

**Supplementary material**

McNew et al. 2020

**No evidence of sex ratio manipulation by Galápagos mockingbirds in response to environmental conditions**

**Supplemental Material**

Table S1. Sex ratio variation among years analyzing complete broods only

<b>cbind(male,female)</b>				
Predictors	Odds Ratios	95% CI	<i>P</i>	Group*
(Intercept)	0.79	0.36 – 1.73	0.549	A
year 2013	1.27	0.45 – 3.59	0.649	A
year 2015	1.65	0.60 – 4.53	0.334	A
year 2016	1.27	0.45 – 3.59	0.649	A
Observations	49			
R <sup>2</sup> Tjur	0.003			

\*Group designation based on post-hoc pairwise comparisons of years using Tukey post hoc tests.

The same letter assignment for all years indicates that no pairwise differences were significant.

Table S2. Sex ratio variation among years assuming all missing nestlings were female.

<b>cbind(male,(female+missing))</b>				
Predictors	Odds Ratios	95% CI	<i>P</i>	Group
(Intercept)	0.45	0.29 – 0.71	<b>0.001</b>	A
year 2013	1.72	0.93 – 3.16	0.082	A
year 2015	1.33	0.73 – 2.44	0.350	A
year 2016	1.50	0.79 – 2.85	0.218	A
Observations	116			
R <sup>2</sup> Tjur	0.000			

Table S2. Sex ratio variation among years assuming all missing nestlings were male.

<b>cbind((male+missing),female)</b>				
Predictors	Odds Ratios	95% CI	<i>P</i>	Group
(Intercept)	1.64	1.06 – 2.52	<b>0.026</b>	A
year 2013	1.18	0.64 – 2.17	0.586	A
year 2015	0.90	0.50 – 1.62	0.730	A
year 2016	0.81	0.44 – 1.52	0.521	A
Observations	116			
R <sup>2</sup> Tjur	0.020			

# Supplementary material

Sabrina McNew

1/17/20

## Missing data

We found no evidence of sex ratio bias in response to environmental conditions (rainfall). However, we did not sex all of the nestlings in the study. Most missing values (19% of total nestlings) are from nestlings that died before blood could be taken, or we were unable to sample blood from that nestling. A small number (4% of total nestlings) did not amplify. Is it possible that sex biases in the missing data could have affected our results?

```
library(dplyr)
library(lme4)
library(lmerTest)
set.seed(1234)
sexnest <- read.csv("./CSVs/sexpernest.csv")
sexnest$hatch.date <- as.Date(sexnest$hatch.date, format= "%m/%d/%y")
sexnest$year <- as.factor(sexnest$year)
sexnest <- sexnest[sexnest$round==1,] #Only include first nesting attempts

sex <- read.csv("./CSVs/nestling-sex2020.csv") #each row is a nestling

head(sexnest) #The "missing" column gives the number of missing nestlings from that nest
```

```
##   yearnest year nest treatment round repeat. male female nestlings
## 1 2012-11 2012 11 P 1 1 0 1 3
## 2 2012-19 2012 19 P 1 1 0 1 4
## 3 2012-2 2012 2 W 1 1 0 1 3
## 4 2012-20 2012 20 W 1 1 1 1 4
## 5 2012-24 2012 24 W 1 1 2 0 3
## 6 2012-25 2012 25 P 1 1 1 0 3
## hatch.date cumulative.rain propfemale complete missing
## 1 2012-03-02 190.5 1.0 0 2
## 2 2012-03-15 190.5 1.0 0 3
## 3 2012-03-24 191.7 1.0 0 2
## 4 2012-03-08 190.5 0.5 0 2
## 5 2012-03-01 190.5 0.0 0 1
## 6 2012-03-08 190.5 0.0 0 2
```

## 1. Simulations

To answer this question we simulated missing data under different scenarios, added those data to our observed data, and asked “Do we still get the same (negative) result?”

How much missing data overall?

```
sum(sexnest$missing)/sum(sexnest$nestlings) #23%
```

```
## [1] 0.2320442
```

## The basic simulation

First, to demonstrate the idea, we will simulate data with a certain probability of being female using the binomial distribution. A probability of 0.5 would give equal ratios of males and females;  $P = 1$  would mean all missing nestlings were female;  $P = 0$  would mean all missing nestlings were male.

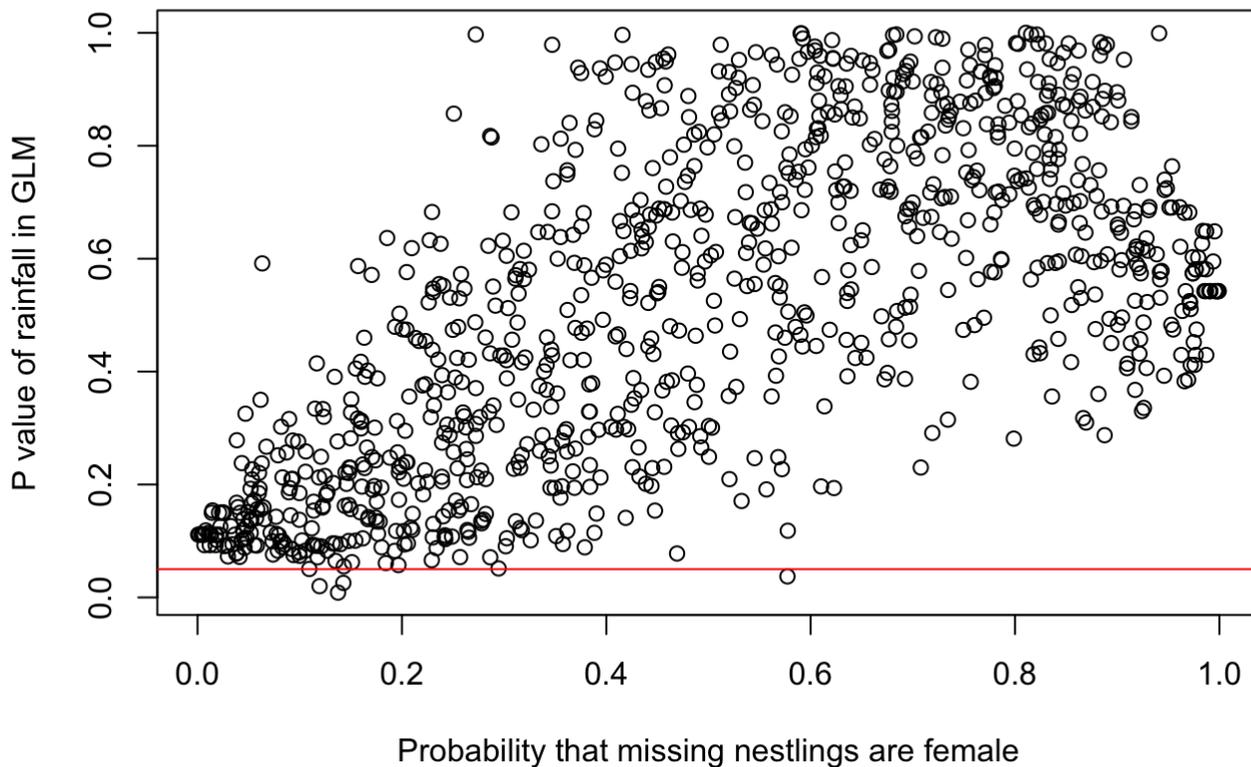
To simulate these data, we will:

- 1) Pick a uniform range of Ps between 0 and 1 to create various scenarios
- 2) Calculate the number of males and females that would have been present in the missing data under each scenario
- 3) Add those “found” nestlings to our observed data
- 4) Use a glm to test whether with this bigger dataset there is a significant relationship between rainfall at each nest and the sex ratio of its nestlings.

```
#Subset data
testdat1 <- select(sexnest, year, male, female, nestlings, missing, cumulative.rain)
#Create a dataframe to save our results
saveframe <- data.frame(iteration = 1:1000, prob.female = NA,
                        pval.rain = NA, prop.missing.female = NA)

#The simulation
for (i in 1:1000) {
  testdat <- testdat1
  prob.female <- runif(1, 0, 1) #Pick a number between 0 and 1, the p(female) for this i
  teration
  n.female <- round(rbinom(nrow(testdat), testdat$missing, prob.female)) #number of miss
  ing nestlings that are female
  n.male <- testdat$missing - n.female #number missing nestlings that are male
  testdat$total.f <- testdat$female + n.female #total number of female nestlings
  testdat$total.m <- testdat$male + n.male #total number of male nestlings
  saveframe$prob.female[i] <- prob.female #save iteration female probability
  saveframe$prop.missing.female[i] <- sum(n.female)/sum(testdat$missing) #proportion of
  missing nestlings that were assigned female
  #run a glm testing for differences in ratio female:male depending on rain and save p v
  alue.
  saveframe$pval.rain[i] <- summary(glm(cbind(total.f,total.m) ~ cumulative.rain,
                                       data=testdat, "binomial"))$coefficients[2,4]
}
```

We can plot the data to see if there's a relationship between the probability that missing nestlings are female and the p value of  $\text{glm}(\text{sex ratio} \sim \text{rainfall})$ . The red line below is  $p = 0.05$ . Points above the line indicate simulations where adding missing data gives us the same result as our observed data alone (no significant effect of rainfall on sex ratio). Any points below that line indicate a simulation where adding missing data leads to a result that there is a significant effect of rainfall on sex ratio.



Out of 1000 simulations, how often was there a significant p value for the effect of rain on sex ratio?

```
filter(saveframe, pval.rain < 0.05) %>% nrow() #4
```

```
## [1] 4
```

Out of 1000 simulations, only 4 resulted in rain having a significant effect on nestling sex ratio. This result suggests that even if nestling mortality is heavily sex biased, it is unlikely to have affected our results about the relationship between rainfall and sex ratio. These results are consistent with the simpler analysis in the main text showing that even assuming all missing nestlings were all male or all female we do not find a significant relationship between rainfall and brood sex ratio.

This is not surprising, of course, because if the missing data are sex biased, but bias is unrelated to our factors of interest (i.e. rainfall) then overall there still will be no difference in sex ratios of nests in different environmental conditions.

The more complicated scenario is when sex ratios of the missing nestlings depends on rainfall.

## Simulation: sex of missing nestlings depend on rainfall

Now, we will simulate data to test whether sex bias in missing nestlings related to rainfall levels created a Type II error.

# Simulating reasonable effect sizes

We have not found any other studies that report effect sizes of rainfall on nestling sex ratios. However, other studies of sex ratios with similar sample sizes report odds ratios between 0.5 and 7.0 between various treatment groups. We will create “effect sizes” that create similar odds ratios between the extremes of our rainfall values. In those papers, the most extreme sex ratio for a group was 0.3, so we will use that as a starting point for the driest nests. Then, based on a randomly chosen odds ratio, we calculate the odds that a missing nestling in the wettest nest was male, and create probabilities that missing nestlings were male along a linear gradient from least to most wet.

For instance, at a fairly extreme odds ratio of 5.0, in the driest nests, the probability that a missing nestling is male will be 0.3, and the probability that a missing nestling is male in the wettest nests can be calculated as follows:

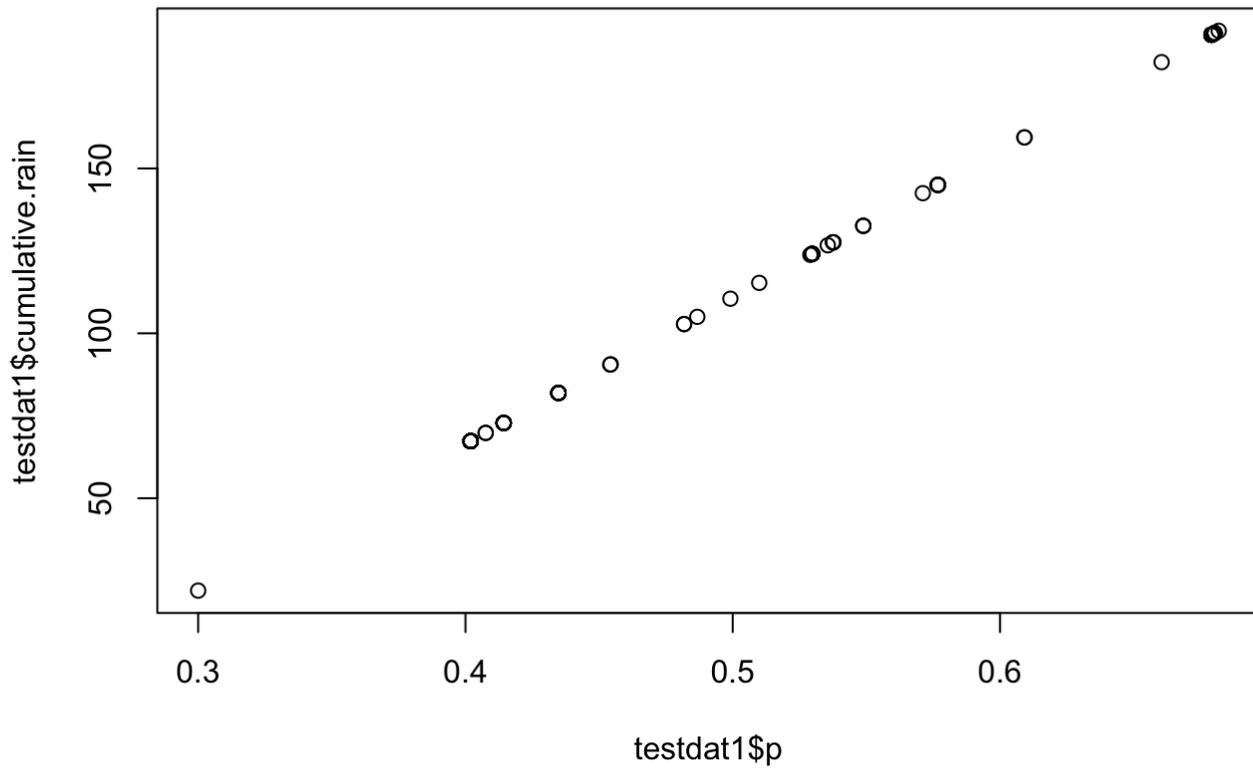
```
low_p <- 0.3 #prob of male in driest nests
low_odds <- low_p/(1-low_p)
odds_ratio <- 5 #An example odds ratio between driest and wettest conditions
high_odds <- low_odds * odds_ratio #odds at the wet extreme
high_p <- high_odds/(1+high_odds) #probability at the wet extreme; 0.68
```

We then can use algebra to create a linear gradient of probabilities based on rainfall

```
unit_increase <- (high_p - low_p) / (max(testdat1$cumulative.rain) - min(testdat1$cumulative.rain)) #calculate the slope of the relationship
intercept <- low_p - (unit_increase * min(testdat1$cumulative.rain)) #calculate intercept
unit_increase*max(testdat1$cumulative.rain) + intercept == high_p #test that the equation of the line recovers our high p at the max rainfall
```

```
## [1] TRUE
```

```
testdat1$p <- unit_increase*testdat1$cumulative.rain + intercept
plot(testdat1$p, testdat1$cumulative.rain)
```



Now the simulation will choose 1000 odds ratios between 0.5 and 7, simulate sexes of the missing nestlings, and test whether rainfall has a significant relationship with sex ratio.

```

results2 <- data.frame(odds_ratio = NA, pval = rep(NA, 1000)) #create a spot to save results
ults

for (i in 1:1000){
testdat1 <- select(sexnest, year, treatment, male, female, nestlings, missing, cumulative.rain)
low_p <- 0.3 #prob of male in driest nests
low_odds <- low_p/(1-low_p)
odds_ratio <- runif(n = 1, min = 0.5, max = 7) #pick a random odds ratio within our range
high_odds <- low_odds * odds_ratio #odds at the wet extreme
high_p <- high_odds/(1+high_odds) #probability at the wet extreme

unit_increase <- (high_p - low_p) / (max(testdat1$cumulative.rain) - min(testdat1$cumulative.rain)) #calculate the slope of the relationship
intercept <- low_p - (unit_increase * min(testdat1$cumulative.rain)) #calculate intercept
unit_increase*max(testdat1$cumulative.rain) + intercept == high_p #test that the equation of the line recovers our high p at the max rainfall

testdat1$p <- unit_increase*testdat1$cumulative.rain + intercept

# now fill in sex of missing nestlings for each nest where the probability that they are male = p

n.male <- round(rbinom(nrow(testdat1), testdat1$missing, testdat1$p)) #calculate number of missing male nestlings
n.female <- testdat1$missing - n.male
testdat1$total.m <- testdat1$male + n.male
testdat1$total.f <- testdat1$female + n.female
results2$pval[i] <- summary(glm(cbind(total.m,total.f) ~ cumulative.rain, data = testdat1, "binomial"))$coefficients[2,4]
results2$odds_ratio[i] <- odds_ratio
}

```

A plot of the results:

```
filter(results2, pval < 0.05) %>% nrow() #How many simulations resulted in significant p values?
```

```
## [1] 173
```

```
filter(results2, odds_ratio > 5, pval < 0.05) %>% nrow() /filter(results2, odds_ratio > 5) %>% nrow() #Proportion of significant simulations when OR > 5:
```

```
## [1] 0.3961039
```

About 20% of our total simulations resulted in significant values. Nearly all these cases were when there were extreme sex biased mortality dependent on rainfall (OR > 5). However, even in these extreme examples, most of the time there was not a significant effect of rainfall on overall brood sex ratio. This result matches the result we get without including missing data.

Do we have any evidence that there was biased mortality of male or female nestlings in different rainfall conditions?

```
summary(glm(fledged ~ sex * cumulative.rain, data = sex, family= "binomial"))
```

```

##
## Call:
## glm(formula = fledged ~ sex * cumulative.rain, family = "binomial",
##      data = sex)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1519  -1.0746   0.4654   0.9927   1.3341
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.760599    0.498490  -3.532 0.000413 ***
## sexM           0.338401    0.707680   0.478 0.632519
## cumulative.rain  0.020796    0.004705   4.420 9.87e-06 ***
## sexM:cumulative.rain -0.001948    0.006605  -0.295 0.768064
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 366.53  on 277  degrees of freedom
## Residual deviance: 319.50  on 274  degrees of freedom
## (68 observations deleted due to missingness)
## AIC: 327.5
##
## Number of Fisher Scoring iterations: 4

```

Of the nestlings that we did sex, they were more likely to fledge in wetter conditions. However, there was no significant interaction between rainfall and sex; i.e. rainfall levels were not associated with mortality of one sex more than the other.

## Conclusions

In conclusion, assuming extreme sex-based differences in mortality dependent on rainfall, our negative result of the effects of rainfall on sex ratio could have been a Type II error. However, this scenario depends on 1) Extremely strong sex-dependent effects of rainfall on early nestling mortality that subsequently disappear. 2) Chance, because even with that sex-biased mortality most simulations still do not find an effect of rainfall on overall brood sex ratio.

Thus we believe that these missing data are unlikely to have affected our conclusions.

## Power analysis

A separate issue is whether we have the power to detect a significant relationship even if it exists. To do a power analysis we will use the same framework. We will create a range of odds ratios that represents possible effect sizes of rain on sex ratios. This time, instead of just simulating missing nestlings we will simulate all nestlings, and then run the GLM to test how often we get a significant result when a true difference does exist given our sample size.

```

results3 <- data.frame(odds_ratio = c(runif(n = 1000, min = 0.5, max = 7)),
                      pval = rep(NA, 1000)) #create a spot to save results

for (i in 1:1000){
testdat1 <- select(sexnest, year, treatment, male, female, nestlings, missing, cumulative.rain)
low_p <- 0.3 #prob of male in driest nests
low_odds <- low_p/(1-low_p)
odds_ratio <- results3$odds_ratio[i]
high_odds <- low_odds * odds_ratio #odds at the wet extreme
high_p <- high_odds/(1+high_odds) #probability at the wet extreme

unit_increase <- (high_p - low_p) / (max(testdat1$cumulative.rain) - min(testdat1$cumulative.rain)) #calculate the slope of the relationship
intercept <- low_p - (unit_increase * min(testdat1$cumulative.rain)) #calculate intercept
unit_increase*max(testdat1$cumulative.rain) + intercept == high_p #test that the equation of the line recovers our high p at the max rainfall

testdat1$p <- unit_increase*testdat1$cumulative.rain + intercept

# now fill in sex of missing nestlings for each nest where the probability that they
# are male = p

n.male <- rbinom(nrow(testdat1), testdat1$nestlings, testdat1$p) #simulate number of male nestlings, keeping brood size the same
n.female <- testdat1$missing - n.male
testdat1$total.m <- rbinom(nrow(testdat1), testdat1$nestlings, testdat1$p)
testdat1$total.f <- testdat1$nestlings - testdat1$total.m
results3$pval[i] <- summary(glm(cbind(total.m,total.f) ~ cumulative.rain, data = testdat1, "binomial"))$coefficients[2,4]
results3$odds_ratio[i] <- odds_ratio

}

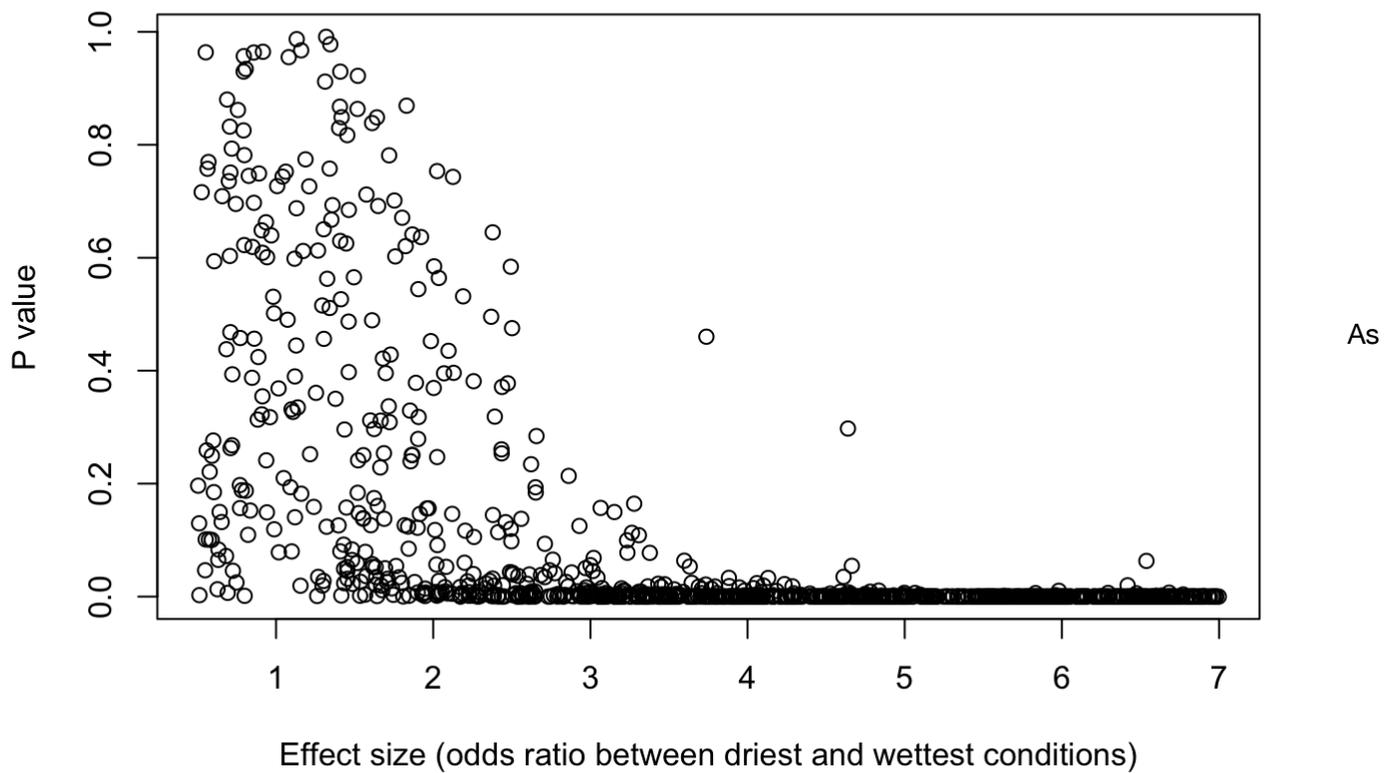
```

Plot the results

```

plot(pval ~ odds_ratio, results3, xlab = "Effect size (odds ratio between driest and wet test conditions)",
     ylab = "P value")

```



the effect size gets bigger, the smaller the p values get.

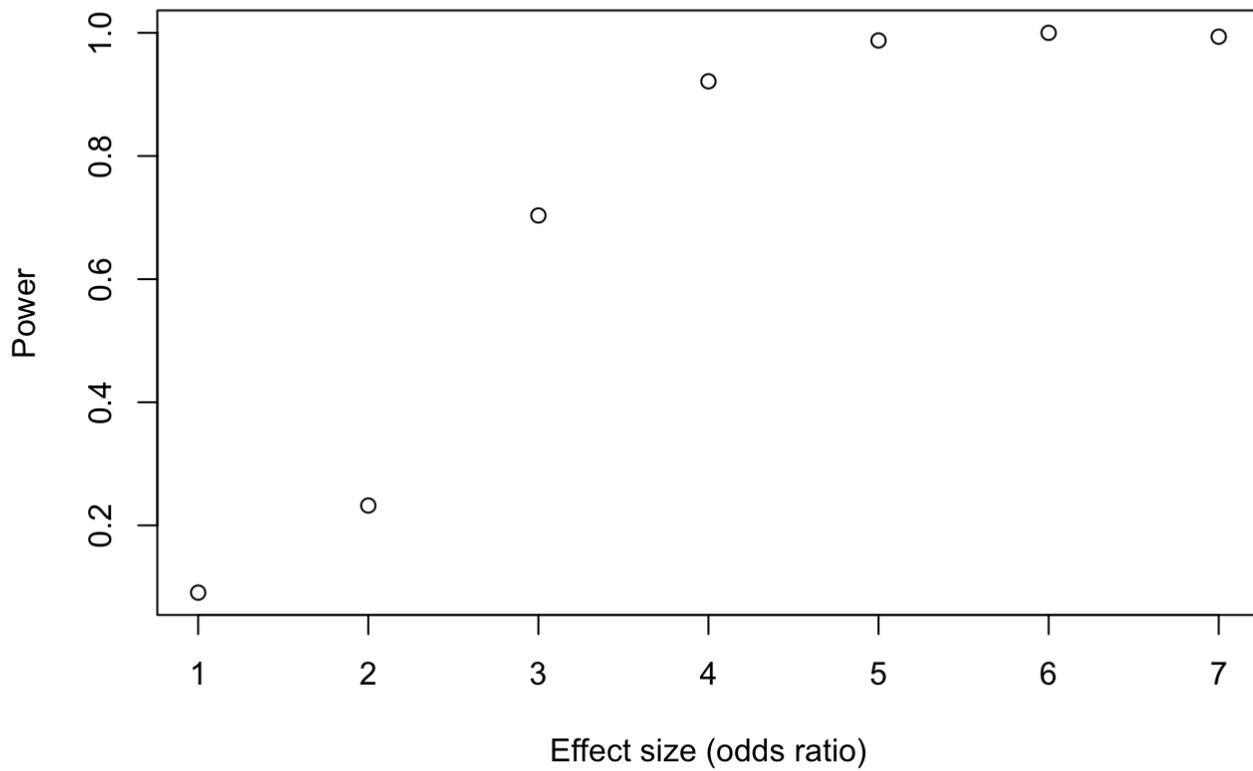
## Achieved power:

```

results3$bin <- cut(results3$odds_ratio, breaks = 0:7, labels = 1:7)

sims <- table(results3$bin) # number of simulations in each category
sig <- filter(results3, pval < 0.05)
pow <- table(sig$bin) / sims
plot (1:7, as.vector(pow), xlab = "Effect size (odds ratio)", ylab = "Power")

```



These simulations demonstrate that our power to detect small effects was low. However, for larger effects that are still well within the range of published brood sex ratio differences we actually had very good power to detect a difference. Thus, we conclude that while we may have missed small effects of rainfall on mockingbird sex ratio, it is unlikely that rainfall has substantial effects on mockingbird sex ratio. These data may be useful for others planning sex ratio studies of birds and seeking to estimate needed sample sizes.