Exercise Intervention Trial in a Population at High Risk for Lung Cancer

Sample size estimates and power calculations

Daniel Spakowicz

5/31/2020

```
# Set parameters
accrue <-  46
dropout.rate <- 0.1
n <- accrue - (accrue * dropout.rate)
n.arm <- floor(n / 2)
conf.level <- 0.95
```

Feasibility outcomes

The primary goal of the exercise intervention is to evaluate the feasibility in a population at of older adults at high risk of developing lung cancer who are undergoing annual screening at the OSU lung cancer screening clinic (LCSC). We will recruit 36 indivduals from the LCSC to account for a conservative drop-out rate of 10%, giving us an evaluable sample size of 32. A follow-up rate of study participation at 12 weeks of 75% (p1) would be desirable; a continuation rate below 40% (p0) would be unacceptable. With 30 eligible and recruited participants, this study is powered to reject acceptability of these measures if the follow-up rate is less than 40% (10% one-sided type one error) and deemed feasible if the follow-up rate is 75% or more (80% power).

```
# Proportion of participants who fully comply
p1 <- 0.75

# Proportion of null
p0 <- 0.5

power.prop1.test(n = n,
                 p1 = p1,
                 p0 = p0,
                 sig.level = 1- conf.level,
                 power = NULL,
                 alternative = c("greater"),
                 cont.corr = TRUE)
```

```
##
##      Power calculation for testing a given proportion (with continuity correction)
##
##               n = 41.4
##           delta = 0.25
##              p1 = 0.75
##              p0 = 0.5
```

```
##       sig.level = 0.05
## exact.sig.level = 0.02177926
##           power = 0.9490866
##     exact.power = 0.9194713
##     alternative = greater
##
## NOTE: n = total sample size
```

The LCSC screens an average of 16 patients per week (8 per clinic, 2 clinics per week) 48 weeks each year, for a total of 768 patients per year. The rate at which patients receive a cancer diagnosis is 10%. Surrogate markers of inflammation in large, prospective cohorts have estimated that exercise reduces risk by 10%

Microbiome outcomes

Alpha diversity difference (normal distribution)

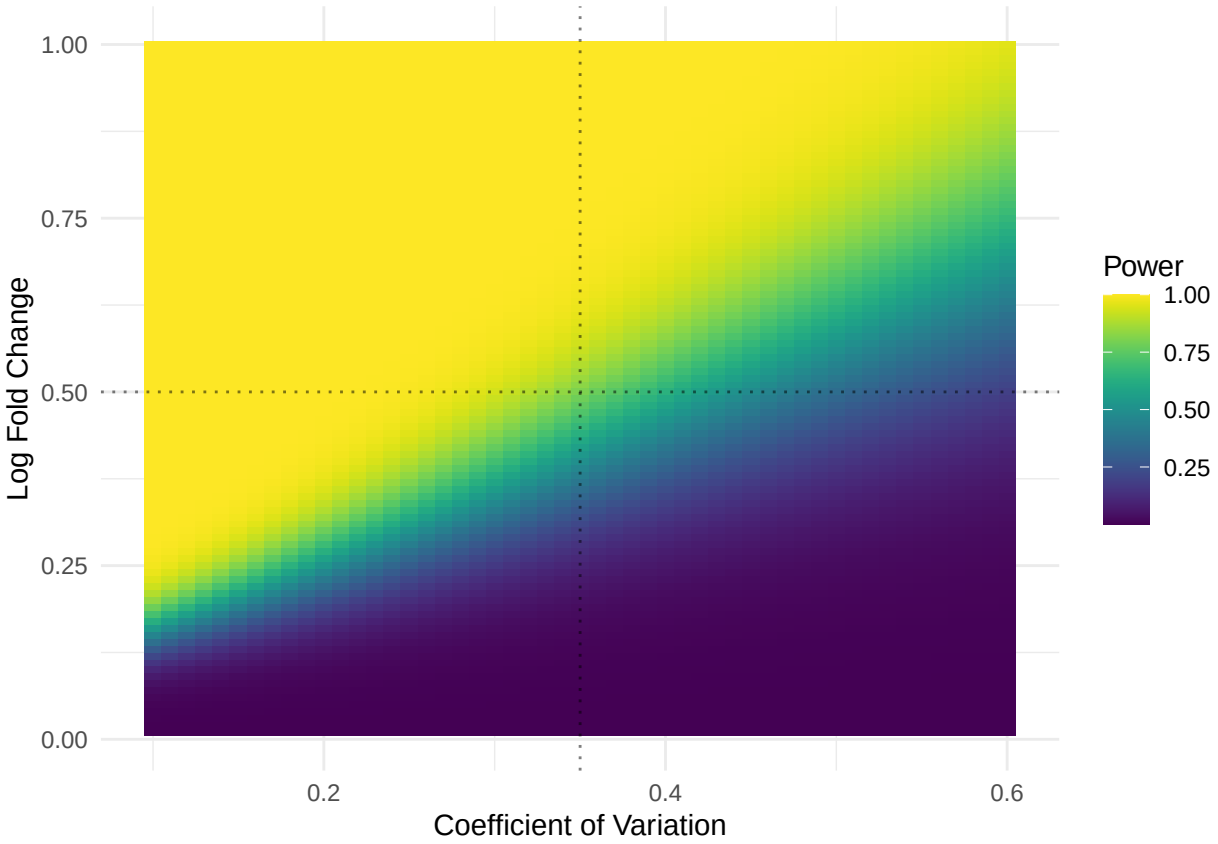
```
# Power
pwr::pwr.t.test(n = n.arm,
                d = NULL,
                sig.level = 0.05,
                power = 0.8,
                type = "paired",
                alternative = "less")
```

```
##
##      Paired t test power calculation
##
##               n = 20
##               d = -0.5769185
##       sig.level = 0.05
##           power = 0.8
##     alternative = less
##
## NOTE: n is number of *pairs*
```

Difference in the relative abundances of microbes (negative binomial)

```
# Define the coefficient of variation and fold change ranges expected of
# microbiome experiments
cv <- seq(0.1, 0.6, by = 0.01)
log.fc <- seq(0.01, 1, by = 0.01)

# Define the number of tests
n.tests <- 100

# Calculate power
power.l <- list()
for (i in 1:length(cv)) {
  power.l[[as.character(cv[i])]] <-
    power.hart(n = n.arm,
               alpha = 0.05 / n.tests,
```

```
                log.fc = log.fc,
                mu = rep(100, length(log.fc)),
                sig = rep(cv[i], length(log.fc)))
}

# Convert to long-format data frame
power.df <-
  power.l %>%
  bind_rows() %>%
  mutate(log.fc = log.fc) %>%
  gather(cv, power, -log.fc) %>%
  mutate(cv = as.numeric(cv)) %>%
  mutate(fold.change = 2^log.fc)

a1.rep <-
  power.df %>%
  filter(power >= 0.8) %>%
  filter(log.fc == 0.5) %>%
  arrange(power) %>%
  head(., 1)

# Plot result
power.df %>%
  ggplot(aes(x = cv, y = log.fc, z = power)) +
  geom_raster(aes(fill = power)) +
  labs(x = "Coefficient of Variation",
       y = "Log Fold Change",
       fill = "Power") +
  theme_minimal() +
  scale_fill_viridis() +
  geom_hline(aes(yintercept = 0.5), linetype = "dotted", alpha = 0.5) +
  geom_vline(xintercept = a1.rep$cv, linetype = "dotted", alpha = 0.5)
```




```
ggsave("../figures/fold-change_v_cv.png", height = 3, width = 4)
```

Inflammatory biomarker outcomes

Difference in CRP, IGF and IL6 (normal distribution)

```
n.tests <- 3

# Power
pwr::pwr.t.test(n = n.arm,
                d = NULL,
                sig.level = 0.05/n.tests,
                power = 0.8,
                type = "paired",
                alternative = "less")
```

```
##
##      Paired t test power calculation
##
##               n = 20
##               d = -0.7068855
##       sig.level = 0.01666667
##           power = 0.8
##     alternative = less
##
```

```
## NOTE: n is number of *pairs*
```