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). For healthy males between 19 to 40 years, normal testosterone is known to fall within the 15.4 to 13 nmol/L range (Kelsey et al. 2014). Healthy levels of hormones for men and women are given in Table, collected from (Matsuzaka et al. 2013)

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

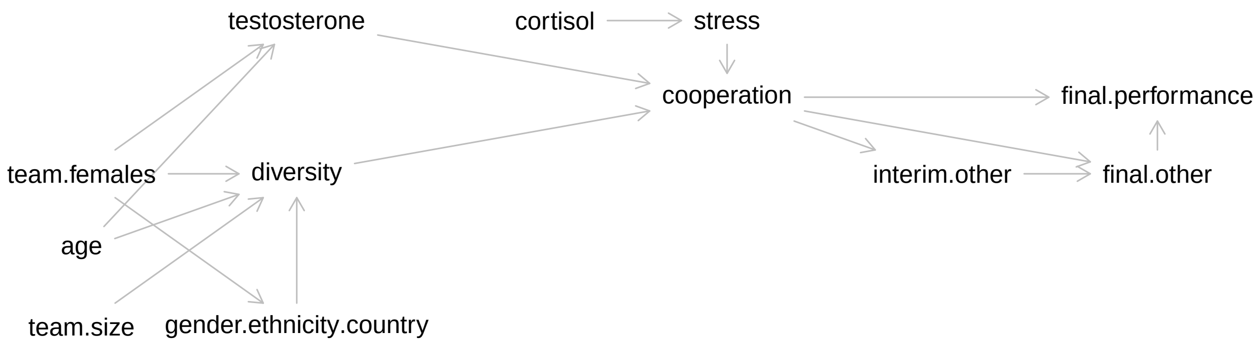


Figure 1: Causal diagram illustrates hypothesized relationships of experimental variables involved in relationship between testosterone and final group performance.

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.

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 aggregate the individual level data. To calculate group level testosterone, cortisol and age we first 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 present in the group. Lastly we calculate proportion of females in the group as the number of females divided by group size.

Model fitting and model selection

We fit linear regression models and plot their fit using the base lm() function in R. For model selection best subsets regression was performed using and model comparison using anova(). To avoid overfitting we used cross validation to check the model performance.

Exploratory Data Analysis & Data Summary

Distribution of hormone levels across individuals and groups

It was clear when for both hormone levels that the log transformed values were distributed with less skew across teams than the raw values and have fewer outlier values. This is preferable so we chose like the authors to use averaged log testosterone per group. Figure 2 shows the distribtuions for testosterone but for cortisol the difference is similar.

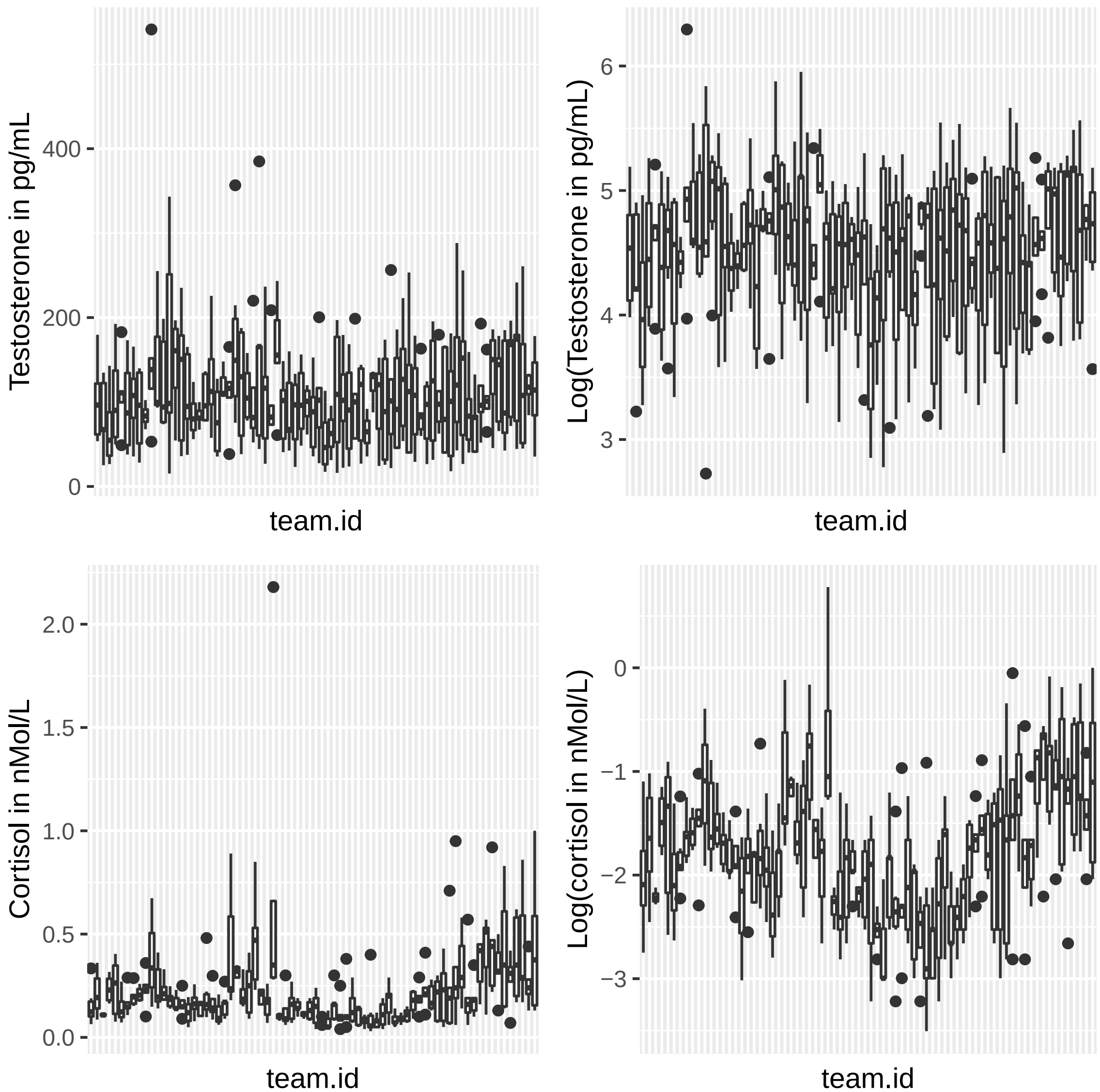


Figure 2: Distributions of testosterone and log testosterone levels in each team

Univariate and pairwise distributions of group level variables

The univariate distributions of the group level variables is given across the diagonal in Figure 3. 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.

We visualized the correlation matrix including the Pearson correlation coefficients (upper right half) between important variables in Figure 3. Based on the correlation coefficients, we do not need to remove variables based on collinearity. Additionally, it seems that

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.

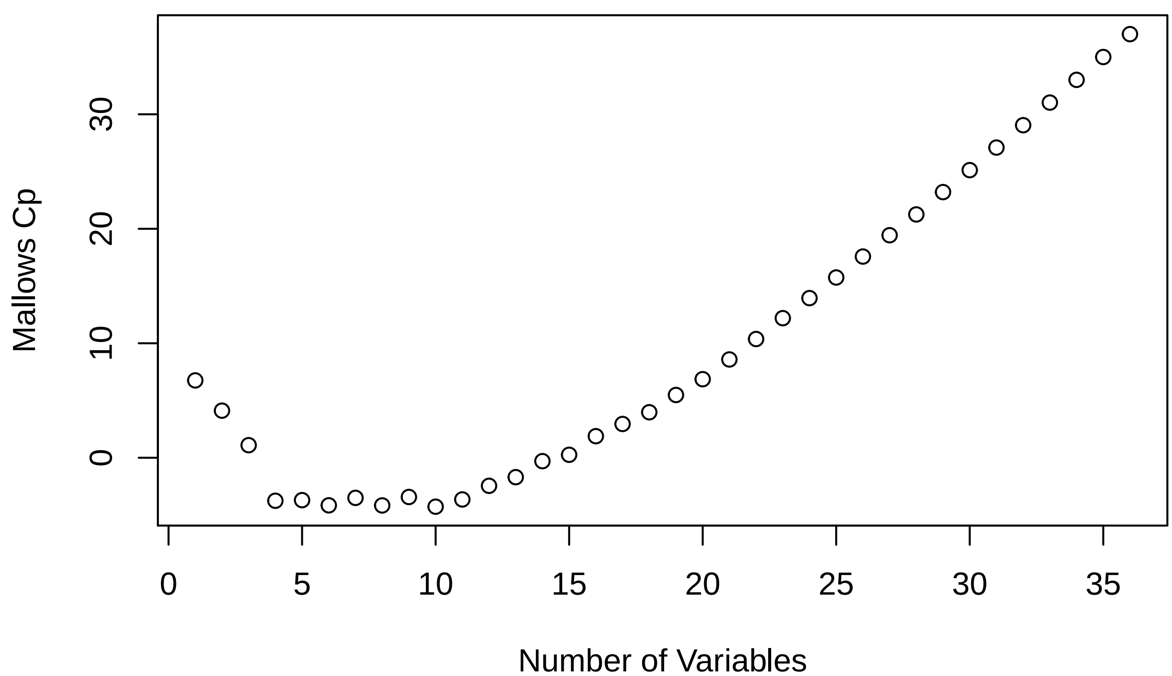
Based on this we knew that in addition to final.performance, avg.log.testosterone, avg.log.cortisol and diversity.score we should consider whether to incorporate the four additional variables proportion.female, avg.age/age.variance, time.of.day, and team.size in our models.

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 $p < .01$

Model selection using AIC




```
## [1] 10
```

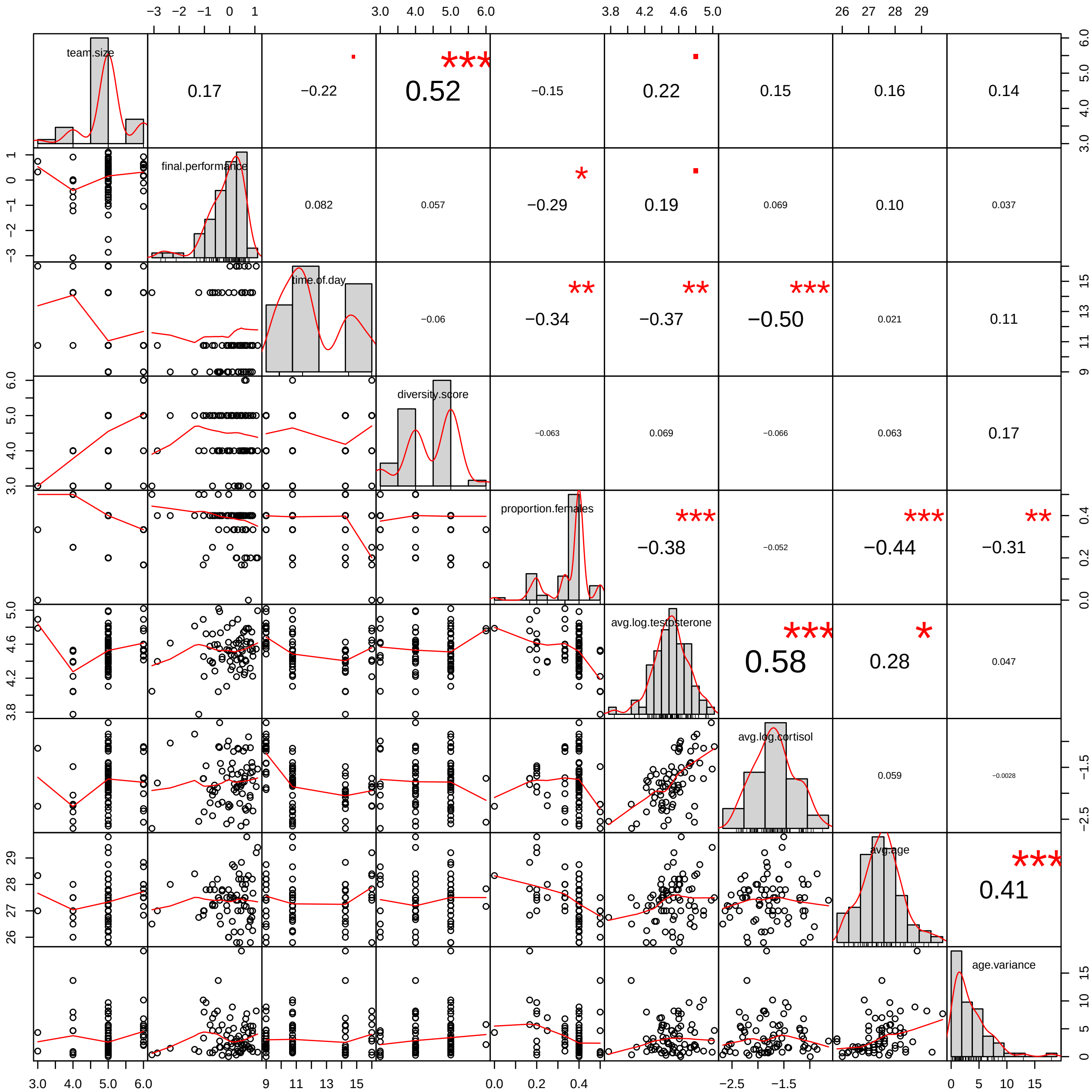


Figure 3: Pairwise correlations of important variables including their Pearson correlation coefficient. Significant correlations are marked by the corresponding number of astericks.

```
##                         (Intercept)                      diversity.score
##                         -35.3451743                            7.0027362
##                  proportion.females        team.size:proportion.females
##                         118.5672900                           -2.0537976
##          team.size:avg.log.cortisol diversity.score:avg.log.testosterone
##                          -0.6330478                           -1.5683704
##  proportion.females:avg.log.cortisol          proportion.females:avg.age
##                           7.1345520                           -3.6526464
##      proportion.females:age.variance        avg.log.testosterone:avg.age
##                           0.7415163                            0.2545245
##        avg.log.cortisol:age.variance
##                           0.1250578
```

```
## Loading required package: MuMIn
```

```
##
## Call:
## lm(formula = final.performance ~ . + avg.log.cortisol:diversity.score +
##     avg.log.testosterone:diversity.score, data = team_dat[, vars])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.64031 -0.36788  0.03976  0.51139  1.38503
##
## Coefficients:
##                                       Estimate Std. Error t value
## (Intercept)                         -37.995228  15.780794  -2.408
## team.size                             0.335997   0.186860   1.798
## time.of.day                           0.055394   0.050851   1.089
## diversity.score                       7.881303   3.249987   2.425
## proportion.females                   -1.098227   1.337972  -0.821
## avg.log.testosterone                  8.787062   3.048956   2.882
## avg.log.cortisol                      1.553163   1.355222   1.146
## avg.age                              -0.026885   0.126924  -0.212
## age.variance                          0.006678   0.029972   0.223
## diversity.score:avg.log.cortisol     -0.371535   0.287205  -1.294
## diversity.score:avg.log.testosterone  -1.898213   0.654714  -2.899
##                                     Pr(>|t|)
## (Intercept)                          0.01899 *
## team.size                            0.07695 .
## time.of.day                          0.28015
## diversity.score                      0.01819 *
## proportion.females                   0.41485
## avg.log.testosterone                 0.00540 **
## avg.log.cortisol                     0.25611
## avg.age                              0.83293
## age.variance                         0.82442
## diversity.score:avg.log.cortisol     0.20052
## diversity.score:avg.log.testosterone  0.00514 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7433 on 63 degrees of freedom
## Multiple R-squared:  0.322,  Adjusted R-squared:  0.2144
## F-statistic: 2.992 on 10 and 63 DF,  p-value: 0.003793
```

```
##                          (Intercept)     time.of.day:avg.log.testosterone
##                           -0.9556385                            0.1315594
##         time.of.day:avg.log.cortisol diversity.score:avg.log.testosterone
##                            0.2766837                           -0.2478487
##     diversity.score:avg.log.cortisol
##                           -0.6204305

##                          (Intercept)     time.of.day:avg.log.testosterone
##                            0.1577122                            0.1084851
##         time.of.day:avg.log.cortisol diversity.score:avg.log.testosterone
##                            0.2423142                           -0.2538802
##     diversity.score:avg.log.cortisol  proportion.females:avg.log.cortisol
##                           -0.6291508                            0.9055138

##                          (Intercept)        team.size:proportion.females
##                           -0.2842658                            0.7205776
##     time.of.day:avg.log.testosterone        time.of.day:avg.log.cortisol
##                            0.1027025                            0.2134798
## diversity.score:avg.log.testosterone    diversity.score:avg.log.cortisol
##                           -0.2805569                           -0.6692384
##  proportion.females:avg.log.cortisol
##                            2.3662279

##                          (Intercept)                      diversity.score
##                          -35.3451743                            7.0027362
##                   proportion.females        team.size:proportion.females
##                          118.5672900                           -2.0537976
##           team.size:avg.log.cortisol diversity.score:avg.log.testosterone
##                           -0.6330478                           -1.5683704
##  proportion.females:avg.log.cortisol           proportion.females:avg.age
##                            7.1345520                           -3.6526464
##      proportion.females:age.variance        avg.log.testosterone:avg.age
##                            0.7415163                            0.2545245
##        avg.log.cortisol:age.variance
##                            0.1250578

##
## Call:
## lm(formula = final.performance ~ . - avg.age - proportion.females -
##     time.of.day + age.variance + avg.log.testosterone:diversity.score,
##     data = team_dat[, vars])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.67055 -0.35597  0.06595  0.42258  1.32606
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          -53.61498   11.40288  -4.702 1.33e-05
## team.size                              0.38299    0.17440   2.196   0.0316
## diversity.score                       10.90565    2.49118   4.378 4.30e-05
## avg.log.testosterone                  11.36147    2.45462   4.629 1.74e-05
## avg.log.cortisol                      -0.31715    0.23382  -1.356   0.1795
## age.variance                           0.01632    0.02667   0.612   0.5426
## diversity.score:avg.log.testosterone  -2.42530    0.55019  -4.408 3.86e-05
##
```

```
## (Intercept)                          ***
## team.size                            *
## diversity.score                      ***
## avg.log.testosterone                 ***
## avg.log.cortisol
## age.variance
## diversity.score:avg.log.testosterone ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7471 on 67 degrees of freedom
## Multiple R-squared:  0.2714, Adjusted R-squared:  0.2062
## F-statistic:  4.16 on 6 and 67 DF,  p-value: 0.001302
```

```
##
## Call:
## lm(formula = final.performance ~ . + avg.log.testosterone:diversity.score,
##     data = team_dat[, vars])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.63577 -0.29005 -0.00661  0.49883  1.28554
##
## Coefficients:
##                                        Estimate Std. Error t value
## (Intercept)                          -50.313765  12.649949  -3.977
## team.size                              0.401820   0.180742   2.223
## time.of.day                            0.051101   0.051008   1.002
## diversity.score                       10.516527   2.545613   4.131
## proportion.females                    -0.875978   1.333858  -0.657
## avg.log.testosterone                  10.903232   2.586462   4.215
## avg.log.cortisol                      -0.169204   0.254148  -0.666
## avg.age                               -0.049446   0.126380  -0.391
## age.variance                           0.008497   0.030096   0.282
## diversity.score:avg.log.testosterone  -2.337221   0.562833  -4.153
##                                      Pr(>|t|)
## (Intercept)                          0.000180 ***
## team.size                            0.029743 *
## time.of.day                          0.320210
## diversity.score                      0.000107 ***
## proportion.females                   0.513712
## avg.log.testosterone                 7.98e-05 ***
## avg.log.cortisol                     0.507953
## avg.age                              0.696914
## age.variance                         0.778595
## diversity.score:avg.log.testosterone 9.92e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7472 on 64 degrees of freedom
## Multiple R-squared:  0.304,  Adjusted R-squared:  0.2061
## F-statistic: 3.106 on 9 and 64 DF,  p-value: 0.003707
```

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

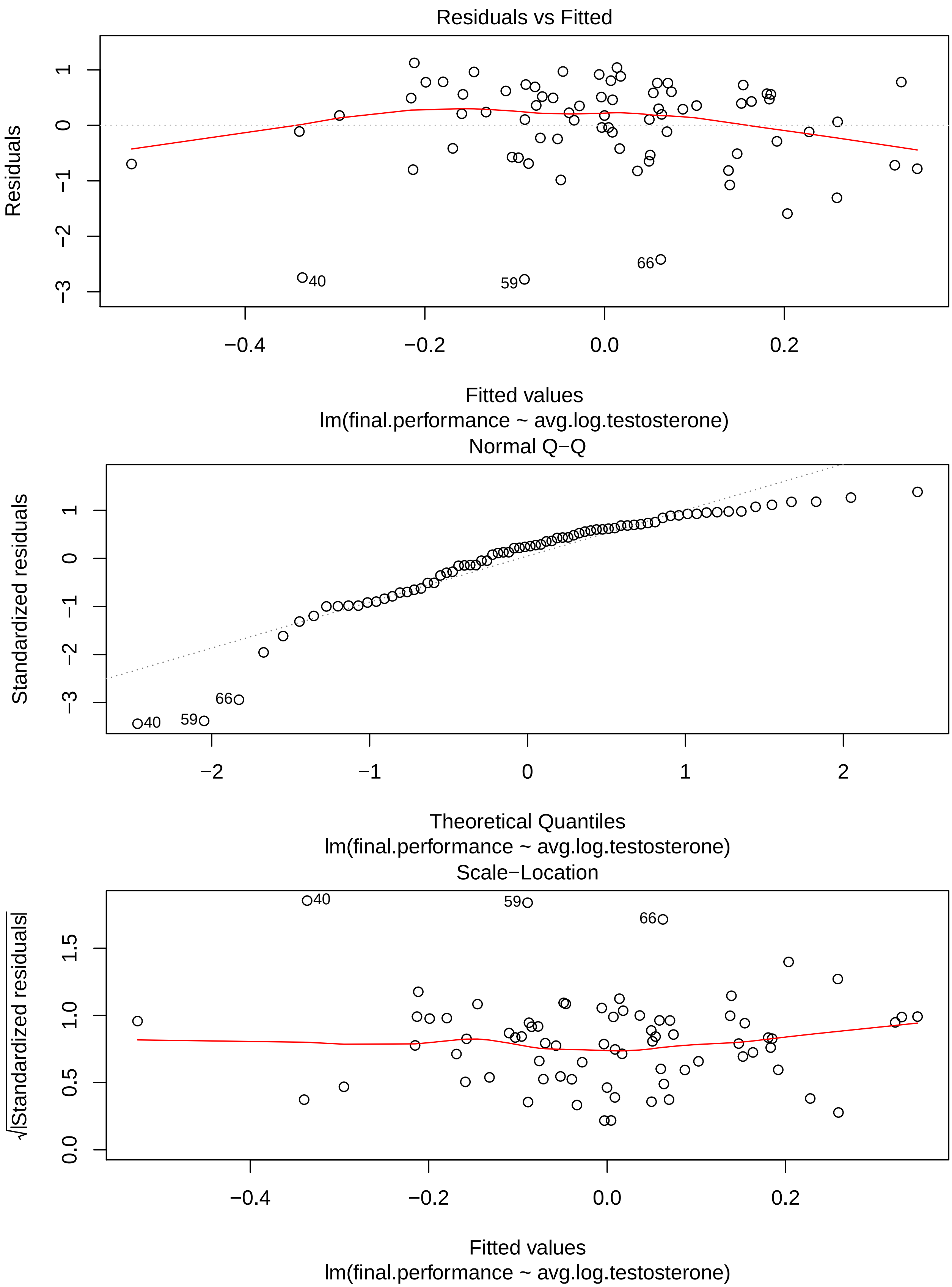
```
## lm(formula = final.performance ~ time.of.day + team.size + diversity.score +
##     avg.log.testosterone + avg.age + +avg.log.cortisol, data = team_dat[,
##     vars])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.71599 -0.42471  0.05258  0.58792  1.25779
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -5.669253   3.714949  -1.526    0.132
## time.of.day           0.072837   0.047882   1.521    0.133
## team.size             0.260160   0.193282   1.346    0.183
## diversity.score      -0.052341   0.160280  -0.327    0.745
## avg.log.testosterone  0.807460   0.556899   1.450    0.152
## avg.age               0.004963   0.124636   0.040    0.968
## avg.log.cortisol      0.021538   0.277950   0.077    0.938
##
## Residual standard error: 0.8341 on 67 degrees of freedom
## Multiple R-squared:  0.09194,    Adjusted R-squared:  0.01062
## F-statistic: 1.131 on 6 and 67 DF,  p-value: 0.3543
```

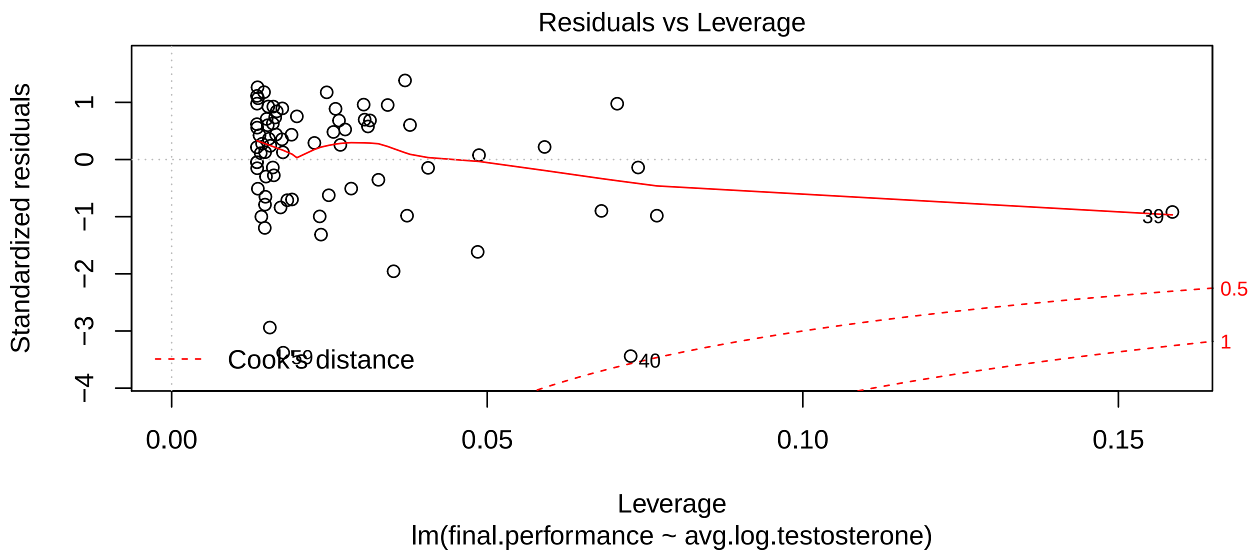
Diversity score and testosterone do not individually significantly predict performance

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.

	x
(Intercept)	-3.181
avg.log.testosterone	0.703

	x
(Intercept)	-0.292
diversity.score	0.066


Residuals vs Fitted
Residuals
Fitted values
lm(final.performance ~ avg.log.testosterone)
40
59
66
Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(final.performance ~ avg.log.testosterone)
Scale−Location
√|Standardized residuals|
Fitted values
lm(final.performance ~ avg.log.testosterone)



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.

```
##
## Call:
## lm(formula = final.performance ~ avg.log.testosterone + diversity.score +
##     team.size, data = team_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8088 -0.4448  0.2347  0.5385  1.1123
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -3.52528    1.97936  -1.781   0.0792 .
## avg.log.testosterone  0.58600    0.43476   1.348   0.1821
## diversity.score      -0.04095    0.15758  -0.260   0.7957
## team.size             0.21123    0.18851   1.121   0.2663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.832 on 70 degrees of freedom
## Multiple R-squared:  0.05604,    Adjusted R-squared:  0.01558
## F-statistic: 1.385 on 3 and 70 DF,  p-value: 0.2545
```

```
##
## 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                ***
## 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
```

	x
(Intercept)	-48.112
avg.log.testosterone	10.274
diversity.score	10.189
team.size	0.353
avg.log.testosterone:diversity.score	-2.258

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 4 which seems weakly linear (adjusted r squared value < 0.2).

	x
(Intercept)	0.216
avg.log.cortisol	0.122

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.

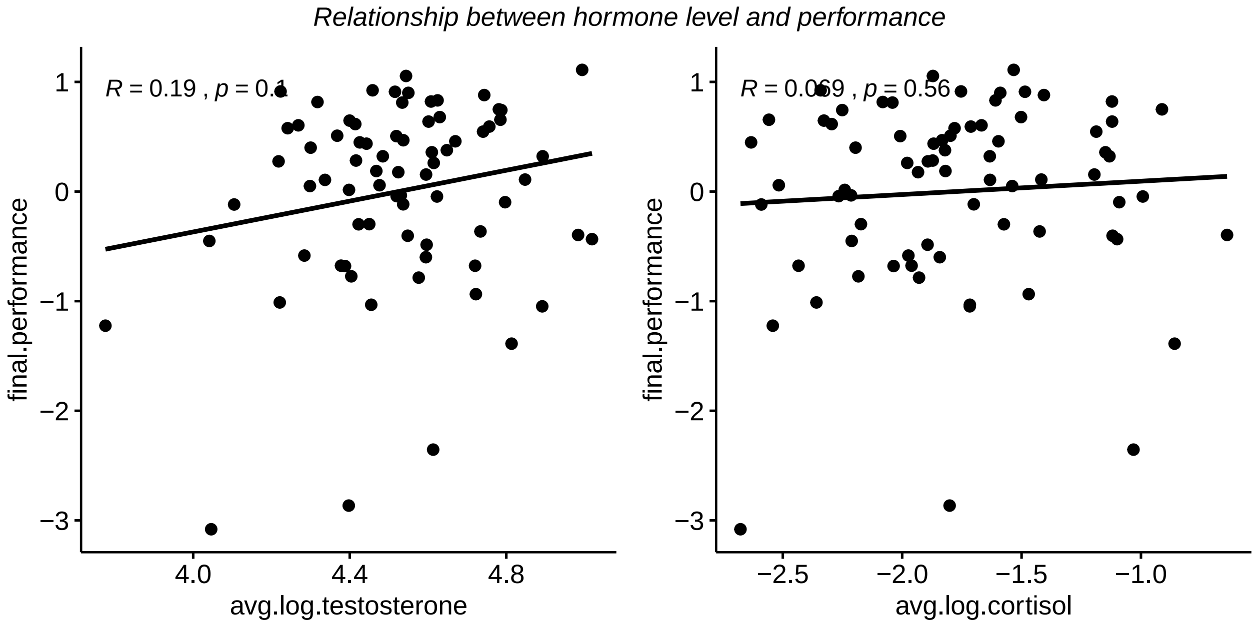


Figure 4: Levels of both testosterone and cortisol correlate with performance

```
##
## 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.

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.

Add: * practical significance i.e. units *

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.

Kelsey, Thomas W., Lucy Q. Li, Rod T. Mitchell, Ashley Whelan, Richard A. Anderson, and W. Hamish B. Wallace. 2014. "A Validated Age-Related Normative Model for Male Total Testosterone Shows Increasing Variance but No Decline after Age 40 Years." Edited by Bin He. *PLoS ONE* 9 (10):e109346. https://doi.org/10.1371/journal.pone.0109346.

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.

Matsuzaka, Hisashi, Hitoshi Maeshima, Sayaka Kida, Hirofumi Kurita, Takahisa Shimano, Yoshiyuki Nakano, Hajime Baba, Toshihito Suzuki, and Heii Arai. 2013. "Gender Differences in Serum Testosterone and Cortisol in Patients with Major Depressive Disorder Compared with Controls." *The International Journal of Psychiatry in Medicine* 46 (2):203–21. https://doi.org/10.2190/PM.46.2.g.

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.