The Effect of Estrogen and Progesterone on Breathing during Exercise in Hypoxia and Normoxia

Client: Sierra Arn

STAT 450 (Case Studies in Statistics)

Instructors: Keegan Korthauer, Rodolfo Lourenzutti, and Melissa Lee

Authors: Thomas Farrell, Licy Li, Lily Xie, and Vicky Lin

Department of Statistics, University of British Columbia

April 11, 2025

Summary

Fluctuations of reproductive hormones, particularly estrogen and progesterone, play a key role in regulating respiratory function throughout the menstrual cycle. This study examines the relationship between hormone levels and ventilation in both premenopausal and postmenopausal females during exercise under different oxygen conditions, with specific focus on menstrual phase differences in younger females and the influence of hormone replacement therapy (HRT) in postmenopausal females. Data come from a total of 31 participants tested in hospital and lab settings.

Exploratory data analysis indicates that individual differences in ventilation outweigh variations due to menstrual phases or oxygen conditions. Linear mixed-effects models with participant-level random effects account for this variability. Hypoxia and higher fitness levels both associate with increased ventilation. In postmenopausal women, HRT appears to lessen the ventilatory effect of hypoxia.

Contents

List of Figures iii

List of Tables iii

1 Introduction 1

2 Data Description 2

3 Methods 4

4 Results 7

5 Conclusions 12

References 13

Appendices 14

A Tables and Figures 14

B R Code 28

List of Figures

1 Interaction plot on phase and condition for ventilation for younger group 3

2 Q-Q plot for the model for older group 9

A1 Boxplot of relationship between ventilation vs. phase and condition for younger group 14

A2 Boxplot of relationship between ventilation vs. HRT and condition for older group 15

A3 Correlation matrix of numeric variables of younger female dataset . . 16

A4 Correlation matrix of numeric variables of older female dataset 17

A5 Interaction plot on HRT and condition for ventilation for older group 18

A6 Plot of residuals against fitted values for younger group 18

A7 Normal Q-Q plot for younger group 19

A8 Plot of residuals against fitted values for older group 20

A9 Plot of residuals against fitted values for group comparison model . . 21

A10 Normal Q-Q plot for group comparison model 22

List of Tables

1 Ranges of calculated p-values for each model term across 1000 imputation runs 11

A1 Summary statistics on key numeric variables for younger group 23

A2 Summary statistics on key numeric variables for older group 23

1 Introduction

Reproductive hormones, particularly estrogen and progesterone, play a vital role in regulating respiratory function across different phases of the menstrual cycle. Prior research suggests that hormonal changes impact ventilation rates, respiratory muscle recruitment, and chemosensitivity, particularly under varying oxygen conditions. Most studies, however, focus on breathing at resting states rather than during exercise. This study investigates the relationship between hormone levels and ventilation in both young and older women under normoxic and hypoxic conditions during exercise. The statistical analysis has three objectives: primarily, to examine the effect on ventilation of varying levels of progesterone and estradiol throughout the menstrual cycle in young women, secondarily, to assesses the impact of hormone replacement therapy (HRT) on ventilation in postmenopausal women. Additional factors influencing respiratory responses are also explored for both groups. The tertiary objective is to compare levels of ventilation between the younger and older groups - for instance, to determine whether younger females may breathe more than older females during the midluteal phase but not the early follicular phase. We perform exploratory data analysis and fit statistical models to identify trends and associations. This report presents a summary of the data, the analytical methods used, and key findings.

2 Data Description

Data for this study are collected from both hospital and lab settings. The experimental protocol consists of three (non-consecutive) days: one for screening and two for testing. During the screening day, participants undergo assessments of maximal oxygen uptake and pulmonary function. This is then followed by two testing days: the younger women complete trials in both the early follicular (EF) and midluteal (ML) phases, while the older women participate only on one testing day.

Before exercise, estrogen and progesterone levels are measured. Participants perform a 5-minute cycling session at 70% maximal intensity, alternating between normoxic (21% oxygen) and hypoxic (15% oxygen) trials. Key physiological measurements include ventilation, ribcage and abdominal contributions, oxygen consumption, and work of breathing (WOB). Additional participant information, such as age, height, weight, BMI, and IUD usage, is collected for further analysis. Older women are categorized based on hormone replacement therapy (HRT) use.

The study includes 13 younger and 18 older female participants from whom two respective datasets are obtained. Both datasets contain missing values in the columns for progesterone and estradiol: some participants have progesterone values reported as "<1 nmol/L" or estradiol values reported as "<90 pmol/L." In particular, 83% of the progesterone measurements and 72% of the estradiol measurements for the older participants are missing. This is due to systematic limitations of the method of measurement used by the hospital. Since exact values are unavailable, we apply uniform imputation to randomly assign values below the measurement threshold, as discussed in the Methods section.

Tables A1 and A2 summarize key numeric variables - including the mean, median, maximum, and minimum for each - computed separately for younger and older groups after imputation. The boxplots in Figures A1 and A2 show differences in ventilation across the two oxygen conditions, highlighting distinctions between the two menstrual phases within younger women, HRT use in older women, and between age groups overall.

The boxplot in Figure A1 illustrates the distribution of ventilation levels for younger female participants across four subgroups based on the combination of menstrual phase and oxygen condition. Ventilation is higher under hypoxia compared to normoxia, regardless of menstrual phase. A slight increase in median ventilation is observed from the early follicular to the midluteal phase under normoxia, although the median ventilation level under hypoxia is lower in the midluteal phase than in the early follicular phase. The boxplot in Figure A2 depicts ventilation levels among older women, grouped by HRT status and oxygen condition. Among participants not using HRT, ventilation is noticeably higher under hypoxia than normoxia. Participants using HRT show lower and more consistent ventilation across oxygen conditions, with a smaller difference between hypoxia and normoxia.

Figures 1 and A5 demonstrate substantial individual variability in baseline ventila-

tion, as evidenced by the wide difference in line ranges across different individuals. This motivates the use of a model with a random effect to account for individual variability. Furthermore, Figure 1 also suggests hypoxia increases ventilation (the lines corresponding to hypoxia are generally above the lines corresponding to normoxia), and menstrual phase may have an effect as indicated by the positive slopes of some of the lines. Analogously for the postmenopausal women, Figure A5 shows that hypoxia increases ventilation, as indicated by the consistently negatively sloped lines. The figure also suggests that HRT may blunt the ventilatory response to hypoxia: the lines for participants on HRT seem to have shallower slopes on average than the lines for participants without HRT.

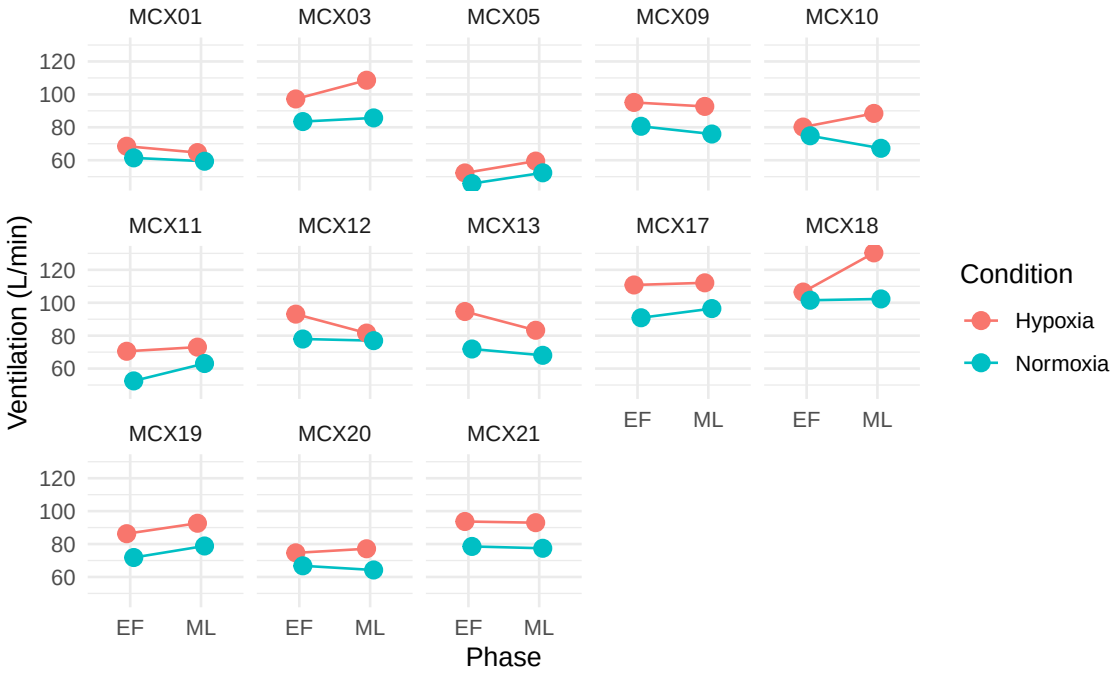

Figure 1: Interaction plot on phase and condition for ventilation for younger group

3 Methods

To examine how hormones affect ventilation across the menstrual cycle in young women, and how HRT influences ventilation in postmenopausal women, we use a linear regression approach. This method provides interpretable estimates of how covariates affect ventilation. Since the EDA reveals greater variation between participants than between menstrual phases or oxygen conditions (Figures 1 & A5), we fit mixed-effects models with separate intercepts per participant.

Imputation Method: Uniform Imputation Within the Known Range

The hormone measurements in the dataset are limited by the detection thresholds of the hospital laboratory's evaluation methods. Specifically, some of the younger women and most of the older women had levels of progesterone and estradiol that were too low to be measured: the assays used could not reliably detect progesterone concentrations below 1 nmol/L or estradiol concentrations below 90 pmol/L. Because these hormone levels are critical to our analysis, imputing the missing values is necessary to proceed with modeling.

While single imputation methods do improve data completeness, they fail to account for uncertainty in the missing values and may lead to overconfident results (Azur et al., 2011). Instead, we apply uniform imputation within the known range. This method replaces missing values with random draws from a uniform distribution between zero and the lower limit of measurement for each hormone.

By introducing randomness, uniform imputation preserves natural variability and avoids overly precise estimates. It also offers a simple, statistically grounded alternative to multivariate imputation by chained equations (MICE), which, while robust, may not be feasible due to computational costs (Azur et al., 2011).

In summary, uniform imputation allows valid and reliable results by preserving uncertainty. We assess its specific suitability for our data in the analysis section.

Model Fitting and Analysis Methods

In order to address the research questions, we followed a model selection and fitting process based on backward selection three times: once for the data on the younger group, once for the data on the older group, and finally for a dataset combining the two groups in order for us to be able to directly compare the groups.

To address the secondary research question specific to the older women, we had to consider the high proportion of missing progesterone and estradiol measurements: out of 18 study participants, only 3 had measurable progesterone levels and only 5 had measurable estradiol levels, so any conclusions drawn specific to the older group

of participants regarding levels of progesterone or estradiol would be of questionable validity no matter what method of imputation was used. The categorical variable regarding the presence of HRT, however, did not have any missing values, and the sizes of the groups with and without HRT (8 and 10 respectively) were not particularly unbalanced. HRT also provides some indication of hormone concentration, so it's reasonable to use it instead of the desired (but mostly missing) variables. We therefore included HRT instead of the progesterone and estradiol measurements in our model selection process for the older group.

For the tertiary research question involving making comparisons between the younger and older groups, we again excluded progesterone and estradiol due to the limited data for the older group. Instead we used two new explanatory variables, which taken together would group all of the study participants into four groups: a binary variable for age, indicating presence in the younger or older group, and a binary variable for higher hormone levels than the individual's baseline (midluteal phase for the younger group and HRT for the older group).

For both groups, our data contains many explanatory variables that are highly correlated with each other (Figures A3 and A4). For instance, predicted VO2max is directly calculated from age and height. Such multicollinearity can introduce issues with both model fitting and inference. In order to avoid these issues, we first excluded some of the explanatory variables based on a combination of calculated variance inflation factor (VIF) values and background knowledge about the variables. For both the younger group and the older group, the variables excluded in this way were BMI, relative VO2max, predicted VO2max, percent of predicted VO2max, and 100% peak power. We also removed the same list of variables for model fitting on the combined dataset.

Backward selection was then used to identify the most relevant subset of the remaining predictors. This process first considers a model that includes all possible covariates and then iteratively removes the least significant variables one at a time using Akaike's Information Criterion (AIC), ensuring that the final model retains only the most impactful predictors. Backward selection was chosen because it does not test a large number of models (compared to, for instance, exhaustive search), which is particularly important given our small sample size. Starting from a full model that included all variables and interaction terms of potential interest, this process thus allowed us to minimize risks of overfitting while reducing the impact of multicollinearity.

We included interaction terms between the main variables of interest (oxygen condition, progesterone, and estradiol for the younger group, and oxygen condition and HRT for the older group) in the full model from which the backward selection process started. Also included was an interaction term between oxygen condition and 70% peak power. Since some of the model terms (including some of the interactions) were judged to be of particular research interest, we prevented the backward selection process from removing them.

To assess the fit of all of our models and determine whether the required model assumptions (that residuals are independent, approximately Normally distributed, and with constant variance) were satisfied, we examined plots of residuals against fitted values, plots of residuals against the explanatory variables of interest, and normal Q-Q plots. Lastly, we performed an additional analysis on the data for the younger group to assess whether our method of randomly imputing the missing progesterone measurements from a uniform distribution between 0 and 1 was appropriate and sufficient for our purposes. We generated 1000 different datasets using this method and fit our selected model to each of them, and then observed the extent of the variation in results across the runs.

4 Results

The following section presents our findings, including the impact of hormone levels, HRT, and general fitness on ventilation, as well as comparisons of ventilation levels across different groups with varying age, menstrual phase, HRT status, and oxygen condition. This section also includes a discussion of the consistency of results across repeated imputation runs.

Model for Younger Group

The terms selected by backward selection were oxygen condition, progesterone, estradiol, 70% peak power, the interaction between oxygen condition and progesterone, the interaction between progesterone and estradiol, and the random intercept per participant.

Table A3 summarizes the backward selection process, where variables marked with "0" were retained in the final model, while variables numbered from 1 to 8 were eliminated in ascending order of importance. The model summary table (Table A5) shows estimates of each predictor in the reduced model. The key findings include the following:

- **Oxygen condition** has the strongest effect on ventilation, and is significant at the 0.1% level ($p = 3.9 \times 10^{-8}$). We found hypoxia to be associated with an average increase in ventilation of 13.0 L/min, all else being constant. It makes intuitive sense that women exercising in a hypoxic environment would breathe more air to compensate for the lower oxygen concentration of that air.
- **70% peak power** also has an effect on ventilation that is significant at the 5% level ($p = 0.02$). An increase of 1 W of 70% peak power while keeping other variables constant is associated with an average increase in ventilation of 0.49 L/min. This appears to indicate that women at a higher fitness level who are able to achieve a higher peak power use more air to achieve that power level than others who only achieve a lower power.
- There is not enough evidence to suggest that **progesterone** ($p = 0.63$), **estradiol** ($p = 0.62$), the interaction between oxygen condition and progesterone ($p = 0.60$) or the interaction between progesterone and estradiol ($p = 0.63$) have an effect on ventilation.
- The random intercepts per **participant** have a standard deviation of 13.0 L/min, indicating that different women with similar values of the explanatory variables may have a relatively wide range of ventilation values.

The 95% confidence intervals for all of these estimates may be found in the appendix (Table A6). The random-effects table (Table A4) indicates the necessity of including

the random intercept per participant. Including this term improves the model fit at the 0.1% significance level ($p = 2.3 \times 10^{-11}$), suggesting that individual differences between participants other than those accounted for by our explanatory variables contribute significantly and meaningfully to ventilation.

The model for younger females suggests that ventilation is significantly influenced by both oxygen condition (hypoxia vs. normoxia) and power output during exercise, with relatively wide variation between individuals. No significant evidence of ventilation varying with the measured hormone levels was found.

The Q-Q plot (Figure A7) shows no clear violation of the assumption of Normality. Most of the points in the plot align well with the reference line, which suggests that the residuals are approximately normally distributed. There are some outliers, but overall, the model is quite a good fit.

Model for Older Group

For the older group, backward selection selected oxygen condition, HRT, measured VO2max, and the interaction between oxygen condition and HRT, as well as the random intercept per participant.

Our results from the model, summarized in Table A9, indicate that ventilation increases by an average of 15.6 L/min under hypoxia for women not receiving HRT (with all other variables kept constant), and that this effect is again significant at the 0.1% level ($p = 2.2 \times 10^{-7}$). Additionally, in contrast to our lack of significant findings relating to hormones in the younger group, women receiving HRT only exhibit a predicted increase of 6.6 L/min in ventilation under hypoxic conditions compared to under normoxic conditions, while keeping all other variables constant, and the difference in the effect of hypoxia depending on the presence of HRT is significant at the 1% level ($p = 0.004$), meaning that there is evidence to suggest that HRT is associated with a lowered effect of hypoxia on ventilation. The association between HRT and ventilation under normoxia was not found to be statistically significant ($p = 0.56$).

We also found that an increase of 1 L/min in measured VO2max is associated with an increase of 21.2 L/min in ventilation, for constant levels of other variables, and this effect is significant at the 5% level ($p = 0.03$). Although the backward selection process selected measured VO2max instead of 70% peak power for the older group, the intuitive explanation for this variable being significant is similar: women with a higher general fitness level are likely to have a higher VO2max measurement, and women with higher VO2max are able to use more oxygen over a given period of time which means they must breathe more in order to make this amount of oxygen available to their bodies.

Although the plot of residuals against fitted values (Figure A8) for this model again showed no clear pattern, the Q-Q plot below shows a few points deviating from the reference line at each end, possibly suggesting a heavy-tailed distribution.

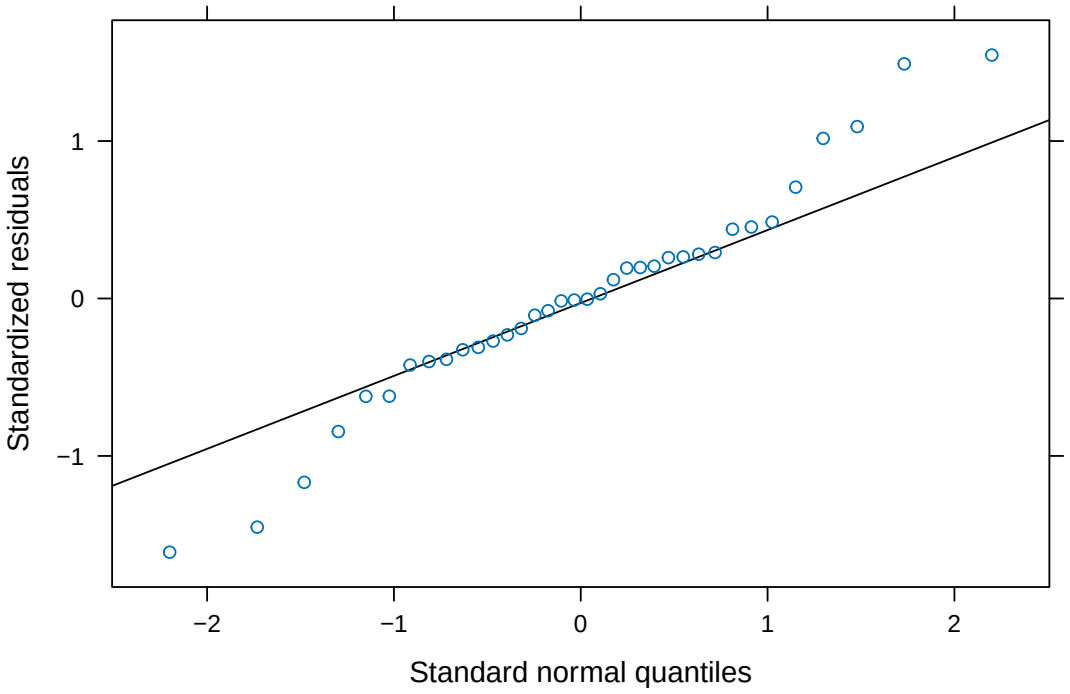

Figure 2: Q-Q plot for the model for older group

Overall, we judge this model to be a reasonable fit, even though the assumption of Normally-distributed residuals is unclear.

Models for Comparison of two groups

The backward selection process on our dataset combining the younger and older groups led to oxygen condition, age group, and hormone group, all possible interaction terms between these three variables, and 70% peak power being selected for the final model.

The model summary table (Table A13) shows that an increase of 1 W of 70% peak power is associated with an average increase of 0.32 L/min in ventilation, for constant values of the other variables, and that this is significant at the 1% level ($p = .006$), which agrees with the results from the model for younger women and reinforces the conclusion that higher general fitness is associated with higher ventilation. The model term for oxygen condition (which in this model corresponds to the difference in ventilation between hypoxia and normoxia for younger women in the early follicular phase–estimates for other groups are obtained by adding model interaction terms) is again significant at the 0.1% level. Of the remaining model terms, however, the only one found to be statistically significant at the 5% level is the three-way interaction term, which is essentially equivalent to the interaction between oxygen condition and HRT from the model for the older group. In particular, the term for age has an associated p-value of 0.66, meaning that we did not find evidence that ventilation differs with age for constant values of the other variables: any differences in mean ventilation between the younger group and the older group can be explained by other variables (likely those measuring general fitness level). We draw a similar conclusion from the term for hormone group: its associated p-value of 0.69 means that we did not find evidence that higher-than-baseline hormone levels (HRT in postmenopausal women or midluteal phase instead of early follicular phase in younger women) is associated with differences in ventilation (except in the case of older women under hypoxia).

Evaluation of imputation method

To evaluate the robustness of our chosen imputation method regarding our results for the younger group, 1,000 datasets were generated by repeatedly imputing the missing progesterone values from a uniform distribution between 0 and 1, and our selected model was fit to each dataset. Across all runs, conclusions remained consistent, with p-values showing minimal variation, confirming that the results are not sensitive to the specific imputed values and supporting the adequacy of uniform imputation in this case.

Table 1: Ranges of calculated p-values for each model term across 1000 imputation runs

Intercept	Condition	Progesterone	Estradiol
Min. :0.91	Min. :0.87	Min. :0.60	Min. :0.61
1st Qu.:0.91	1st Qu.:0.87	1st Qu.:0.62	1st Qu.:0.63
Median :0.91	Median :0.87	Median :0.63	Median :0.65
Mean :0.91	Mean :0.87	Mean :0.63	Mean :0.65
3rd Qu.:0.91	3rd Qu.:0.87	3rd Qu.:0.64	3rd Qu.:0.67
Max. :0.91	Max. :0.88	Max. :0.66	Max. :0.69

70% peak	Condition × Progesterone	Condition × 70% peak
Min. :0.034	Min. :0.56	Min. :0.21
1st Qu.:0.034	1st Qu.:0.58	1st Qu.:0.22
Median :0.034	Median :0.58	Median :0.22
Mean :0.034	Mean :0.58	Mean :0.22
3rd Qu.:0.034	3rd Qu.:0.58	3rd Qu.:0.22
Max. :0.034	Max. :0.60	Max. :0.22

5 Conclusions

For young women, our results show no statistically significant differences in ventilation due to levels of progesterone or estradiol. We confirmed that hypoxic conditions lead to a significant rise in ventilation, with an estimated average increase of 13.0 L/min compared to normoxia, and individuals with higher general fitness are also likely to exhibit higher ventilation. However, substantial individual variability remains, indicating a strong random effect.

For older women, our findings again confirm that compared to normoxia, hypoxia significantly increases ventilation. We also estimate that women receiving HRT show somewhat higher average ventilation relative to those not on HRT, but this estimate is not statistically significant. However, the significant interaction term suggests that HRT may lower the ventilatory response to hypoxia. Similar to the younger group, individual variability remains considerable, emphasizing the need for further investigation.

A regression model was also used to compare the younger and older groups, but the lack of new statistically significant findings showed that any differences in ventilation between the two groups can be adequately explained by variables other than age, such as those measuring general fitness level.

Our analysis has some limitations. Participants in the study tend to have higher-than-average fitness levels, and this biased sample restricts the generalizability of the results. It is likely not reasonable to attempt to apply the results to the population as a whole. The small sample size limits the study's ability to detect smaller effects on ventilation, such as those that we might expect to find from varying hormone levels. Additionally, because of the small sample size, we did not split the data, but this has increased the risk of Type I error due to double-dipping. The younger participants were also only tested in two of the four menstrual phases, and since the midluteal phase tends to be associated with increased levels of both progesterone and estradiol compared to the early follicular phase, these two variables have somewhat high correlation in this dataset. This means that it can be difficult for statistical methods to isolate the effect of one or the other. The inability to precisely measure the hormone levels for most of the older group meant that we could only reasonably attempt to directly model their effect on ventilation for the younger group. To better understand how productive hormones influence ventilation under varying physiological conditions, we recommend the following: avoiding sampling bias, including larger cohorts of participants, testing participants at a wider variety of menstrual phases, and measuring hormone levels more precisely.

References

Pelizzo, G., Calcaterra, V., Baldassarre, P., Marinaro, M., Taranto, S., Ceresola, M., Capelo, G., Gazzola, C., & Zuccotti, G. (2024). The impact of hormones on lung development and function: an overlooked aspect to consider from early childhood. Frontiers in Endocrinology, 15, 1425149. https://doi.org/10.3389/fendo.2024.1425149

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work?. International journal of methods in psychiatric research, 20(1), 40–49. https://doi.org/10.1002/mpr.329

Appendices

A Tables and Figures

Figures

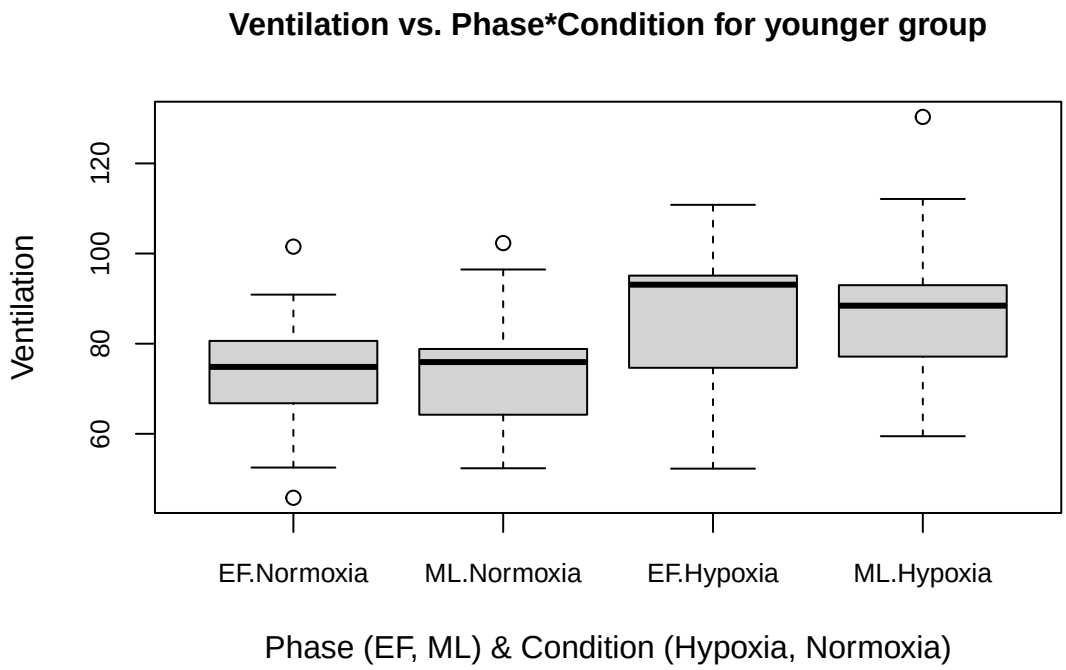

Figure A1: Boxplot of relationship between ventilation vs. phase and condition for younger group

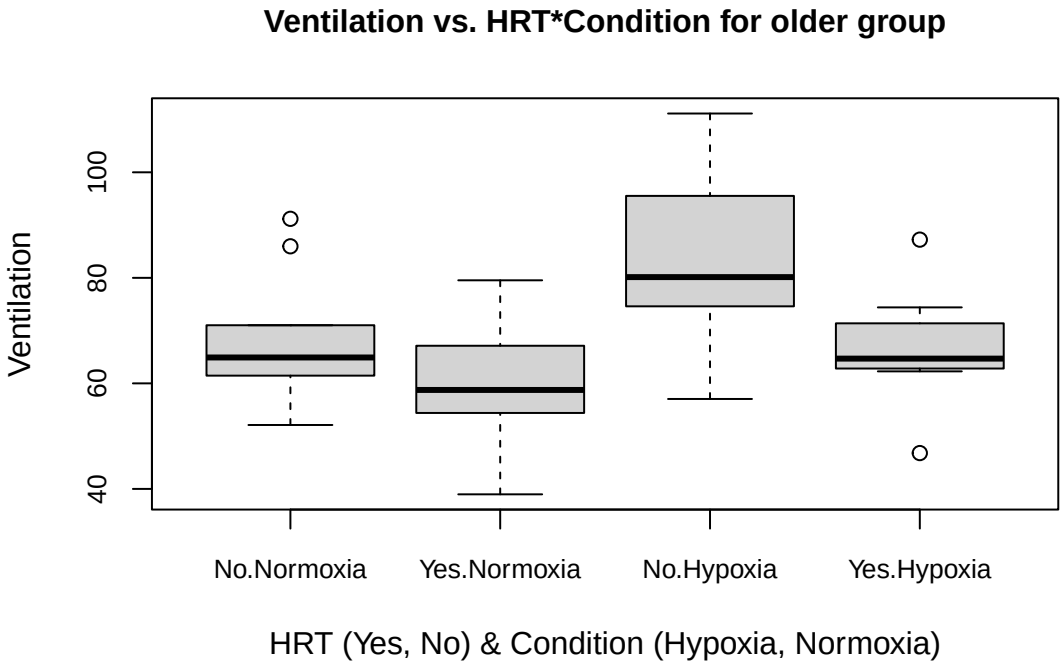

Figure A2: Boxplot of relationship between ventilation vs. HRT and condition for older group

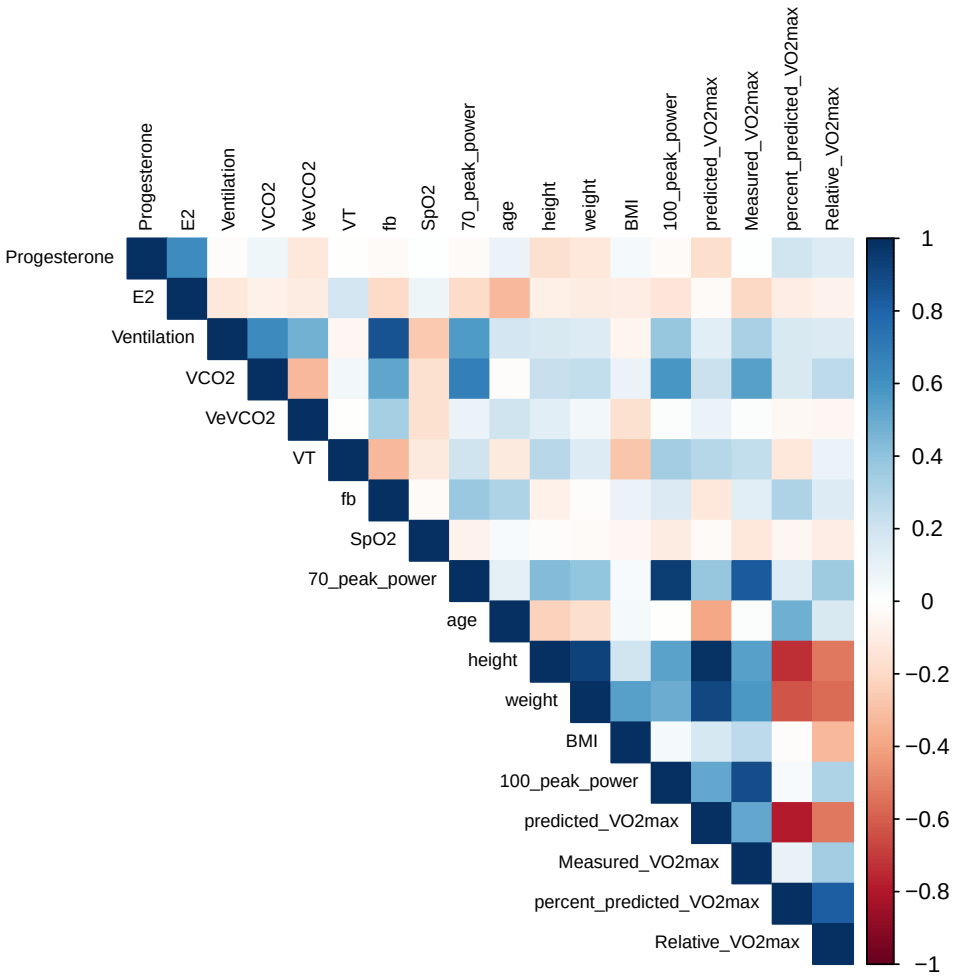

Figure A3: Correlation matrix of numeric variables of younger female dataset

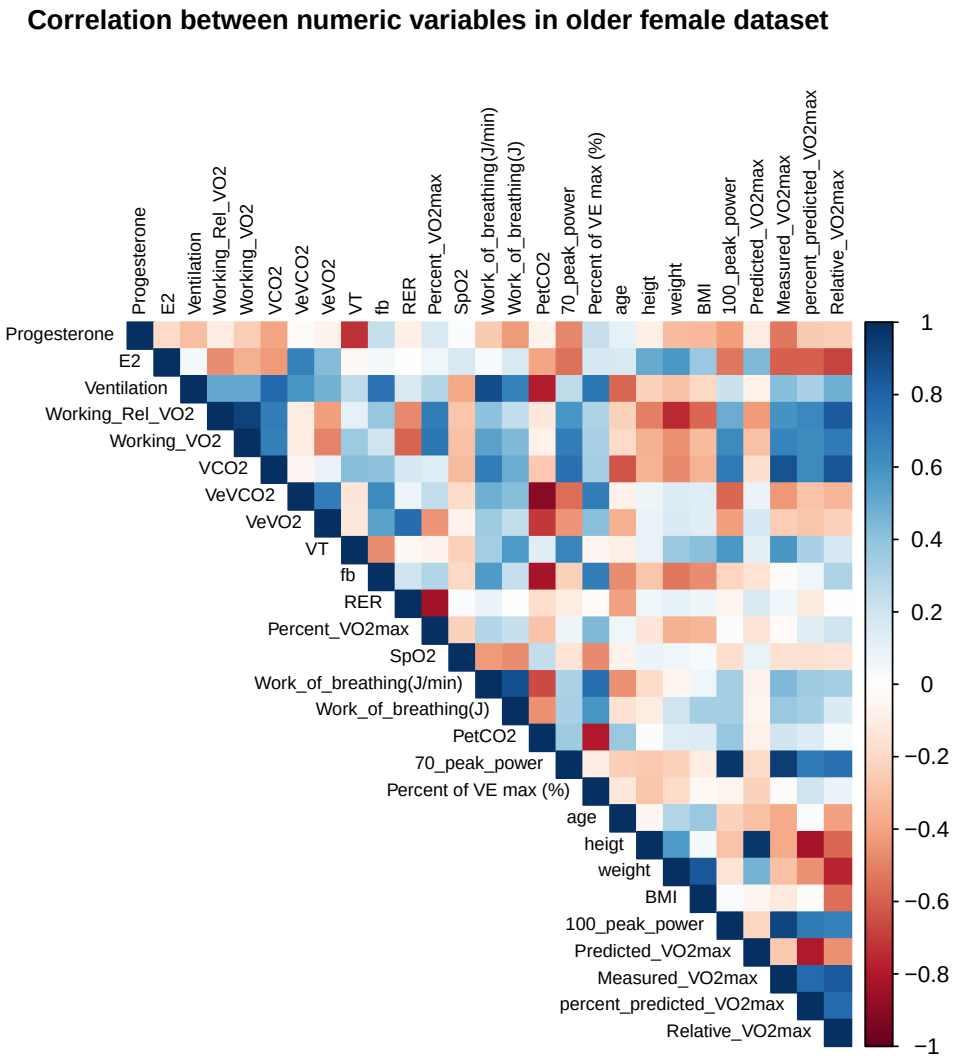

Figure A4: Correlation matrix of numeric variables of older female dataset

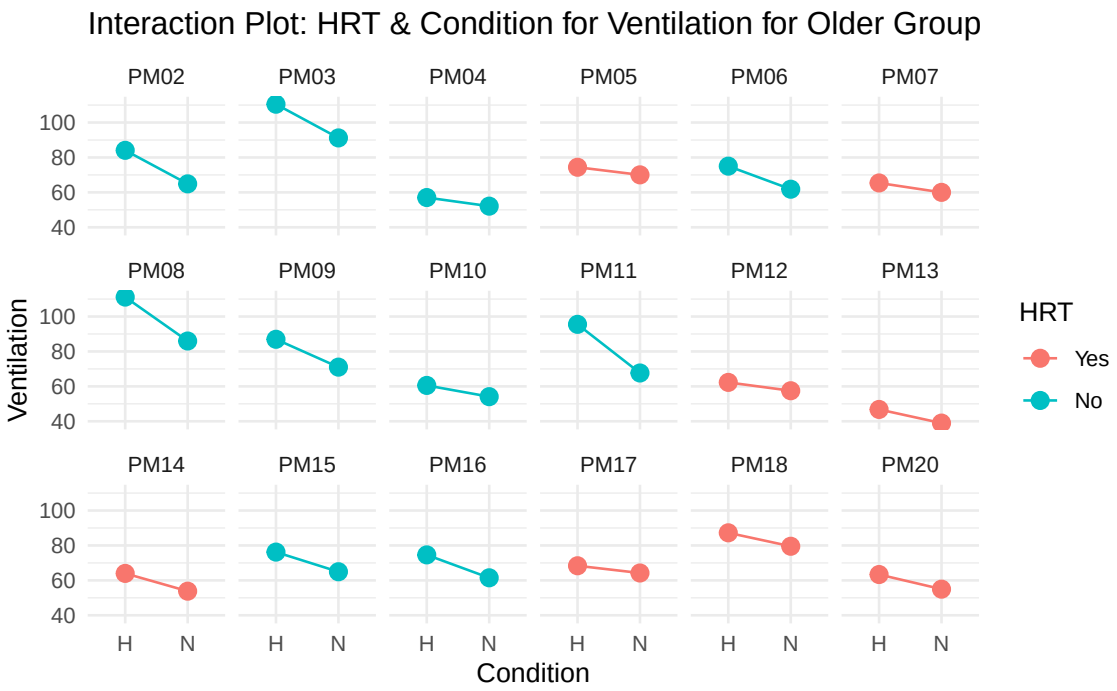

Figure A5: Interaction plot on HRT and condition for ventilation for older group

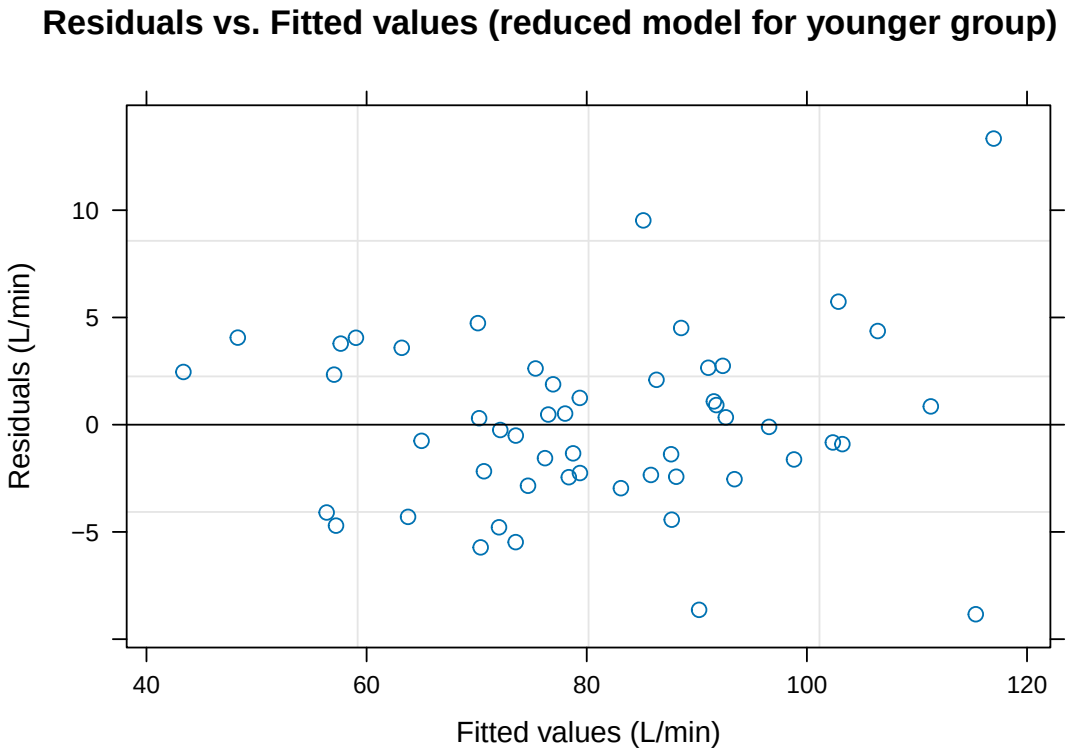

Figure A6: Plot of residuals against fitted values for younger group

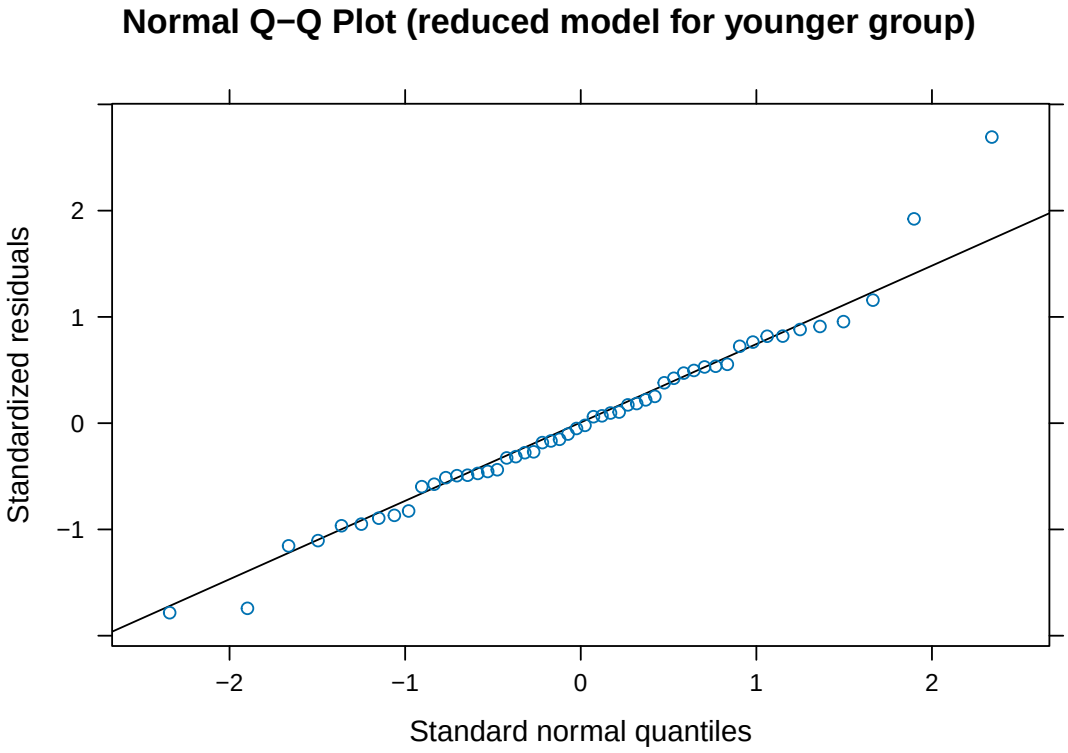

Figure A7: Normal Q-Q plot for younger group

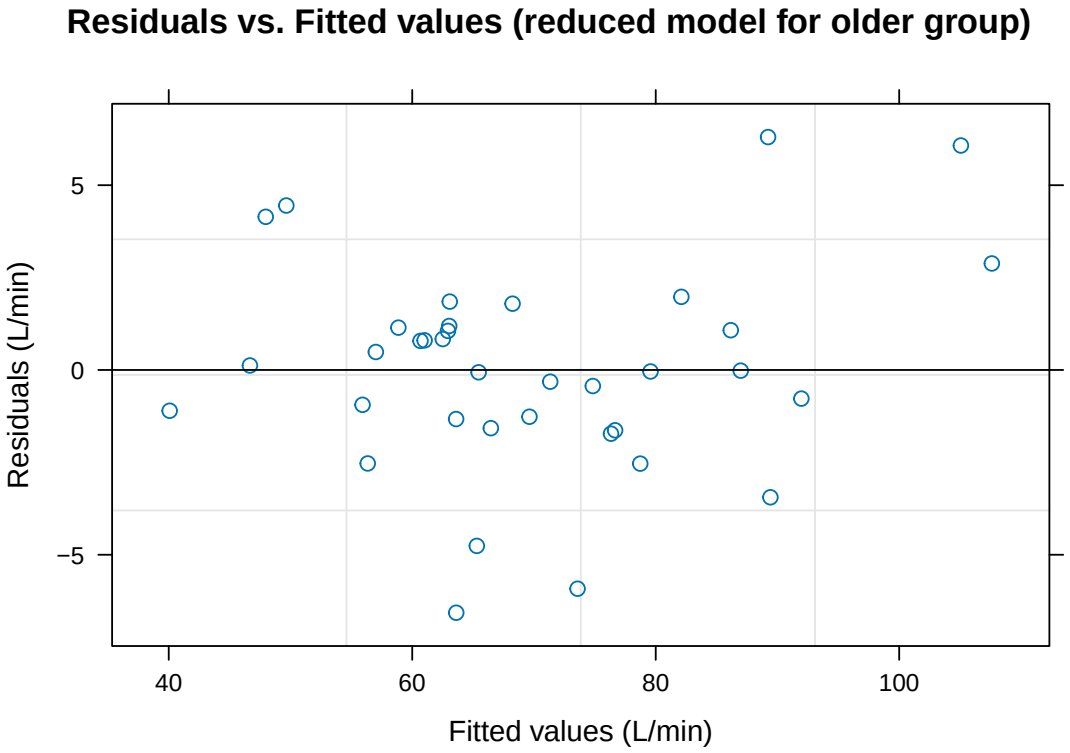

Figure A8: Plot of residuals against fitted values for older group

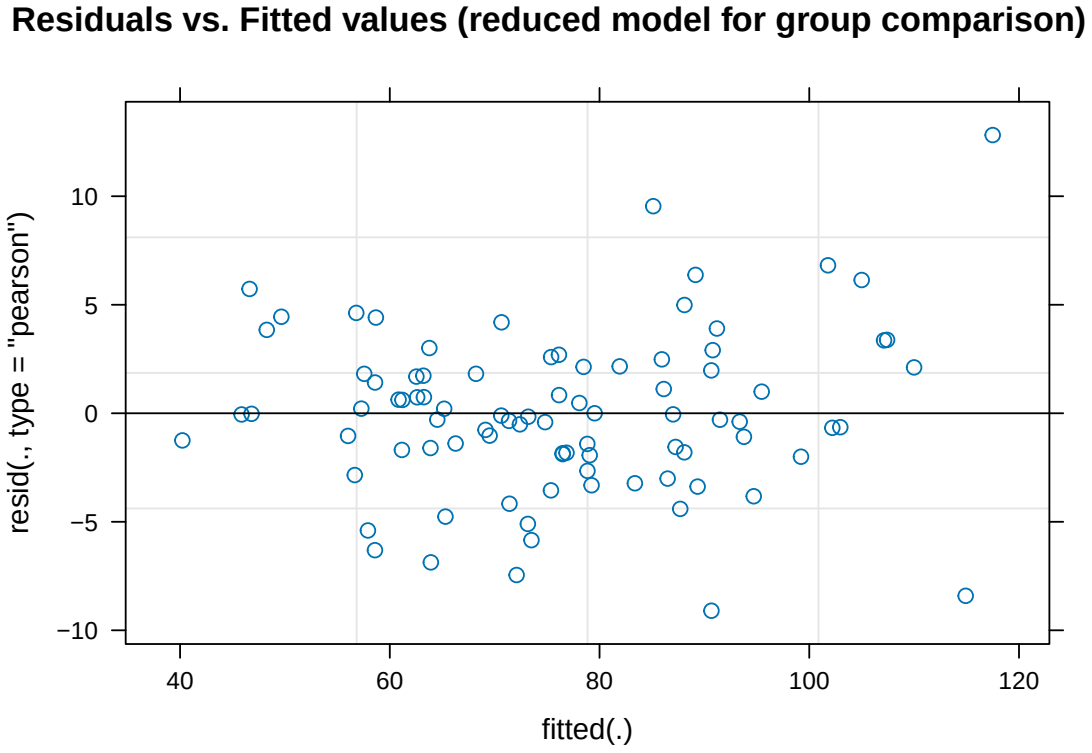

Figure A9: Plot of residuals against fitted values for group comparison model

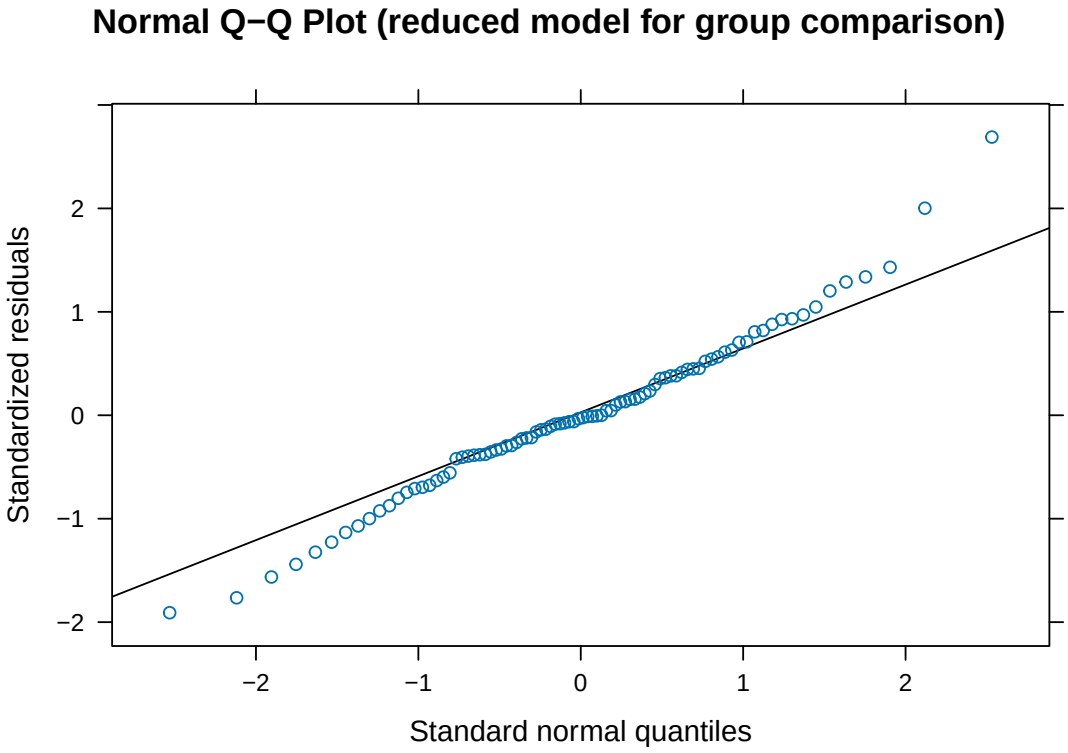

Figure A10: Normal Q-Q plot for group comparison model

Tables

Table A1: Summary statistics on key numeric variables for younger group

Progesterone	E2	Ventilation	age
Min. : 0.294	Min. : 86.7	Min. : 45.8	Min. :19.0
1st Qu.: 0.698	1st Qu.: 135.0	1st Qu.: 68.4	1st Qu.:22.0
Median : 7.000	Median : 240.5	Median : 78.7	Median :22.0
Mean :15.147	Mean : 327.6	Mean : 80.9	Mean :24.2
3rd Qu.:31.000	3rd Qu.: 465.0	3rd Qu.: 93.0	3rd Qu.:28.0
Max. :58.000	Max. :1040.0	Max. :130.3	Max. :29.0

Table A2: Summary statistics on key numeric variables for older group

Progesterone	E2	Ventilation	age
Min. : 0.0774	Min. : 1.81	Min. : 39.0	Min. :52.0
1st Qu.: 0.1714	1st Qu.: 17.30	1st Qu.: 60.4	1st Qu.:57.0
Median : 0.6703	Median : 55.01	Median : 65.2	Median :59.0
Mean : 1.8154	Mean : 93.66	Mean : 69.9	Mean :58.5
3rd Qu.: 0.9640	3rd Qu.:107.00	3rd Qu.: 77.0	3rd Qu.:60.0
Max. :19.0000	Max. :497.00	Max. :111.1	Max. :64.0

Table A3. Fixed-effects table for younger group

	Eliminated	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
age	1	0.16	0.16	1	6.08	0.01	0.94
height	2	1.59	1.59	1	6.97	0.06	0.81
weight	3	8.25	8.25	1	7.98	0.34	0.58
condition:`70_peak_power`	4	13.58	13.58	1	31.05	0.55	0.46
condition:progesterone:estradiol	5	45.51	45.51	1	32.05	1.88	0.18
condition:estradiol	6	15.59	15.59	1	33.05	0.63	0.43
IUD	7	57.14	57.14	1	8.93	2.33	0.16
measured_VO2max	8	48.56	48.56	1	9.91	1.98	0.19
`70_peak_power`	0	175.59	175.59	1	11.02	7.15	0.02
progesterone:estradiol	0	5.83	5.83	1	36.63	0.24	0.63
condition:progesterone	0	6.97	6.97	1	34.01	0.28	0.60

Table A4. Random-effects table for younger group

	Eliminated	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	NA	17	-183.5	401.1	NA	NA	NA
(1 | participant)	0	16	-205.9	443.8	44.7	1	2.31e-11

Table A5. Model summary for younger group

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-11.3	32.1	11.1	-0.352	0.731
conditionhypoxia	13	1.84	34	7.05	3.88e-08
progesterone	0.0484	0.0997	34.9	0.486	0.63
estradiol	-0.00373	0.00738	37.7	-0.506	0.616
`70_peak_power`	0.485	0.181	11	2.67	0.0216
progesterone:estradiol	9.24e-05	0.000189	36.6	0.487	0.629
conditionhypoxia:progesterone	0.0431	0.0808	34	0.533	0.597

Table A6. Model parameter confidence intervals for younger group

	2.5 %	97.5 %
sd_(Intercept)|participant	8.24208	18.68303
sigma	3.76576	5.89092
(Intercept)	-73.51155	50.77415
conditionhypoxia	9.51305	16.42845
progesterone	-0.13737	0.23737
estradiol	-0.01743	0.01032
`70_peak_power`	0.13460	0.83730
progesterone:estradiol	-0.00027	0.00044
conditionhypoxia:progesterone	-0.10877	0.19494

Table A7. Fixed-effects table for older group

	Eliminated	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
age	1	0.00	0.00	1	11	0.00	1.00
weight	2	1.59	1.59	1	12	0.10	0.76

	Eliminated	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
height	3	9.61	9.61	1	13	0.58	0.46
condition:70_peak_power	4	17.95	17.95	1	15	1.09	0.31
70_peak_power	5	18.65	18.65	1	14	1.12	0.31
measured_VO2max	0	93.36	93.36	1	15	5.62	0.03
condition:HRT	0	182.21	182.21	1	16	10.97	0.00

Table A8. Random-effects table for older group

	Eliminated	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	NA	12	-108.6	241.2	NA	NA	NA
(1 | participant)	0	11	-121.3	264.6	25.4	1	4.76e-07

Table A9. Model summary for older group

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	9.33	24.9	15	0.375	0.713
conditionhypoxia	15.6	1.82	16	8.58	2.2e-07
HRTYes	4.72	7.91	15.9	0.596	0.559
measured_VO2max	21.2	8.94	15	2.37	0.0316
conditionhypoxia:HRTYes	-9.06	2.73	16	-3.31	0.00441

Table A10. Model parameter confidence intervals for older group

	2.5 %	97.5 %
sd_(Intercept)|participant	7.829	15.864
sigma	2.862	5.544
(Intercept)	-37.639	56.298
conditionhypoxia	12.091	19.204
HRTYes	-10.200	19.638
measured_VO2max	4.306	38.098
conditionhypoxia:HRTYes	-14.390	-3.721

Table A11. Fixed-effects table for group comparison model

	Eliminated	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
measured_VO2max	1	0.07	0.07	1	22.78	0.00	0.96
height	2	0.26	0.26	1	23.93	0.01	0.91
weight	3	2.04	2.04	1	25.11	0.09	0.76
IUD	4	7.75	7.75	1	25.18	0.35	0.56
condition:70_peak_power	5	54.21	54.21	1	51.01	2.45	0.12
70_peak_power	0	204.49	204.49	1	27.13	9.00	0.01
condition:hormone_group:age_group	0	155.94	155.94	1	52.01	6.87	0.01

Table A12. Random-effects table for group comparison model

	Eliminated	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	NA	16	-280.2	592.4	NA	NA	NA
(1 | participant)	0	15	-322.0	674.1	83.7	1	5.74e-20

Table A13. Model summary for group comparison model

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	17.7	19	27.3	0.933	0.359
conditionhypoxia	12.7	1.87	52	6.8	1.02e-08
hormone_groupHigher	0.759	1.87	52	0.406	0.686
age_groupOlder	2.91	6.46	30.5	0.451	0.655
70_peak_power	0.319	0.106	27.1	3	0.00572
hormone_groupHigher:age_groupOlder	1.09	7.41	34.6	0.147	0.884
conditionhypoxia:hormone_groupHigher	1.81	2.64	52	0.687	0.495
conditionhypoxia:age_groupOlder	2.93	2.83	52	1.03	0.306
conditionhypoxia:hormone_groupHigher:age_groupOlder	-10.9	4.15	52	-2.62	0.0115

Table A14. Model parameter confidence intervals for group comparison model

	2.5 %	97.5 %
sd_(Intercept)|participant	9.242	15.716
sigma	3.828	5.534
(Intercept)	-18.132	53.676
conditionhypoxia	9.157	16.275
hormone_groupHigher	-2.799	4.318
age_groupOlder	-9.262	15.078
70_peak_power	0.118	0.519
hormone_groupHigher:age_groupOlder	-12.840	15.016
conditionhypoxia:hormone_groupHigher	-3.218	6.847

	2.5 %	97.5 %
conditionhypoxia:age_groupOlder	-2.466	8.328
conditionhypoxia:hormone_groupHigher:age_groupOlder	-18.767	-2.973

B R Code

Code used for exploratory data analysis:

```
# ---------------------- Read Data --------------------------
# identify excel file path and list all sheet names
file_path <- "Desktop/STAT450/Sierra_Arn_Data.xlsx"
sheet_names <- excel_sheets(file_path)
print(sheet_names)

# Loop through each sheet name
for (sheet in sheet_names) {
  # Create a data frame for the sheet
  df <- read_excel(file_path, sheet = sheet)

  # Dynamically assign the data frame to a variable named after sheet
  assign(sheet, df)
}

# Print the names of the data frames
print(ls())

# Assign shorter name and access each data frame
day1_young_rawdata <- `Day 1, young females`
day1_old_rawdata <- `Day 1, older females `
young_rawdata <- `Young Females (menstrual cycle)`
old_rawdata <- `Oder (PostMenopausal) Females `

# ------------------ Data Cleansing ---------------------------
# rename columns in younger female dataset
names(young_rawdata) <-
  c("Participant", "Phase", "Condition", "Progesterone",
    "E2", "Ventilation", "Working_VO2", "Working_Rel_VO2",
    "VCO2","VeVCO2","VeVO2","VT","fb","RER",
    "Percent_VO2_max","SpO2","Work_of_breathing(J/min)",
    "Work_of_breathing(J)","percent_ribcage",
    "percent abdomen", "PetCO2","70_peak_power","age",
    "height","weight","BMI","100_peak_power",
    "predicted_VO2max","Measured_VO2max",
    "percent_predicted_VO2max","Relative_VO2max","IUD")

# check data type
sapply(young_rawdata,class)

# impute "<1" Progesterone column
young_rawdata_impute <- young_rawdata %>%
  mutate(Progesterone = ifelse(Progesterone == "<1",
                               runif(n(), min = 0, max = 1),
                               as.numeric(Progesterone)))

# change datatype
young_rawdata_impute$Phase <- as.factor(young_rawdata_impute$Phase)
str(young_rawdata_impute$Phase)
```

```r
young_rawdata_impute$Condition <-
  as.factor(young_rawdata_impute$Condition)
str(young_rawdata_impute$Condition)
young_rawdata_impute$IUD <- as.factor(young_rawdata_impute$IUD)
str(young_rawdata_impute$IUD)

young_rawdata_impute$Progesterone <-
  as.numeric(young_rawdata_impute$Progesterone)
str(young_rawdata_impute$Progesterone)

# ------------------------------------------------------------------
# rename columns in older female dataset
names(old_rawdata) <-
  c("Participant", "HRT", "Condition", "Progesterone","E2",
    "Ventilation","Working_Rel_VO2","Working_VO2","VCO2",
    "VeVCO2","VeVO2","VT","fb","RER","Percent_VO2max",
    "SpO2","Work_of_breathing(J/min)","Work_of_breathing(J)",
    "PetCO2","70_peak_power","age","heigt","weight","BMI",
    "100_peak_power","Predicted_VO2max","Measured_VO2max",
    "percent_predicted_VO2max","Relative_VO2max")

# check data type
sapply(old_rawdata,class)

# impute "<1" in Progesterone column
old_rawdata_impute <- old_rawdata %>%
  mutate(Progesterone = ifelse(Progesterone == "<1",
                               runif(n(), min = 0, max = 1),
                               as.numeric(Progesterone)))

# impute "<90" in E2 column
old_rawdata_impute <- old_rawdata_impute %>%
  mutate(E2 = ifelse(E2 == "<90",
                     runif(n(), min = 0, max = 90),
                     as.numeric(E2)))
old_rawdata_impute$E2

# change datatype
old_rawdata_impute$HRT <- as.factor(old_rawdata_impute$HRT)
str(old_rawdata_impute$HRT)
old_rawdata_impute$Condition <-
  as.factor(old_rawdata_impute$Condition)
str(old_rawdata_impute$Condition)

old_rawdata_impute$Progesterone <-
  as.numeric(old_rawdata_impute$Progesterone)
str(old_rawdata_impute$Progesterone)
old_rawdata_impute$E2 <- as.numeric(old_rawdata_impute$E2)
str(old_rawdata_impute$E2)
old_rawdata_impute$`Work_of_breathing(J/min)` <-
  as.numeric(old_rawdata_impute$`Work_of_breathing(J/min)`)
str(old_rawdata_impute$`Work_of_breathing(J/min)`)
```

```r
old_rawdata_impute$`Work_of_breathing(J)` <-
  as.numeric(old_rawdata_impute$`Work_of_breathing(J)`)
str(old_rawdata_impute$`Work_of_breathing(J)`)

young_rawdata_clean <- young_rawdata_impute %>%
  select(where(is.numeric) | where(is.factor))

# Select only numeric columns
numeric_data_young <- young_rawdata_clean %>%
  select(where(is.numeric))


old_rawdata_clean <- old_rawdata_impute %>%
  select(where(is.numeric) | where(is.factor))

# Select only numeric columns
numeric_data_old <- old_rawdata_clean %>%
  select(where(is.numeric))

# ---------------------------- EDA -------------------------------
# Summary statistics
table1.1 <- numeric_data_young %>%
  select(Progesterone,E2,Ventilation,SpO2) %>% summary()
kable(table1.1)

table1.2 <- numeric_data_old %>%
  select(Progesterone,E2,Ventilation,SpO2) %>% summary()
kable(table1.2)

# Creating Boxplots
# Younger female
young_rawdata$PhaseCondition <-
  interaction(young_rawdata$Phase,
              young_rawdata$Condition, sep = ".")
# Reorder factor levels in the desired order
young_rawdata$PhaseCondition <-
  factor(young_rawdata$PhaseCondition,
         levels = c("1.2", "2.2", "1.1", "2.1"),
         labels = c("EF.Normoxia", "ML.Normoxia",
                    "EF.Hypoxia","ML.Hypoxia"))
# Boxplot for younger female
boxplot(Ventilation ~ PhaseCondition, data = young_rawdata,
        main = "Ventilation vs. Phase*Condition for younger female",
        #xlab = "Phase (1=EF, 2=ML) ; Condition (1=H, 2=N)",
        xlab = "Phase (EF, ML) & Condition (Hypoxia, Normoxia)",
        ylab = "Ventilation", border = "black",
        cex.main = 1.0,
        cex.axis = 0.8)

# Older female
old_rawdata$HRTCondition <- interaction(old_rawdata$HRT,
                                old_rawdata$Condition, sep = ".")
```

```r
# Reorder factor levels in the desired order
old_rawdata$HRTCondition <-
  factor(old_rawdata$HRTCondition,
         levels = c("2.2", "1.2", "2.1", "1.1"),
         labels = c("No.Normoxia", "Yes.Normoxia",
                    "No.Hypoxia","Yes.Hypoxia"))
# Boxplot for older female
boxplot(Ventilation ~ HRTCondition, data = old_rawdata,
        main = "Ventilation vs. HRT*Condition for older female",
        #xlab = "HRT (1=Yes, 2=No) ; Condition (1=H, 2=N)",
        xlab = "HRT (Yes, No) & Condition (Hypoxia, Normoxia)",
        ylab = "Ventilation", border = "black",
        cex.main = 1.0,
        cex.axis = 0.8)

# Creating correlation matrix
# Younger female
# Compute the correlation matrix
cor_matrix <- cor(numeric_data_young, use = "complete.obs")

# Display the correlation matrix
print(cor_matrix)

# Visualize the correlation matrix with a heatmap
library(corrplot)
corrplot(cor_matrix, method = "color", type = "upper",
         tl.cex = 0.7, tl.col = "blue")

# Older female
# Compute the correlation matrix
cor_matrix_old <- cor(numeric_data_old, use = "complete.obs")

# Display the correlation matrix
print(cor_matrix_old)

# Visualize the correlation matrix with a heatmap
library(corrplot)
corrplot(cor_matrix_old, method = "color", type = "upper",
         tl.cex = 0.7, tl.col = "black")

# Interaction plots showing random effects
# Younger female
ggplot(young_rawdata_impute,
  aes(x = Phase, y = Ventilation,
      color = Condition, group = Condition)) +
  # Lines for conditions
  geom_line(aes(group = Condition), position = position_dodge(0.2)) +
  # Points for conditions
  geom_point(position = position_dodge(0.2), size = 3) +
  # Facet by participant
  facet_wrap(~ Participant) +
  ggtitle("Phase & Condition for Ventilation for Younger Female") +
  xlab("Phase") +
```

```
  ylab("Ventilation") +
  theme_minimal() +
  scale_x_discrete(
    labels = c("1" = "EF", "2" = "ML"))+
  scale_color_discrete(
    name = "Condition",  # Change legend title
    labels = c("1" = "Hypoxia", "2" = "Normaxia")) #Rename categories

# Older female
ggplot(old_rawdata_impute,
       aes(x = Condition, y = Ventilation,
           color = HRT, group = HRT)) +
  # Lines for conditions
  geom_line(aes(group = HRT), position = position_dodge(0.2)) +
  # Points for conditions
  geom_point(position = position_dodge(0.2), size = 3) +
  # Facet by participant
  facet_wrap(~ Participant) +
  ggtitle("HRT & Condition for Ventilation for Older Female") +
  xlab("Condition") +
  ylab("Ventilation") +
  theme_minimal() +
  scale_x_discrete(
    labels = c("1" = "H", "2" = "N"))+
  scale_color_discrete(
    name = "HRT",  # Change legend title
    labels = c("1" = "Yes", "2" = "No") # Rename categories
  )
```

Code for the model fitting process with ventilation as the response variable:

```
set.seed(450) # for reproducibility of results

# Read and clean data, impute missing hormone measurements
data_young <- read_excel("data/Sierra_Arn_Data_Updated_5.xlsx",
                         "Young Females (menstrual cycle)",
                         na = "NA") |>
  dplyr::select(c(1:18, 21:32))
colnames(data_young) <-
  c("participant", "phase", "condition", "progesterone", "estradiol",
    "ventilation", "working_VO2", "relative_VO2", "VCO2", "Ve/VCO2", "Ve/VO2",
    "VT", "fb", "RER", "percent_VO2max", "SpO2", "work_breathing_min",
    "work_breathing", "PetCO2", "70_peak_power", "age", "height", "weight",
    "BMI", "100_peak_power", "predicted_VO2max", "measured_VO2max",
    "percent_predicted", "relative_VO2max", "IUD")
data_young <- data_young |>
  mutate(progesterone = as.numeric(ifelse(progesterone == "<1",
                                          rep(runif(n()/2, 0, 1), 2),
                                          progesterone)),
         estradiol = as.numeric(ifelse(estradiol == "<90",
                                       rep(runif(n()/2, 0, 90), 2),
                                       estradiol)),
         participant = as.factor(participant),
         phase = as.factor(ifelse(phase == 1, "EF", "ML")),
         condition = factor(ifelse(condition == 1, "hypoxia", "normoxia"),
                            levels = c("normoxia", "hypoxia")),
         IUD = as.factor(ifelse(IUD == "Y", "Yes", "No")))
data_older <- read_excel("data/Sierra_Arn_Data_Updated_5.xlsx",
                         "Oder (PostMenopausal) Females ",
                         na = "NA") |>
  dplyr::select(c(1:20, 22:30))
colnames(data_older) <-
  c("participant", "HRT", "condition", "progesterone", "estradiol",
    "ventilation", "working_VO2", "relative_VO2", "VCO2", "Ve/VCO2", "Ve/VO2",
    "VT", "fb", "RER", "percent_VO2max", "SpO2", "work_breathing_min",
    "work_breathing", "PetCO2", "70_peak_power", "age", "height", "weight",
    "BMI", "100_peak_power", "predicted_VO2max", "measured_VO2max",
    "percent_predicted", "relative_VO2max")
data_older <- data_older |>
  mutate(progesterone = as.numeric(ifelse(progesterone == "<1",
                                          rep(runif(n()/2, 0, 1), 2),
                                          progesterone)),
         estradiol = as.numeric(ifelse(estradiol == "<90",
                                       rep(runif(n()/2, 0, 90), 2),
                                       estradiol)),
         participant = as.factor(participant),
         HRT = as.factor(ifelse(HRT == 1, "Yes", "No")),
         condition = factor(ifelse(condition == 1, "hypoxia", "normoxia"),
                            levels = c("normoxia", "hypoxia")))

# Check VIF to detect multicollinearity
vif(lmer(ventilation ~ condition + progesterone + estradiol + age + height +
```

```
          weight + BMI + `70_peak_power` + `100_peak_power` + predicted_VO2max +
          measured_VO2max + percent_predicted + relative_VO2max + IUD +
          (1 | participant),
        data = data_young))
# Remove predicted_VO2max (calculated from height and age)
sqrt(vif(lmer(ventilation ~ condition + progesterone + estradiol + age + height +
               weight + BMI + `70_peak_power` + `100_peak_power` +
               measured_VO2max + percent_predicted + relative_VO2max + IUD +
               (1 | participant),
             data = data_young)))
# Remove BMI (calculated from height and weight)
sqrt(vif(lmer(ventilation ~ condition + progesterone + estradiol + age + height +
               weight + `70_peak_power` + `100_peak_power` + measured_VO2max +
               percent_predicted + relative_VO2max + IUD + (1 | participant),
             data = data_young)))
# Remove relative_VO2max (calculated from measured_VO2max and weight)
sqrt(vif(lmer(ventilation ~ condition + progesterone + estradiol + age + height +
               weight + `70_peak_power` + `100_peak_power` + measured_VO2max +
               percent_predicted + IUD + (1 | participant),
             data = data_young)))
# Remove percent_predicted (calculated from measured_VO2max, height, and age)
sqrt(vif(lmer(ventilation ~ condition + progesterone + estradiol + age + height +
               weight + `70_peak_power` + `100_peak_power` + measured_VO2max +
               IUD + (1 | participant),
             data = data_young)))
# Remove `100_peak_power` (ventilation was measured at 70% peak power)
sqrt(vif(lmer(ventilation ~ condition + progesterone + estradiol + age + height +
               weight + `70_peak_power` + measured_VO2max + IUD +
               (1 | participant),
             data = data_young)))
# all values under 5, try adding interactions
sqrt(vif(lmer(ventilation ~ condition*(progesterone*estradiol + `70_peak_power` +
                                       measured_VO2max) +
               age + height + weight + `70_peak_power` + measured_VO2max + IUD +
               (1 | participant),
             data = data_young)))
# Remove condition:measured_VO2max interaction
sqrt(vif(lmer(ventilation ~ condition*(progesterone*estradiol + `70_peak_power`) +
               age + height + weight + `70_peak_power` + measured_VO2max + IUD +
               (1 | participant),
             data = data_young)))
# some relatively high values remain but proceeding anyway because those model
# terms are of specific research interest

# Fit models and show summary tables
full_model_young <- lmer(ventilation ~ condition*(progesterone*estradiol +
                                                  `70_peak_power`) +
                           age + height + weight + `70_peak_power` +
                           measured_VO2max + IUD + (1 | participant),
                         data = data_young)
summary(full_model_young)
selection_young <- step(full_model_young,
                        keep = c("condition:progesterone",
```

```
                                    "progesterone:estradiol"))
selection_young
reduced_model_young <- get_model(selection_young)
summary(reduced_model_young)
confint(reduced_model_young, oldNames = FALSE)

# Model diagnostic plots
plot(reduced_model_young,
     main = "Residuals vs. Fitted values (model for younger group)",
     xlab = "Fitted values (L/min)", ylab = "Residuals (L/min)")
boxplot(resid(reduced_model_young) ~ data_young$condition,
        main = "Residuals vs. Condition (model for younger group)",
        xlab = "Oxygen condition", ylab = "Residuals (L/min)")
plot(reduced_model_young, resid(.) ~ progesterone,
     main = "Residuals vs. Progesterone (model for younger group)",
     xlab = "Progesterone (nmol/L)", ylab = "Residuals (L/min)")
plot(reduced_model_young, resid(.) ~ estradiol,
     main = "Residuals vs. Estradiol (model for younger group)",
     xlab = "Estradiol (pmol/L)", ylab = "Residuals (L/min)")
qqmath(reduced_model_young, main = "Normal Q-Q Plot (model for younger group)")

# Check VIF for older group
vif(lmer(ventilation ~ condition + HRT + age + height + weight + BMI +
           `70_peak_power` + `100_peak_power` + predicted_VO2max +
           measured_VO2max + percent_predicted + relative_VO2max +
           (1 | participant),
         data = data_older))
# Remove predicted_VO2max (calculated from height and age)
sqrt(vif(lmer(ventilation ~ condition + HRT + age + height + weight + BMI +
                `70_peak_power` + `100_peak_power` + measured_VO2max +
                percent_predicted + relative_VO2max + (1 | participant),
              data = data_older)))
# Remove BMI (calculated from height and weight)
sqrt(vif(lmer(ventilation ~ condition + HRT + age + height + weight +
                `70_peak_power` + `100_peak_power` + measured_VO2max +
                percent_predicted + relative_VO2max + (1 | participant),
              data = data_older)))
# Remove percent_predicted (calculated from measured_VO2max, height, and age)
sqrt(vif(lmer(ventilation ~ condition + HRT + age + height + weight +
                `70_peak_power` + `100_peak_power` + measured_VO2max +
                relative_VO2max + (1 | participant),
              data = data_older)))
# Remove `100_peak_power` (ventilation was measured at 70% peak power)
sqrt(vif(lmer(ventilation ~ condition + HRT + age + height + weight +
                `70_peak_power` + measured_VO2max + relative_VO2max +
                (1 | participant),
              data = data_older)))
# Remove relative_VO2max (calculated from measured_VO2max and weight)
sqrt(vif(lmer(ventilation ~ condition + HRT + age + height + weight +
                `70_peak_power` + measured_VO2max + (1 | participant),
              data = data_older)))
# all values under 5, try adding interactions
```

```r
sqrt(vif(lmer(ventilation ~ condition*(HRT + `70_peak_power` + measured_VO2max) +
                age + height + weight + `70_peak_power` + measured_VO2max +
                (1 | participant),
              data = data_older)))
# Remove condition:measured_VO2max interaction
sqrt(vif(lmer(ventilation ~ condition*(HRT + `70_peak_power`) + age + height +
                weight + `70_peak_power` + measured_VO2max + (1 | participant),
              data = data_older)))
# proceeding, multicollinearity concerns are mostly alleviated

# Fit models and show summary tables
full_model_older <- lmer(ventilation ~ condition*(HRT + `70_peak_power`) + age +
                           height + weight + `70_peak_power` + measured_VO2max +
                           (1 | participant),
                         data = data_older)
summary(full_model_older)
selection_older <- step(full_model_older,
                        keep = c("condition:HRT"))
selection_older
reduced_model_older <- get_model(selection_older)
summary(reduced_model_older)
confint(reduced_model_older, oldNames = FALSE)

# Model diagnostic plots
plot(reduced_model_older,
     main = "Residuals vs. Fitted values (model for older group)",
     xlab = "Fitted values (L/min)", ylab = "Residuals (L/min)")
boxplot(resid(reduced_model_older) ~ data_older$condition,
        main = "Residuals vs. Condition (model for older group)",
        xlab = "Oxygen condition", ylab = "Residuals (L/min)")
boxplot(resid(reduced_model_older) ~ data_older$HRT,
        main = "Residuals vs. HRT (model for older group)",
        xlab = "HRT", ylab = "Residuals (L/min)")
qqmath(reduced_model_older, main = "Normal Q-Q Plot (model for older group)")

# Creating plot of predicted means for older group
summary(data_older$measured_VO2max)
x_values <- seq(1.5, 3.2, 0.1)
n_values <- length(x_values)
estimates <- summary(reduced_model_older)$coefficients[,1]
predicted <- tibble(condition = factor(c(rep("normoxia", n_values*2),
                                         rep("hypoxia", n_values*2)),
                                       levels = c("normoxia", "hypoxia")),
                    HRT = c(rep(c(rep("No", n_values),
                                  rep("Yes", n_values)),
                                2)),
                    measured_VO2max = rep(x_values, 4),
                    ventilation = estimates["(Intercept)"] +
                      ifelse(condition == "hypoxia",
                             estimates["conditionhypoxia"],
                             0) +
                      ifelse(HRT == "Yes",
```

```
                              ifelse(condition == "hypoxia",
                                     estimates["conditionhypoxia:HRTYes"] +
                                       estimates["HRTYes"],
                                     estimates["HRTYes"]),
                              0) +
                       estimates["measured_VO2max"]*measured_VO2max)
ggplot(predicted,
       aes(x = measured_VO2max,
           y = ventilation,
           colour = factor(
             str_to_lower(interaction(condition, HRT)) |>
               str_replace_all(fixed("."), ", "),
             levels = c("normoxia, no", "hypoxia, no",
                        "normoxia, yes", "hypoxia, yes")))) +
  geom_line(size = 1) +
  scale_colour_manual(values = c("#ff9933", "#cc0033", "#00cccc", "#0033ff")) +
  geom_point(data = data_older, size = 3) +
  #scale_x_continuous(breaks = seq(60, 180, 20)) +
  labs(x = "Measured VO2max (L/min)", y = "Ventilation (L/min)",
       colour = "Oxygen condition,\npresence of HRT") +
  theme(text = element_text(size = 20))
ggsave("docs/older_predicted_means.png", width = 11, height = 6)
```

Code for fitting model to combined dataset:

```
set.seed(450) # for reproducibility of results

# Read and clean data, impute missing hormone measurements
data_young <- read_excel("data/Sierra_Arn_Data_Updated_5.xlsx",
                         "Young Females (menstrual cycle)",
                         na = "NA") |>
  dplyr::select(c(1:18, 21:32))
colnames(data_young) <-
  c("participant", "phase", "condition", "progesterone", "estradiol",
    "ventilation", "working_VO2", "relative_VO2", "VCO2", "Ve/VCO2", "Ve/VO2",
    "VT", "fb", "RER", "percent_VO2max", "SpO2", "work_breathing_min",
    "work_breathing", "PetCO2", "70_peak_power", "age", "height", "weight",
    "BMI", "100_peak_power", "predicted_VO2max", "measured_VO2max",
    "percent_predicted", "relative_VO2max", "IUD")
data_young <- mutate(data_young,
                     progesterone = as.numeric(ifelse(progesterone == "<1",
                                                      rep(runif(n()/2, 0, 1), 2),
                                                      progesterone)),
                     estradiol = as.numeric(ifelse(estradiol == "<90",
                                                   rep(runif(n()/2, 0, 90), 2),
                                                   estradiol)))
data_older <- read_excel("data/Sierra_Arn_Data_Updated_5.xlsx",
                         "Oder (PostMenopausal) Females ",
                         na = "NA") |>
  dplyr::select(c(1:20, 22:30))
colnames(data_older) <-
  c("participant", "HRT", "condition", "progesterone", "estradiol",
    "ventilation", "working_VO2", "relative_VO2", "VCO2", "Ve/VCO2", "Ve/VO2",
    "VT", "fb", "RER", "percent_VO2max", "SpO2", "work_breathing_min",
    "work_breathing", "PetCO2", "70_peak_power", "age", "height", "weight",
    "BMI", "100_peak_power", "predicted_VO2max", "measured_VO2max",
    "percent_predicted", "relative_VO2max")
data_older <- mutate(data_older,
                     progesterone = as.numeric(ifelse(progesterone == "<1",
                                                      rep(runif(n()/2, 0, 1), 2),
                                                      progesterone)),
                     estradiol = as.numeric(ifelse(estradiol == "<90",
                                                   rep(runif(n()/2, 0, 90), 2),
                                                   estradiol)))

# Combine datasets
data_combined <- bind_rows(data_young, data_older) |>
  mutate(participant = as.factor(participant),
         age_group = factor(ifelse(is.na(phase), "Older", "Younger"),
                            levels = c("Younger", "Older")),
         hormone_group = factor(ifelse(is.na(phase),
                                       ifelse(HRT == 1, "Higher", "Lower"),
                                       ifelse(phase == 1, "Lower", "Higher")),
                                levels = c("Lower", "Higher")),
         condition = factor(ifelse(condition == 1, "hypoxia", "normoxia"),
                            levels = c("normoxia", "hypoxia")),
```

```
         IUD = as.factor(ifelse(is.na(IUD),
                                "No",
                                ifelse(IUD == "Y", "Yes", "No"))),
         .keep = "unused")

# Fit models and show summary tables
# (removed the same variables as in VIF checks for other models)
full_model_combined <- lmer(ventilation ~ condition*(hormone_group*age_group +
                                                       `70_peak_power`) +
                              hormone_group + age_group + height + weight +
                              `70_peak_power` + measured_VO2max + IUD +
                              (1 | participant),
                            data = data_combined)
summary(full_model_combined)
selection_combined <- step(full_model_combined)
selection_combined
reduced_model_combined <- get_model(selection_combined)
summary(reduced_model_combined)
confint(reduced_model_combined, oldNames = FALSE)

# Model diagnostic plots
plot(reduced_model_combined,
     main = "Residuals vs. Fitted values (model for group comparison)",
     xlab = "Fitted values (L/min)", ylab = "Residuals (L/min)")
boxplot(resid(reduced_model_combined) ~ data_combined$condition,
        main = "Residuals vs. Condition (model for group comparison)",
        xlab = "Oxygen condition", ylab = "Residuals (L/min)")
boxplot(resid(reduced_model_combined) ~
          data_combined$age_group*data_combined$hormone_group,
        main = "Residuals vs. Group (model for group comparison)",
        xlab = "Grouped age and hormone levels", ylab = "Residuals (L/min)")
qqmath(reduced_model_combined,
       main = "Normal Q-Q Plot (model for group comparison)")
```

Code for fitting models with other response variables:

```
set.seed(450) # for reproducibility of results

# Read and clean data, impute missing hormone measurements
data_young <- read_excel("data/Sierra_Arn_Data_Updated_5.xlsx",
                         "Young Females (menstrual cycle)",
                         na = "NA") |>
  dplyr::select(c(1:18, 21:32))
colnames(data_young) <-
  c("participant", "phase", "condition", "progesterone", "estradiol",
    "ventilation", "working_VO2", "relative_VO2", "VCO2", "Ve/VCO2", "Ve/VO2",
    "VT", "fb", "RER", "percent_VO2max", "SpO2", "work_breathing_min",
    "work_breathing", "PetCO2", "70_peak_power", "age", "height", "weight",
    "BMI", "100_peak_power", "predicted_VO2max", "measured_VO2max",
    "percent_predicted", "relative_VO2max", "IUD")
data_young <- data_young |>
  mutate(progesterone = as.numeric(ifelse(progesterone == "<1",
                                          rep(runif(n()/2, 0, 1), 2),
                                          progesterone)),
         estradiol = as.numeric(ifelse(estradiol == "<90",
                                       rep(runif(n()/2, 0, 90), 2),
                                       estradiol)),
         participant = as.factor(participant),
         phase = as.factor(ifelse(phase == 1, "EF", "ML")),
         condition = factor(ifelse(condition == 1, "hypoxia", "normoxia"),
                            levels = c("normoxia", "hypoxia")),
         IUD = as.factor(ifelse(IUD == "Y", "Yes", "No")))

# Function to fit full model for a given response variable
fit_full_model <- function(response) {
  lmer(response ~ condition*(progesterone*estradiol + `70_peak_power` +
                               measured_VO2max) + age + height + weight + IUD +
         (1 | participant),
       data = data_young)
}

# SpO2
full_model_SpO2 <- fit_full_model(data_young$SpO2)
summary(full_model_SpO2)
selection_SpO2 <- step(full_model_SpO2)
selection_SpO2
reduced_model_SpO2 <- get_model(selection_SpO2)
summary(reduced_model_SpO2)
confint(reduced_model_SpO2, oldNames = FALSE)

# selection removed the random effect term so we use different plotting methods
plot(resid(reduced_model_SpO2) ~ fitted.values(reduced_model_SpO2),
     main = "Residuals vs. Fitted values (model for SpO2)",
     xlab = "Fitted values (% oxygen saturation)",
     ylab = "Residuals (% oxygen saturation)")
boxplot(resid(reduced_model_SpO2) ~ data_young$condition,
        main = "Residuals vs. Condition (model for SpO2)",
```

```
        xlab = "Oxygen condition", ylab = "Residuals (% oxygen saturation)")
plot(resid(reduced_model_SpO2) ~ data_young$progesterone,
     main = "Residuals vs. Progesterone (model for SpO2)",
     xlab = "Progesterone (nmol/L)", ylab = "Residuals (% oxygen saturation)")
plot(resid(reduced_model_SpO2) ~ data_young$estradiol,
     main = "Residuals vs. Estradiol (model for SpO2)",
     xlab = "Estradiol (pmol/L)", ylab = "Residuals (% oxygen saturation)")
plot(reduced_model_SpO2, 2)

# Work of Breathing (J/min)
# Removing rows with missing values
data_young <- drop_na(data_young, work_breathing_min)
full_model_work_min <- fit_full_model(data_young$work_breathing_min)
summary(full_model_work_min)
selection_work_min <- step(full_model_work_min)
selection_work_min
reduced_model_work_min <- get_model(selection_work_min)
summary(reduced_model_work_min)
confint(reduced_model_work_min, oldNames = FALSE)

plot(reduced_model_work_min,
     main = "Residuals vs. Fitted values (model for work of breathing)",
     xlab = "Fitted values", ylab = "Residuals")
boxplot(resid(reduced_model_work_min) ~ data_young$condition,
        main = "Residuals vs. Condition (model for work of breathing)",
        xlab = "Oxygen condition", ylab = "Residuals")
plot(reduced_model_work_min, resid(.) ~ progesterone,
     main = "Residuals vs. Progesterone (model for work of breathing)",
     xlab = "Progesterone (nmol/L)", ylab = "Residuals")
plot(reduced_model_work_min, resid(.) ~ estradiol,
     main = "Residuals vs. Estradiol (model for work of breathing)",
     xlab = "Estradiol (pmol/L)", ylab = "Residuals")
qqmath(reduced_model_work_min,
       main = "Normal Q-Q Plot (model for work of breathing)")
```

Code for analysis and determining adequacy of imputation method:

```
set.seed(450) # for reproducibility of results

# Read and clean data
data_young <- read_excel("data/Sierra_Arn_Data_Updated_5.xlsx",
                         "Young Females (menstrual cycle)",
                         na = "NA") |>
  dplyr::select(c(1:18, 21:32))
colnames(data_young) <-
  c("participant", "phase", "condition", "progesterone", "estradiol",
    "ventilation", "working_VO2", "relative_VO2", "VCO2", "Ve/VCO2", "Ve/VO2",
    "VT", "fb", "RER", "percent_VO2max", "SpO2", "work_breathing_min",
    "work_breathing", "PetCO2", "70_peak_power", "age", "height", "weight",
    "BMI", "100_peak_power", "predicted_VO2max", "measured_VO2max",
    "percent_predicted", "relative_VO2max", "IUD")
data_young <- data_young |>
  mutate(progesterone = as.numeric(ifelse(progesterone == "<1", 0, progesterone)),
         estradiol = as.numeric(ifelse(estradiol == "<90", 0, estradiol)),
         participant = as.factor(participant),
         phase = as.factor(ifelse(phase == 1, "EF", "ML")),
         condition = factor(ifelse(condition == 1, "hypoxia", "normoxia"),
                            levels = c("normoxia", "hypoxia")),
         IUD = as.factor(ifelse(IUD == "Y", "Yes", "No")))

# Generate 1000 imputed datasets
data_young_imputed <- list()
for (i in 1:1000) {
  data_young_imputed[[i]] <- mutate(data_young,
                                    progesterone =
                                      ifelse(progesterone == 0,
                                             rep(runif(n()/2, 0, 1), 2),
                                             progesterone),
                                    estradiol =
                                      ifelse(estradiol == 0,
                                             rep(runif(n()/2, 0, 90), 2),
                                             estradiol))
}

# Fit final model to each dataset and summarize p-values
models <- lapply(data_young_imputed,
                 \(x) lmer(ventilation ~
                             condition*(progesterone + `70_peak_power`) +
                             progesterone*estradiol + (1 | participant),
                           data = x))
results <- lapply(models, \(x) as.data.frame(t(lmerTest:::get_coefmat(x)))[5,]) |>
  bind_rows()
summary(results)
```