Testosterone, diversity, and group project performance

Cathy Su

9/10/2019

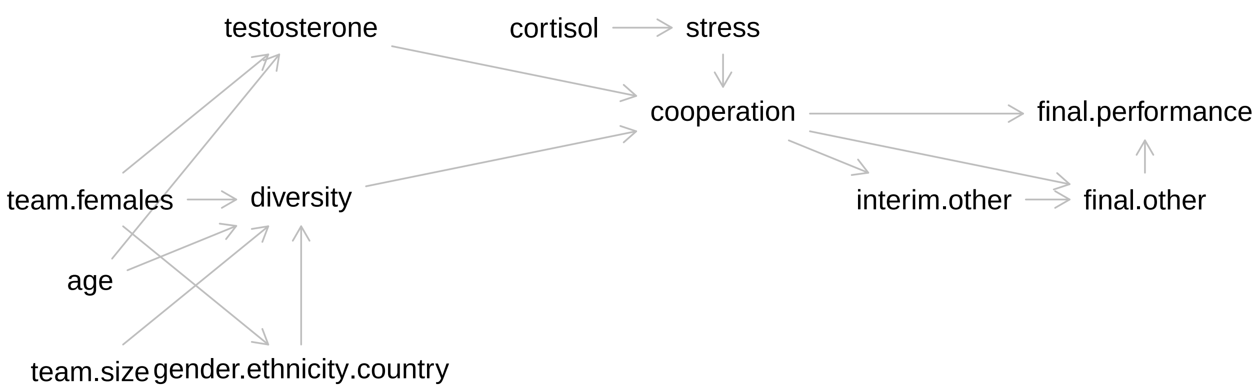
Introduction

We have both individual-level and team-level data for MBA students working on a competitive task (Akinola et al. 2018), processed at Nifty Datasets. There are 370 individuals organized into 74 teams. Our objective in the project is to further examine the relationship between testosterone and group performance, adopting the authors' hypothesis that:

> diversity is beneficial for performance, but only if group-level testosterone is low; diversity has a negative effect on performance if group-level testosterone is high.

Causal diagram

Based on the preamble from (Akinola et al. 2018) we may guess that the effects of testosterone and diversity on performance are mediated by their opposite effects on 'cooperation' (not directly measured) in the group. Furthermore cortisol levels largely unevaluated by the study may influence performance through affecting group 'stress' (not directly measured). Putting this together with the measured variables, our hypothesized causal diagram follows:



Methods

Handling missing data

Before calculating additional team level statistics, we saw there were <10 individuals with partly missing data (shown below). We preferred to not remove any individuals since we are trying to look at team level performance. For these individuals, not everything was missing so we calculated group average measurements, e.g. average hormone measurements, from other members. When we did this, there were no teams with missing data aside from in the 'interim' variables.

```
##      ID team.id Age Gender   Ethnicity Cortisol Testosterone log.cortisol
## 22  140      16  26 Female       Black       NA           NA           NA
## 45  218       2  NA   Male       Other    0.126       122.44    -2.071473
## 53  236       3  NA   Male       Other    0.086        66.54    -2.453408
## 95  346      36  28   Male       White       NA           NA           NA
## 113 420      47  30   Male South Asian       NA           NA           NA
## 180 546      97  NA   Male       Asian    0.350       132.65    -1.049822
##     log.testosterone Country
## 22                NA     USA
## 45          4.807621
## 53          4.197803
## 95                NA  Canada
## 113               NA   India
## 180         4.887714   Korea
```

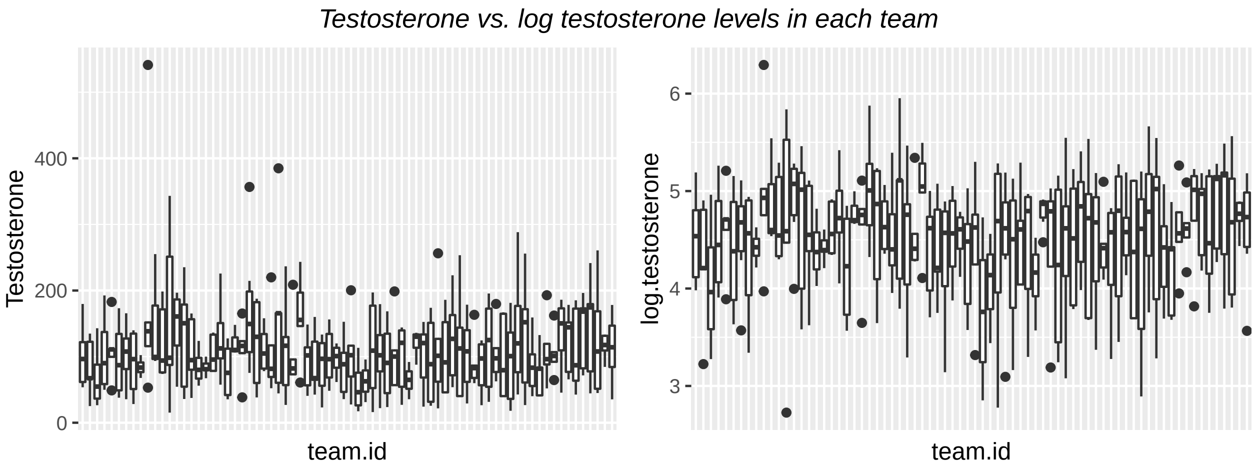
Calculation of group level variables from individual level variables

We are interested in doing our analysis at the group level therefore we needed to calculate the group level values for the diversity score, average log testosterone level, average log cortisol level, average age and average age variance. Note that unlike in the original study, we have calculated group diversity score as the number of unique gender-ethnicity-country combinations (normalized by group size). For the sake of making the units clear we show the code below. With this we can proceed to data exploration.

```
#  calculate the number of unique gender-ethnicity-country combinations
ind_dat$combo <-paste(ind_dat$Gender, ind_dat$Ethnicity, ind_dat$Country)
team_dat$score<- unlist(lapply(team_dat$team.id,
               function(x){length(unique(ind_dat$combo[ind_dat$team.id == x]))}))
team_dat$diversity.score <- team_dat$score/team_dat$team.size #suggested by Q4
team_dat$proportion.females <- team_dat$females/team_dat$team.size
#  calculate the average testosterone level for each group.
# some have missing testosterone data, this means we need to average.
team_dat$avg.log.testosterone<- unlist(lapply(team_dat$team.id,
               function(x){mean(ind_dat$log.testosterone[(ind_dat$team.id == x)], na.rm = TRUE)}))

#  calculate the average cortisol level for each group.
team_dat$avg.log.cortisol<- unlist(lapply(team_dat$team.id,
               function(x){mean(ind_dat$log.cortisol[ind_dat$team.id == x], na.rm = TRUE)}))

#  calculate the average age for each group.
team_dat$avg.age<- unlist(lapply(team_dat$team.id,
               function(x){mean(ind_dat$Age[ind_dat$team.id == x], na.rm = TRUE)}))

team_dat$age.variance<- unlist(lapply(team_dat$team.id,
               function(x){var(ind_dat$Age[ind_dat$team.id == x], na.rm = TRUE)}))
```

Log vs raw values of testosterone and cortisol

Since we are interested in how the the levels of these two hormones may be related to the other variables, we first wanted to check whether we should use their log vs raw values in our EDA.

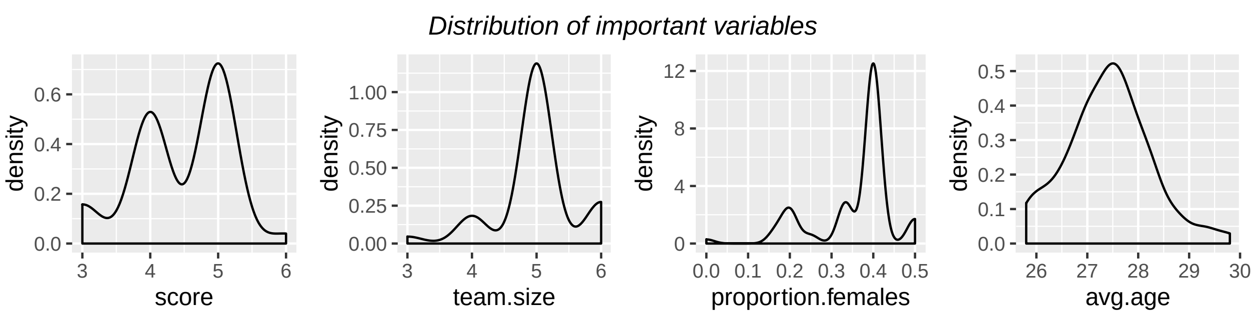


(Cortisol looks similar to the above, except for scale.) It was clear when for both hormones that the log transformed values were distributed with less skew across teams which would be our preference. The authors also chose to use averaged log testosterone per group.

Exploratory Data Analysis & Data Summary

Distributions of variables

We checked the distribution of other key variables besides hormone level in our model, and saw that in particular, our diversity score and team size appear bimodal. Although our score is differently calculated, this agrees with (Akinola et al. 2018) which classified diversity score into two bins.



Variable selection

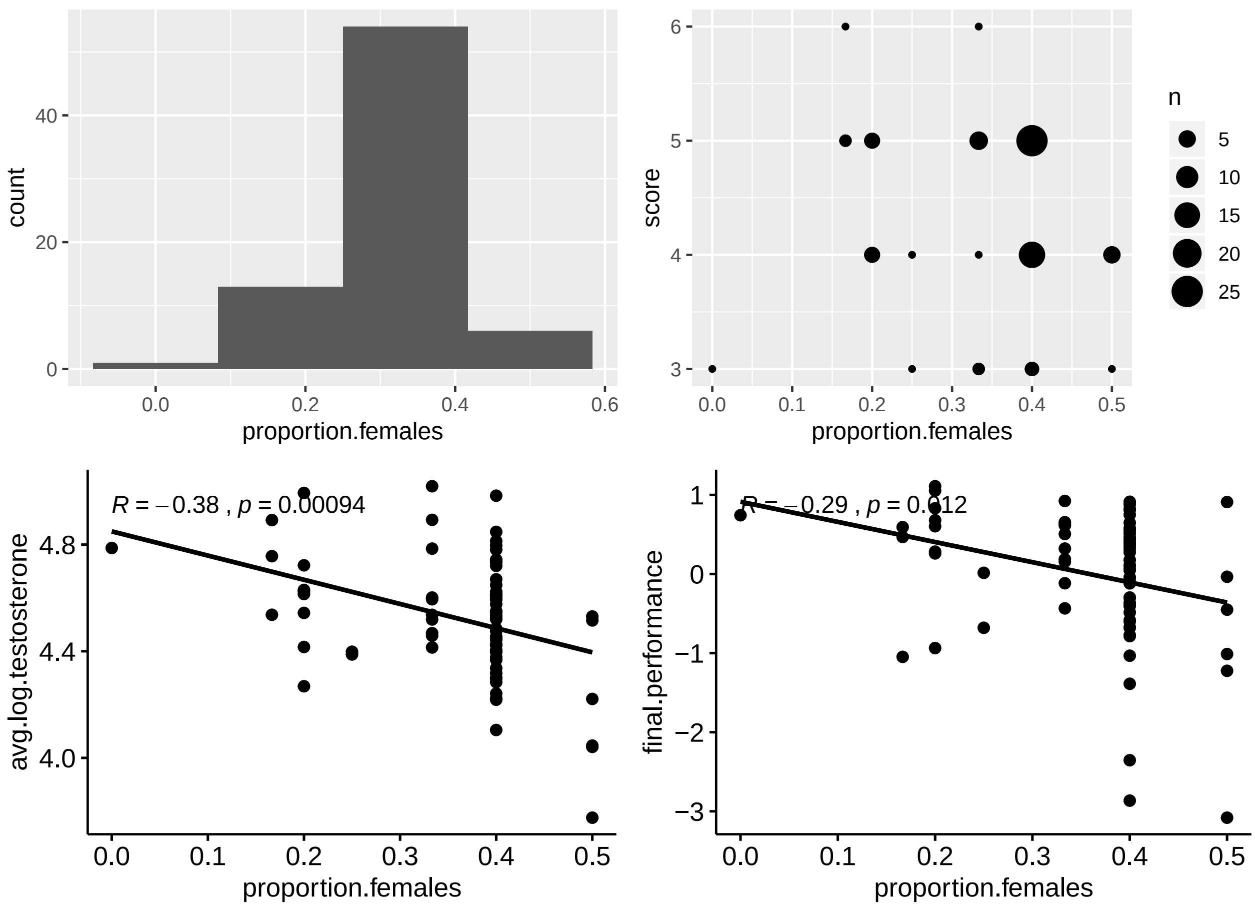
To choose what other variables to include in further analysis, we visually inspected pairwise correlations. We removed the following:

- time.of.day : didn't seem to have a straightforward relationship with other variables.
- 'interim'variables: contain missing data for many teams.
- other 'final' variables besides final.performance: these variables are generally correlated, but not in a straightforward way. The original study states that these measures were standardized and then averaged to form the final.performance score. Since the judges who assigned the final.performance are also domain experts, we have decided to discard the other variables.

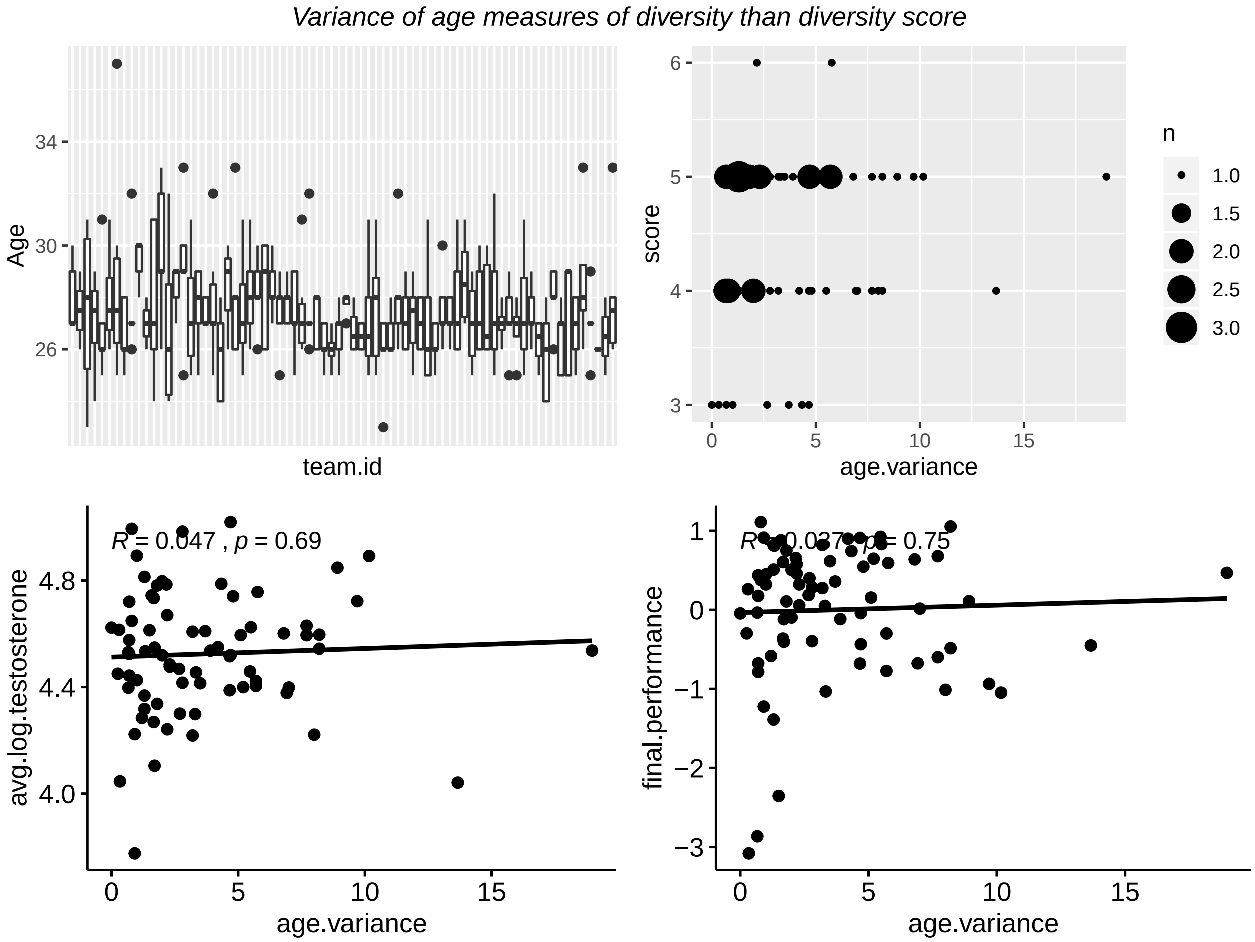
We then looked more closely at the variables 'females' and 'age' which may contribute to diversity outside of diversity score.

Gender is incorporated in the diversity score, but the variable females was measured separately. Since we are looking at average hormone scores which are known to differ between genders, we examined the proportion of females per group. As seen below, proportion of females shows little correlation with diversity score. However, there is similar negative correlation with both testosterone and team.size. Our hypothesized causal graph suggested proportion of females should influence performance through both diversity and through affecting testosterone level. In contrast the similar pearson correlation coefficients suggest that proportion of females may not play a large role in determining performance outside of its effects on testosterone.

Proportion of females is negatively correlated with both testosterone and performance



A similar situation occurs for age. When we plotted the spread of age across groups (below left), it seemed mean age is not too different between groups. Since according to our causal graph, the effects of age may influence performance through changing cooperation, we further plotted the variance of age since different ages in a group could also lead to conflict. It also doesn't correlate clearly with diversity score but does show a similar weak positive correlation with both testosterone and performance (see pearson correlation scores). This suggests that age may not play a large role in determining performance outside of its effects on testosterone.



Based on the above, we decided to keep discard both age and proportion of females as variables since they may not necessarily tell us more about the relationship between testosterone and performance. Furthermore this simplifies the model.

Results

Here we perform statistical analysis to verify our hypothesis. The results presented by the original study (Akinola et al. 2018) include that:

- when group diversity was low, group testosterone significantly positively predicted performance at p < .01
- when group diversity was relatively high, group testosterone significantly negatively predicted performance p < .01

Effect of diversity score and testosterone individually on performance

To start we want to check the simplest assumptions that 1) diversity score positively predicts performance and 2) testosterone negatively predicts performance.

```
mod_t = lm( final.performance ~avg.log.testosterone, data = team_dat)
mod_d = lm( final.performance ~score, data = team_dat)
#summary(mod_d)
#summary(mod_t)
```

For these simple models, the coefficient of diversity score (0.065) and average log testosterone (0.7032) are not particularly large in magnitude or significant (p >0.05) indicating that neither of these two predictors alone shows a clear relationship with group performance. This agrees with what the authors found.

Q4: Effect of group diversity on relationship between testosterone and performance

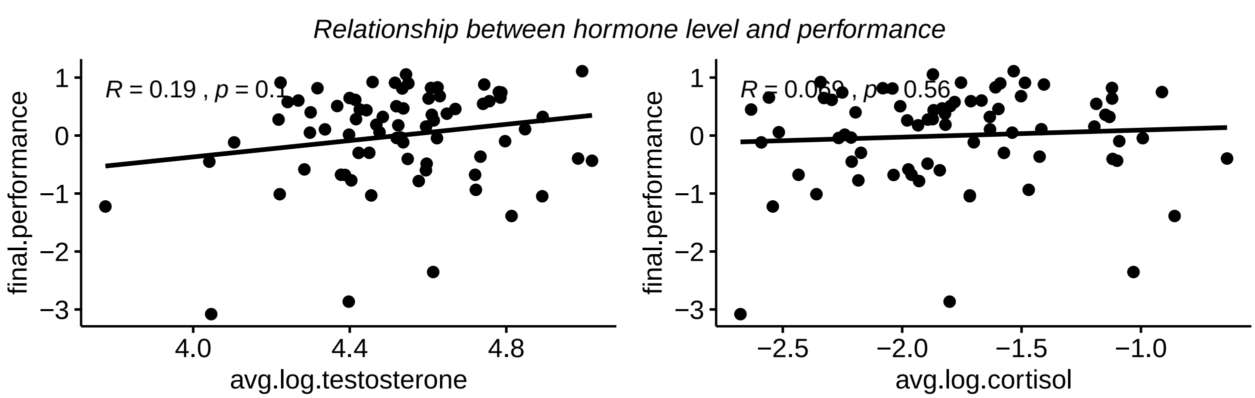
Next we fit a model to examine whether the interaction between them could predict performance.

```
##
## Call:
## lm(formula = final.performance ~ avg.log.testosterone + score +
##     team.size + avg.log.testosterone:score, data = team_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7440 -0.4649  0.1188  0.5049  1.2373
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -48.1116    10.7961  -4.456 3.14e-05 ***
## avg.log.testosterone        10.2743     2.3466   4.378 4.16e-05 ***
## score                       10.1889     2.4472   4.163 8.91e-05 ***
## team.size                    0.3533     0.1729   2.043   0.0448 *
## avg.log.testosterone:score  -2.2577     0.5392  -4.187 8.20e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7483 on 69 degrees of freedom
## Multiple R-squared:  0.2473, Adjusted R-squared:  0.2037
## F-statistic: 5.667 on 4 and 69 DF,  p-value: 0.0005281
```

We found a significant positive effect of testosterone on performance when controlling for diversity score and team size (coefficient = 10.2743, $p < 0.001$). Further we find a significant positive effect of diversity score on performance when controlling for testosterone, team size and their interaction (coefficient = 10.1889, $p < 0.001$). The interaction term has a negative coefficient. This suggests that whereas each of testosterone and diversity aids performance, their interaction works against these effects. Our results are in line with those of the original study.

Q5: Effect of cortisol on relationship between diversity and performance

We then examined whether stressed groups have better or worse performance by looking at pearson correlation. In the diagnostic plot, the correlation of cortisol with final.performance seems weaker than the correlation of testosterone with final.performance as shown below. However the r squared value shows that there is still some linearity in this relationship.



When we fit the very simplest model of final.performance ~ avg.log.cortisol, we find a positive (0.1217) but not significant (p-value 0.56) coefficient as our scatterplots above may suggest.

Model with interaction of cortisol and diversity score

Next, we tested whether cortisol levels could change the relationship between diversity score and performance with a model containing each of these variables and their three way interaction.

```
##
## Call:
## lm(formula = final.performance ~ avg.log.cortisol + score + team.size +
##     avg.log.cortisol:score, data = team_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8542 -0.4321  0.2252  0.5555  1.5956
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)              5.1777     2.2746   2.276  0.02593 *
## avg.log.cortisol         3.3481     1.1353   2.949  0.00435 **
## score                   -1.3627     0.4747  -2.870  0.00544 **
## team.size                0.1803     0.1826   0.987  0.32692
## avg.log.cortisol:score  -0.7483     0.2549  -2.935  0.00452 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7997 on 69 degrees of freedom
## Multiple R-squared:  0.1403, Adjusted R-squared:  0.09049
## F-statistic: 2.816 on 4 and 69 DF,  p-value: 0.0317
```

Here we found that stress seems to positively impact performance (coefficient of 3.348 units, $p < 0.01$) when controlling for diversity score, team size and the interaction between cortisol and diversity score. However, the diversity score is estimated here to have a negative effect on performance (coefficient of -6.8455 units, $p < 0.05$). Furthermore the interaction term also has a weak negative effect (coefficient of -0.7483 units, $p < 0.05$).

This suggests stressed groups have better performance and stress changes the effect of diversity to negatively impact performance.

Incorporating both hormone measurements

This suggests cortisol should also be in our model, so we lastly fit a full model with both hormone measurements and their three way interaction with diversity.

```
##
## Call:
## lm(formula = final.performance ~ avg.log.cortisol + avg.log.testosterone +
##     score + team.size + avg.log.cortisol:score:avg.log.testosterone,
##     data = team_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8533 -0.4548  0.1692  0.5490  1.2781
##
## Coefficients:
##                                             Estimate Std. Error t value
## (Intercept)                                  8.11591    5.37545   1.510
```

```
## avg.log.cortisol                              2.95961    1.18450   2.499
## avg.log.testosterone                          -0.72231    0.74998  -0.963
## score                                         -1.23930    0.46886  -2.643
## team.size                                      0.12317    0.18626   0.661
## avg.log.cortisol:avg.log.testosterone:score -0.14901    0.05573  -2.674
##                                               Pr(>|t|)
## (Intercept)                                    0.13572
## avg.log.cortisol                               0.01489 *
## avg.log.testosterone                           0.33891
## score                                          0.01019 *
## team.size                                      0.51066
## avg.log.cortisol:avg.log.testosterone:score  0.00939 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8013 on 68 degrees of freedom
## Multiple R-squared:  0.1494, Adjusted R-squared:  0.08681
## F-statistic: 2.388 on 5 and 68 DF,  p-value: 0.04687
```

Here the coefficients on cortisol is positive (2.95) but the coefficient on diversity score is negative (-1.23) at $p < 0.05$, which is what we would expect. However the coefficient on testosterone is not significantly nonzero contrary to our hypothesis that group testosterone is important. Hoever the model could be a bit too complex since it has a lower R squared value than the model without cortisol.

Caveats of the analysis

Although we had some similar results to teh original study, our results may not be directly comparable because of the following:

- their model includes many more terms such as proportion of females which we chose to discard.
- We built our hypothesis according to the causal graph, but we do not have sufficient domain knowledge so some of our assumptions may be wrong (e.g. does proportion of females only affect testosterone and diversity, as we assumed in order to discard this variable in EDA?)
- our diversity score was calculated differently than their faultline analysis.

Furthermore, we have shown that stress has a significant effect as measured by cortisol levels which was not discussed in depth in their original analysis.

Bibliography

Akinola, Modupe, Elizabeth Page-Gould, Pranjal H. Mehta, and Zaijia Liu. 2018. "Hormone-Diversity Fit: Collective Testosterone Moderates the Effect of Diversity on Group Performance." *Psychological Science* 29 (6):859–67. https://doi.org/10.1177/0956797617744282.