

# Power calculations JOP submission

This version: June 2023

```
rm(list=ls())
pacman::p_load(here, Hmisc, dplyr, gridExtra, grid, stargazer, hrbrthemes, quanteda, readtext,
               tidyverse, knitr, papeR, tidyr, kableExtra, ggpubr, printr, tab, lfe, sjPlot,
               clusterSEs, BBmisc, miceadds, mediation, estimatr, sensemakr, SentimentAnalysis,
               ggplot2, ggridges, viridis, randomizr, doParallel, sandwich, ggplot2, reshape2,
               kableExtra, ggpattern)
```

## JOP Registered Study: peer praise effect on inclusion towards racial/ethnic outgroups

Here we conduct power calculations for the effects of **Peer Praise** on following outcomes (all for where the target is an outgroup member; we are focusing on a White American receiving peer praise versus control):

- **(H1 DV Behavioral Empathy Task)** Measure the effect of the treatment (praise) on the behavioral empathy task (choosing the FEEL task), when target is an outgroup member.
- **(H2 DV Self-reported Empathy)** Measure if the effect of treatment (praise) on the outcome (reporting feeling empathy for racial/ethnic outgroups), when target is an outgroup member.
- **(H3 DV Donate)** Measure if the effect of treatment (praise) on the outcome (donate bonus amount), when target is an outgroup member.
- **(H4 DV Letter)** Measure if the effect of treatment (praise) on the outcome (write supportive letter), when target is an outgroup member.
- **(H5 DV Social Distance)** Measure if the effect of treatment (praise) on the outcome (social distance scale), when target is an outgroup member.
- **(H6 DV Thermometer)** Measure if the effect of treatment (praise) on the outcome (thermometer rating), when target is an outgroup member.

## Assumptions

Power calculations require two main inputs: estimates of effects sizes and standard deviations. Below, we detail where our estimates for those parameters come from. Whenever possible, we use information from our extensive series of pilots for these values. Where that's not possible, we base our estimates on studies that are as similar to ours as possible along key dimensions (for example, in using the same dependent measure). The table below highlights the source for these values that are subsequently used in the power calculations.

Here we make several key assumptions for our simulations: 1) effect sizes for peer praise on the six outcome DVs for outgroup members and 2) their standard deviations.

1. For H1, we take effect and sd information from Pilot Study 6, which similarly utilized peer praise on the behavioral choice outcome.
2. For H2 Self reported empathy, as we do not have a piloted peer praise study to directly look for effect sizes, we look at findings from Williamson et al. 2021, which had a similar self-reported empathy outcome but utilized a family history perspective as its intervention.

3. For H3 similarly, we use the same study, from Williamson et al. 2021, and borrow from their outcome of expressed support for more open policies; as we expect donation might be harder to move on than support for policies, we reduce by 75
4. For H4 Letter we look at a similar tested outcome from Adida Lo Platas 2018 Table S1—a letter to the White House—and utilize the same treatment effect sizes and sds.
5. For H5 Social distance scores we do not have an equivalent study to lean on for priors on effect sizes, so we assume an increment of 0.1 rank move with a similar sized sd.
6. For H6 Thermometer, we also turn to similar tested outcomes from Adida Lo Platas 2018 Table S1 – a thermometer towards an outgroup (though in this case refugees) and use the same treatment effect size and sds estimated.

We calculate wages and time as well, using the highest min wage in 2022 (California) \$15/hr.

Finally we keep track of our 6 hypotheses in this peer praise on outgroup section for MH adjustments later.

```
set.seed(123)
# n sizes we simulate through
possible.ns <- c(seq(from=100, to=800, by=50))
# n of trials (here we do only 1)
n_trials<-c(1)
n_k<-expand.grid(possible.ns,n_trials)
names(n_k)<-c("N","K")
# we calculate time to take study and pay
n_k$time<-8+(n_k$K*1.0)
n_k$pay<-(((n_k$time/60)*15)+1)*n_k$N
# alpha level, sims set, number of hypotheses to adjust
alpha <- 0.05
sims <- 1000
n.hyp = 6

# Population assumptions for respondents
# based on prior proportions in pilots
#proportion of White, proportion of Black in sample
#mean(c(61.3,70.4,54.9,70, 64.5,72.9)) White
prop_white<-66
#mean(c(10.4,13,14.6,13,22.2,15.4))
prop_black<-15

# Effects: treatment main effects for each of H1-H6
sds<-rep(NA,n.hyp)
#H1: effect and sd information from Pilot Study 6
beta1=0.17
sds[1]=0.0655
#H2: effect and sd from Williamson et al. 2021, which had a
#similar self-reported empathy outcome
beta2=0.171 #H2
sds[2]=0.042
#H3: effect and sd from Williamson et al. 2021, and borrow from
#their outcome of expressed support for more open policies;
#as we expect donation might be harder to move on than
#support for policies, we reduce by 75% the treatment effect
#size from the study (and conservatively keep the same sd).
beta3=0.397*0.25 #H3
sds[3]=0.112
```

```

#H4: effect and sd from similar tested outcome from
#Adida Lo Platas 2018 Table S1---a letter to the White House
#and utilize the same treatment effect sizes and sds.
beta4=0.165
sds[4]=0.02
#H5 Social distance scores we do not have an equivalent study
#to lean on for priors on effect sizes, so we assume an
#increment of 0.1 rank move with a similar sized sd.
beta5=0.1
sds[5]=0.1
#H6: effect and sd from similar tested outcomes from
#Adida Lo Platas 2018 Table S1 -- a thermometer towards an
#outgroup (though in this case refugees) and use the same
#treatment effect size and sds estimated.
beta6=4.622
sds[6]=0.08

```

## Generate simulations

One potential complication in our power calculations (that makes using “off the shelf” solutions difficult) is that we must consider the overall number of subjects as well as the “usable” number of subjects,” since our main registered hypotheses are for White subjects (and some proportion of respondents recruited will not be White). The code below takes this into account in the following way:

Because we’d like to account for the probabilistic aspects of 1) receiving respondents who are not all White Americans (for which our main registered hypotheses are on) and 2) whether they are randomized to see an outgroup image (as the platform does not allow for sampling conditional on race), we simulate 1000 experiments for each N size with a generated set of data in each simulation and estimate the six tests (with registered Benjamini and Hochberg MH adjustments) and store. From here we can find how often, under each N scenario, the six treatment effects are able to be recovered — the level of power. We aim for a power level above 0.95 per JOP suggestions.

```

# Data holders: p-values and means
set.seed(123)
c.main1 <- c.main2 <- c.main3 <- c.main4 <- c.main5 <- c.main6 <-rep(NA, sims)
p.main1 <- p.main2 <- p.main3 <- p.main4 <- p.main5 <- p.main6 <-rep(NA, sims)
obs <- rep(NA, sims)
#set up parallel cores ii 185-191
no_clusters<-detectCores()-2
if (Sys.getenv("RSTUDIO") == "1" && !nzchar(Sys.getenv("RSTUDIO_TERM"))) &&
  Sys.info()["sysname"] == "Darwin" && getRversion() >= "4.0.0") {
  parallel::setDefaultClusterOptions(setup_strategy = "sequential")
}
clust <- parallel::makeCluster(no_clusters)
registerDoParallel(clust)
t<-Sys.time()
mydata<-vector("list",nrow(n_k))
#### Outer loop to vary the overall number of subjects ####
dat <- foreach(j=1:nrow(n_k), .combine=rbind,.packages = c("lmtest")) %dopar% {
  N <- n_k$N[j]
  for (i in 1:sims){#inner loop
    ### (I) ###
    # Generate Respondent Race, Treatment/Image, Respondent ID
    RespondentRace<-sample(c("Black","White","Other"),N,prob=c(prop_black

```

```

      , prop_white,(100-prop_black-prop_white)), replace=TRUE)
Treatment<- sample(rep(c("Peer Praise","Control"),each=N/2),N,replace=F)
ID<-1:N
Image_name<-sample(c("Black", "Hispanic", "White"),N,prob=c(0.45, 0.45, 0.1),replace=T)
ImageBH<-ifelse(Image_name=="White","Ingroup","Outgroup")

#Bind the X data
X<-data.frame(ID=ID,RespondentRace=RespondentRace,Treatment=as.factor(Treatment),
              ImageBH=as.factor(ImageBH),Image_name=as.factor(Image_name))
X$ControlvPraise<-ifelse(X$Treatment=="Control",0,1)

X <- within(X, ImageBH <- relevel(ImageBH, ref = "Ingroup"))
X <- within(X,Treatment <- relevel(Treatment, ref="Peer Praise"))
d <- subset(X, RespondentRace=="White") #Powered hypotheses on White resp
### (II) ###
# Generate Ys
PeerPraise<-d$ControlvPraise
ImageBH<-ifelse(d$ImageBH=="Ingroup",0,1)
d$Y1 = beta1*PeerPraise+ rnorm(length(PeerPraise),0,sds[1]) #H1
d$Y2 = beta2*PeerPraise+ rnorm(length(PeerPraise),0,sds[2]) #H2
d$Y3 = beta3*PeerPraise+ rnorm(length(PeerPraise),0,sds[3]) #H3
d$Y4 = beta4*PeerPraise+ rnorm(length(PeerPraise),0,sds[4]) #H4
d$Y5 = beta5*PeerPraise+ rnorm(length(PeerPraise),0,sds[5]) #H5 distance
d$Y6 = beta6*PeerPraise+ rnorm(length(PeerPraise),0,sds[6]) #H6
dout<-subset(d,ImageBH=="Outgroup")
### (III) ###
## Estimate models: H1-H6
main1 <- lm(formula = Y1 ~ ControlvPraise, data = dout)
main2 <- lm(formula = Y2 ~ ControlvPraise, data = d)
main3 <- lm(formula = Y3 ~ ControlvPraise, data = d)
main4 <- lm(formula = Y4 ~ ControlvPraise, data = d)
main5 <- lm(formula = Y5 ~ ControlvPraise, data = d)
main6 <- lm(formula = Y6 ~ ControlvPraise, data = d)
### (IV) ###
## Store coefficients & Benjamini-Hochberg adjusted pvals
obs[i] <- nrow(model.frame(main1))
c.main1[i] <- coef(main1)[2] #H1
p.main1[i] <- p.adjust(summary(main1)$coefficients[2,4],"BH",n=n.hyp)
c.main2[i] <- coef(main2)[2] #H2
p.main2[i] <- p.adjust(summary(main2)$coefficients[2,4],"BH",n=n.hyp)
c.main3[i] <- coef(main3)[2] #H3
p.main3[i] <- p.adjust(summary(main3)$coefficients[2,4],"BH",n=n.hyp)
c.main4[i] <- coef(main4)[2] #H4
p.main4[i] <- p.adjust(summary(main4)$coefficients[2,4],"BH",n=n.hyp)
c.main5[i] <- coef(main5)[2] #H5
p.main5[i] <- p.adjust(summary(main5)$coefficients[2,4],"BH",n=n.hyp)
c.main6[i] <- coef(main6)[2] #H6
p.main6[i] <- p.adjust(summary(main6)$coefficients[2,4],"BH",n=n.hyp)
}
## Collect power info
powerAll<- mean(abs(c.main1)>0.0 & p.main1<alpha & abs(c.main2)>0.0 & p.main2<alpha
               & abs(c.main3)>0.0 & p.main3<alpha & abs(c.main4)>0.0 & p.main4<alpha
               & abs(c.main5)>0.0 & p.main5<alpha & abs(c.main6)>0.0 & p.main6<alpha)

```

```

## Collect remaining obs
Obs <- mean(obs)
data.frame(Obs=Obs,powerAll=powerAll)
}
stopCluster(clust)
print(Sys.time()-t)

dat<-cbind(n_k,dat)

saveRDS(dat,file=here::here("power_JOP_v2.rds"))

```

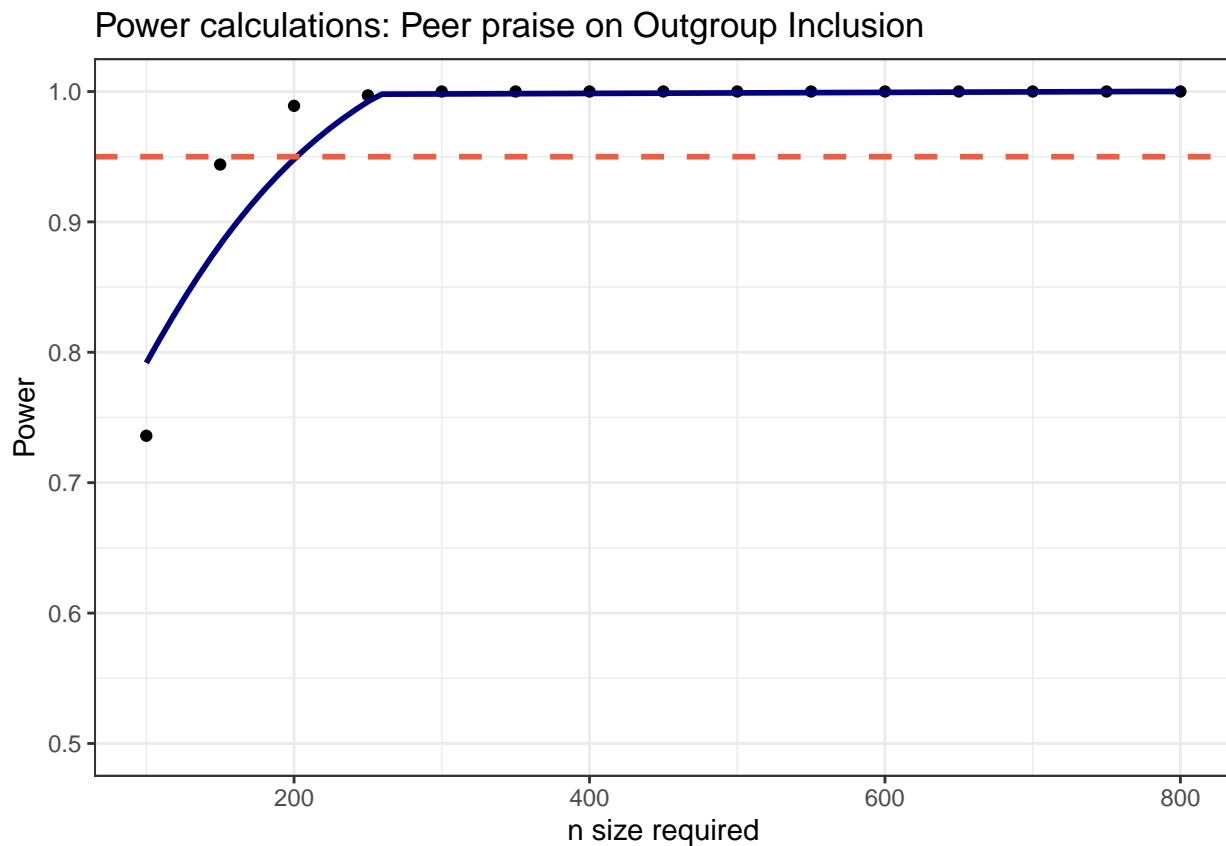
## Plots

```

dat<-readRDS(file=here::here("power_JOP_v2.rds"))

#power
plot1<- ggplot(data=dat,aes(x=N,y=powerAll))+
  geom_point(alpha=1)+
  labs(x="n size required",y="Power") + ylim(0.5,1) +
  theme_bw()+
  geom_smooth(se = FALSE, method = "loess", color="navy")+
  geom_hline(yintercept=0.95,linetype="dashed",color = "tomato2", size=1) +
  ggtitle("Power calculations: Peer praise on Outgroup Inclusion")
plot1

```



```
ggsave(plot1,file=here::here("Figures","powerJOP v2.pdf"), width=8, height=7)
```

To power around 0.95 for H1-H6 we require around  $N=200$  observations; we conservatively plan to field 400 observations.

## Co-partisan peer praise effect on inclusion towards racial/ethnic outgroups.

We also consider the same set up as the above but utilize praise from co-partisan peers on MTurk for our parallel “Co-Partisan Peer Praise”; that is we design and field a separate study that looks at the effects of peer praise (from peers who share partisan identity with the respondent) on inclusion towards racial/ethnic outgroups. Again, we have the same six outcomes, though this time, we generate peer praise such that each time a respondent is randomized to receive treatment, they receive praise from a co-partisan group of peers (e.g. a respondent who self-identifies as Republican receives peer praise from a group of Republican peers).

## Assumptions

As it’s unclear whether co-partisan identity creates a peer praise effect that is stronger or weaker compared to peer praise from a general peer, we make the same assumptions of peer praise effect sizes as before and therefore similarly expect in powering around 0.95 for the six tests, we require another  $N=400$ .

## Mock ups

Code to generate mock up fig.

```
set.seed(12345)
#Set up data for Mock-up
N<-400
# Generate Respondent Race, Treatment/Image, Respondent ID
RespondentRace<-sample(c("Black", "White", "Other"),N,prob=c(prop_black
                    ,prop_white,(100-prop_black-prop_white)), replace=TRUE)
Treatment<- sample(rep(c("Peer Praise","Control"),each=N/2),N,replace=F)
ID<-1:N
Image_name<-sample(c("Black", "Hispanic", "White"),N,prob=c(0.45, 0.45, 0.1),replace=T)
ImageBH<-ifelse(Image_name=="White","Ingroup","Outgroup")
#Bind the X data
X<-data.frame(ID=ID,RespondentRace=RespondentRace,Treatment=as.factor(Treatment)
              ,ImageBH=as.factor(ImageBH),Image_name=as.factor(Image_name))
X$ControlvPraise<-ifelse(X$Treatment=="Control",0,1)
X <- within(X, ImageBH <- relevel(ImageBH, ref = "Ingroup"))
X <- within(X,Treatment <- relevel(Treatment, ref="Peer Praise"))
d <- subset(X, RespondentRace=="White") #Primary powered hypotheses on White respondents
# Generate Y1 (pick some reasonable baseline Y0 values to plot)
PeerPraise<-d$ControlvPraise
ImageBH<-ifelse(d$ImageBH=="Ingroup",0,1)
d$Y1 = 0.3 + beta1*PeerPraise + rnorm(length(PeerPraise),0,sds[1])
dout<-subset(d,ImageBH=="Outgroup")
## Estimate models: H1 (mock just do one)
main1 <- lm(formula = Y1 ~ ControlvPraise, data = dout)
#p.adjust(summary(main1)$coefficients[2,4], "BH",n=n.hyp)
## Graphics
mock<-data.frame(Treatment=c("Control","Peer Praise"),
                 Mean=c(main1$coefficients[1],sum(main1$coefficients)))
# For tweaking position of brackets
e = 0.3
```

```

r = 0.3
w = 0.025
cp1<-0.525
cp2<-0.6
nudgeup<-0.02
mockplot<-ggplot(mock,aes(x=Treatment, y=Mean, fill=Treatment)) + theme_bw() +
  geom_bar(stat="identity") +
  ylim(0,0.8) + ylab("Behavioral Choice Task: choose empathy") + xlab("") +
  scale_fill_viridis(option="magma",discrete=T,begin=0.15,end=0.7,alpha=0.75) +
  #Mean labels
  geom_text(x=1,y=mock$Mean[1]+nudgeup,label=paste(round(mock$Mean[1],3),color="gray40") +
  geom_text(x=2,y=mock$Mean[2]+nudgeup,label=paste(round(mock$Mean[2],3),color="gray40") +
  #Control v Peer Praise
  geom_text(x=1.5,y=cp1+r*e+w,label=paste("
      Difference in means=0.179305 (MH adj p=4.414082e-48)",color="gray30") +
  geom_segment(x=1,xend=2,y=cp1+r*e,yend=cp1+r*e,size=0.3,color="gray30") +
  geom_segment(x=1,xend=1,y=cp1+r*e,yend=cp1+r*e-0.08,size=0.3,color="gray30") +
  geom_segment(x=2,xend=2,y=cp1+r*e,yend=cp1+r*e-0.08,size=0.3,color="gray30")
#mockplot
ggsave(mockplot,file=here::here("Figures","mockplot v2.pdf"),width=8,height=7.5)

```