Testosterone, diversity, and group project performance

Cathy Su

9/10/2019

Executive Summary

In this report, we analyzed a demographic data set collected by *(Akinola et al. 2018)* and explore the relationship between a set of variables that contribute to group performance on a competetive task. The data comprises individual level and group level statistics collected from groups of MBA students completing a 7-day group project. We use exploratory data analysis and regression models to mainly explore how **diversity, cortisol and testosterone** levels affect **final.performance**. The f-test shows that diversity score, cortisol and testosterone individually do not significantly affect final.performance. Then, we fit several linear regression models and found that the model with the highest adjusted R squared value predicts performance as a function of average log testosterone, diversity and group size. In this model, if we hold group size constant, indeed diversity has a positive effect on performance, but only if group-level testosterone is low.

Introduction

Diversity and conflict are considered important factors which influence how well we work in groups (Knippenberg and Schippers 2007). As the working world becomes more connected across the globe and thus the diversity of organizational groups increases, it is important to characterize the effect of diversity on group performance. Previous work by (Akinola et al. 2018) suggests that both diversity and group hormone levels will influence how well groups perform on a competetive task. In their study, they considered levels of the two hormones testosterone and cortisol. Testosterone is involved in dominance and competition related behaviour in individuals and is produced at a higher level in males than females, while cortisol is a hormone released during physical and psychological stress (Mehta and Prasad 2015).

In their work, (Akinola et al. 2018) collected both demographic data and hormone measurements from 370 MBA students organized into 74 groups who partcipated in a competitive week long project where their goal was to outperform other groups. There were 370 individuals randomly organized into 74 groups. Based on their demographic and hormone measurement data, the authors concluded that diversity is beneficial for performance, but only if group-level testosterone is low; and diversity has a negative effect on performance if group-level testosterone is high. However, the authors did not mention analyzing cortisol even though cortisol levels is suggested to have an effect testosterone's role in status-relevant behavior (Mehta and Prasad 2015).

To validate the author's hypothesis and additionally examine the specific role of cortisol, we have obtained the (Akinola et al. 2018) dataset which has been processed by Nifty Datasets into separate individual level and group level datasets. Here we test the interactions between the hormone profiles of both cortisol and testosterone by modelling their effect on performance in the context of the demographic variables collected and the group diversity.

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 Figure 1. Here 'interim.other' describes other interim measurements of group performance which were in the dataset and 'final.other' describes the measurements of group performance at the conclusion of the task which contribute to the **final.performance** score. This diagram helps set the context for reasoning about which regression models we should try.

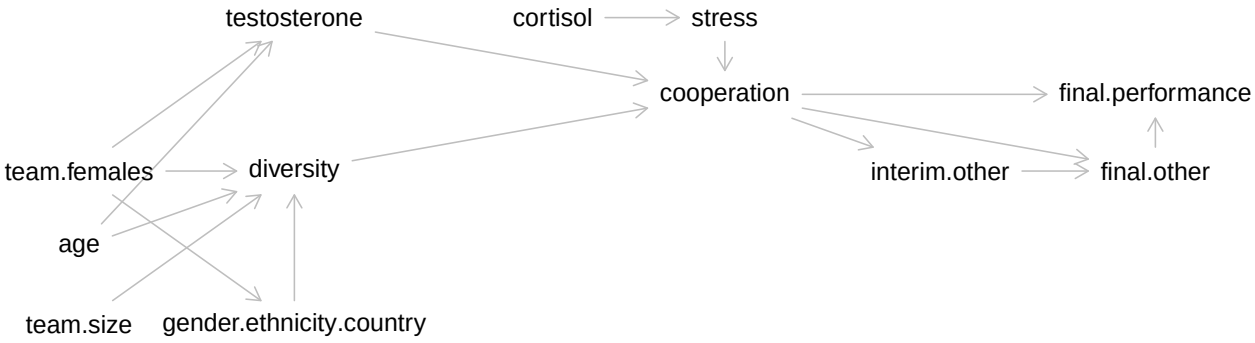

Figure 1: Causal diagram illustrates hypothesized relationships of experimental variables involved in relationship between testosterone and final group performance.

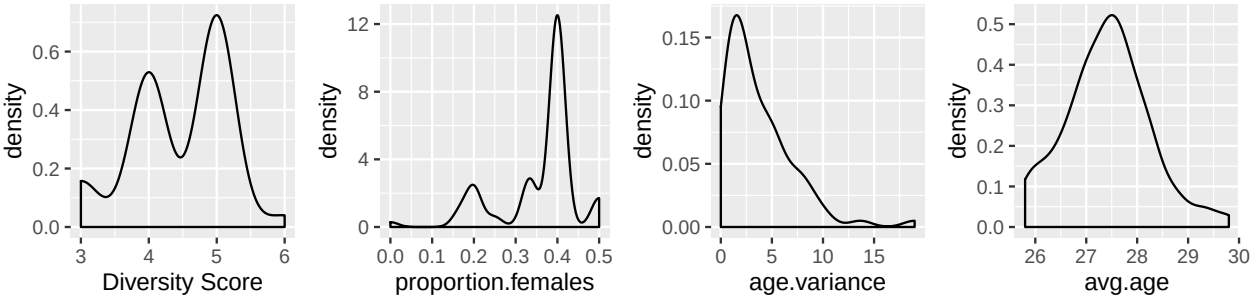

Figure 2: Univariate distribution of important group level variables in the analysis.

Methods

Handling missing data

Before calculating additional team level statistics, we saw there were <10 individuals with partly missing data. Since we are trying to look at team level performance, we did not remove any individuals. For these individuals, not everything was missing so we calculated group average measurements, e.g. average hormone measurements, from other members.

From this we obtain a complete group level dataset where only measurements in the 'interim' variables are missing. Given that it's unclear how the multiple interim measurements may relate to the final score and they contain many missing values, we removed these variables.

Calculation of other 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 e.g. the average age and average age variance. For each of these we have averaged the corresponding individual-level statistics, ignoring missing cases.

Additionally, we have calculated group diversity score as the number of unique gender-ethnicity-country combinations (normalized by group size). Lastly we calculate proportion of females in the group as the number of females divided by group size.

The univariate distributions of these new group level variables is given in Figure 2. We saw that in particular, our diversity score appears bimodal. Although our score is calculated differently, (Akinola et al. 2018) classified diversity score into two bins in their faultline analysis. This suggests that our diversity score is reasonable since it may also reflect some intrinsic bimodality present in the data.

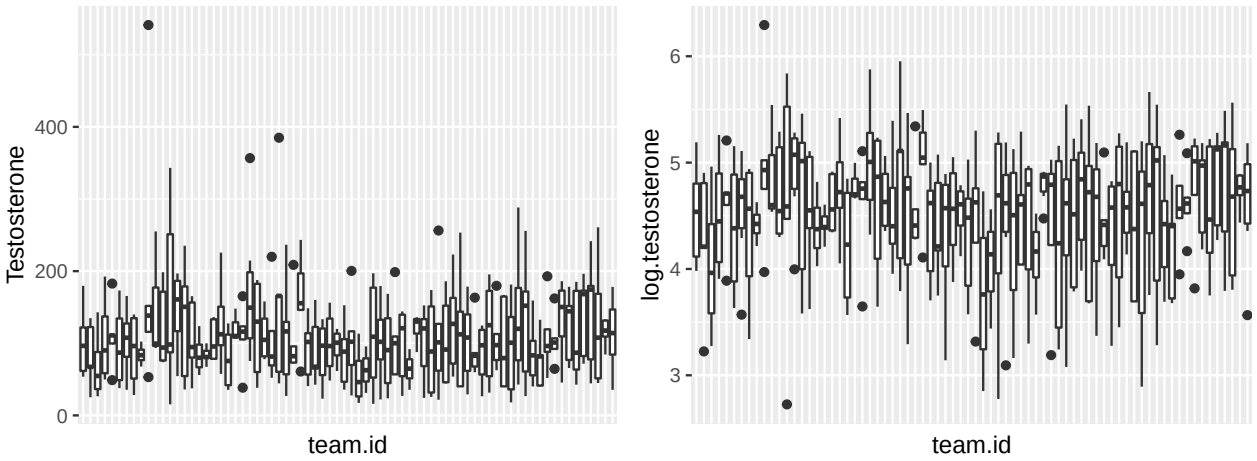

Figure 3: Distributions of testosterone and log testosterone levels in each team

Exploratory Data Analysis & Data Summary

Log vs raw values of testosterone and cortisol

Since we are interested in the levels of these two hormones, we wanted to check whether we should use their log or raw values in our EDA. It was clear when for both hormones that the log transformed values were distributed with less skew across teams. This is preferable so we chose like the authors to use averaged log testosterone per group. Figure 3 shows the distribtuions for testosterone but for cortisol the difference is similar.

Pairwise distributions of variables

We visualized the correlation matrix including the Pearson correlation coefficients (upper right half) between important variables in Figure 4. Right away we can make the following observations about the key variables:

- performance appears correlated with proportion of females and testosterone.
- testosterone appears correlated with cortisol, average age, proportion of females, time of day, performance and team size.
- diversity score appears correlated with team size.

Based on this we knew that in addition to final.performance, avg.log.testosterone, avg.log.cortisol and diversity.score we should consider whether to incorporate the four additional variables proportion.female, avg.age/age.variance, time.of.day, and team.size in our models.

Variable selection

Based on the causal graph we had made, we reasoned that out of the four additional variables we had chosen to investigate we could expect 'proportion.female' and 'avg.age' to contribute to performance through their biological effect on the diversity score and on testosterone levels. However, Figure4 shows that for both of these variables, their correlation with diversity is not significant. Since we already had hormone levels in the model, we wanted to investigate whether they would still be useful to include in a model of the final group performance.

Gender is incorporated in the diversity score, but the variable females was measured separately. As seen in Figure5, proportion.females shows similar weakly linear correlation (r squared < 0.2) with both testosterone and team.size. This in combination with the insignificant pearson correlation coefficients suggest that proportion.females may not be useful in determining performance outside of its effects on testosterone.

For age, we have two variables: the variance of ages (age.variance) and the average age (avg.age) in the group. As seen in Figure6 both however show very weak linear relationship with both testosterone and performance

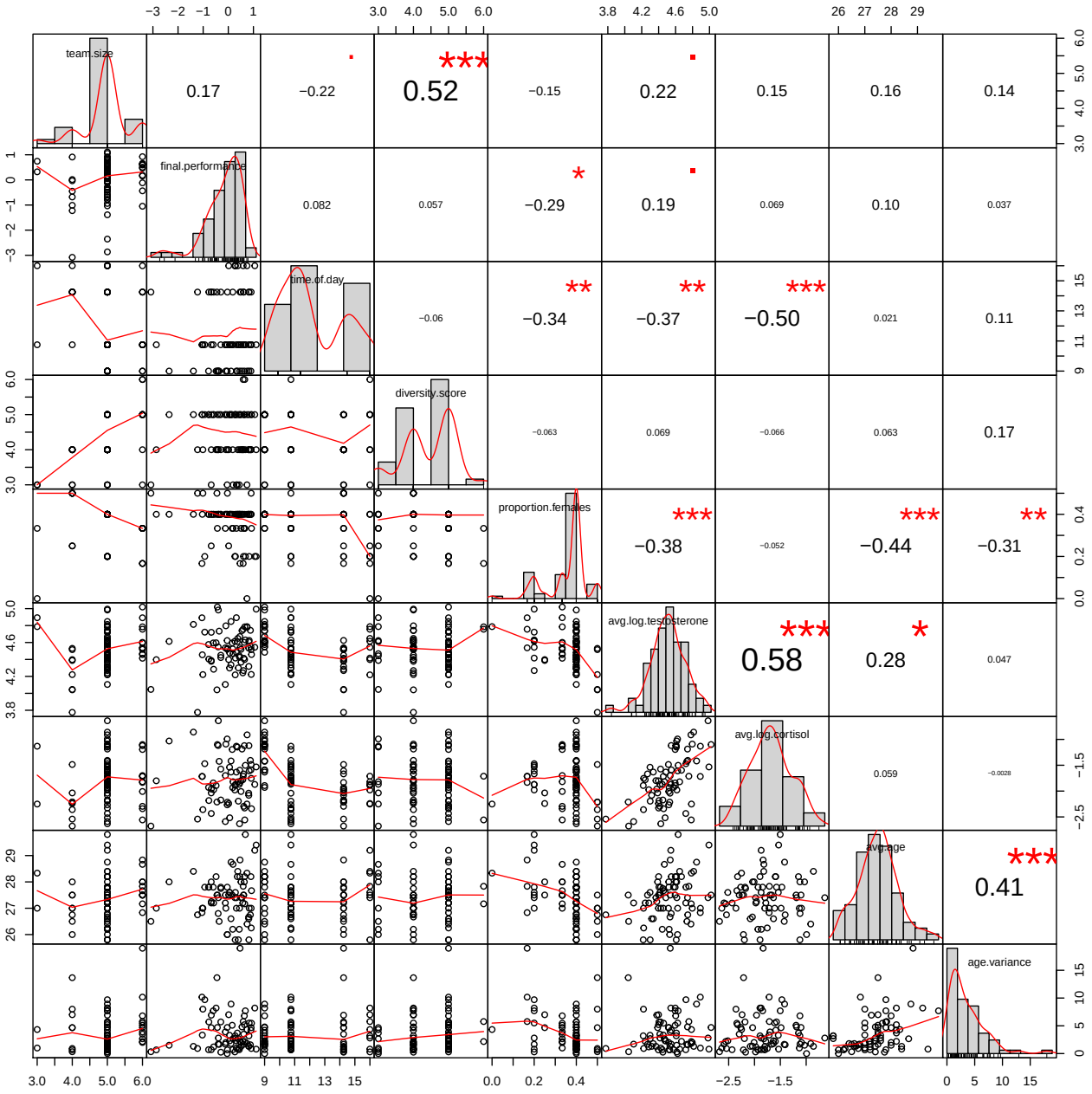

Figure 4: Pairwise correlations of important variables including their Pearson correlation coefficient. Significant correlations are marked by the corresponding number of astericks.

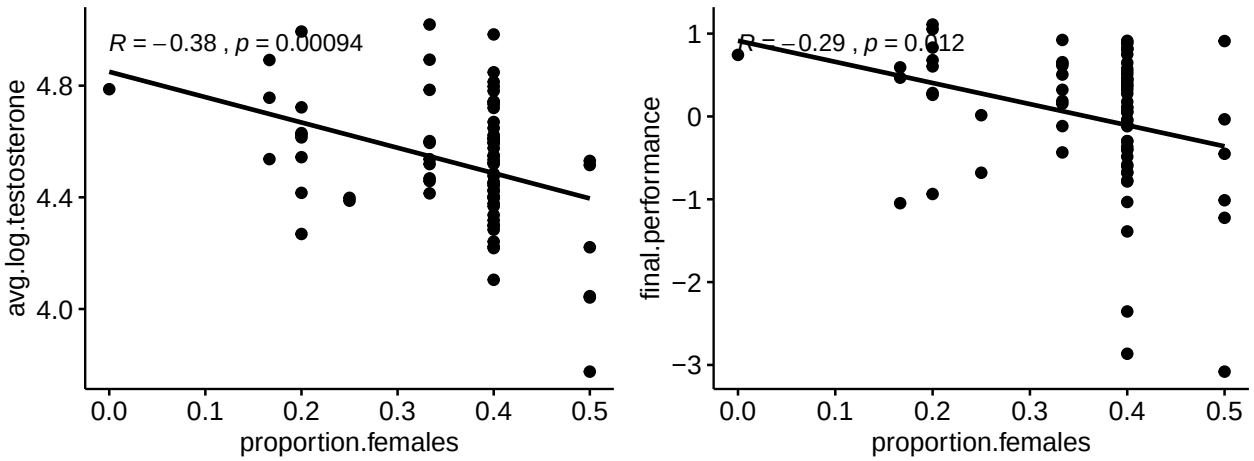

Figure 5: Proportion of females is weakly negatively correlated with both testosterone and performance.

(r squared < 0.1 and < 0.01) respectively.

Based on the above reasoning, 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 diversity, testosterone and performance. Furthermore this simplifies the model.

Results

The results discussed by the original study (Akinola et al. 2018) include that:

- considered in isolation, group diversity and testosterone are not significantly correlated with performance.
- 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 $p < .01$

Diversity score and testosterone do not individually significantly predict performance

To start we want to check the simplest assumptions from (Akinola et al. 2018) that diversity score and testosterone do not significantly predict performance on their own. We use the F-test to compare this null hypothesis to the alternative hypothesis where including these variables significantly improves model fit to the data.

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

```
##
## Call:
## lm(formula = final.performance ~ diversity.score, data = team_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9857 -0.4162  0.2151  0.5781  1.1392
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.29196    0.61360  -0.476    0.636
```

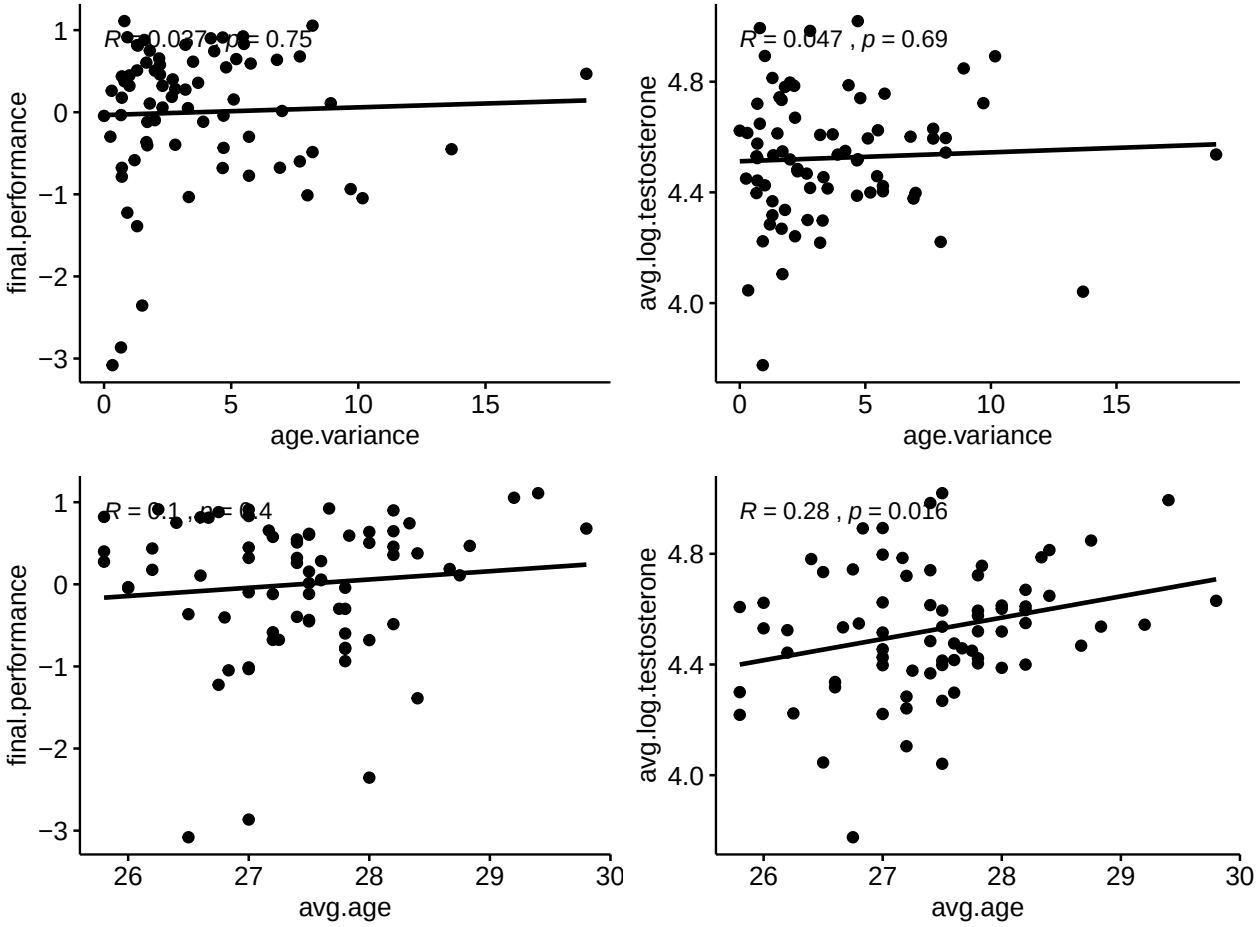

Figure 6: Variance of age and average age of group members show very weak correlation with final performance and testosterone levels.

```
## diversity.score  0.06567    0.13624   0.482    0.631
##
## Residual standard error: 0.843 on 72 degrees of freedom
## Multiple R-squared:  0.003216,   Adjusted R-squared:  -0.01063
## F-statistic: 0.2323 on 1 and 72 DF,  p-value: 0.6313
```

```
summary(mod_t)
```

```
##
## Call:
## lm(formula = final.performance ~ avg.log.testosterone, data = team_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7757 -0.4887  0.2030  0.5663  1.1249
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -3.1813     1.9095  -1.666   0.1001
## avg.log.testosterone   0.7032     0.4215   1.668   0.0996 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8285 on 72 degrees of freedom
## Multiple R-squared:  0.03721,    Adjusted R-squared:  0.02384
## F-statistic: 2.783 on 1 and 72 DF,  p-value: 0.09963
```

For these simple models, the coefficient of diversity score (0.065) and average log testosterone (0.7032) are not large in magnitude or significant (p >0.05) indicating that we do not reject the null hypothesis. This agrees with what the authors found.

Group diversity on relationship between testosterone and performance

Next we fit a model to examine whether including an interaction between them could predict performance. We have controlled for team.size by including it in the model. It has the Residual standard error of 0.75 on 69 degrees of freedom.

```
##
## Call:
## lm(formula = final.performance ~ avg.log.testosterone + diversity.score +
##     team.size + avg.log.testosterone:diversity.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
## diversity.score                     10.1889     2.4472   4.163 8.91e-05
## team.size                            0.3533     0.1729   2.043   0.0448
## avg.log.testosterone:diversity.score  -2.2577     0.5392  -4.187 8.20e-05
##
## (Intercept)                        ***
## avg.log.testosterone               ***
```

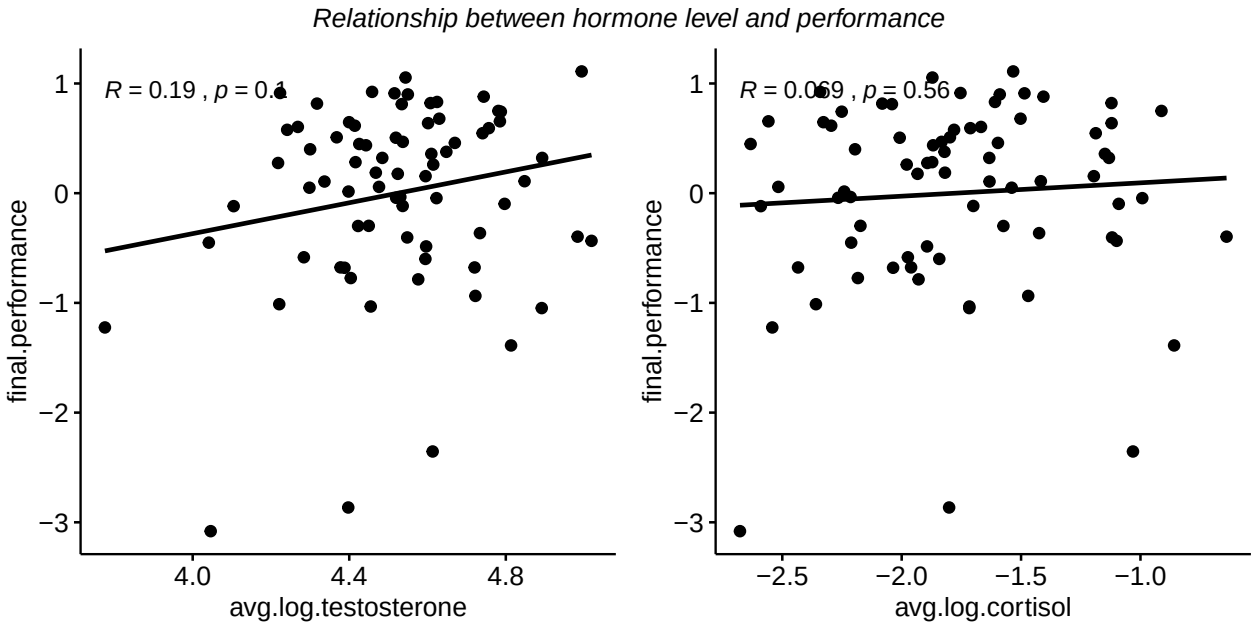

Figure 7: Levels of both testosterone and cortisol correlate with performance

```
## diversity.score                      ***
## team.size                            *
## avg.log.testosterone:diversity.score ***
## ---
## 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 replot the correlation of cortisol with final.performance in Figure 7 which seems weakly linear (adjusted r squared value < 0.2).

Accordingly, 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. Again we are controlling for team size by including it as a term in the model.

```
##
## Call:
```

```
## lm(formula = final.performance ~ avg.log.cortisol + diversity.score +
##     team.size + avg.log.cortisol:diversity.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 **
## diversity.score                  -1.3627     0.4747  -2.870  0.00544 **
## team.size                         0.1803     0.1826   0.987  0.32692
## avg.log.cortisol:diversity.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
```

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

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 +
##     diversity.score + team.size + avg.log.cortisol:diversity.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
## (Intercept)                                            8.11591    5.37545
## avg.log.cortisol                                       2.95961    1.18450
## avg.log.testosterone                                  -0.72231    0.74998
## diversity.score                                       -1.23930    0.46886
## team.size                                              0.12317    0.18626
## avg.log.cortisol:avg.log.testosterone:diversity.score -0.14901    0.05573
##                                                       t value Pr(>|t|)
## (Intercept)                                             1.510  0.13572
## avg.log.cortisol                                        2.499  0.01489 *
## avg.log.testosterone                                   -0.963  0.33891
## diversity.score                                        -2.643  0.01019 *
```

```
## team.size                                                0.661  0.51066
## avg.log.cortisol:avg.log.testosterone:diversity.score  -2.674  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.

Conclusion

Here we have analyzed demographic data and hormone measurements from groups of MBA students performing a competetive project, previously published by (Akinola et al. 2018). We sought to investigate the authors' hypothesis that group diversity has a testosterone-dependent effect on group performance and also to check whether cortisol levels had an effect on this relationship.

By building linear models of performance and testing the significance of the terms with an F-test, we have shown that although testosterone and diversity score alone do not predict performance, when they are both included in the model interaction between diversity and testosterone has a significant negative effect on performance ($p < 0.01$) implying that high diversity and high testosterone are antagonizing factors. Although stressed groups did not have significantly different performance, we also found that when controlling for diversity cortisol has similar effects. The interaction between cortisol and diversity also has a significant negative effect on performance ($p < 0.05$) implying that higher diversity and higher cortisol counteract each other. When looking at both hormone measurements simultaneously with diversity score, surprisingly we found that when accounting for cortisol, testosterone levels do not seem to have a significant effect on performance. Rather only the interaction of cortisol and testosterone together has a slight negative effect on performance ($p < 0.01$). However, the model we tested containing both hormones has a lower adjusted R squared than the model containing just testosterone. Overall, we do find that diversity is beneficial for performance, in the presence of low group-level testosterone. Additionally this analysis suggests that perhaps, stress has a role in group performance as well.

Although we had some similar findings to the original study when examining diversity and testosterone, our results may not be directly comparable because of some differences in our methodology. Most prominently, (Akinola et al. 2018) have used a faultline analysis to evaluate diversity whereas we have constructed a diversity score. As well, we have not included some of the variables that are present in the models which they tested e.g. proportion of females. We chose to discard these variables based upon our EDA and our reasoning about the relationship between variables collected in the study. Lastly we cannot compare our findings about cortisol because this 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.

Knippenberg, Daan van, and Michaéla C. Schippers. 2007. "Work Group Diversity." *Annual Review of Psychology* 58 (1):515–41. https://doi.org/10.1146/annurev.psych.58.110405.085546.

Mehta, Pranjal H, and Smrithi Prasad. 2015. "The Dual-Hormone Hypothesis: A Brief Review and Future Research Agenda." *Current Opinion in Behavioral Sciences* 3:163–68. https://doi.org/https://doi.org/10.1016/j.cobeha.2015.04.008.