

# survSNP: Power and Sample Size Calculations for SNP Association Studies with Censored Time to Event Outcomes

Kouros Owzar      Zhiguo Li      Nancy Cox      Sin-Ho Jung  
Chanhee Yi

August 6, 2012

## 1 Introduction

This vignette serves as a tutorial for using the `survSNP` extension package for conducting power and sample size calculations for SNP association studies with censored time to event outcomes. We begin by loading the package.

```
> library(survSNP)
```

## 2 Example 1

In this example, we conduct power calculations for a SNP for which the relative allelic frequency for the risk allele  $B$  is  $q = 0.1$ . The validation data set is assumed to consist of  $n = 500$  patients. The event rate (e.g., death rate) is assumed to be  $\rho = 0.75$  (i.e., 375 events among 500 patients). We hypothesize an effect size (genotype hazard ratio)  $\Delta = 1.25$ . Finally, we assume that the median in the population is 1 unit of time (i.e.,  $P(\tilde{T} > t) = 0.5$ ). To find the asymptotic power, at the two-sided  $\alpha = 0.05$  level, the `sim.snp.expsurv.power` function is used.

```
> res1<-sim.snp.expsurv.power(1.25, n=500, raf=0.1, erate=0.75,  
+                               pilm=0.5, lm=1, B=0,  
+                               model="additive", test="additive", alpha=0.05)  
>
```

Below, we show some of the relevant columns from the output.

```
> res1[,c("n", "GHR", "erate", "raf", "B", "alpha", "pow0", "pow", "powB")]
```

```
      n  GHR erate raf B alpha      pow0 pow powB
power 500 1.25  0.75 0.1 0  0.05 0.4706994 NA  NA
```

The asymptotic power, at the two-sided  $\alpha = 0.05$  level, is 0.47. Note that we are assuming that the median for the survival function in the population is 1 unit of time (that is why we have set `pilm=0.5` and `lm=1`). If we had desired to set power the study based on a population whose 0.7 quantile is say 2 units of time, we would have set `pilm=0.7` and `lm=2`. The current version of the package supports power calculations for additive models.

By default, the asymptotic power based on the approximate asymptotic variance formula is computed. The corresponding column name is `pow0` in the output shown above. The asymptotic power based on the exact variance formula (column `pow`) or the empirical power (column `powB`) can be computed by setting the arguments `exactvar=TRUE` or `B=b` where `b` is a positive integer. Be sure to set a random number generation seed when calculating the empirical or the asymptotic power based on the exact variance formula to get reproducible results. We illustrate these features next. This example is based on  $B = 10$  simulation replicates.

```
> set.seed(123)
> res1b<-sim.snp.expsurv.power(1.25, n=500, raf=0.1, erate=0.75,
+                               pilm=0.5, lm=1,
+                               exactvar=TRUE, B=B,
+                               model="additive", test="additive", alpha=0.05)
>
```

The results are shown below.

```
> res1b[,c("n", "GHR", "erate", "raf", "B", "alpha", "pow0", "pow", "powB")]
```

```
      n  GHR erate raf      B alpha      pow0      pow      powB
power 500 1.25  0.75 0.1 10000  0.05 0.4706994 0.4667144 0.4691
```

### 3 Example 2

For power calculations for SNP association studies with censored outcomes, it is generally desired to vary the effect size, sample size, event rate and the relative minor allele frequency. These are denoted by the variable names `GHRs`, `ns`, `rafs` and `erates` in this example.

```
> GHRs<-seq(1.05,1.5,by=0.05)
> ns<-c(100,500,700)
> rafs<-c(0.1,0.3,0.5)
> erates<-c(0.5,0.7,0.9)
```

The function `survSNP.power.table` can be used to generate power calculations for this combination of parameters. This is a wrapper function for `sim.snp.expsurv.power`.

```
> res2<-survSNP.power.table(GHRs,ns,rafs,erates,
+                           pilm=0.5,lm=1,
+                           model="additive",test="additive",
+                           alpha=0.05)
```

We print selected columns from the first three rows of the previous object next.

```
> res2[1:3,c("n","GHR","erate","raf","pow0","pow","powB")]
```

	n	GHR	erate	raf	pow0	pow	powB
power	100	1.05	0.5	0.1	0.05719290	NA	NA
power1	100	1.05	0.7	0.1	0.05819468	NA	NA
power2	100	1.05	0.9	0.1	0.05917498	NA	NA

Next, we will consider illustrating the previous set of results using the `lattice` package. The power is illustrated in Figure 1. A revised version of this illustration limiting the presentation to  $n=100$  and displaying the type I error rate  $\alpha$  is shown in Figure 2.

## 4 Example 3

We can also use the `xtable` package to summarize the results in a table. For this illustration, we consider a subset of the rows from Example 2.

```
> cols<-c("n","GHR","erate","raf","pow0")
> res3<-subset(res2,GHR==1.25&raf==0.3&n==500,select=cols)
> res3
```

	n	GHR	erate	raf	pow0
power314	500	1.25	0.5	0.3	0.6407236
power1114	500	1.25	0.7	0.3	0.7694030
power2114	500	1.25	0.9	0.3	0.8515087

The corresponding table generated by `xtable` is shown in Table~1.

```
> print(xtable(res3,digits=c(0,0,1,1,1,3)),
+       include.rownames=FALSE,floating=FALSE)
```

n	GHR	erate	raf	pow0
500	1.2	0.5	0.3	0.641
500	1.2	0.7	0.3	0.769
500	1.2	0.9	0.3	0.852

Table 1: Tabular summary of the results from Example 2

## 5 Miscellaneous

The `tikzDevice` and `latticeExtra` packages can be used to considerably refine the illustrations. The software development repository provides some examples [bitbucket.org/kowzar/survsnr/](https://bitbucket.org/kowzar/survsnr/).

The session information is provided in Table~2

```

> KEY=paste("q=",levels(factor(res2$raf)),sep="")
> KEY<-list(lines=list(col=1:length(KEY),lty=1:length(KEY)),
+           text=list(labels=paste("q=",levels(factor(res2$raf)),sep="")),
+           column=3)
> print(xyplot(pow0~GHR|factor(erate)*factor(n),group=factor(raf),
+           data=res2,type="l",lty=KEY$lines$lty,col=KEY$lines$col,
+           key=KEY,
+           xlab="Genotype Hazard Ratio",ylab="Power"))

```

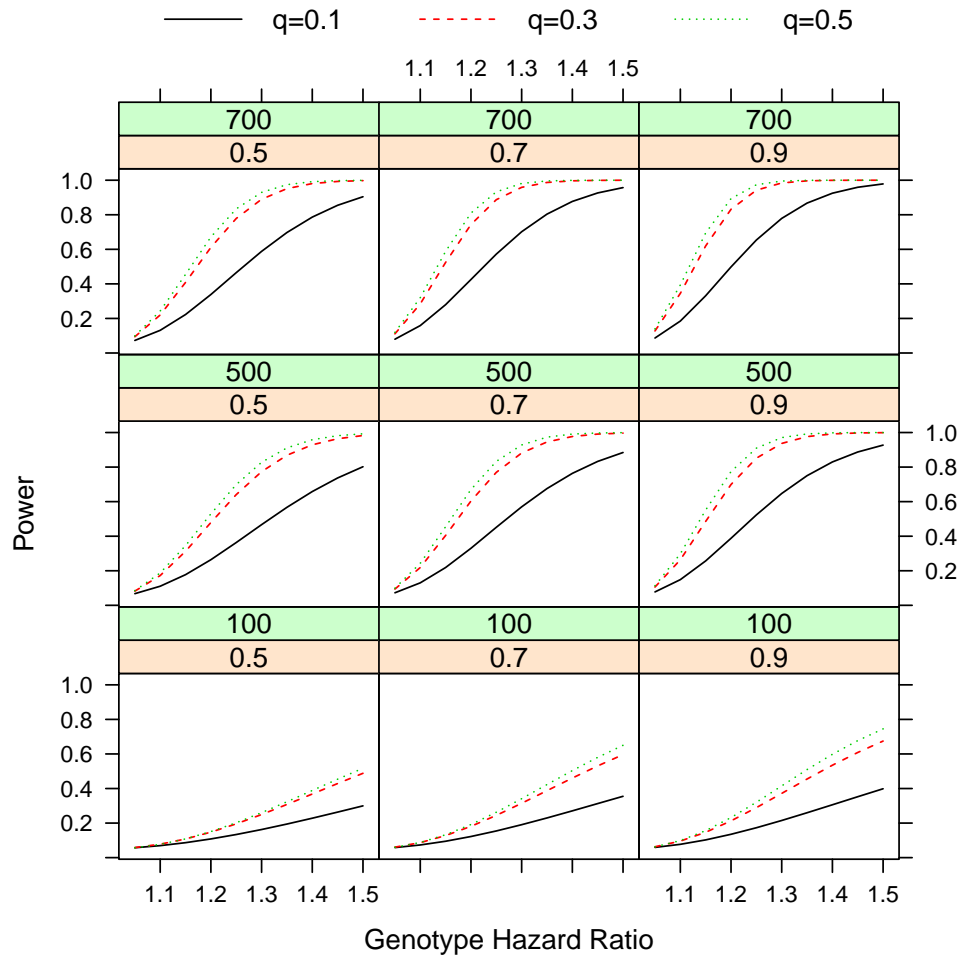


Figure 1: Power Illustration for Example 1.

```

> print(xyplot(pow0~GHR/factor(erate),group=factor(raf),
+           data=subset(res2,n==ns[1]),
+           type="l",lty=KEY$lines$lty,col=KEY$lines$col,
+           key=KEY,
+           xlab="Genotype Hazard Ratio",ylab="Power",
+           sub=paste("n=",ns[1],", alpha=",round(unique(res2$alpha),2))))

```

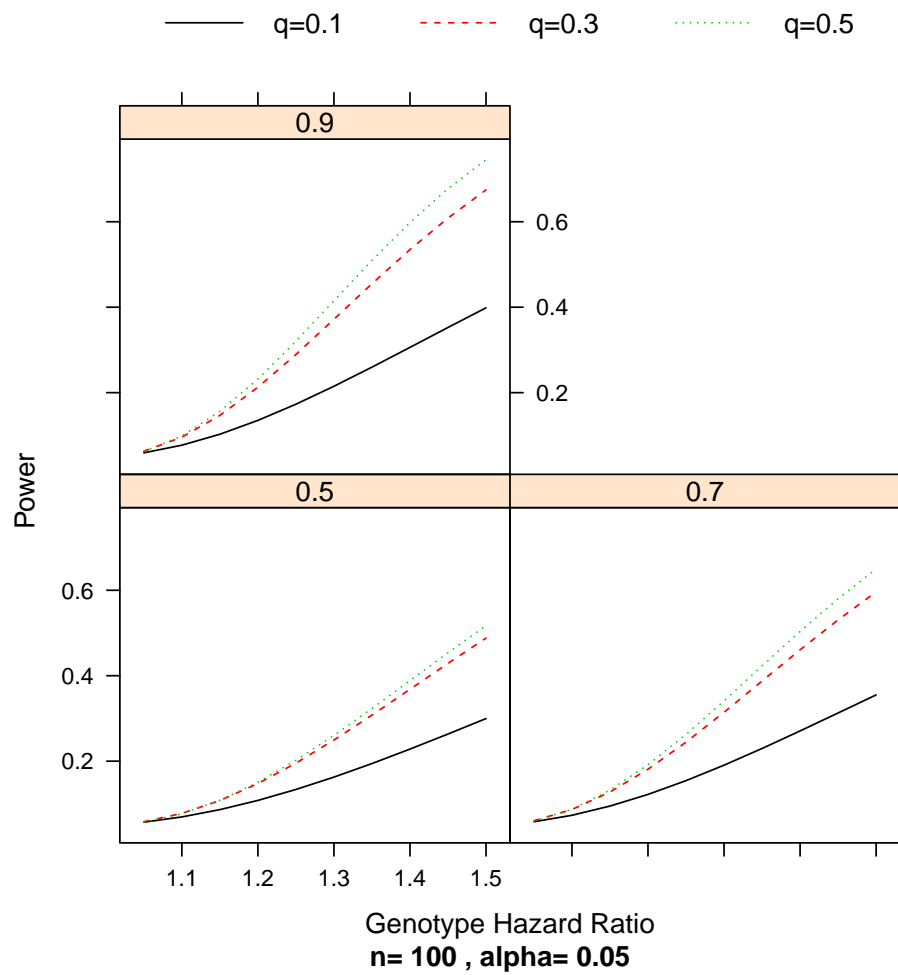


Figure 2: Power Illustration for Example 1 (restricted to  $n = 100$ ).

- R version 2.15.1 (2012-06-22), x86\_64-pc-linux-gnu
- Locale: LC\_CTYPE=en\_US.UTF-8, LC\_NUMERIC=C, LC\_TIME=en\_US.UTF-8, LC\_COLLATE=C, LC\_MONETARY=en\_US.UTF-8, LC\_MESSAGES=en\_US.UTF-8, LC\_PAPER=C, LC\_NAME=C, LC\_ADDRESS=C, LC\_TELEPHONE=C, LC\_MEASUREMENT=en\_US.UTF-8, LC\_IDENTIFICATION=C
- Base packages: base, datasets, grDevices, graphics, methods, splines, stats, utils
- Other packages: Rcpp~0.9.13, foreach~1.4.0, lattice~0.20-6, survSNP~0.21-5, survival~2.36-14, xtable~1.7-0
- Loaded via a namespace (and not attached): codetools~0.2-8, compiler~2.15.1, grid~2.15.1, iterators~1.0.6, tools~2.15.1

Table 2: Session Information