

サンプルサイズ計算について

一宮西病院呼吸器内科 城下彰宏 (PGY6)

2021/01/10

0.1 総論

サンプルサイズ計算の要素

- **Sample size**

- **Minimum detectable difference** : H_0 と H_A それぞれの statistics の意味のある差。 - **Power** (検出力) : 実際に差があった場合に、「差がある」と判断する確率。 $1 - \beta$ 、つまり、alternative hypothesis (H_A) の下で critical value を越える確率 (rejection region)。

- **One-sided or two-sided**

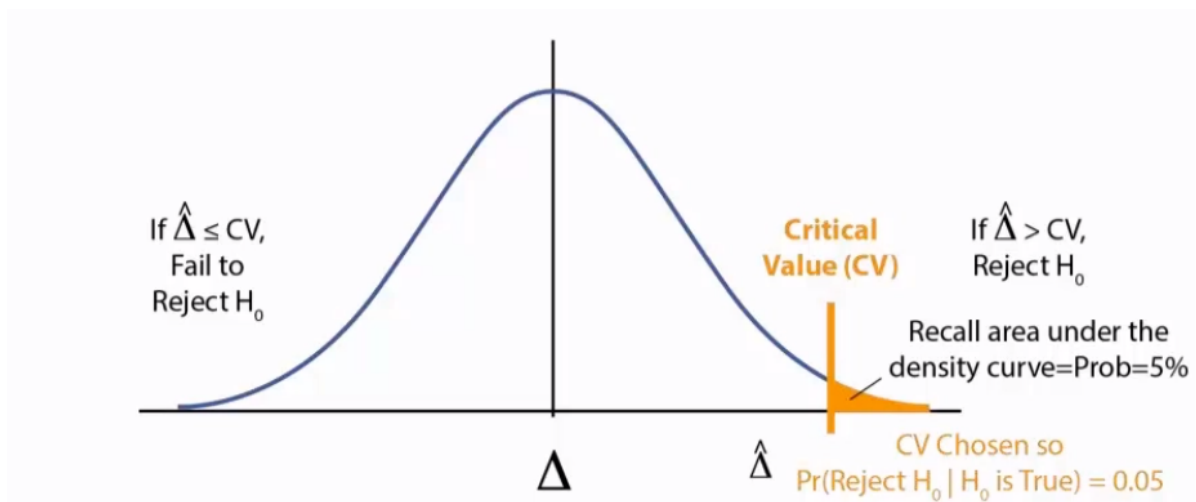
- **Allocation ratio** Target population からの random sampling が study sample であり、Null hypothesis (H_0) が正しいという仮定の下で sample から推定された statistic が「本当の (target population の) statistic」と違うのかを検討する。

Critical value よりも大きい場合にそれは起こりうる可能性の低い仮定となり、null hypothesis は棄却され、 H_A が採択される。

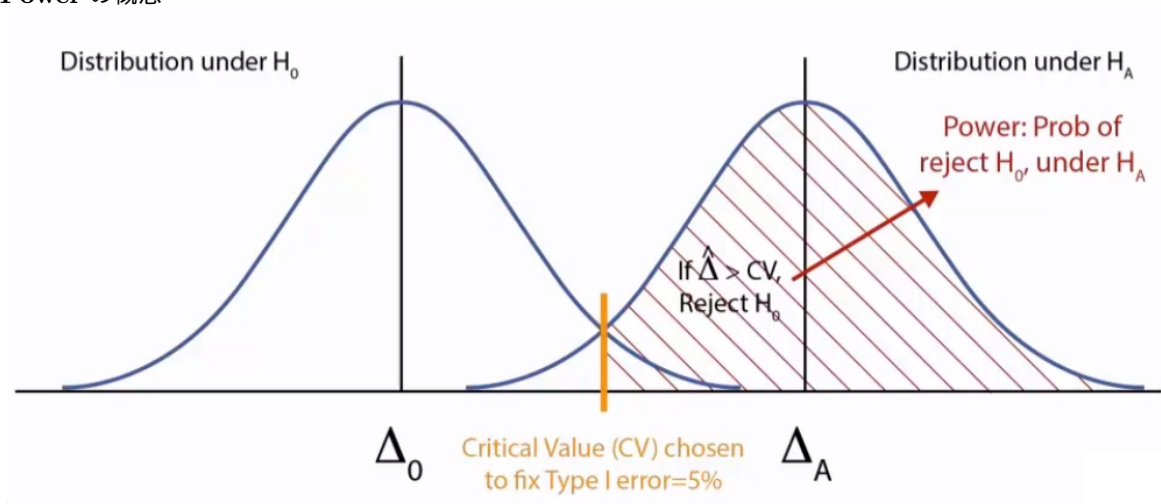
Type I error (α) : H_0 が正しいときに間違って H_0 を棄却してしまう確率

Type II error (β) : H_A が正しいときに間違って H_0 を棄却し損なう確率
仮説検定の場合は通常 Type error のみを考慮しているため、サンプルサイズ計算が特に重要ということになる。それぞれの概念を図式化して確認してみる。

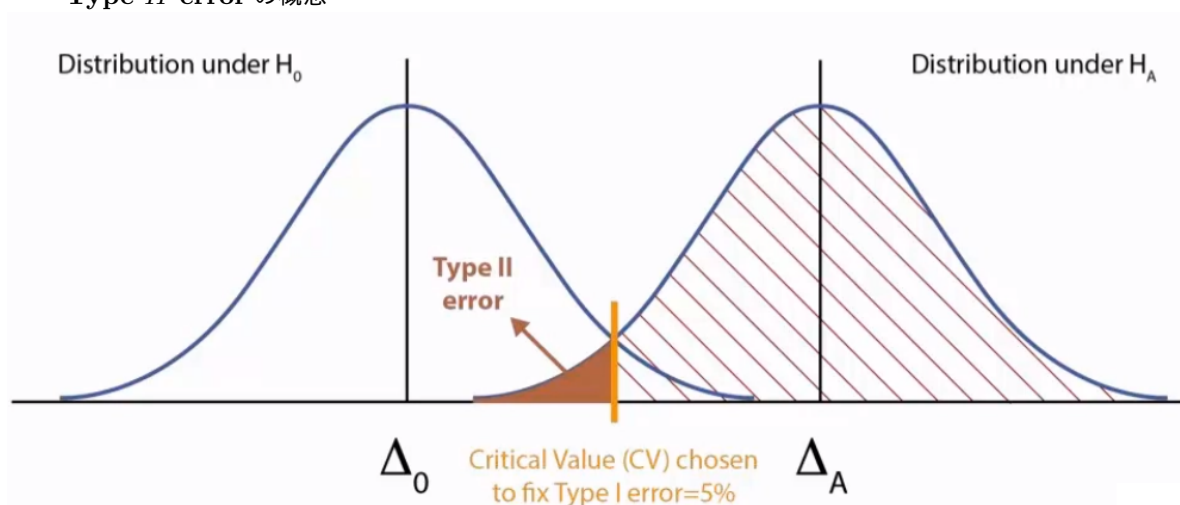
Type I error の概念



Power の概念



Type II error の概念



R では単純なサンプルサイズ計算は pwr package で行うことができる。

0.2 One sample mean (cross-sectional aspect)

$H_0 : \mu = \mu_0, H_A : \mu > \mu_1$ となる。Sample standard deviation = $s = \sqrt{\frac{1}{N-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ であり、population standard deviation = $\sigma = \frac{s}{\sqrt{N}}$ である。Statistics は $T = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ (T は σ が既知の場合は正規分布に、 σ が s によって推定される場合は t 分布に従う) であるため、 $T > Z_\alpha$ (two-sided の場合は $|T| \geq Z_{\alpha/2}$) の場合に H_0 が棄却される。実際は $\alpha=0.05$ であり、one-sided の場合は $Z_\alpha = 1.64$ である。

```
library(pwr)
#TO CALCULATE SAMPLE SIZE
power.t.test(delta=3, sd=1, power=0.8, type="one.sample", alternative="one.sided")
```

```
##
##      One-sample t test power calculation
##
##              n = 2.552324
##              delta = 3
##              sd = 1
##              sig.level = 0.05
##              power = 0.8
##      alternative = one.sided
```

```
#TO CALCULATE DELTA
power.t.test(n=50, sd=1, power=0.8, type="one.sample", alternative="one.sided")
```

```
##
##      One-sample t test power calculation
##
##              n = 50
##              delta = 0.3566013
##              sd = 1
##              sig.level = 0.05
##              power = 0.8
##      alternative = one.sided
```

```
#TO CALCULATE POWER
power.t.test(n=50, delta=3, sd=1, type="one.sample", alternative="one.sided")
```

```
##
##      One-sample t test power calculation
```

```
##
##          n = 50
##        delta = 3
##          sd = 1
##      sig.level = 0.05
##          power = 1
##  alternative = one.sided
```

0.3 Two sample mean (cross-sectional aspect)

Population 1 と Population 2 からそれぞれの sampling population を考える。

$T = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\bar{x}_1 - \bar{x}_2}}$ であるが、2つの population の分散が等しく既知の場合と等しく未知の場合、等しくなく未知の場合に分かれる（数式は割愛）。考え方は one sample mean と同様である。

ちなみに paired two sample comparison の場合は one sample case と同様に扱うことができる。

以下は 2 つの population の分散が等しい場合の計算である。

```
#TO CALCULATE POWER
d<-1/1.4
pwr.t.test(d=d, n=30, sig.level=0.05, type="two.sample", alternative="two.sided")
```

```
##
##      Two-sample t test power calculation
##
##          n = 30
##          d = 0.7142857
##      sig.level = 0.05
##          power = 0.7764889
##  alternative = two.sided
##
## NOTE: n is number in each group
```

```
#n is the total sample size. assume equal allocation between the two groups.
#d is the standardized effect size. i.e. (mu1-mu0)/SD
```

```
#TO CALCULATE SAMPLE SIZE
pwr.t.test(d=d, power=0.8, sig.level=0.05, type="two.sample", alternative="two.sided")
```

```
##
##      Two-sample t test power calculation
```

```
##
##          n = 31.75708
##          d = 0.7142857
##      sig.level = 0.05
##          power = 0.8
##      alternative = two.sided
##
## NOTE: n is number in each group
```

2つの population の分散が未知の等しくない場合も参考までに掲載する。

```
#TO CALCULATE POWER
powerfunc<-function(alpha, mu1, mu2, sigma1, sigma2, n1, n2){
  delta<-mu1-mu2
  sat.df<-(sigma1^2/n1+sigma2^2/n2)^2/((1/(n1-1))*(sigma1^2/n1)^2+(1/(n2-1))*(sigma2^2/n2)^2) #Sat
  non.cent<-delta/(sqrt(sigma1^2/n1+sigma2^2/n2)) #Non-centrality parameter
  #Two sided alternative
  power<-pt(qt(p=(1-alpha/2), df=sat.df, lower.tail=TRUE),df=sat.df, ncp=non.cent, lower.tail=F)+
    pt(qt(p=(alpha/2), df=sat.df, lower.tail=TRUE),df=sat.df, ncp=non.cent, lower.tail=T)
  return(power)
}
powerfunc(alpha=0.05, mu1=5, mu2=3, sigma1=1, sigma2=2, n1=40,n2=50)
```

```
## [1] 0.999982
```

```
#TO CALCULATE SAMPLE SIZE
n<-seq(50,80,1)
n1<-n
n2<-n
power.result<-c()
for (i in 1:length(n)){
  power.result[i]<-powerfunc(alpha=0.05, mu1=5, mu2=3, sigma1=1, sigma2=2, n1=n1[i],n2=n2[i])
}
power.result
```

```
## [1] 0.9999906 0.9999930 0.9999947 0.9999961 0.9999971 0.9999978 0.9999984
## [8] 0.9999988 0.9999991 0.9999993 0.9999995 0.9999996 0.9999997 0.9999998
## [15] 0.9999998 0.9999999 0.9999999 0.9999999 1.0000000 1.0000000 1.0000000
## [22] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [29] 1.0000000 1.0000000 1.0000000
```

0.4 Correlation coefficient, simple linear regression, and multiple linear regression

Correlation coefficient である r は skewed data なので、Fisher's Z transformation を行わないといけない (r : $F(r) = \frac{1}{2} \ln(\frac{1+r}{1-r})$)。 $F(r)$ は正規分布に従うことが知られている。そこで $T = \frac{F(r) - F(\rho_0)}{\sigma_{F(r)}}$ となり、 $\sigma_{F(r)} = \frac{1}{\sqrt{N-3}}$ からこれまでと同様に検定について考えることができる。

ただし、R では simple linear regression $Y = \beta_0 + \beta_1 X + \epsilon$ 、つまり、 $\hat{\beta}_1 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{SS_x}$ ($SS_x = \sum_{i=1}^N (x_i - \bar{x})^2$) において、 $H_0: \beta_1 = 0$ とすることで計算する。 $T = \frac{\hat{\beta}_1 - 0}{\sigma_{\hat{\beta}_1}}$ が F 分布 (df=1, N-2) に従うことから計算するということである。そこで、 SS_x と residual variance が必要になる。

R では multiple linear regression も対応している `pwr.f2.test` function を使用する。他のパッケージもあるので、色々試してみると良いかもしれない。

```
sd.y<-1
sd.x<-6
b<-0.1
sd.e<-sqrt(sd.y^2-0.1^2*sd.x^2) #This is the standard deviation of residuals.
r.sq<-1-(sd.e^2/sd.y^2) #This is the R_square.
r.sq
```

```
## [1] 0.36
```

```
pwr.f2.test(u=1, f2=r.sq/(1-r.sq), power=0.8)
```

```
##
##      Multiple regression power calculation
##
##              u = 1
##              v = 14.12059
##              f2 = 0.5625
##      sig.level = 0.05
##      power = 0.8
```

```
#u is the DF of numerator,
#since this is simple linear regression testing for one variable, u=1.
#v is left unspecified but power= is specified, so this is requesting output for n.
#v: the DF of the denominator
#v=N-2 -> N=v+2.
```

0.5 One sample proportion (cross-sectional aspect)

Binary outcome つまり、2*2 表の書ける outcome について。Y=Y₁ + Y₂ +Y_n で、Y は二項分布 binomial(n,p) に従う (p=prevalence of characteristics)。ここで p を推定したいので、実際のデータから求めら $\hat{p} = y/n$ であるので、 $H_0 : p = p_0$ を仮定することで検定ができる。サンプルサイズが「十分に」大きくなると sampling distribution は正規分布に近似できることを用いて、 $p\hat{p}$ は平均 \hat{p} で、分散 $n * \hat{p} * p$ の正規分布に従うことになる。 $T = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$ が statistics になり、これまで通りの検定ができる。ちなみに「十分に」というのは $p_0 * (1 - p_0) * n > 5$ を目安にしている。近似できない (サンプルサイズが小さい) 場合は arc-sin transformation を行なうという選択肢がある。 $2\text{srccsin}(\sqrt{\hat{p}}) \sim \text{Normal}(2\text{arcsin}\sqrt{p}, 1/n)$ を用いて検定する。R の pwr package は arcsine transformation を用いていることに注意が必要である。

```
#TO CALCULATE POWER
```

```
h <- ES.h(0.02, 0.03)
```

```
pwr.p.test(h=h, n=300, sig.level=0.05, power=NULL, alternative="two.sided")
```

```
##
```

```
##      proportion power calculation for binomial distribution (arcsine transformation)
```

```
##
```

```
##              h = 0.06437191
```

```
##              n = 300
```

```
##      sig.level = 0.05
```

```
##              power = 0.2001055
```

```
##      alternative = two.sided
```

```
#To CALCULATE SAMPLE SIZE
```

```
h <- ES.h(0.02, 0.03)
```

```
pwr.p.test(h=h, n=NULL, sig.level=0.05, power=.8, alternative="two.sided")
```

```
##
```

```
##      proportion power calculation for binomial distribution (arcsine transformation)
```

```
##
```

```
##              h = 0.06437191
```

```
##              n = 1894.147
```

```
##      sig.level = 0.05
```

```
##              power = 0.8
```

```
##      alternative = two.sided
```

0.6 Two sample proportion (cross-sectional aspect)

2つの population を想定し、Fisher's exact test で検定する。Cumulative incidence の場合も 2*2 表を作成できるので同様である（ただし、loss to follow-up を考慮しなくてはならない）。

2つの sample size が等しい場面を想定した R のコードを記載した。ここでは両群のサンプルサイズが異なっても使用できる `pwr.2p2n.test` function を使用したが、`pwr.p.test` でも計算できる。

```
#TO CALCULATE POWER
```

```
h <- ES.h(0.4, 0.5)
```

```
pwr.2p2n.test(h, 200, 300)
```

```
##
```

```
##      difference of proportion power calculation for binomial distribution (arcsine transformation)
```

```
##
```

```
##              h = 0.2013579
```

```
##              n1 = 200
```

```
##              n2 = 300
```

```
##      sig.level = 0.05
```

```
##      power = 0.5970976
```

```
##      alternative = two.sided
```

```
##
```

```
## NOTE: different sample sizes
```

```
#To CALCULATE SAMPLE SIZE
```

```
h <- ES.h(0.4, 0.5)
```

```
pwr.2p.test(h, power=0.8)
```

```
##
```

```
##      Difference of proportion power calculation for binomial distribution (arcsine transformation)
```

```
##
```

```
##              h = 0.2013579
```

```
##              n = 387.1677
```

```
##      sig.level = 0.05
```

```
##      power = 0.8
```

```
##      alternative = two.sided
```

```
##
```

```
## NOTE: same sample sizes
```



```
pwr.2p2n.test(h, n1=1000, n2=NULL, sig.level=0.05, power=0.8, alternative="two.sided")
```

```
##
##      difference of proportion power calculation for binomial distribution (arcsine transformation)
##
##          h = 0.2013579
##          n1 = 1000
##          n2 = 240.0546
##      sig.level = 0.05
##          power = 0.8
##      alternative = two.sided
##
## NOTE: different sample sizes
```

0.7 Two sample difference, ratios, and odds ratios (cross-sectional aspect)

Two sample proportion とほとんど同じである。Statistics が difference は $Z_\delta = \frac{\hat{\delta}}{SE(\hat{\delta})}$ 、ratios は $Z_\rho = \frac{\ln \hat{\rho}}{SE(\ln \hat{\rho})}$ 、odds-ratio は $Z_\theta = \frac{\ln \hat{\theta}}{SE(\ln \hat{\theta})}$ であるのが違いとなる。Statistics を正規分布に近似していることも proportion の場合と同じである。

0.8 Incidence rate (longitudinal aspect)

Incidence rate (mortality) は at-risk time を考慮しないとイケない ($I=y/T$ =Total number of events/Total person-time at risk)。まず、仮定として、ある時点で event の起こる可能性はどの患者でも等しいとする。Y はポアソン分布に従い ($Poisson(\lambda, T)$ に従う)、T は研究デザインから固定されているとすると、 $y/T = T\lambda/T = \hat{\lambda}$ が incidence rate となる。 $H_0: \lambda = \lambda_0$ 、 $H_1: \lambda \neq \lambda_0$ (two-sided) となる。 H_0 を仮定し、 $T\lambda > 30$ の場合はポアソン分布は正規分布に近似することができる $Y \sim N(T\lambda, T\lambda)$ ことから、分散が既知の場合と同様の検定ができる。ちなみに、 $W = \sqrt{Y}$ とし、 $W \sim N(\sqrt{T\lambda}, 1/4)$ や、Event 数が少ないときは exact methods も使用される。

まずは one sample incidence のコードから掲載する。

```
#TO CALCULATE POWER
rate.power.one <- function(rnull,ralt,ttt,alpha,sss) {
  xnull <-sqrt(rnull*ttt) #rnull: null rate, ttt:person time
  xalt <- sqrt(ralt*ttt) #anull: alternative rate
  int1 <-(xnull-xalt)/sqrt(0.25)
  zalp <- qnorm(1-alpha/sss) #sss=1, 2 sided
  int2 <- 1-pnorm(zalp+int1)
```

```

int2}
#TO CALCULATE SAMPLE SIZE
rate.ss.one <- function(rnull,ralt,pow,alpha,sss) {
xnull <-sqrt(rnull)
xalt <- sqrt(ralt)
zalpha <- qnorm(1-alpha/sss)
zbeta <- qnorm(pow)
numerator <- (zalpha + zbeta)
denom <- xalt - xnull
final <- 0.25 * (numerator/denom)**2
final
}

```

次に two sample incidence の比較についてのコードを記載する。Incidence rate ratio や incidence rate difference を statistics とし、 H_0 : IRR=1, IRD=0 とすることで、それぞれ正規分布に従う $\ln(IRR) = \ln(\frac{Y_A}{T_A} / \frac{Y_B}{T_B})$ 、 $\sqrt{Y_A/T_A} - \sqrt{Y_B/T_B} = W_A - W_B$ から検定を行う。。IRR と IRD のどちらも使用しても大差はないだろう。

```

#TO CALCULATE POWER
rate.power.twosqrt <- function(r1,r2,n1,n2,alpha,sss) {
num <- sqrt(r1) - sqrt(r2)
den <- .5 * sqrt(1/n1 + 1/n2)
zalpha <- qnorm(1 - alpha/sss)
int1 <- num / den
pnorm(int1 - zalpha)
}

#TO CALCULATE SAMPLE SIZE
rate.ss.twosqrt <- function(r1,r2,r,alpha,power,sss) { #r = ratio of sample sizes t2/t1
term1 <- 1+r
zalpha <- qnorm(1 - alpha/sss)
zbeta <- qnorm(power)
term2 <- 0.25 * (zalpha+zbeta)**2
term3 <- (sqrt(r1) - sqrt(r2))**2
ss1 <- term1 * term2 / term3
ss2 <- ss1*r
list(ss1,ss2)
}

```

##Continuous outcome (longitudinal aspect)

Continuous outcome が経時的に変化する場合については paired t-test や two-sample paired t-test を行うことを想定する。

Binary outcome (longitudinal aspect) Binary outcome が経時的に変化する場合については、

0.9 Survival outcomes (longitudinal aspect)

Log-rank test を想定した time-to-event outcome のサンプルサイズ計算については、censoring について（特に right censoring）について考慮しないといけない。ここでは Schoenfeld method を紹介する。Survival function $S(t)=\Pr(T>t)=1-\Pr(T\leq t)$ は nonparametric method の Kaplan-Meier curve で推定できる。一方、hazard function は $\lambda(t) = P(T = t|T \geq t) = \frac{P(T=t)}{P(T\geq t)}$ (T:discrete)、つまり instantaneous failure rate (t まで生存した患者が t で event を発生する risk) を表す。Log-rank test は $H_0: \lambda_A(t) = \lambda_B(t)$ (または $S_A(t) = S_b(t)$)、 $H_A: \lambda_A(t) \neq \lambda_B(t)$ (two-sided) と仮定する検定であり、それぞれの failure time で 2*2 表を作成してまとめるイメージである。そこで、 $H_0: \theta = 0$ ($\theta = \log(\frac{\lambda_1}{\lambda_0})$) と変換することができ (hazard ratio)、statistics を $\chi^2(df = 1)$ を考えることができる。ここから、必要な event 数は $m = \frac{Z_{\alpha/2} + Z_{\beta}}{\theta^2 \pi(1-\pi)^2} (\pi \text{ allocation proportion to first group event } \$: \text{enrollment period, f : フォロアップ期間, proportional hazards model を仮定したもとの survival function の情報を合わせて計算する。それぞれの群の survival function から平均を算出して } (\bar{S}(t) = \pi * \bar{S}_0(t) + (1-\pi) * \bar{S}_1(t)), \text{ 試験終了までに event の起こる確率を計算する } (\bar{F}(a+f) = 1 - 1/6[\bar{S}(f) + 4\bar{S}(0.5a+f) + \bar{S}(a+f)]。最終的に } N = \frac{m}{\bar{F}(a+f)} \text{ からサンプルサイズが求められる。R では Hmisc package を使用する。}$

```
#TO CALCULATE POWER
library(Hmisc)
tref<-2.5      #Specify a time about the survival probability
n<-4000       #total sample size
mc<-0.45      #Probability of having a event by tref, i.e. 1-Survial prob at tref.
hr<-1.13      #Hazard ratio.
r<-(1-((1-(1-mc)^hr)/mc))*100 #Reduction of risk
#1- the ratio of event probability in intervention group compared to control group
#at tref. Can be specified directly without useing
accrual<-2     #Length of accrual period
tmin<-3        #Minimum length of follow-up
pwr.res<-cpower(tref=tref, n=n, mc=mc, r=r, accrual=accrual, tmin=tmin)

##
## Accrual duration: 2 y  Minimum follow-up: 3 y
##
## Total sample size: 4000
##
## Alpha= 0.05
```

```
##
## 2.5-year Mortalities
##      Control Intervention
##      0.4500000      0.4911265
##
## Hazard Rates
##      Control Intervention
##      0.2391348      0.2702223
##
## Probabilities of an Event During Study
##      Control Intervention
##      0.6121072      0.6565620
##
## Expected Number of Events
##      Control Intervention
##      1224.2      1313.1
##
## Hazard ratio: 1.13
## Standard deviation of log hazard ratio: 0.03972899
```

```
pwr.res
```

```
##      Power
## 0.8678575
```

```
#TO CALCULATE SAMPLESIZE
Nsim<-seq(1000,3000, 100)
pwr.res<-c()
for (i in 1:length(Nsim) ){
  pwr.tmp<-cpower(tref=tref, n=Nsim[i], mc=mc, r=r, accrual=accrual, tmin=tmin,pr=FALSE)
  pwr.res[i]<-pwr.tmp["Power"]
}
pwr.res.all<-cbind(Nsim, pwr.res)
pwr.res.all
```

```
##      Nsim  pwr.res
## [1,] 1000 0.3368116
## [2,] 1100 0.3645668
## [3,] 1200 0.3917863
## [4,] 1300 0.4184145
```

```
## [5,] 1400 0.4444043
## [6,] 1500 0.4697172
## [7,] 1600 0.4943217
## [8,] 1700 0.5181931
## [9,] 1800 0.5413128
## [10,] 1900 0.5636678
## [11,] 2000 0.5852497
## [12,] 2100 0.6060549
## [13,] 2200 0.6260835
## [14,] 2300 0.6453390
## [15,] 2400 0.6638283
## [16,] 2500 0.6815605
## [17,] 2600 0.6985475
## [18,] 2700 0.7148029
## [19,] 2800 0.7303421
## [20,] 2900 0.7451818
## [21,] 3000 0.7593401
```