

Online Appendix:

```
# R script for calculating samples sizes for a case-control and familial
# studies for a rare genetic
# variant of unknown significance.

#
# Reference: Fleiss, Levin, and Paik. Statistical Methods for Rates and
# Proportions, 3rd ed.
# Hoboken, New Jersey: Wiley-Interscience, 2003. Section 4 (pp. 75-77)
#

# Returns carrier frequency in cases (p1) corresponding to
# relative risk (rr) and carrier frequency in the population (p0)
#
rr2p1 <- function(rr, p0) {
  p1 <- (rr*p0)/(rr*p0 + 1 - p0)
}

# Returns penetrance in non-carriers corresponding to
# relative risk (rr), minor allele frequency (maf)
# population incidence (inc)

nc.penetance <- function(rr, maf, inc) {
  cf <- 2*maf*(1-maf) + maf*maf # carrier frequency in population under HWE
  inc/(rr*cf + (1 - cf))
}

# Calculate total sample size (ss) for one-sided test with type 1 error
# (alpha)
# and type 2 error (beta) from two populations for which the carrier
# frequencies are p0 and p1 in the first and second populations
# and when the number of individuals sampled from the second population is
# r times the number of individuals sampled from the first population.
#
ss <- function(p0, p1, alpha, beta, r) {
  a <- qnorm(1 - alpha)
  b <- qnorm(1 - beta)
  pmean <- (p0 + r*p1)/(r+1)

  num1 <- a*sqrt((r+1)*pmean*(1-pmean))
  num2 <- b*sqrt(r*p0*(1 - p0) + p1*(1 - p1))
  num <- (num1 + num2)^2
  den <- r*((p1 - p0)^2)

  m.prime <- num/den
  m <- corrected.m(m.prime, p0, p1, r)

  total.size <- round((1+r)*m)
}

# Returns control sample size after incorporating a continuity correction.
#
corrected.m <- function(m.prime, p0, p1, r) {
  x <- sqrt( 1 + ( 2*(r+1) / (m.prime*r*abs(p1-p0)) ) )
  m <- (m.prime/4) * ( 1 + x )^2 )
}
```

```

}

# Prints total sample size for one-sided test for specified
# relative risk (rel.risk), population minor allele frequency (maf),
# type 1 error (alpha), type 2 error (beta), when there are equal
# numbers of cases and population controls.
#
print.sample.sizes <- function(rel.risk, maf, alpha, beta) {
  r = 1 # equal numbers of case and controls
  p0 <- 2*maf*(1-maf) + maf*maf # carrier frequency in population under HWE
  results <- NULL
  for (rr in rel.risk) {
    p1 <- rr2p1(rr, p0)
    results <- rbind(results, ss(p0, p1, alpha, beta, r))
  }
  rownames(results) <- rel.risk
  colnames(results) <- maf
  writeLines(" ")
  writeLines(paste("alpha=", alpha, " beta=", beta, sep=" "))
  print(results)
}

# Prints total sample size for one-sided test for specified
# relative risk (rel.risk), population minor allele frequency (maf),
# disease incidence (inc), type 1 error (alpha), and
# type 2 error (beta), when the ratio of carriers to non-carriers is 0.6
#
print.family.sample.sizes <- function(rel.risk, inc, maf, alpha, beta) {
  r = 0.6 # assume ratio of carriers/non-carriers to be 0.60 in family data.
  results <- NULL
  for (rr in rel.risk) {
    p0 <- nc.penetance(rr, maf, inc)
    p1 <- rr*p0
    results <- rbind(results, ss(p0, p1, alpha, beta, r))
  }
  rownames(results) <- rel.risk
  colnames(results) <- maf
  writeLines(" ")
  writeLines(paste("alpha=", alpha, " beta=", beta, sep=" "))
  print(results)
}

alpha <- 0.10 # Type 1 error (one-tailed test)
beta <- 0.20 # Type 2 error (one-tailed test)
maf <- c(0.001, 0.0001, 0.00001) # population minor allele frequency
bc.rel.risk <- c(12, 6, 3, 1.5) # relative risk - breast cancer
cc.rel.risk <- c(20, 10, 5, 2.5) # relative risk - colon cancer
bc.inc <- 0.08 # incidence - breast cancer
cc.inc <- 0.03 # incidence - colon cancer

writeLines(" "); writeLines("Breast Cancer - unrelateds")
print.sample.sizes(bc.rel.risk, maf, alpha, beta)

writeLines(" "); writeLines("Colon Cancer - unrelateds")
print.sample.sizes(cc.rel.risk, maf, alpha, beta)

```

```
writeLines(" "); writeLines("Breast Cancer - relateds")
print.family.sample.sizes(bc.rel.risk, bc.inc, maf, alpha1, beta)

writeLines(" "); writeLines("Colon Cancer - relateds")
print.family.sample.sizes(cc.rel.risk, cc.inc, maf, alpha1, beta)
```