Some Sample Size and Power Considerations for Clinical Studies

Some factors increasing the power to detect a difference in outcomes between two treatments, and lowering the sample size needed to achieve a given power:

1. Large total number of subjects
2. Large true treatment effect
3. Binary outcome incidence close to ½
      Exclude low-risk patients from study
      Exclude doomed patients
4. Large total number of events in a follow-up study (probability of event × number of subjects)
      Can be achieved by longer follow-up
5. Use an ordinal or continuous endpoint instead of a binary one
      For ordinal responses, the more categories that commonly occur the better
6. Equal allocation ratio, for
      Continuous responses with equal variance in both treatment groups
      Binary endpoints that don't have extreme distributions, when treatment effect is large
7. Low variability in responses across subjects (for continuous responses)
8. Higher type I error
9. Not having to adjust α downward for multiple comparisons
      Prioritize endpoints, report both "positive" and "negative" results, or be Bayesian
10. Low proportion of dropouts (non-compliers) and drop-ins (crossover from placebo to active arm)
11. Pattern of treatment effect over time matches that assumed by the statistical model to be used in the primary analysis
      Cox-logrank survival analysis method assumes equal effect of treatment over time
      Log-normal survival model assumes treatment effect decreases over time
      Some drugs take time to begin working
      Some treatments wear off (e.g., coronary bypass surgery – in long-term patients' native disease advances)
      For ordinal clinical outcome scales, think about whether the treatment is likely to have the same effect for all levels of the outcome scale.
12. Covariable adjustment for strong prognostic factors

Methods for estimating power and sample size:

1. Simple model, willing to assume that the model will fit the data (e.g., 2-sample t-test, Cox-logrank, 2-sample binomial)
      Exact or approximate formulas (S-Plus Statistics menu for 1- and 2-sample t-test, S-Plus Hmisc library `bpower, bpower.sim, bsamsize, cpower, popower, ballocation` functions)
2. Allowance for model not to fit data perfectly, including problems with compliance and delayed treatment effect
      Approximations based on modifying "pure" formulas (`cpower` function)
      Simulation (`spower` function)

Recommendations for including power and sample size estimates in proposals and study plans:

1. Above all, get an estimate of the treatment effect that is clinically worthwhile to detect and an estimate of variance and compliance **before** showing any investigators any of the calculations
2. There is a strong tendency for investigators to exaggerate the expected treatment effect to lower the sample size needed
   Don't choose a treatment effect that is **more than clinically important** to patients
   Result would be an inability to detect a clinically meaningful effect
   Confidence interval for treatment effect will be so wide as to be useless
3. Instead of computing one power or one sample size, compute them for a plausible **range** of treatment effects, compliance, variance, etc. Results are best presented as graphs rather than tables.

Example 1: Acute study, binary endpoint. Probability of a clinical endpoint in control arm is 0.1. Compute the power to detect various treatment effects (odds ratio .2, .3, …, .9) for sample sizes from 10 to 1000 by increments of 10.

```
n <- seq(10, 1000, by=10)
OR <- seq(.2, .9, by=.1)
pow <- lapply(OR, function(or,n)
               list(x=n, y=bpower(p1=.1, odds.ratio=or, n=n)), n=n)
names(pow) <- format(OR)
labcurve(pow, pl=T, xlab='n', ylab='Power')
abline(h=.9, lty=2)
```
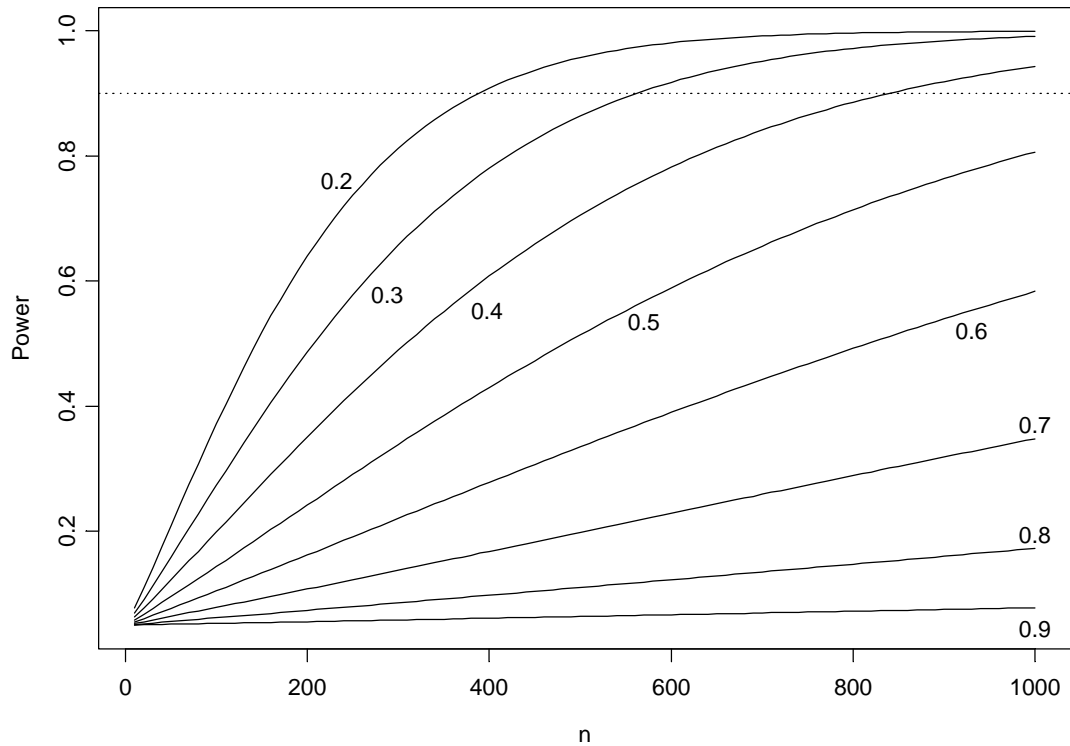
Another useful graph is a contour plot for various probabilities of outcome in the control group, fixing the treatment effect (odds ratio), here at .8 ([p2/(1-p2) / p1/(1-p1)] = .8). We also allow n to vary.

```
p1 <- seq(.01,.99,by=.01)
n  <- seq(100,5000,by=250)
pow <- outer(p1, n, function(p1,n) bpower(p1, n=n, odds.ratio=.8))
# This forms a length(p1)*length(n) matrix of power estimates
contour(p1, n, pow)
```
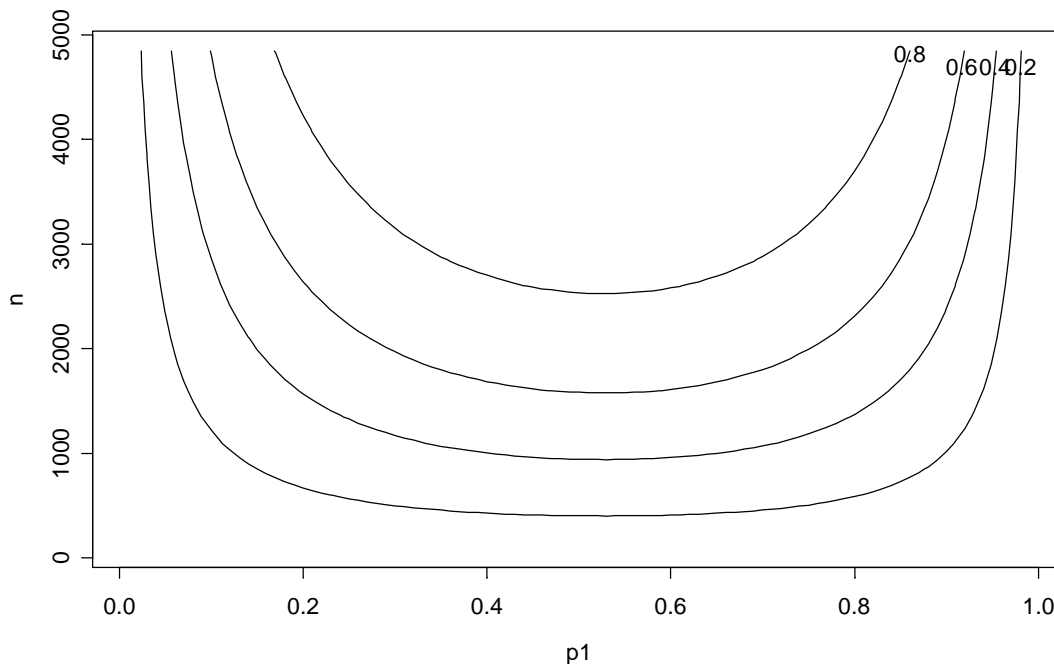
Example 2: Long-term follow-up study, endpoint is time until death, all causes. In this example, 4 plots are drawn on one page, one plot for each combination of noncompliance percentages. Within a plot, the 5-year mortality % in the control group is on the x-axis, and separate curves are drawn for several % reductions in mortality with the intervention. The accrual period is 1.5y, with all patients followed at least 5y and some 6.5y. There is a total of 14,000 patients in the study.

```
par(mfrow=c(2,2),oma=c(3,0,3,0))

morts <- seq(10,25,length=50)  # 5-year control mortalities
red <- c(10,15,20,25)              # Hypothesized % reductions in mortality

for(noncomp in c(0,10,15,-1)) {
  if(noncomp>=0) nc.i <- nc.c <- noncomp else {nc.i <- 25; nc.c <- 15}
  z <- paste("Drop-in ",nc.c,"%, Non-adherence ",nc.i,"%",sep="")
  plot(0, 0, xlim=range(morts), ylim=c(0,1),
       xlab="5-year Mortality in Control Patients (%)",
       ylab="Power", type="n")
  title(z)
  lty <- 0
  for(r in red) {
        lty <- lty+1
        power <- morts
        i <- 0
        for(m in morts) {
```
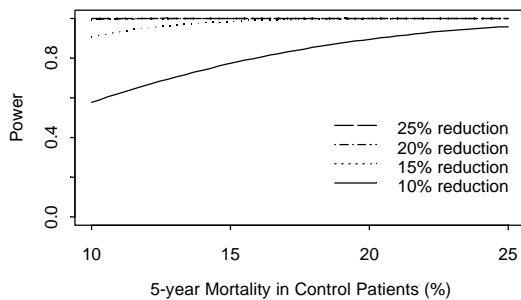
```
            i <- i+1
            power[i] <- cpower(5, 14000, m/100, r, 1.5, 5,
                               nc.c, nc.i, pr=F)
        }
        lines(morts, power, lty=lty)
    }
    if(noncomp==0)legend(18,.55,rev(paste(red,"% reduction",sep="")),
                        lty=4:1,bty="n")
}
mtitle("Power vs Non-Adherence for Main Comparison",
       ll="alpha=.05, 2-tailed, Total N=14000",cex.l=.8)
```
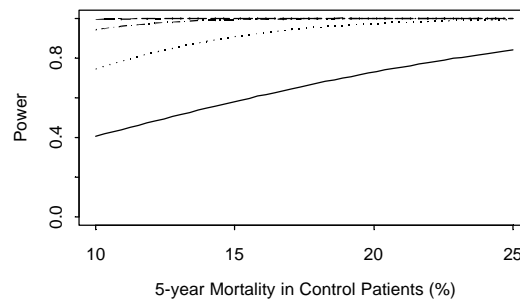
Power vs Non-Adherence for Main Comparison
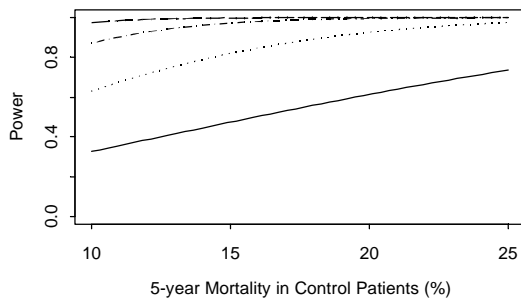
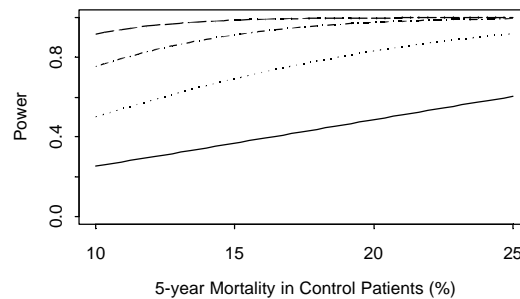Drop-in 0%, Non-adherence 0%   Drop-in 10%, Non-adherence 10%

Drop-in 15%, Non-adherence 15%   Drop-in 15%, Non-adherence 25%

alpha=.05, 2-tailed, Total N=14000   Wed Feb 17 14:27:20 1999

Example 3: Using simulation to estimate exact power in a complex situation.

Here we use the `spower` function to simulate a 2-arm 5-year follow-up study (350 patients per arm) for which the control group's survival distribution is Weibull with 1-year survival of .95 and 3-year survival of 0.7. All subjects are followed at least one year, and patients enter the study with linearly increasing probability after that. Assume there is no chance of drop-in for the first 6 months, then the probability increases linearly up to .15 at 5 years. Assume there is a linearly increasing chance of dropout up to .3 at 5 years. Assume that the treatment has no effect for the first 9 months, then it has a constant effect (hazard ratio of .75).

```
# First find the right Weibull distribution for
# compliant control patients
```

```
sc <- Weibull2(c(1,3), c(.95,.7))

# Specify censoring distribution
rcens <- function(n) 1 + (5-1) * (runif(n) ^ .5)

# Put it all together

f <- Quantile2(sc,
      hratio=function(x)ifelse(x<=.75, 1, .75),
      dropin=function(x)ifelse(x<=.5, 0, .15*(x-.5)/(5-.5)),
      dropout=function(x).3*x/5)

par(mfrow=c(2,2))
plot(f, 'all', label.curves=list(keys='lines'))
```
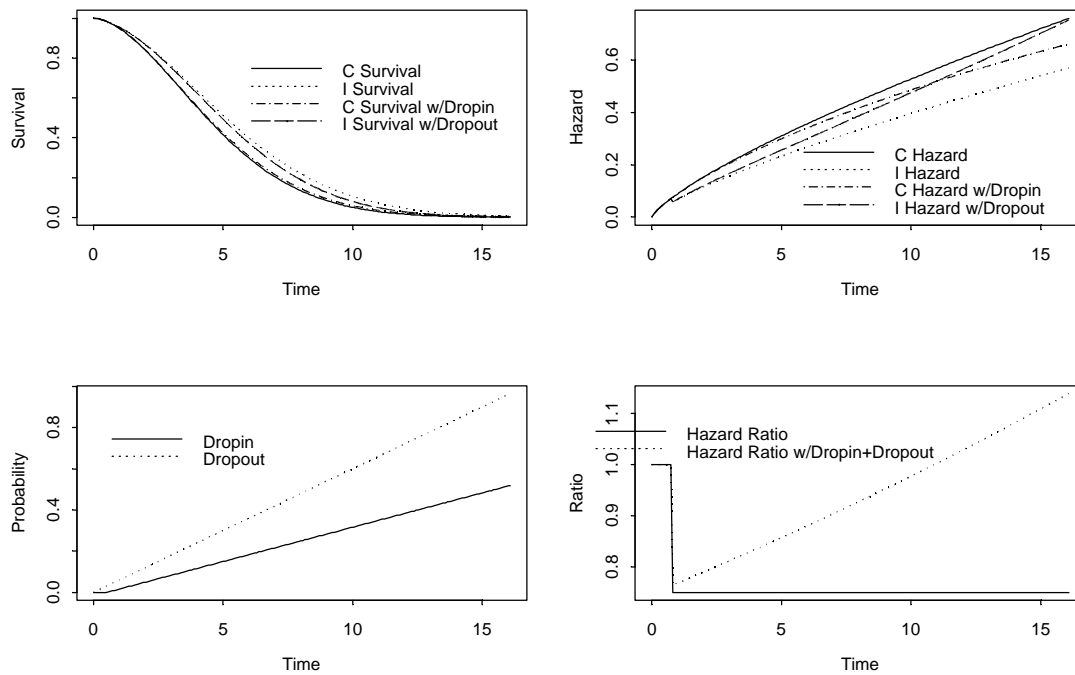


```
rcontrol <- function(n) f(n, 'control')
rinterv  <- function(n) f(n, 'intervention')

set.seed(211)
spower(rcontrol, rinterv, rcens, nc=350, ni=350,
      test=logrank, nsim=300)
[1] 0.4033333
```

The power for an 0.05-level Cox-logrank test for this situation is thus 0.40.