross) if HR=0.75	0.0293	0.1274
Final	Z	1.9602	1.9602
N: 702	p (1-sided)	0.0250	0.0250
Events: 459	HR at bound	0.8327	0.8327
Month: 36	P(Cross) if HR=1	0.0199	0.9801
	P(Cross) if HR=0.75	0.8000	0.2000

Table 2: Asymmetric two-sided group sequential design with non-binding futility bound, 2 analyses, time-to-event outcome with sample size 701 and 459 events required, 80 percent power, 2.5 percent (1-sided) Type I error to detect a hazard ratio of 0.75. Enrollment and total study durations are assumed to be 24 and 36 months, respectively. Efficacy bounds derived using a Hwang-Shih-DeCani spending function with gamma=-10. Futility bounds derived using a Hwang-Shih-DeCani spending function with gamma=2.