Statistical power

Federica Grosso 904701

2024-08-06

```
library(ggplot2)
```

Power of MR analysis

Load of IV

```
path <- "C:\\Users\\feder_phxiw9d\\Desktop\\Tesi\\Materiale_Serena\\Results\\Results_MacularD
eg\\Harm_data_E_GCST90027737_O_MACDEG.txt"
# path <- "C:\\Users\\feder_phxiw9d\\Desktop\\Tesi\\Materiale_Serena\\Results\\Results_Macula
rDeg\\Harm_data_E_GCST90027663_O_MACDEG.txt"
```

```
data <- read.table(path, sep = "\t")
```

Calculation of explained variance R^2

```
eaf = data$eaf.exposure
MAF <- ifelse(eaf <= 0.5, eaf, 1-eaf)
beta = data$beta.exposure
se =data$se.exposure
N=7738
R2 = 2 * beta^2 * MAF * (1 - MAF) / (2 * beta^2 * MAF * (1 - MAF) + se^2 * 2 * N * MAF * (1 -
MAF))
R_2=sum(R2)
R_2
```

```
## [1] 0.01859125
```

Calculation of power

```
expit <- function(x) { return(exp(x)/(1+exp(x))) }

rsq = R_2 # squared correlation
b1 = 0.2 # causal effect (log odds ratio per SD
#b1 = log(1.2) # or log of OR per SD)
sig = 0.05 # significance level (alpha)
pow = 0.8 # power level (1-beta)
ratio = 361 # ratio of cases:controls = 1:ratio
n = 7738 # Sample size
```

Continuous outcome

```
cat("Power of analysis with ", n, "participants: ",
    pnorm(sqrt(rsq*b1^2*n)-qnorm(1-sig/2)))
```

```
## Power of analysis with  7738 participants:  0.6696191
```

Binary outcome

```
cat("Power of analysis with ", n, "participants: ",
    pnorm(sqrt(n*rsq*(ratio/(1+ratio))*(1/(1+ratio)))*b1-qnorm(1-sig/2)))
```

```
## Power of analysis with  7738 participants:  0.03332263
```

Sample size

```
cat("Sample size required for ", pow*100, "% power: ",
    (qnorm(1-sig/2)+qnorm(pow))^2/b1^2/rsq/(ratio/(1+ratio))/(1/(1+ratio)))
```

```
## Sample size required for  80 % power:  3831325
```

F statistics

```
k=nrow(data)
F <- R_2*(N-1-k)/((1-R_2)*k)
F
```

```
## [1] 24.40861
```

Plot of effect size vs power

```
# Effect size grid
beta_values <- seq(0.05, 0.60, by=0.05)
```

```r
beta_values <- seq(0.05, 0.60, by=0.05)
ratio_values <- c(361, 14.37, 6.5, 38)
# Funzione ipotetica per calcolare la potenza in base a beta e sample size
calculate_power <- function(b1, n, ratio, sig, rsq) {
  power <- pnorm(sqrt(n*rsq*(ratio/(1+ratio))*(1/(1+ratio)))*b1-qnorm(1-sig/2))*100
  return(power)
}


# Creare un data frame per contenere i risultati
results <- expand.grid(Beta = beta_values, SampleSize = n, Ratio = ratio_values, Sigma = 0.0
5, Rsq = R_2 )
results$Power <- mapply(calculate_power, results$Beta, results$SampleSize, results$Ratio, res
ults$Sigma, results$Rsq)
```

Plot of power of AMD

```r
ggplot(results, aes(x = Beta, y = Power, color = as.factor(Ratio))) +
  geom_point(size = 2.5) +
  geom_line(lwd = 1) +
  scale_color_manual(name = "Ratio of Cases to Controls",
                     values = c("#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#9467bd"),
                     labels = c("ratio = 6.5 (GCST010723)", "ratio = 14.37 (GCST90086112)","r
atio = 38 (Finngen study)", "ratio = 361 (GCST90043776 -\nmain analysis)")) +
  labs(x = expression(Causal~Effect~(beta)),
       y = "Power (%)") +
  theme_minimal() +
  theme(
    panel.border = element_rect(color = "darkgrey", fill = NA, size = 1),
    axis.line = element_line(color = "darkgrey", size = 0.5),
    axis.ticks = element_line(color = "darkgrey"),
    axis.text.x = element_text(size = 12),  # Modifica la dimensione dei numeri sull'asse x
    axis.text.y = element_text(size = 12)   # Modifica la dimensione dei numeri sull'asse y
  )
```

```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```
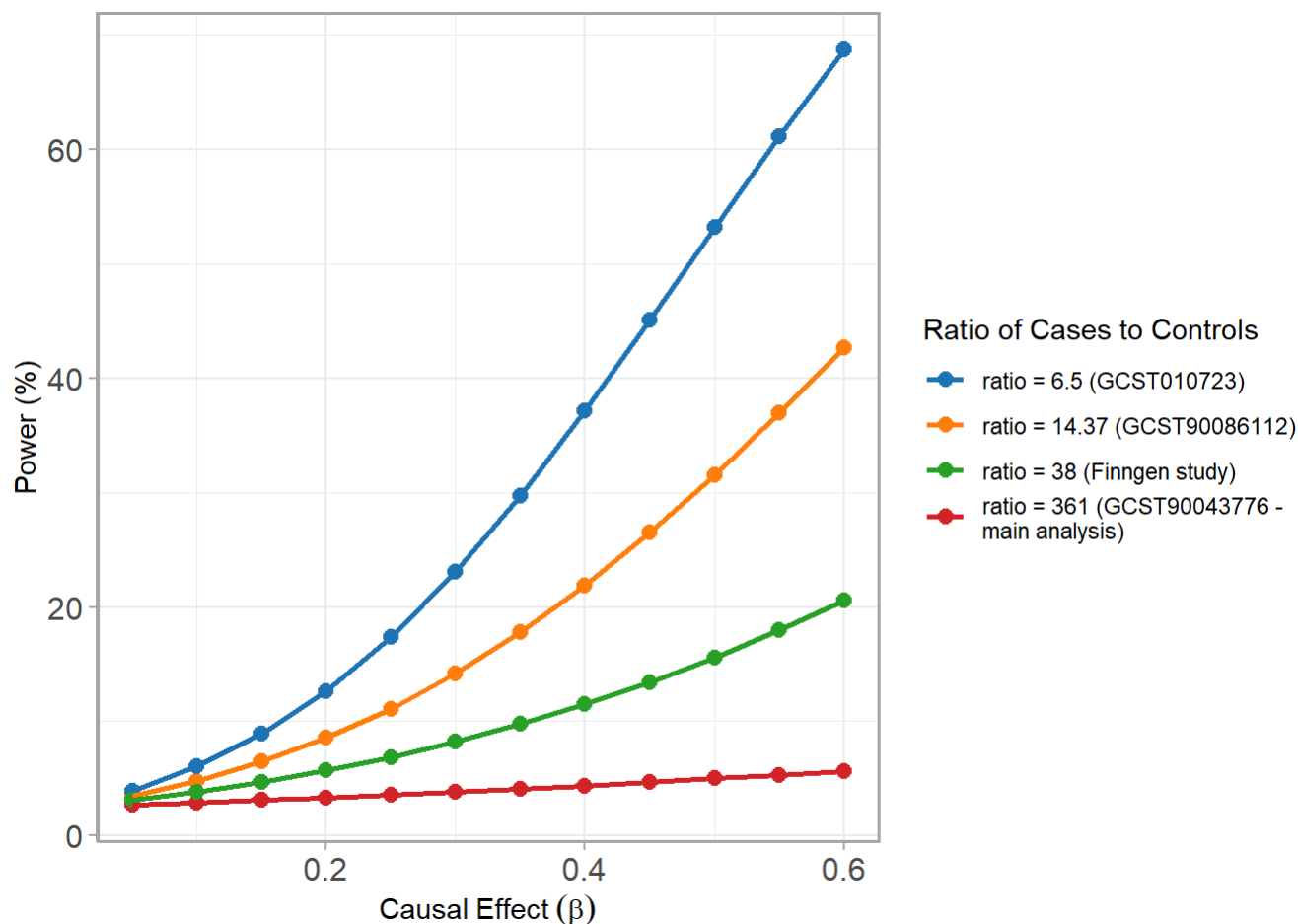
LACT vs TRAIL

Load of IV

```
#path <-"C:\\Users\\feder_phxiw9d\\Desktop\\Tesi\\Materiale_Serena\\Results\\Other_analysis
\\Results_LACT_TRAIL\\Harm_data_E_GCST90027488_O_TRAIL.txt"
path <- "C:\\Users\\feder_phxiw9d\\Desktop\\Tesi\\Materiale_Serena\\Results\\Results_IP_all
\\Harm_data_E_GCST90027488_O_GCST90274843.txt"
#path <-"C:\\Users\\feder_phxiw9d\\Desktop\\Tesi\\Materiale_Serena\\Results\\Results_IP_all
\\Harm_data_E_GCST90027754_O_GCST90274846.txt"
```

```
data <- read.table(path, sep = "\t")
```

Calculation of explained variance R^2

```
eaf = data$eaf.exposure
MAF <- ifelse(eaf <= 0.5, eaf, 1-eaf)
beta = data$beta.exposure
se =data$se.exposure
N=7738
R2 = 2 * beta^2 * MAF * (1 - MAF) / (2 * beta^2 * MAF * (1 - MAF) + se^2 * 2 * N * MAF * (1 -
MAF))
R_2=sum(R2)
R_2
```

```
## [1] 0.03150409
```

Calculation of power

```
expit <- function(x) { return(exp(x)/(1+exp(x))) }

rsq = R_2 # squared correlation
b1 = 0.2 # causal effect (log odds ratio per SD
#b1 = log(1.2) # or log of OR per SD)
sig = 0.05 # significance level (alpha)
pow = 0.8 # power level (1-beta)
#ratio = 361 # ratio of cases:controls = 1:ratio
n = 7738 # Sample size
```

Continuous outcome

```
cat("Power of analysis with ", n, "participants: ",
    pnorm(sqrt(rsq*b1^2*n)-qnorm(1-sig/2)))
```

```
## Power of analysis with  7738 participants:  0.8775281
```

F statistics

```
k=nrow(data)
F <- R_2*(N-1-k)/((1-R_2)*k)
F
```
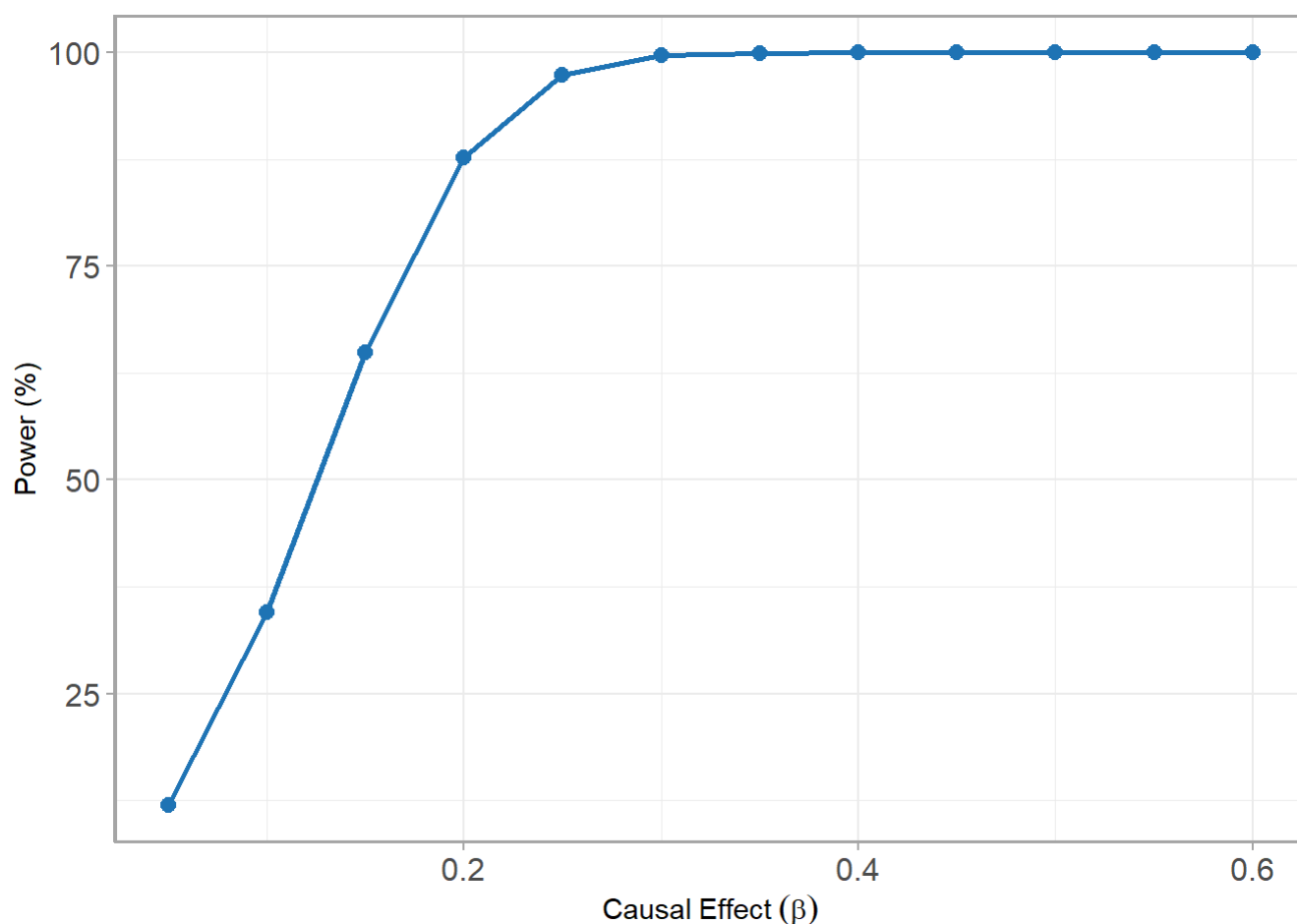
```
## [1] 27.93147
```

Plot of power of TRAIL

```
# Effect size grid
beta_values <- seq(0.05, 0.60, by=0.05)
```

```
# Funzione ipotetica per calcolare la potenza in base a beta e sample size
calculate_power <- function(b1, n, sig, rsq) {
  power <- pnorm(sqrt(rsq*b1^2*n)-qnorm(1-sig/2))*100
  return(power)
}

# Creare un data frame per contenere i risultati
results <- expand.grid(Beta = beta_values, SampleSize = n, Sigma = 0.05, Rsq = R_2 )
results$Power <- mapply(calculate_power, results$Beta, results$SampleSize,results$Sigma, resu
lts$Rsq)
```

```
ggplot(results, aes(x = Beta, y = Power)) +
  geom_point(size = 2.5, color= "#1f77b4") +
  geom_line(lwd = 1, color="#1f77b4") +
  labs(x = expression(Causal~Effect~(beta)),
       y = "Power (%)") +
  theme_minimal() +
  theme(
    panel.border = element_rect(color = "darkgrey", fill = NA, size = 1),
    axis.line = element_line(color = "darkgrey", size = 0.5),
    axis.ticks = element_line(color = "darkgrey"),
    axis.text.x = element_text(size = 12),   # Modifica la dimensione dei numeri sull'asse x
    axis.text.y = element_text(size = 12)   # Modifica la dimensione dei numeri sull'asse y
  )
```


Bifidobacterium adolescentis vs TNFSF12

Load of IV

```
path <-"C:\\Users\\feder_phxiw9d\\Desktop\\Tesi\\Materiale_Serena\\Results\\Results_IP_all\\H
arm_data_E_GCST90027754_O_GCST90274846.txt"
path1 <- "C:\\Users\\feder_phxiw9d\\Desktop\\Tesi\\Materiale_Serena\\Results\\Other_analysis
\\Results_ado_TNFSF12\\Harm_data_E_GCST90032220_O_TNFSF12.txt"
```

```
data <- read.table(path, sep = "\t")
data1 <- read.table(path1, sep = "\t")
```

Calculation of explained variance R^2

```
eaf = data$eaf.exposure
MAF <- ifelse(eaf <= 0.5, eaf, 1-eaf)
beta = data$beta.exposure
se =data$se.exposure
N=7738
R2 = 2 * beta^2 * MAF * (1 - MAF) / (2 * beta^2 * MAF * (1 - MAF) + se^2 * 2 * N * MAF * (1 -
MAF))
R_2=sum(R2)
R_2
```

```
## [1] 0.02709539
```

```
eaf1 = data1$eaf.exposure
MAF1 <- ifelse(eaf1 <= 0.5, eaf1, 1-eaf1)
beta1 = data1$beta.exposure
se1 =data1$se.exposure
N1=5959
R21 = 2 * beta1^2 * MAF1 * (1 - MAF1) / (2 * beta1^2 * MAF1 * (1 - MAF1) + se1^2 * 2 * N1 * M
AF1 * (1 - MAF1))
R_21=sum(R21)
R_21
```

```
## [1] 0.06742903
```

Calculation of power

```
expit <- function(x) { return(exp(x)/(1+exp(x))) }

rsq = R_2 # squared correlation
b1 = 0.2 # causal effect (log odds ratio per SD
#b1 = log(1.2) # or log of OR per SD)
sig = 0.05 # significance level (alpha)
pow = 0.8 # power level (1-beta)
#ratio = 361 # ratio of cases:controls = 1:ratio
n = 7738 # Sample size

rsq1 = R_21 # squared correlation
b11 = 0.2 # causal effect (log odds ratio per SD
#b1 = log(1.2) # or log of OR per SD)
sig1 = 0.05 # significance level (alpha)
pow1 = 0.8 # power level (1-beta)
#ratio = 361 # ratio of cases:controls = 1:ratio
n1 = 5959 # Sample size
```

Continuous outcome

```
cat("Power of analysis with ", n, "participants: ",
    pnorm(sqrt(rsq*b1^2*n)-qnorm(1-sig/2)))
```

```
## Power of analysis with  7738 participants:  0.8253615
```

```
cat("\nPower of analysis with ", n1, "participants: ",
    pnorm(sqrt(rsq1*b11^2*n1)-qnorm(1-sig1/2)))
```

```
##
## Power of analysis with  5959 participants:  0.9797726
```

F statistics

```
k=nrow(data)
F <- R_2*(N-1-k)/((1-R_2)*k)
F
```

```
## [1] 26.90658
```

Plot of power of TNFSF12

```
# Effect size grid
beta_values <- seq(0.05, 0.60, by=0.05)
```

```
# Funzione ipotetica per calcolare la potenza in base a beta e sample size
calculate_power <- function(b1, n, sig, rsq) {
  power <- pnorm(sqrt(rsq*b1^2*n)-qnorm(1-sig/2))*100
  return(power)
}

# Creare un data frame per contenere i risultati
results <- expand.grid(Beta = beta_values, SampleSize = n, Sigma = 0.05, Rsq = R_2 )
results$Power <- mapply(calculate_power, results$Beta, results$SampleSize,results$Sigma, results$Rsq)

results1 <- expand.grid(Beta = beta_values, SampleSize = n1, Sigma = 0.05, Rsq = R_21 )
results1$Power <- mapply(calculate_power, results1$Beta, results1$SampleSize,results1$Sigma, results1$Rsq)
```
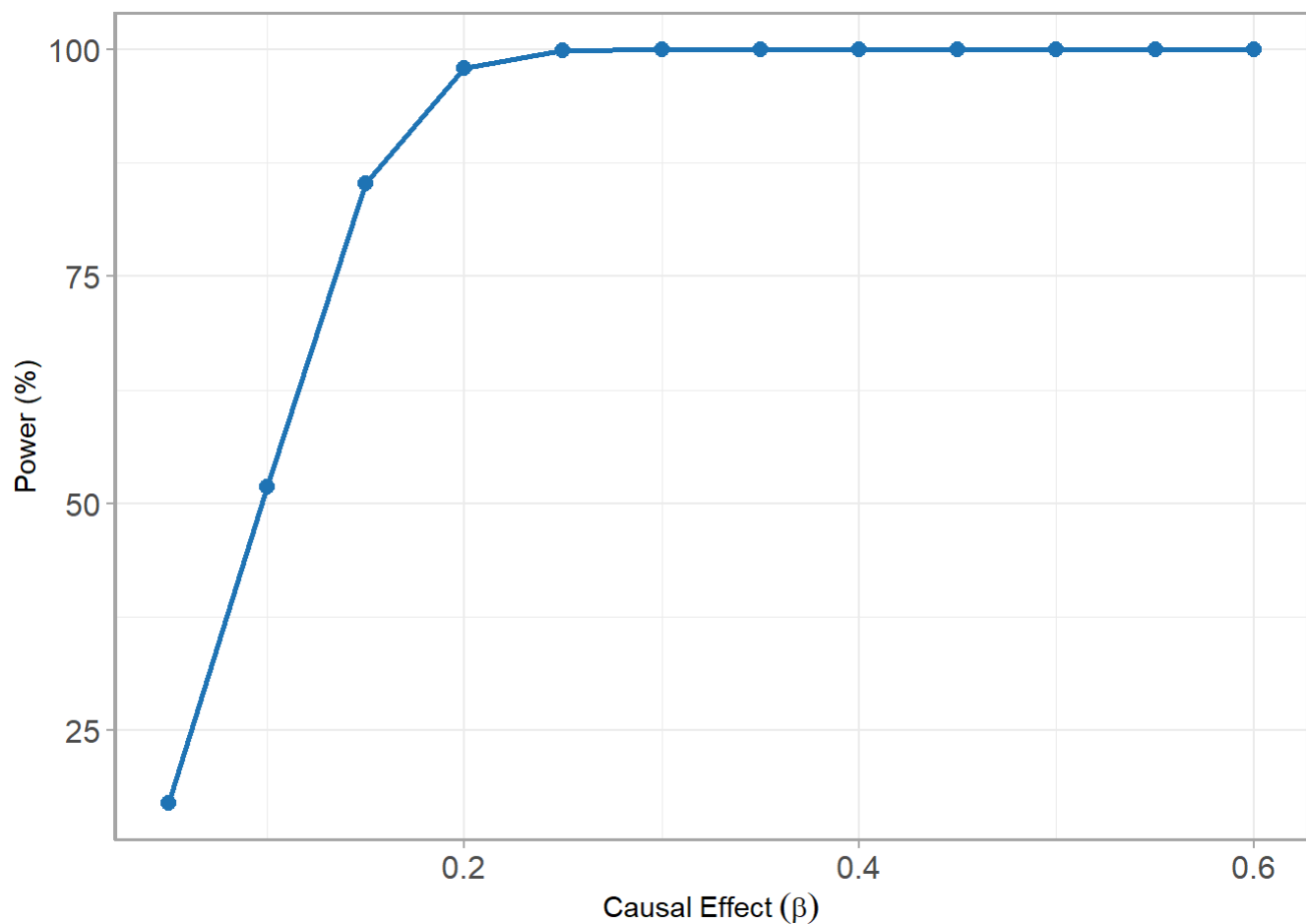
```
ggplot(results, aes(x = Beta, y = Power)) +
  geom_point(size = 2.5, color= "#1f77b4") +
  geom_line(lwd = 1, color="#1f77b4") +
  labs(x = expression(Causal~Effect~(beta)),
       y = "Power (%)") +
  theme_minimal() +
  theme(
    panel.border = element_rect(color = "darkgrey", fill = NA, size = 1),
    axis.line = element_line(color = "darkgrey", size = 0.5),
    axis.ticks = element_line(color = "darkgrey"),
    axis.text.x = element_text(size = 12),   # Modifica la dimensione dei numeri sull'asse x
    axis.text.y = element_text(size = 12)    # Modifica la dimensione dei numeri sull'asse y
  )
```
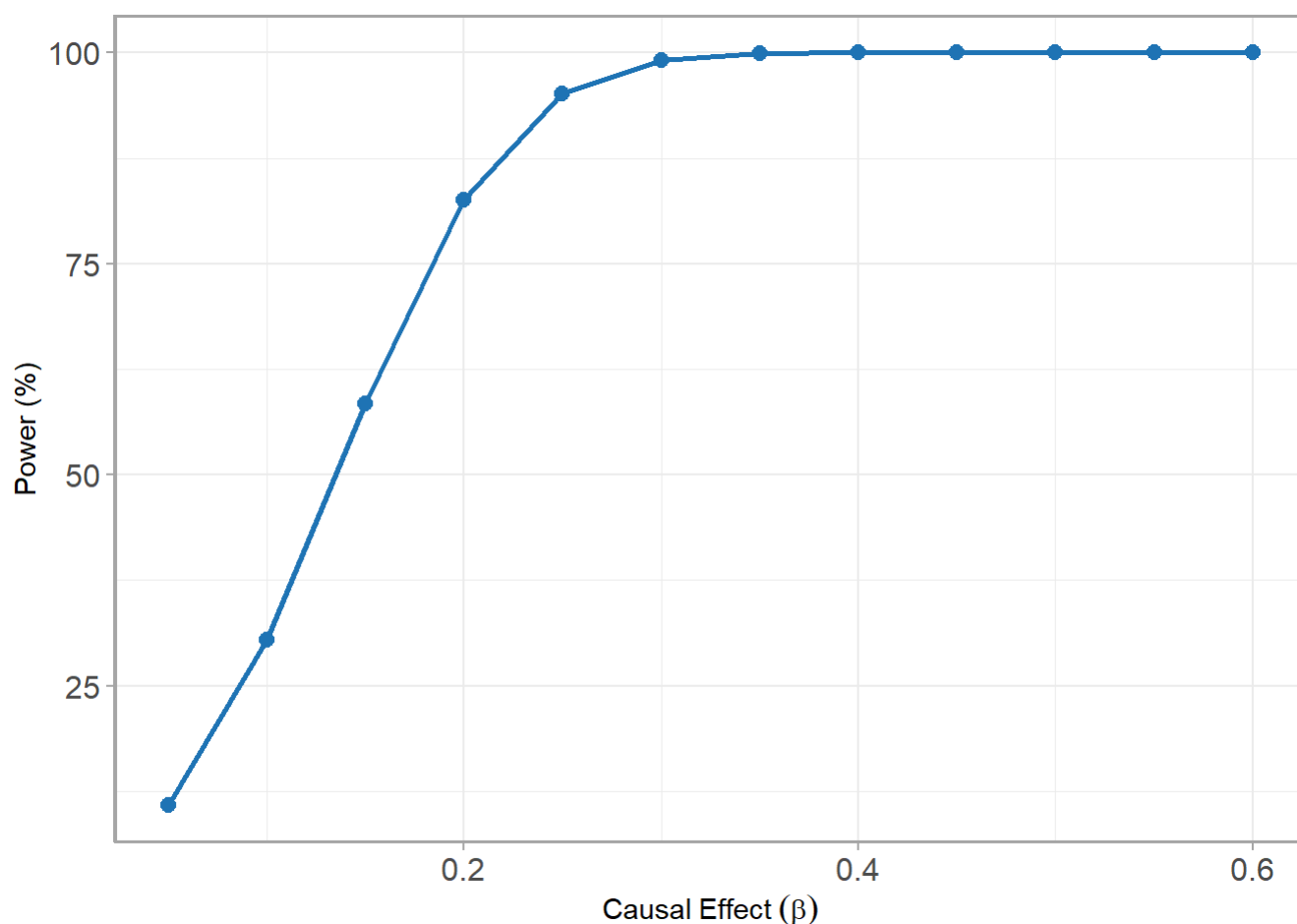


```
ggplot(results1, aes(x = Beta, y = Power)) +
  geom_point(size = 2.5, color= "#1f77b4") +
  geom_line(lwd = 1, color="#1f77b4") +
  labs(x = expression(Causal~Effect~(beta)),
       y = "Power (%)") +
  theme_minimal() +
  theme(
    panel.border = element_rect(color = "darkgrey", fill = NA, size = 1),
    axis.line = element_line(color = "darkgrey", size = 0.5),
    axis.ticks = element_line(color = "darkgrey"),
    axis.text.x = element_text(size = 12),   # Modifica la dimensione dei numeri sull'asse x
    axis.text.y = element_text(size = 12)    # Modifica la dimensione dei numeri sull'asse y
  )
```

```
# Assuming results and results1 are your datasets
results$Group <- 'GCST90027754'
results1$Group <- 'GCST90032220'

# Combine the datasets
combined_results <- rbind(results, results1)

# Plot
ggplot(combined_results, aes(x = Beta, y = Power, color = Group, group = Group)) +
  geom_point(size = 2.5) +
  geom_line(lwd = 1) +
  scale_color_manual(values = c("GCST90027754" = "#1f77b4", "GCST90032220" = "#ff7f0e")) +
  labs(x = expression(Causal~Effect~(beta)),
       y = "Power (%)",  color = "Exposure") +
  theme_minimal() +
  theme(
    panel.border = element_rect(color = "darkgrey", fill = NA, size = 1),
    axis.line = element_line(color = "darkgrey", size = 0.5),
    axis.ticks = element_line(color = "darkgrey"),
    axis.text.x = element_text(size = 12),
    axis.text.y = element_text(size = 12)
  )
```
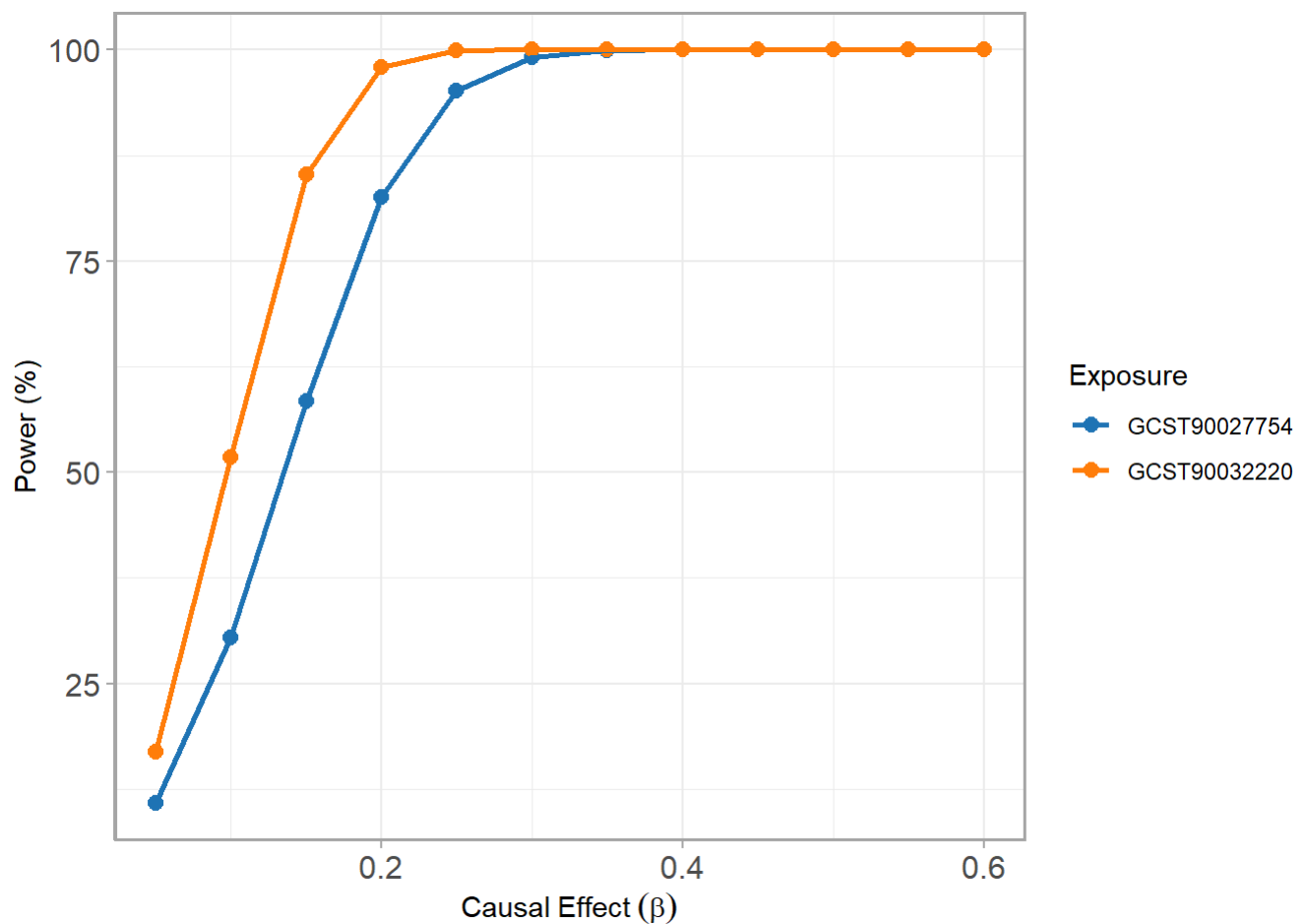
LACT vs In15

Load of IV

```
path <- "C:\\Users\\feder_phxiw9d\\Desktop\\Tesi\\Materiale_Serena\\Results\\Other_analysis
\\Results_LACT_In15\\Harm_data_E_GCST90027488_O_In15.txt"
data <- read.table(path, sep = "\t")
```

Calculation of explained variance R^2

```
eaf = data$eaf.exposure
MAF <- ifelse(eaf <= 0.5, eaf, 1-eaf)
beta = data$beta.exposure
se =data$se.exposure
N=7738
R2 = 2 * beta^2 * MAF * (1 - MAF) / (2 * beta^2 * MAF * (1 - MAF) + se^2 * 2 * N * MAF * (1 -
MAF))
R_2=sum(R2)
R_2
```

```
## [1] 0.03150409
```

Calculation of power

```
expit <- function(x) { return(exp(x)/(1+exp(x))) }


rsq = R_2 # squared correlation
b1 = 0.2 # causal effect (log odds ratio per SD
#b1 = log(1.2) # or log of OR per SD)
sig = 0.05 # significance level (alpha)
pow = 0.8 # power level (1-beta)
#ratio = 361 # ratio of cases:controls = 1:ratio
n = 7738 # Sample size
```

Continuous outcome

```
cat("Power of analysis with ", n, "participants: ",
    pnorm(sqrt(rsq*b1^2*n)-qnorm(1-sig/2)))
```

```
## Power of analysis with  7738 participants:   0.8775281
```

F statistics

```
k=nrow(data)
F <- R_2*(N-1-k)/((1-R_2)*k)
F
```

```
## [1] 27.93147
```

Plot of power of TNFSF12

```
# Effect size grid
beta_values <- seq(0.05, 0.60, by=0.05)
```

```
# Funzione ipotetica per calcolare la potenza in base a beta e sample size
calculate_power <- function(b1, n, sig, rsq) {
  power <- pnorm(sqrt(rsq*b1^2*n)-qnorm(1-sig/2))*100
  return(power)
}

# Creare un data frame per contenere i risultati
results <- expand.grid(Beta = beta_values, SampleSize = n, Sigma = 0.05, Rsq = R_2 )
results$Power <- mapply(calculate_power, results$Beta, results$SampleSize,results$Sigma, resu
lts$Rsq)
```

```
ggplot(results, aes(x = Beta, y = Power)) +
  geom_point(size = 2.5, color= "#1f77b4") +
  geom_line(lwd = 1, color="#1f77b4") +
  labs(x = expression(Causal~Effect~(beta)),
       y = "Power (%)") +
  theme_minimal() +
  theme(
    panel.border = element_rect(color = "darkgrey", fill = NA, size = 1),
    axis.line = element_line(color = "darkgrey", size = 0.5),
    axis.ticks = element_line(color = "darkgrey"),
    axis.text.x = element_text(size = 12),   # Modifica la dimensione dei numeri sull'asse x
    axis.text.y = element_text(size = 12)    # Modifica la dimensione dei numeri sull'asse y
  )
```