Integrating AI for Predicting Stress Outcomes in Chiropractic Research

Name: Mannat Ahuja

Student ID: 22173443

Supervisor's Name: Dr. Hammed Fatoyinbo

Industry Supervisor: Dr. Imran Niazi & Dr. Imran Amjad

Organization: New Zealand College of Chiropractic

Paper Code: STAT995

Integrating AI for Predicting Stress Outcomes in Chiropractic Research

Abstract

This study examined the effects of chiropractic care on stress, using both physiological (cortisol) and psychological (DASS-21) measures in adults and children over 12 weeks. Linear mixed-effects models showed that baseline cortisol and DASS scores were stronger predictors of stress outcomes than subluxation scores or treatment groups, which was supported by the results from machine learning models (Random Forest, Gradient Boosting, and SVR). While cortisol levels increased post-intervention, DASS scores declined, suggesting differing physiological and psychological responses. However, the results for the children's data were inconclusive. Overall results suggested that cortisol response may be more biologically driven and not easily altered through short-term chiropractic care, highlighting the need for longer-term studies and broader biomarker inclusion.

Contents

Section 1: Introduction

This research investigates the predictive relationship between vertebral subluxation, balance, and stress using artificial intelligence (AI) and machine learning. This study aims to enhance chiropractic care strategies through advanced data-driven insights. It is conducted by the Centre for Chiropractic Research at the New Zealand College of Chiropractic, which is recognised internationally for its contributions to chiropractic science. This interdisciplinary study uses AI-based bio-signal and neuromuscular data analytics to explore how chiropractic adjustments influence brain function, biomechanics, and overall health. The study integrates pre-collected longitudinal data, including postural sway and gait characteristics (IMU data), stress indicators (DASS scores, salivary and hair cortisol levels), and chiropractic intervention timestamps collected over 12 weeks.

Section 2: Literature Review

Stress is a widespread public health concern that affects a person's physiological, psychological, and neurological well-being. Chronic stress has been directly linked to adverse health outcomes, elevated cortisol, autonomic nervous system imbalances, and muscle tension.

Random Forest has been widely used in stress prediction studies due to its ability to handle high-dimensional data and avoid overfitting. It performs well with physiological and numerical inputs like EEG signals and heart rate. It is also favoured for its balance between predictive performance and interpretability (Mentis et al., 2024). Hadi et al. (2019) demonstrated that the Random Forest classifier, when combined with feature selection and oversampling, achieved superior performance in predicting automobile drivers' stress levels. Among several classifiers tested, RF provided the most stable and accurate metrics. Meanwhile, Ahuja & Banga (2019) proposed machine-learning methods to predict stress, using RF, Linear Regression, and SVMs, wherein SVMs provided the highest accuracy. Ding et al. (2023) use tree-based models like RF and Gradient Boosting Machine for stress prediction, which showed high accuracy. They proposed a hybrid model combining RF and GB, which performed better than other models. These findings support the suitability of LR, RF, GB, and SVR for stress prediction tasks, especially when enhanced with appropriate data. Tree-based models and hybrid approaches therefore offer superior performance for stress detection, with careful feature selection being essential to improving accuracy.

There are several biomarkers for stress, salivary cortisol being one. While salivary cortisol measurement is relatively more straightforward, non-invasive, and useful for large-scale research, it does not consider hormonal changes that may affect cortisol levels, as Hellhammer et al. (2009) stated. The study concludes that the relationship between self-reported stress and salivary cortisol is somewhat weak due to complex neurobiological mechanisms of stress regulation. Similar results were demonstrated by Steudte et al. (2011), where the relationship between anxiety and salivary cortisol levels was inconclusive. Interestingly, a study by Takai et al. (2004) concluded that salivary amylases are more responsive as a biomarker for psychological stress than salivary cortisol. The results regarding salivary cortisol were inconclusive due to the shorter scale of the study and the limited time frame.

Measuring hair cortisol as a biomarker for stress is a more recent approach and arguably a better approach than measuring salivary cortisol. Several studies show mixed results. Some suggest that low hair cortisol levels are linked with anxiety disorder (Steudte et al., 2011). However, the study does not consider comorbidities like depression and has a small sample size. Other studies suggest that hair cortisol concentration has a positive relationship with work stress levels (Herr et al., 2018) but consider a sample of only men from one company, thus reducing generalizability to women and other occupations.

DASS (Depression Anxiety Stress Scales) have been proven to have excellent psychometric properties. Brown et al. (1996) strongly support and validate the psychometric abilities of DASS Scores in justifying its use in clinical anxiety and mood disorder samples, concluding that further research is required to assess its sensitivity to

clinical changes. While previous research supports the longer DASS-42 studies, studies on DASS-21 remain limited. While a study by Sinclair et al. (2012) confirms the internal consistency of DASS-21, potential biases and a lack of diversity in the sample due to online data collection raised concerns about the ability of DASS-21 to differentiate between stress, anxiety, and depression.

Chiropractic care is a healthcare practice wherein the basic idea is that every organ and cell in the body is controlled by the nervous system. It aims to diagnose and treat musculoskeletal disorders, particularly those related to the spine, by aligning the spine, optimising nerve flow, and contributing towards overall health and wellness by improving the nervous system. Several studies found that chiropractic care improved heart rate variability (Zhang et al., 2006) and chronic low back pain (Haas et al., 2004). However, the extent to which chiropractic care can improve these conditions is yet to be determined for long-term visits. Several other studies show mixed results when it comes to chiropractic care.

Rubinstein et al. (2007) examine both the positives and negatives of chiropractic care, stating that while adverse events and worsening of neck pain were common initially, 48% of patients recovered after the fourth visit, and 65% showed continued improvement after 3 months, implying that benefits outweigh the risks. The literature on chiropractic care and its effect on cortisol levels is limited. Chiropractic care has been proven to show improvements in neck pain, back pain, and headaches. At the same time, further studies need to be conducted to analyse its effect on other conditions like menstrual pain, hypertension, and lung disease (Kaptchuk et al., 1998). Tuchin (1998) stated a non-

significant relationship between salivary cortisol levels and chiropractic care. The results from the study were inconclusive due to a high drop-out rate (50%) and loss of interest due to the study's prolonged duration, concluding that future studies should account for such limitations for the validity of results.

Chiropractic adjustments are thought to affect the autonomic nervous system (ANS) specifically by enhancing parasympathetic activity (rest and digest) and reducing sympathetic dominance (fight or flight response). Brockman (2004) examines whether chiropractic intervention (active or sham) affects psychological well-being and whether any observed effects can be attributed to manipulation itself. While initially, both groups showed improvements in depression, fatigue, and tension, thus supporting the above claim, there was no significant difference between the two groups, implying that factors other than manipulation contributed to the improvements. The study was limited to only 20 samples, thus limiting generalizability. While existing research suggests that chiropractic care may influence physiological and psychological health outcomes, its direct impact on stress biomarkers remains unclear.

Studies on salivary cortisol and chiropractic care have yielded inconclusive results, and hair cortisol as a long-term stress marker has not been widely explored in this context. Some studies do indicate that chiropractic adjustments may enhance autonomic nervous system function. However, it remains unclear whether these improvements are due to manipulation itself or other factors such as placebo effects. Addressing these gaps by incorporating long-term stress biomarkers and validated psychological stress measures

(such as DASS) and using models like RF, GB, and SVR will provide a clearer understanding of the effect of chiropractic care on stress biomarkers.

Section 3: Objectives

This research's primary focus would be to answer the questions, "What is the predictive relationship between spinal dysfunction and stress indicators (DASS scores, salivary cortisol, and hair cortisol levels)?" and "Does chiropractic care influence stress levels over time, and if so, how?"

The scope of the research would be to answer the above questions using pre-collected data. This study will not analyse balance outcomes. The analysis will be performed on existing data sets; no additional data will be used. Additionally, this study would not investigate conditions unrelated to stress (for example, pain relief, posture, etc), nor would it investigate any suggestions or remedies for managing these conditions.

Section 4: Methodology

This study focuses on longitudinal observation design with a predictive modelling approach, using artificial intelligence (AI) and machine learning techniques to examine the relationship between spinal dysfunction and stress outcomes. This study utilises pre-collected data to assess whether chiropractic care influences stress indicators over time and whether baseline spinal dysfunction can predict treatment response.

Section 4.1: Participants and Data Collection

Participants were observed over a period of 12 weeks and data was collected at multiple points:

1. **Baseline (Week 0)** – Before any intervention
2. **Midway (Week 6)** – During chiropractic care
3. **Post-Intervention (Week 12)** – After chiropractic care
4. **Follow-Up (Week 16)** – To assess long-term effects (for the Active group only)

Participants were assigned to two groups: the Intervention Group, which received chiropractic care for over 12 weeks, and the Control Group, which did not receive chiropractic care. Spinal dysfunction data was collected for all participants regardless of group assignment.

Section 4.2: Data Source and Variables

1. **Spinal Dysfunction Data:** Clinical subluxation records documented by chiropractic clinicians that require extraction, preprocessing and preparation for analysis.
2. **Stress Outcome Measures:** Depression, Anxiety, and Stress Scale (DASS) scores, which were self-reported, and Physiological stress markers (salivary cortisol levels and hair cortisol levels) collected at key time points.
3. **Intervention Timing Data:** Chiropractic care timestamps for each participant to allow for longitudinal comparisons.

Section 4.3: Preprocessing

Data cleaning and structuring are performed to prepare the spinal dysfunction and stress datasets for analysis with the help of Feature Engineering. This includes handling missing values, encoding categorical variables, extracting features to calculate session-wise subluxation scores, calculating weighted scores, and aggregating or transforming stress indicators for meaningful features.

Section 4.4: Methods for Exploratory Analysis

A. **Descriptive Statistics:** Calculate the key descriptive statistics (mean, median, maximum, and minimum) for the stress variables and visualise them using boxplots.

B. Correlation Analysis: We will use Pearson's correlation analysis to assess the relationship between the stress variables and intervention scores at key time points. It ranges from **-1 to +1**:

+1 = perfect increasing relationship

-1 = perfect decreasing relationship

0 = no relationship

Section 4.5: Methods for Predictive Modelling

We will use the following methods for predicting post-intervention stress outcomes using subluxation scores from various time-points and stress variables:

A. **Random Forest Regression:** Random Forest Regression is a machine learning technique used for the prediction of continuous variables. It is based on the Random Forest algorithm, which builds a group of decision trees and averages their outputs to make a prediction. The regression involves data sampling, which includes training a dataset which is randomly sampled with replacement, training a decision tree on each subset, and making predictions using each tree, the final output of which is the average of all the predictions.

B. **Gradient Boosting:** Gradient Boosting is a machine learning technique used for both regression and classification problems. It builds a predictive model in an iterative manner by combining weak learners (usually decision trees) into a strong ensemble model.

It is known for its high prediction accuracy and its ability to handle complex data, although if not tuned properly, it may lead to overfitting.

C. **Support Vector Regression:** Support Vector Regression (SVR) is a type of Support Vector Machine (SVM) used for predicting continuous outcomes. While SVMs are commonly used for classification, SVR adapts the same principles for regression tasks. SVR tries to fit the best line or hyperplane that stays within a margin from the actual data points. Errors within this margin are not penalised. It aims to minimise the model complexity while allowing for some deviations.

Section 4.6: Methods for Model Evaluation

A. R-Squared: R-Squared measures how well the model explains the variability of the target variable. It ranges from 0 to 1. 0 implies that the model does not explain any variability, and 1 implies that the model perfectly explains the variability.

$$R^2 = 1 - \frac{Sum\ of\ Squared\ Residuals\ (SSR)}{Total\ Sum\ of\ Squares\ (TSS)}, \quad \textbf{(1)}$$

where, SSR = ∑(actual values – predicted values)^2 and TSS = ∑(actual values – mean(actual values))^2.

i. *Train R-Squared:* It is a metric that measures how well our model explains the variability of the target variable on the training dataset.

ii. *Test R-Squared:* It is a metric that measures how well our model explains the variability of the target variable on the test dataset.

iii. *Marginal R-Squared:* It measures the proportion of variance explained by the fixed effects.

iv. *Conditional R-Squared: Measures the proportion of variance explained by both fixed and random effects*.

B. Root Mean Squared Error (RMSE): RMSE is a standard way to measure the error of a regression model in predicting quantitative data. It tells us how far the

predictions are, on average, from the actual values, in the same unit as the target variable.

A low RMSE is generally considered better.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad \textbf{(2)}$$

where, y_i = actual values, $\hat{y}_i$ = predicted values, *and* n = number of observations.

ID	Enrolment Date	Age Group	Initial Care Plan	Time/Session 1	Vertebral Subluxation Before Care	Chiro Adjustment
39	19/4/2022	Adults	3x4	10:30am 19/4	PL Sx +++, L5 BL +++, T12 BR ++, T6 BR +++, C2 BR +++, C1 PIL ++	PL Sx, T6 BR, C2 BR, C1 PIL
76	19/4	Adults	3x4	11:40am 19/4	PR Sx +++, T6 anterior +++, C1 PIR +++, C2 BL ++	PR Sx, C1 PIR, C2 BL
24	19/4	Adults	3x4	12:20pm 19/4	R PI +++, PL SX +++, T12 BR ++, T2 anterior +++, C1 PIL +++, C6 BR +++	R PI, PL Sx, T2 anterior, C1 PIL, C6 BR

Table 1A: *Raw Data*

Section 5: Pre-processing

Table 1A is a snippet of the raw data for one session. We are mainly interested in the 'Vertebral Subluxation Before Care' column. The column contains subluxation records collected by the clinicians over the course of the study.

The first step is to understand what each record means. We know that the human spine consists mainly of five regions: cervical (C1 to C7), thoracic (T1 to T12), lumbar (L1 to L5), sacral (S1 to S5) and coccyx (fused vertebrae) and subluxation might be Posterior, Anterior, on the right or the left. For example, we have 'C2 BR +++'. Here, 'C2' means

Vertebral Subluxation Before Care for Session 1						
PL Sx +++, L5 BL +++, T12 BR ++, T6 BR +++, C2 BR +++, C1 PIL ++						
Session 1 PLSX	**Session 1 L5BL**	**Session 1 T12BR**	**Session 1 T6BR**	**Session 1 C2BR**	**Session 1 C1PIL**	**Session 1 PRSX**
3	3	2	3	3	2	0

Table 1B: *Processed Subluxation Scores*

'Cervical Vertebra 2', 'BR' means 'Base Right Subluxation', and '+++' implies the severity level, which is '3' in this case.

The subluxation scores were extracted using Python libraries like numpy, pd, etc, as demonstrated by Table 1B.

Each subluxation score also needed to be multiplied by specified weights. For C1 and C2, each score would be multiplied by 3. For C3 to C7, each score would be multiplied by 2. The rest will remain the same. The scores were then added session-wise. The resulting table is Table 1C. The 'Session 1 Weighted Sum' is the weighted subluxation score for Session 1 for each participant ID. We will use these scores for further analysis.

ID	Enrolment Date	Initial Care Plan	Age Group	Active/Passive	Session 1	Session 1 Weighted Sum
39	19/4/2022	3x4	Adults	Active	10:30am 19/4	16
76	19/4	3x4	Adults	Active	11:40am 19/4	11
24	19/4	3x4	Adults	Active	12:20pm 19/4	17

Table 1C: *Processed Data with Weighted Scores*

Since most participants stopped attending before the last session, we calculated two additional variables. This helps us align the subluxation scores with the key time points of the stress variables:

Mid Intervention Score: The weighted subluxation scores mid-way through the treatment.

Last Intervention Score: The weighted subluxation score for the last session each participant attended, before they stopped attending.

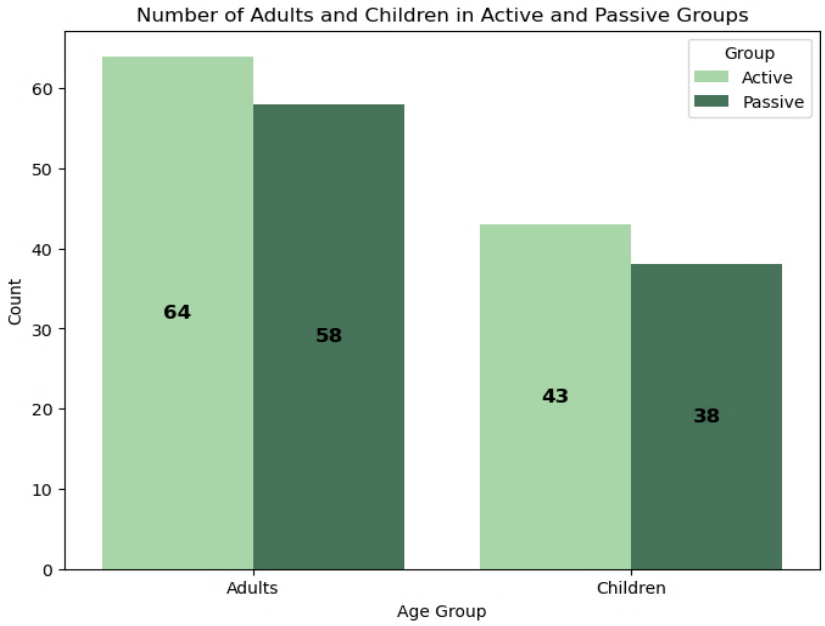

Figure 1: *Number if Adults and Children in Active and Passive Groups*

Section 6: Results

This section discusses the results from the Exploratory Analysis, Correlation Analysis, and Predictive Modelling.

Section 6.1: Results from the Exploratory Analysis

A. Number of Adults and Children

Figure 1 visualises the number of adults and children in the study in the active and passive groups through a bar graph. There are a total of 112 adults in the study (64 in the active group and 58 in the passive group) and 81 children (43 in the active group and 38 in the passive group). The number of adults and children is balanced within the two age groups.

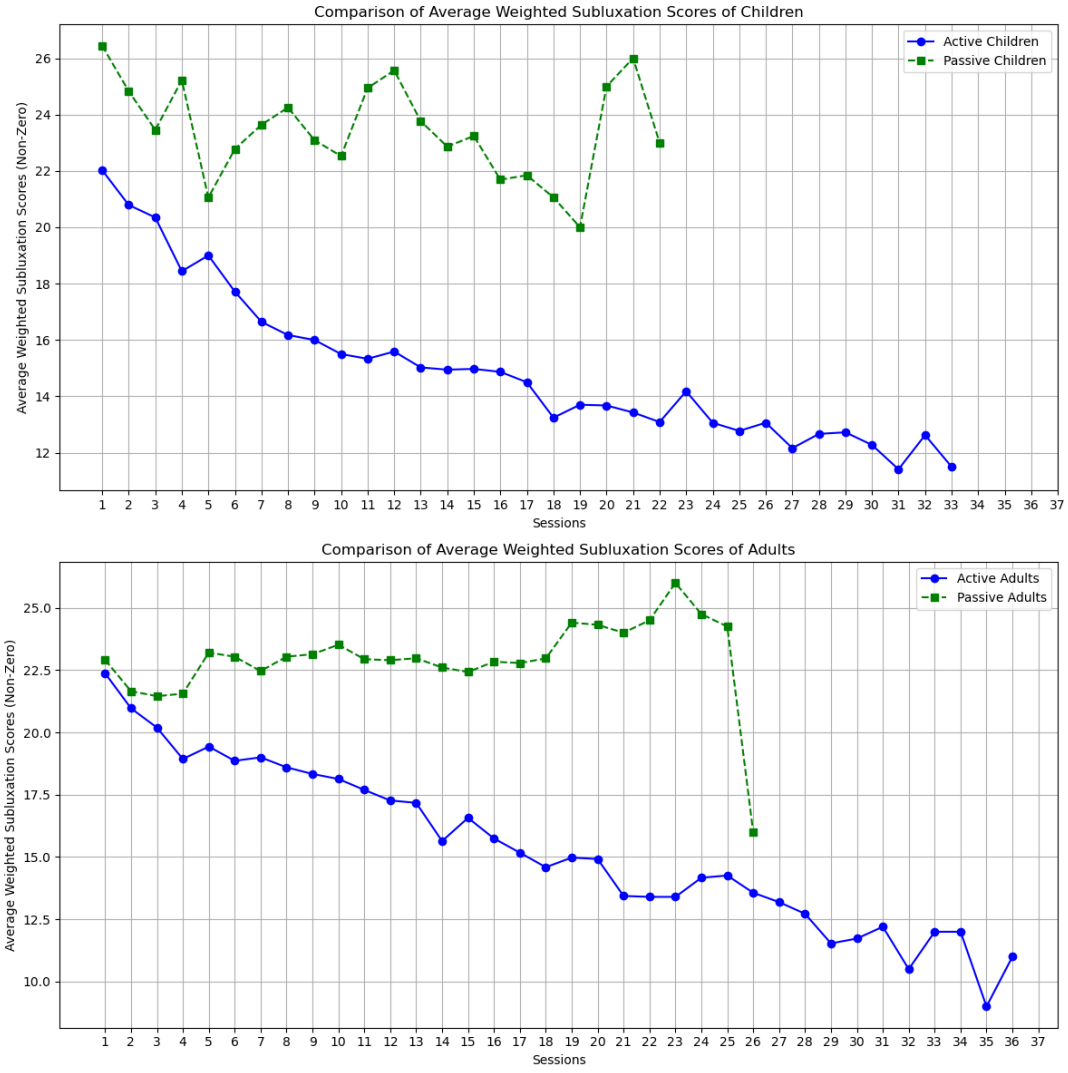

Figure 2: *Trends in Average Weighted Subluxation Scores*

B. Trends in Weighted Subluxation Scores

Figure 2 visualises the trends in the average weighted subluxation scores for Adults and Children of the Active and Passive groups. There is a gradual decline in the subluxation scores for Active and Passive groups, although the scores for the Passive group appear higher relative to the Active group, indicating that those who received treatment had improved subluxation scores over time, compared to those who received "sham" treatment.

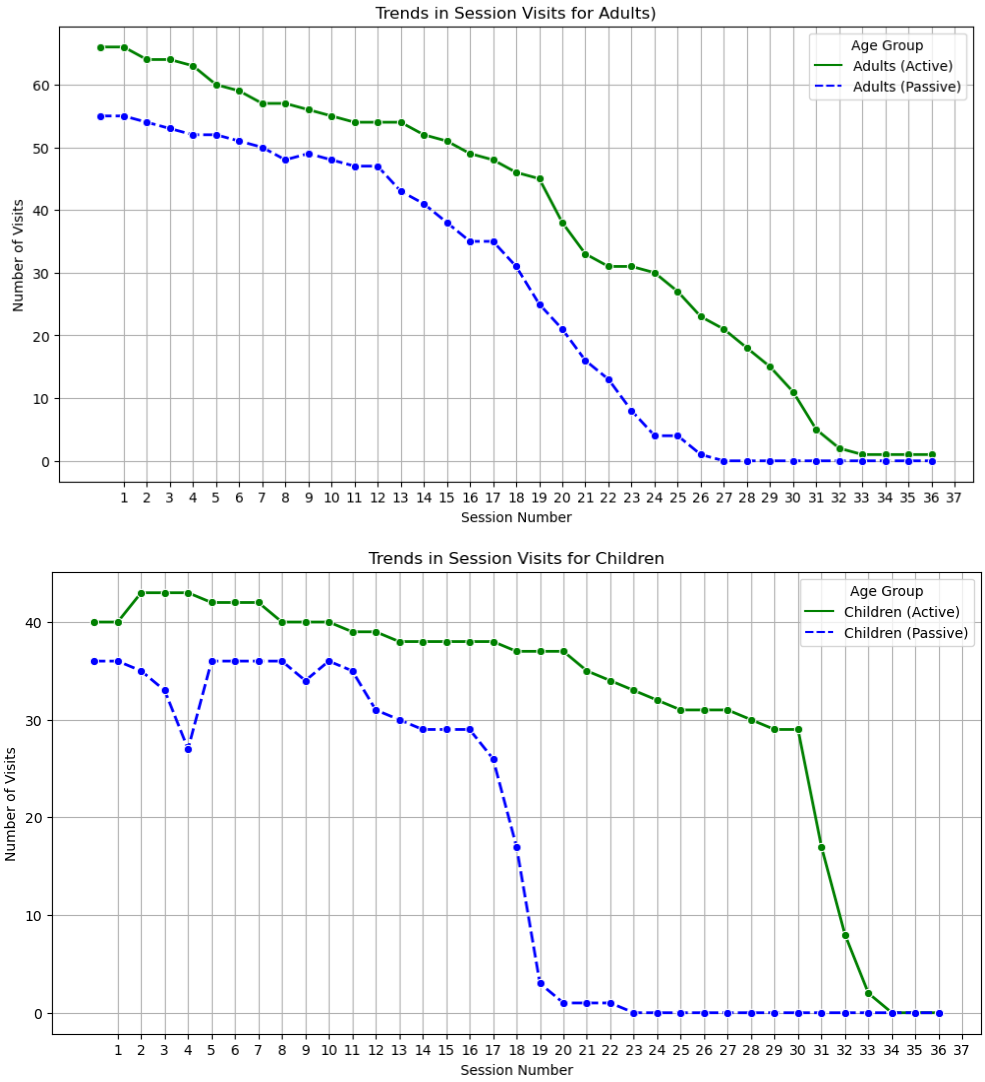

Figure 3: *Trends in Session Visits for Adults and Children*

C. Trends in Session Visits

The average session visits for adults and children in both Active and Passive groups gradually falls, although the number of visits for children drastically falls, specifically for the Passive group. This is reflected in the massive decline of the subluxation scores of the Passive group in Figure 3.

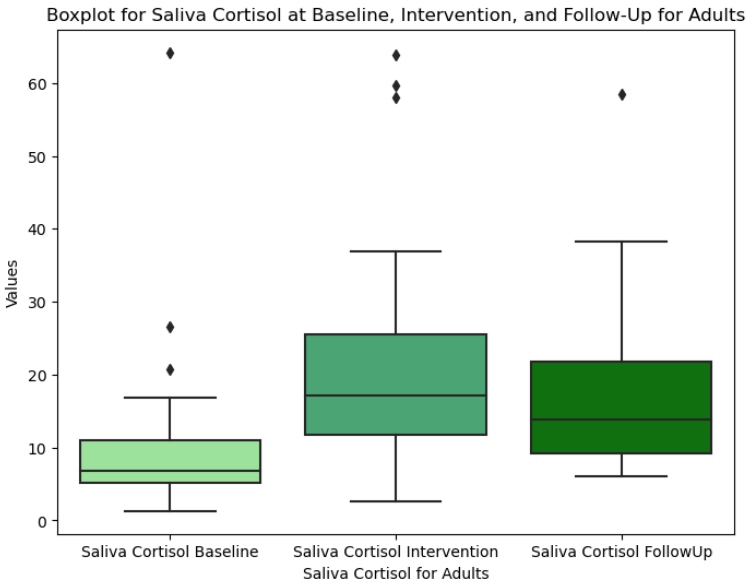

Figure 4A: *Boxplots of Saliva Cortisol at Key Time Points (Adults)*

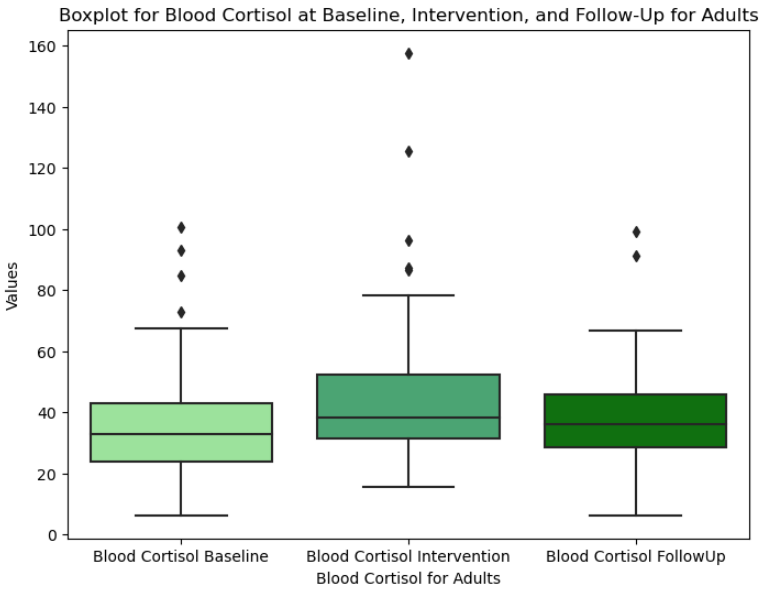

Figure 4B: *Boxplots of Blood Cortisol at Key Time Points (Adults)*

D. Boxplots

Cortisol in Adults: Figures 4A and 4B visualise the distribution of saliva cortisol and blood cortisol for adults. The box plots indicate that the distribution of both variables is positively skewed. Median baseline saliva cortisol is around 7 units. We observe that the post-intervention levels increase to about 19 units, followed by a slight decline to 14 units in the follow-up session. Similar observations are made in blood cortisol levels, with median baseline levels being around 30 units, post-intervention levels being around 40 units, and the follow-up levels being around 38 units. We also observe a few outliers.

Cortisol in Children: Figures 4C and 4D visualise the distribution of saliva cortisol and hair cortisol for children. The box plots indicate that the distribution of both variables is positively skewed. Median baseline saliva cortisol is around 7 units. We observe that the post-

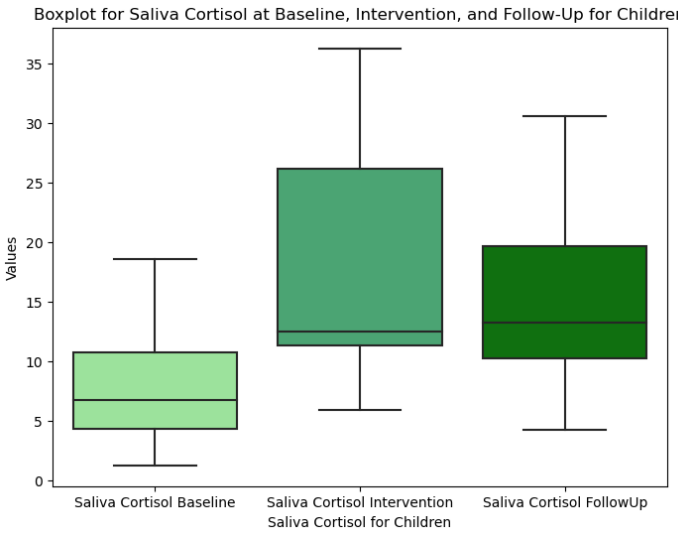

Figure 4C: *Boxplots of Saliva Cortisol at Key Time Points (Children)*

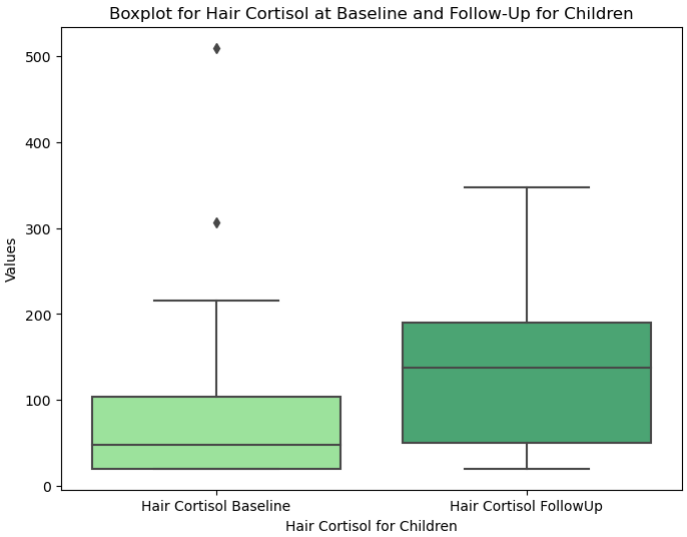

Figure 4D: *Boxplots of Hair Cortisol at Key Time Points (Children)*

intervention levels increase to about 13 units, followed by a slight increase to 14 units in the follow-up session. Similar observations are made in hair cortisol levels, with the data for children, where the median baseline levels were around 40 units, and the follow-up levels were around 150 units. We also observe a few outliers in the case of hair cortisol.

DASS Scores in Adults: Median DASS scores are mostly consistent throughout the timepoints, between 5-7 points. They are positively skewed, and we observe many outliers (Figure 4E).

E. Correlation Analysis

No significant correlation was found between the stress variables for both adults and children. Baseline and Mid-Intervention Subluxation scores were positively correlated with each other for both adults and children, with Pearson's correlation coefficients being +0.86 and +0.66, respectively.

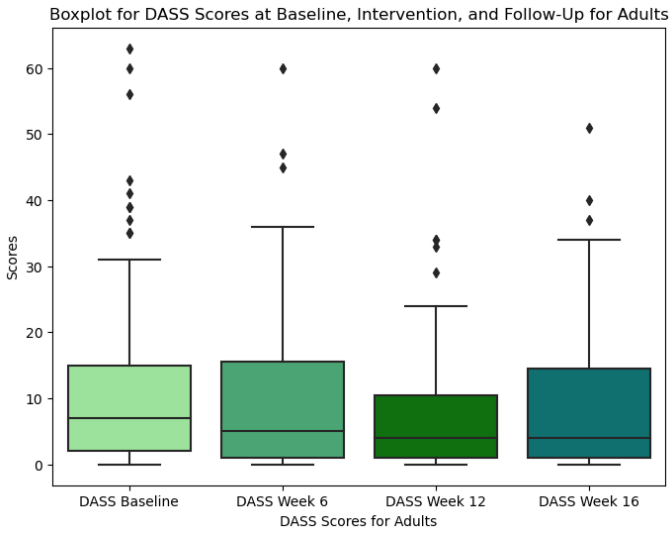

Figure 4E: *Boxplots of DASS scores at Key Timepoints (Adults)*

Section 6.2: Results from the Linear Mixed-Effects Models

In this section, we will examine the key predictors of stress variables for adults and children and the statistical significance and effect sizes by using Linear Mixed Effects Models. We will examine what these results allow us to infer about the sample.

To address the positive skewness and to normalise the distribution, the stress variables have been log-transformed for the LMM models. The following variables have been log-transformed: Baseline Saliva Cortisol, Post-Intervention Saliva Cortisol (for both adults and children), Baseline Blood Cortisol, Post-Intervention Blood Cortisol (for adults), and Baseline Hair Cortisol (for children).

A. Determining How Blood Cortisol Changes Over Time in Adults

We fitted a Linear Mixed-Effects Model with fixed effects for Timepoint, log-transformed DASS scores at baseline, log-transformed salivary cortisol at baseline, Mid-Intervention Score, and Last Intervention Score. To account for individual differences in baseline blood cortisol levels, participant ID was included as a random intercept. Additionally, we included random slopes for Timepoint to allow each participant to have their own cortisol response trajectory from baseline to post-intervention. The following is the estimated equation:

$$\begin{aligned} &\boldsymbol{log}\left(\boldsymbol{BloodCortisol_{ij}}\right) \\ &\quad = \mathbf{0.948} + \boldsymbol{b_{0i}} + (\mathbf{0.175} + \boldsymbol{b_{1i}}) \cdot \boldsymbol{Timepoint}\left(\boldsymbol{PostIntervention_{ij}}\right) \\ &\quad + \mathbf{0.223} \cdot \boldsymbol{Passive_i} + \mathbf{0.236} \cdot \mathbf{log}\left(\boldsymbol{DASSBaseline_{ij}}\right) - \mathbf{0.024} \\ &\quad \cdot \boldsymbol{MidInterventionScore_{ij}} - \mathbf{0.014} \cdot \boldsymbol{LastInterventionScore_{ij}} + \mathbf{1.290} \\ &\quad \cdot \mathbf{log}\left(\boldsymbol{SalivaCortBaseline_{ij}}\right) + \boldsymbol{\varepsilon_{ij}}, \end{aligned} \quad \textbf{(3)}$$

where, i is the participant, $b_{0i} \sim N(0, \sigma^2_{intercept})$ is the participant-specific random intercept $b_{1i} \sim N(0, \sigma^2_{Timepoint\ slope})$ is the participant-specific random slope for Timepoint, and slope is ≈ 0 and $\varepsilon_{ij} \sim N(0, \sigma^2_{Residual})$.

Table 4 summarises the results of the regression analysis and interprets the coefficients. For residual plots and a detailed result summary, refer to Appendix A.

Predictor	Coefficient	p-value	Interpretation
Intercept	0.983	0.043	This is the estimated average blood cortisol at baseline for the Active group, when all predictors are at 0. (about 2.672 units). It is statistically significant.
Timepoint: PostIntervention	+0.175	< 0.001	Blood cortisol significantly increases by 19% after the intervention, holding other variables constant.
Passive vs Active (Passive)	+0.321	0.153	Passive participants tend to have higher blood cortisol (about 37% higher), but not significantly so ($p > 0.05$).
DASSBaseline_log	+0.204	0.001	A 1% increase in the Baseline DASS scores increases Blood Cortisol by about 0.204%.
Mid_Intervention Score	-0.013	0.474	Both subluxation scores had non-significant associations with salivary cortisol. This suggests that spinal subluxation scores during or after treatment do not strongly relate to short-term stress hormone levels.
Last_Intervention_Score	-0.031	0.052	
SalivaBaseline_log	+1.320	< 0.001	A 10% increase in Saliva Cortisol at Baseline is associated with a 13.2% increase in Blood Cortisol.
Group Variance	0.746	-	Random effect variance shows modest variation in baseline blood cortisol between participants (on log scale).
Intercept-Slope Covariance	-0.079		Slight negative correlation: participants with higher baseline blood cortisol may tend to show less increase in post-intervention, but the effect is small.
Slope Variance (Timepoint)	0.009		There is some variability across participants in their change in cortisol from baseline to post-intervention.

Table 4: *Interpreting Coefficients for LMM (Blood Cortisol)*

B. Determining How Saliva Cortisol Changes Over Time in Adults

We fitted a Linear Mixed-Effects Model with fixed effects for Timepoint, log-transformed DASS scores at baseline, log-transformed saliva cortisol at baseline, Mid-Intervention Score, and Last Intervention Score. To account for individual differences in baseline blood cortisol levels, participant ID was included as a random intercept. Additionally, we included random slopes for Timepoint to allow each participant to have their own cortisol response trajectory from baseline to post-intervention. The following is the estimated equation:

$$\begin{aligned} &\boldsymbol{log}\left(\boldsymbol{SalivaCortisol_{ij}}\right) \\ &\quad = -\mathbf{0.447} + \boldsymbol{b_{0i}} + (\mathbf{0.573} + \boldsymbol{b_{1i}}) \cdot \boldsymbol{Timepoint}\left(\boldsymbol{PostIntervention_{ij}}\right) \\ &\quad + \mathbf{0.148} \cdot \boldsymbol{Passive_i} + \mathbf{0.025} \cdot \boldsymbol{MidInterventionScore_{ij}} - \mathbf{0.017} \\ &\quad \cdot \boldsymbol{LastInterventionScore_{ij}} + \mathbf{0.647} \cdot \mathbf{log}\left(\boldsymbol{BloodCortBaseline_{ij}}\right) + \boldsymbol{\varepsilon_{ij}}, \end{aligned} \quad \textbf{(4)}$$

where, i is the participant, $b_{0i} \sim N(0, \sigma^2_{intercept})$ is the participant-specific random intercept, $b_{1i} \sim N(0, \sigma^2_{Timepoint\ slope})$ is the participant-specific random slope for Timepoint, and $\varepsilon_{ij} \sim N(0, \sigma^2_{Residual})$.

Table 5 summarises the results of the regression analysis and interprets the coefficients. For residual plots and a detailed result summary, refer to Appendix B.

Predictor	Coefficient	p-value	Interpretation
Intercept	-0.447	0.162	Not statistically significant. Represents estimated saliva cortisol at baseline for the Active group, when all other predictors are 0.
Timepoint: PostIntervention	+0.573	< 0.001	Saliva cortisol significantly increases by 77% after intervention, across participants. A strong and reliable effect.
Passive vs Active (Passive)	+0.148	0.296	Passive group has slightly higher saliva cortisol (15% higher) than Active, but not statistically significant.
Last_Intervention_Score	-0.017	0.103	It shows a negative association between final subluxation score and saliva cortisol, but non-significant (p = 0.103).
Mid_Intervention_Score	+0.025	0.035	Significant positive association: A 1-unit increase in Mid-subluxation is associated with about 2.5% increase in post intervention Saliva Cortisol.
BloodBaseline_log	+0.647	< 0.001	Very strong positive association: 10% increase in Blood Cortisol at Baseline is associated with a 6.48% increase in Saliva Cortisol post-intervention.
Group Var	0.142	-	Random effect variance shows small to moderate variation in baseline saliva cortisol between participants (on log scale).
Intercept-Slope Covariance	-0.068		Slight negative correlation: participants with higher baseline saliva cortisol may tend to show less increase in post-intervention, but the effect is small.
Slope Variance (Timepoint)	0.142		There is some variability across participants in their change in cortisol from baseline to post-intervention.

Table 5: *Interpreting Coefficients for LMM (Saliva Cortisol ADULTS)*

C. Determining How DASS Scores Changes Over Time in Adults

We fitted a Linear Mixed-Effects Model with fixed effects for Timepoint, log-transformed Anxiety scores at baseline (which is a DASS subcategory), Mid-Intervention Score, Last Intervention Score, and Last Intervention Score's interaction with Treatment Groups. To account for individual differences in baseline DASS scores, participant ID was included as a random intercept. While random slopes allow the model to account for individual variability in how predictors affect the outcome, they are only appropriate when there is meaningful within-group variation in those predictors. In the present study, a random slope for Last_Intervention_Score was initially included to test whether participants differed in their response to changes in this variable.

However, model results showed that the variance of the random slope for Last_Intervention_Score was effectively zero (0.000), and its standard error was relatively large. This indicates that there was no substantial variation across participants in how Last_Intervention_Score was associated with DASS scores. Additionally, including this random slope increased model complexity without improving fit or interpretability. We also attempted to include random slope for Timepoint, however, the insights on treatment groups and subluxation scores were limited and non-significant.

Therefore, for parsimony and model stability, the final model includes **only a random intercept per participant**. This approach appropriately accounts for individual baseline differences while avoiding overparameterization.

The following is the estimated equation:

$$\begin{aligned} log(DASS_{ij}) = {} & 1.317 + b_{0i} + 0.440 \cdot Timepoint(PostIntervention_{ij}) - 0.107 \\ & \cdot Week6_{ij} - 1.544 \cdot Passive_i + 0.765 \cdot AnxietyBaseline_i - 0.006 \\ & \cdot MidInterventionScore_{ij} - 0.023 \cdot LastInterventionScore_{ij} + 0.071 \\ & \cdot (LastInterventionScore_{ij} \times Passive_i) + \epsilon_{ij}, \end{aligned} \quad (3)$$

where i is the participant, $b_{0i} \sim N(0, \sigma^2_{intercept})$ is the participant-specific random intercept and $\varepsilon_{ij} \sim N(0, \sigma^2_{Residual})$.

Table 6 summarises the results of the regression analysis and interprets the coefficients. For residual plots and a detailed result summary, refer to Appendix C.

D. Determining How Saliva Cortisol Changes Over Time in Children

We fitted a Linear Mixed-Effects Model with fixed effects for Timepoint, log-transformed hair cortisol at baseline, Mid-Intervention Score, and Last Intervention Score. To account for individual differences in baseline saliva cortisol levels in children, participant ID was included as a random intercept. Additionally, we included random slopes for Timepoint to allow each participant to have their own cortisol response trajectory from baseline to post-intervention. The following is the estimated equation:

$$\begin{aligned} & log(SalivaCortisol_{ij}) \\ & = 2.180 + b_{0i} + (0.821 + b_{1i}) \cdot Timepoint(PostIntervention_{ij}) \\ & - 0.163 \cdot Passive_i - 0.005 \cdot MidInterventionScore_{ij} + 0.012 \\ & \cdot LastInterventionScore_{ij} - 0.041 \cdot \log(HairCortBaseline_{ij}) + \varepsilon_{ij}, \end{aligned} \quad (4)$$

Predictor	Coef	p-value	Interpretation
Intercept	1.317	0.006	Estimated DASS at baseline for Active group, when other covariates = 0.
Timepoint: PostIntervention	−0.440	< .001	Compared to baseline, DASS scores significantly decreased post-intervention (35.6%).
Timepoint: Week6	−0.107	0.321	Compared to baseline, DASS scores decreased slightly in Week 6, but not significantly.
Passive vs Active (Passive)	−1.544	0.011	Passive participants had significantly lower baseline DASS scores compared to active participants (about 78.6% lower).
Anxiety_Baseline_log	+0.765	<0.001	Higher anxiety at baseline is associated with higher DASS scores across all timepoints (about 114% higher).
Mid_Intervention_Score	-0.006	0.693	Not significant (p = 0.693); mid-point subluxation score is not predictive of DASS.
Last_Intervention_Score	-0.023	0.226	No significant effect overall.
Score × Passive	+0.071	0.012	In the passive group, a higher last subluxation score is significantly associated with higher DASS (7.3% higher than Active group).
Group Variance	0.307	-	Random effect variance shows small to moderate variation in DASS scores between participants (on log scale).

Table 6: *Interpreting Coefficients for LMM (DASS Scores)*

where, i is the participant, $b_{0i} \sim N(0, \sigma^2_{intercept})$ is the participant-specific random intercept, $b_{1i} \sim N(0, \sigma^2_{Timepoint\ slope})$ is the participant-specific random slope for Timepoint, and $\varepsilon_{ij} \sim N(0, \sigma^2_{Residual})$.

Table 7 summarises the results of the regression analysis and interprets the coefficients. For residual plots and a detailed result summary, refer to Appendix D.

Predictor	Coefficient	p-value	Interpretation
Intercept	2.180	< 0.001	The intercept is statistically significant, meaning the baseline estimate is meaningfully different from zero.
Timepoint: PostIntervention	+0.821	< 0.001	Salivary cortisol levels were significantly higher (127%) after the intervention compared to baseline.
Passive vs Active (Passive)	−0.163	0.173	Passive group (sham treatment) participants had, on average, lower cortisol than the Active group, but this effect is not statistically significant.
Hair_Baseline_log	-0.041	0.417	Baseline hair cortisol has a very small, non-significant negative association with salivary cortisol.
Mid_Intervention_Score	-0.005	0.665	Both subluxation scores had non-significant associations with salivary cortisol. This suggests that spinal subluxation scores during or after treatment do not strongly relate to short-term stress hormone levels.
Last_Intervention_Score	+0.012	0.320	
Group Variance	0.132	-	The between-participant variance is moderate, indicating minor differences in baseline cortisol levels across individuals on log scale.
Slope Variance (Timepoint)	0.108		There is considerable variation in how children's cortisol changes over time (i.e., some increase more than others).
Intercept-Slope Covariance	-0.100		A negative relationship between baseline levels and change over time. Children with higher baseline cortisol increased less, and those with lower baseline increased more.

Table 7: *Interpreting Coefficients for LMM (Saliva Cortisol CHILDREN)*

Model fit statistics (Table 8) suggest that the models fit fairly well on the data for adults, as the R-squared for both fixed effects and the fixed + random effects are

	Blood Cortisol	Saliva Cortisol (Adults)	DASS Scores	Saliva Cortisol (Children)
R Squared (Fixed Effects)	0.777	0.750	0.432	0.440
R Squared (Fixed + Random Effects)	0.970	0.836	0.614	0.516
Residuals	The Histograms and the Q-Q plots satisfy the normality assumptions; however, the scatter plots show some bias despite log-transforming the stress variables. The residuals are **not** perfectly met.			

Table 8: *Model Fit*

reasonable, although the model for DASS scores does not perform as well compared to blood cortisol and saliva cortisol, as the predictors only explain 43.2% variability (for fixed effects) and 61.4% variability (for fixed + random effects). The model for saliva cortisol (children) does not capture the full variability in the dependent variable, as it only explains 44% and 51.6% of variability, for fixed effects and fixed + random effects respectively. The residuals are not perfectly met for any of the models.

Section 6.3: Results from Predictive Modelling

Features for the models were selected based on their biological relevance, psychological impact, and availability in the dataset. This included *Psychological* (DASS Scores at Baseline and Week 6), *Physiological* (Cortisol Levels at Baseline; Blood and Saliva for Adults, Hair and Saliva for Children), *Treatment-related* (Active/Passive) and *Clinical progress markers* (Subluxation Scores at Mid-Intervention and Post-Intervention).

Post-model analysis confirmed that the most theoretically justified features were also the most statistically important.

Model	Train R-Squared	Train RMSE	Test R-Squared	Test RMSE
SVR	0.65	14.61	0.48	17.67
Random Forest	0.78	13.39	0.50	15.55
Gradient Boosting	**0.87**	**10.35**	**0.64**	**13.07**

Table 9: *Model Comparison for Blood Cortisol Prediction*

A. Predicting Post-Intervention Blood Cortisol (for adults)

We fit SVR, Random Forest, and Gradient Boosting models to predict post-intervention blood cortisol levels in adults. Comparing the three models (Table 9), we find that Gradient Boosting performed the best as it provides the best balance between bias and variance, achieving higher accuracy without much overfitting, suggesting better generalization than other models. The model performs well on the training set, explaining about 87% of the variability in the response variable. The performance slightly falls on the test set, but 64% of the variability explained is reasonable.

Residuals indicate some heteroscedasticity, but the spread of residuals is reasonable, and the model generally makes unbiased predictions. For residual plots, refer to Appendix E.

The scatter points (Figure 5) generally follow the red 1:1 reference line, indicating the model makes good predictions for most cases. The overall pattern suggests that the model captures the trend well, but performance slightly decreases at the higher end of cortisol levels. Further feature engineering could help reduce the residual variance in those outlier predictions

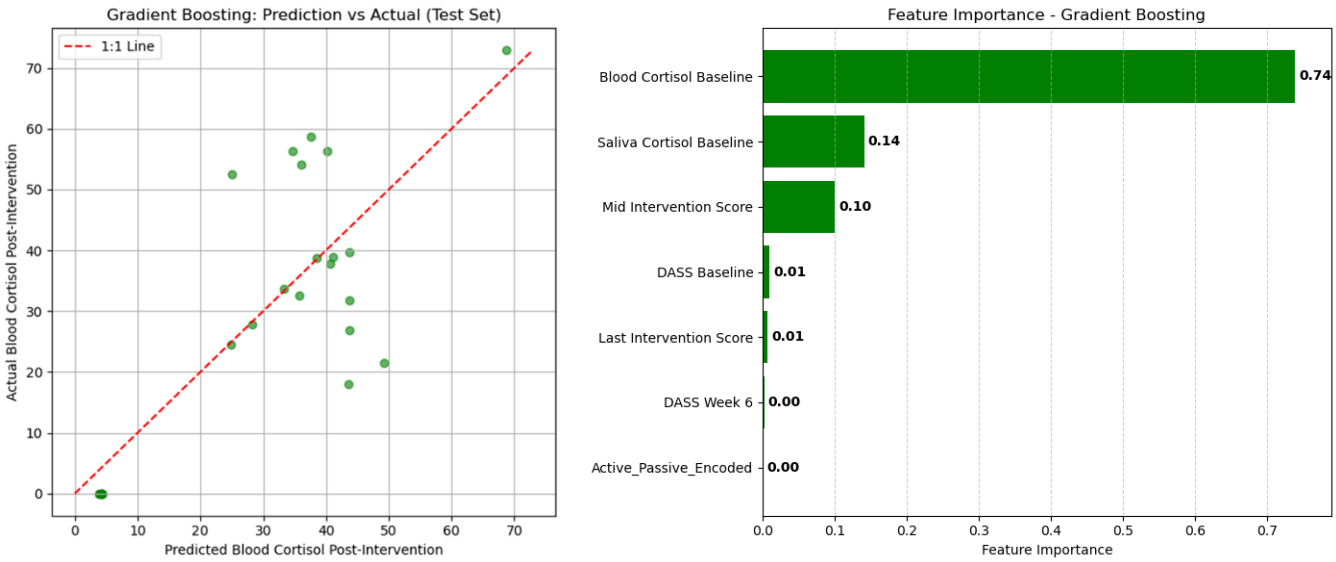

Figure 5: *Predictions of Gradient Boosting and Feature Importance Plot*

The Feature Importance plot (Figure 5) indicates that baseline blood cortisol levels were the strongest predictor of Week 12 blood cortisol, contributing 74% to the model's predictive power. Saliva cortisol baseline and mid-point subluxation scores also contributed, though to a much lesser extent. Psychosocial variables (DASS) and treatment groups had negligible predictive value.

B. Predicting Post-Intervention Saliva Cortisol (for adults)

We fit SVR, Random Forest, and Gradient Boosting models to predict post-intervention saliva cortisol levels in adults. Comparing the three models (Table 10), we find that Random Forest provides the best balance between bias and variance. The model achieves high accuracy on the training set, explaining 73% of variability in the response factor, but the performance on the test-set is minimal, explaining 47% of the variability on

Model	Train R-Squared	Train RMSE	Test R-Squared	Test RMSE
SVR	0.64	7.604	0.35	10.93
Random Forest	**0.73**	**6.48**	**0.47**	**10.76**
Gradient Boosting	0.55	8.35	0.33	12.02

Table 10: *Model Comparison for Saliva Cortisol (ADULTS) Prediction*

unseen data. This may suggest overfitting. The residuals indicate heteroscedasticity, and there is obvious bias. (Refer to Appendix F for the residual plots)

The scatter points (Figure 6) suggest that the model predicts lower cortisol levels well but underestimates saliva cortisol at higher levels. The overall pattern suggests that the model struggles to capture full variability in actual saliva cortisol responses.

The Feature Importance plot (Figure 6) suggests that baseline saliva cortisol levels were the strongest predictor of post-intervention saliva cortisol, contributing 68% to the model's predictive power. DASS at baseline contributed to 13%, and last-intervention subluxation scores contributed to only 7%. Mid-Intervention scores and treatment groups had negligible predictive value.

C. Predicting Post-Intervention DASS Scores (for adults)

We fit SVR, Random Forest, and Gradient Boosting models to predict post-intervention DASS scores in adults. Comparing the three models (Table 11), we find that Random Forest performs the best out of the three. Even though Gradient Boosting provides a higher performance on the train set, Random Forest provides the smallest RMSE without

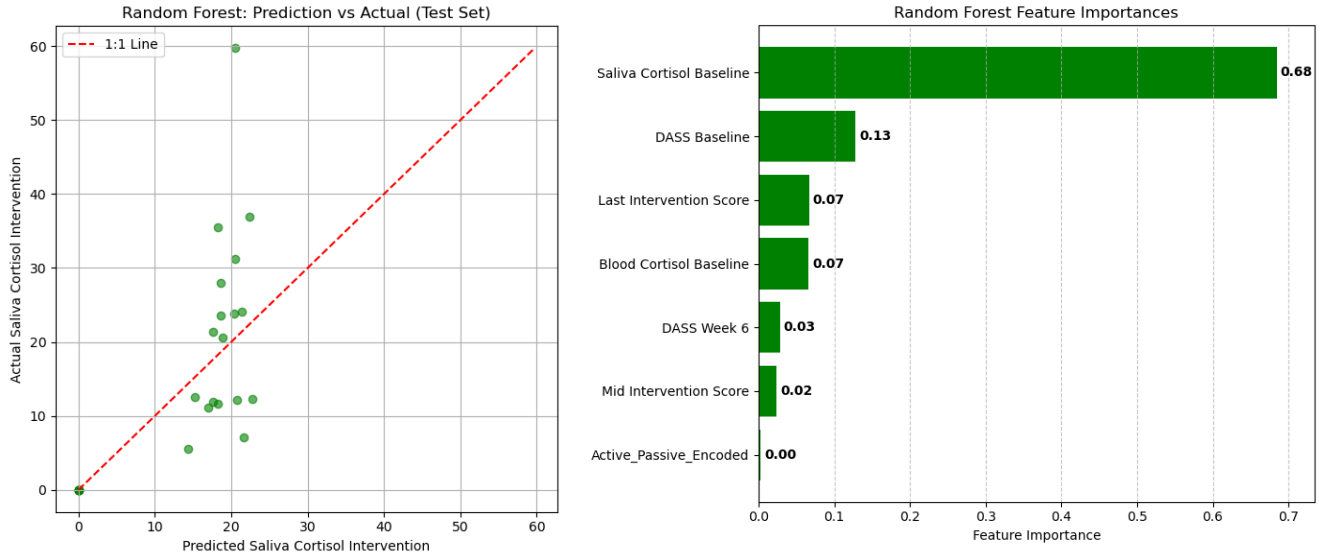

Figure 6: *Predictions of Random Forest and Feature Importance Plot*

much overfitting, making it more reliable. It explains 62% of variability on the training set, but only 47% on unseen data, which leaves room for more improvement.

The scatter plot (Figure 7) suggests that the model predicts higher DASS scores well but underpredicts the values at lower levels. The overall pattern suggests that the model struggles to capture full variability in actual post-intervention DASS scores (For residuals, refer to Appendix G).

The Feature Importance plot (Figure 7) suggests that Week 6 DASS Scores were the strongest predictor of Week 12 DASS scores, contributing 68% to the model's predictive power. DASS at baseline contributed to 19%. However, Mid-Intervention and Last Intervention scores and treatment groups had negligible predictive value.

Model	Train R-Squared	Train RMSE	Test R-Squared	Test RMSE
SVR	0.52	6.97	0.09	8.54
Random Forest	**0.62**	**6.2**	**0.47**	**6.53**
Gradient Boosting	0.72	5.3	0.43	6.79

Table 11: *Model Comparison for DASS Scores Prediction*

D. Predicting Post-Intervention Saliva Cortisol (for children)

We fit SVR, Random Forest, and Gradient Boosting models to predict post-intervention saliva cortisol levels in children. However, we find that none of the models were reliable in predicting Week 12 saliva cortisol for children. All three models fail to explain the variance in the outcome for unseen data. All models show higher R^2 on the training set but dramatically worse performance on the test set, indicating overfitting**,** with high RMSE (Table 12).

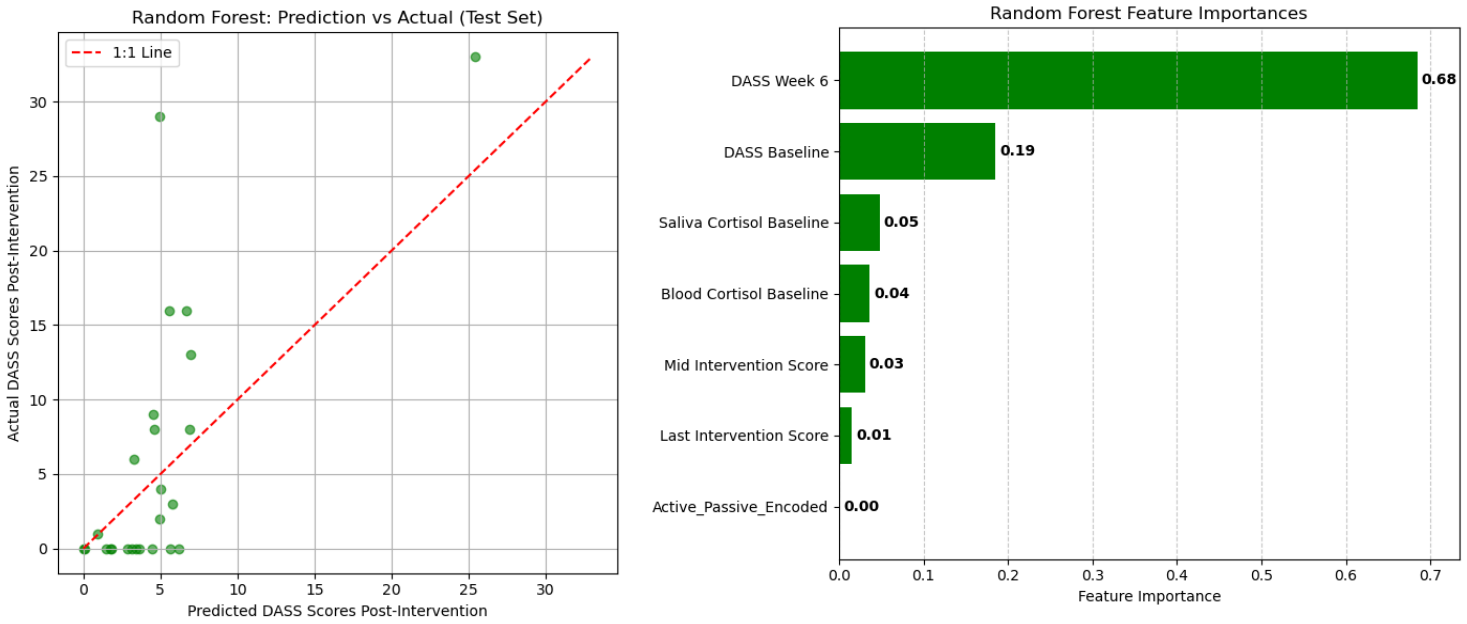

Figure 7: *Predictions of Random Forest and Feature Importance Plot*

Section 7: Implications

A. Implications from Linear Mixed Effects Models

The findings of the Linear Mixed Effects models suggest that subluxation may influence physiological stress indirectly rather than directly driving the cortisol levels.

While DASS scores (indicator of psychological stress) did improve post-intervention (they reduced by 35.6%), the physiological response was more nuanced. Random intercepts in the models revealed considerable individual variability In baseline cortisol levels, highlighting that participants began the study with markedly different physiological stress baseline.

Model	Train R-Squared	Train RMSE	Test R-Squared	Test RMSE
SVR	0.54	5.031	-0.61	11.62
Random Forest	0.53	5.11	-0.32	10.51
Gradient Boosting	0.42	5.69	-0.24	10.22

Table 12: *Model Comparison for Saliva Cortisol (CHILDREN) Prediction*

Moreover, by including random slopes for Timepoint in the cortisol models, we accounted for individual differences in physiological response trajectories over time. This indicates that while cortisol levels (saliva and blood) generally increased post-intervention, the degree of change varied across individuals, possibly reflecting personalised stress adaptation or regulation patterns.

Initial anxiety levels, particularly baseline anxiety scores, were strong predictors of overall stress, emphasising the clinical importance of psychological screening prior to care. Interestingly, there was no consistent difference between the Active and Passive treatment groups in terms of cortisol outcomes, suggesting that the type of care delivered may not be the primary determinant of physiological change. Although, the interaction observed between unresolved subluxation and group type, where Passive participants with higher subluxation towards the end reported greater stress, suggesting that in some cases, physical indicators may still influence perceived stress when not sufficiently addressed.

In children, the results remained inconclusive, likely due to smaller sample sizes and developmental variability. Biomarkers like saliva and hair cortisol may not fully capture the stress experience of younger populations.

Overall, these findings underline the complexity of stress physiology and the need for multifactorial assessment in evaluating chiropractic intervention outcomes.

B. Implications from the Predictive Models

We observe that baseline variables are the strongest predictors of post-intervention variables. For example, baseline blood cortisol was the strongest predictor of post-intervention blood cortisol. Subluxation scores and treatment groups had negligible predictive value, implying that the physiological state of the participants played a primary role in determining how their cortisol levels responded after treatment. This raises important questions about whether short-term spinal adjustments can meaningfully alter cortisol regulation or if these responses are largely biologically predetermined.

Variables like subluxation and treatment groups, despite being central to the intervention logic, did not have strong or measurable effects on cortisol levels in this sample, which may indicate that the biological effects of spinal adjustments, if they exist, may take longer to manifest, may be subtle, or may depend on individual-level differences which are not captured here. Because cortisol regulation is a slow-changing, complex physiological process, a short-term 12-week intervention may not be enough to observe measurable effects on post-intervention cortisol levels.

Another factor to consider is that the study was conducted when the COVID-19 pandemic was ongoing, which was a period of widespread uncertainty and chronic stress. These environmental stressors may have overwhelmed or blunted the effects of chiropractic care. Specifically, cortisol levels are highly responsive to both acute and

chronic psychological stress, and the possibility of the pandemic introducing noise to the data should not be ignored. Participants may have been under elevated stress levels regardless of treatment, making it difficult to observe any measurable changes in the stress variables.

Section 8: Discussion and Future Work

Chronic stress is well-established to elevate cortisol levels (Hellhammer et al., 2009), and in this study, both saliva and blood cortisol levels increased in Week 12 post-intervention, suggesting the presence of a physiological stress response that may be indirectly associated with chiropractic care (Tuchin, 1998). In contrast, self-reported DASS scores decreased by 35.6%, indicating substantial psychological improvement following intervention, validating its psychometric abilities (Brown et al., 1996; Sinclair et al., 2012).

Interestingly, Passive participants began with lower baseline DASS scores, however, unresolved subluxation in this group were associated with poorer psychological outcomes, suggesting a potential mind-body interaction. Mid-intervention subluxation scores had a small association with increased saliva cortisol at Week 12 (2.5%), but there was no significant association between Last-intervention subluxation scores and saliva cortisol at Week 12, aligning with previous findings (Tuchin 1998).

While machine learning models such as Random Forest and Gradient Boosting were able to capture some variance in stress-related outcomes, their predictive accuracy was limited, likely due to inherently subjective and multifactorial nature of stress (Mentis et al., 2024). Existing literature presents mixed evidence regarding the effects of chiropractic care on cortisol regulation. Although musculoskeletal improvements are consistently supported, effects on cortisol remain inconclusive (Tuchin 1998), which also aligns with this study's findings that treatment groups (Active and Passive) did not significantly influence cortisol levels (Brockman 2004).

Despite the methodological efforts and statistical approaches employed, several limitations may affect the interpretation and generalizability of the findings.

First, there is an inconsistency in participant attendance. Many participants stop attending the sessions way before the last session, which reduces the model generalizability as it may not produce reliable results. The stress variables measured at 2-3 time points further make it challenging to generalise the model predictions. The session's irregular timing and possible non-linearity over time were not deeply explored, as only scores from the mid and last intervention sessions were considered for each participant.

Second, variables like sleep, Heart Rate Variability, or comorbidities are not considered in this study. These factors are known to significantly influence physiological stress markers (e.g., cortisol levels) and psychological well-being. Their omission limits the comprehensiveness of the model and increases the risk of omitted variable bias.

Third, the DASS scores used in the study were self-reported, introducing potential biases related to social desirability or individual differences in introspective capacity. As a result, these scores may not objectively reflect true psychological stress levels, potentially weakening associations with physiological stress markers. Even well-fitted models may fail to generalise stress response across a diverse population due to the inter-individual variability of stress.

Fourth, analyses involving the paediatric subgroup yielded weak or inconsistent associations between intervention and stress outcomes. This may be attributed to smaller sample sizes, greater developmental variability in stress psychology, and potential

measurement challenges in younger participants. As such, conclusions drawn from the children's data should be treated with caution.

Fifth, the predictive models, particularly those using machine learning and fixed effects approaches, were prone to overfitting due to relatively small sample sizes and a high number of predictors. Although model performance on training data appeared strong, this may not translate as well to unseen data. Further, diagnostic plots of the residuals suggested systematic bias, indicating models may not fully capture the underlying data structure, thus reducing the robustness and reliability of the findings.

Still, an increase in cortisol levels specifically may not necessarily mean that chiropractic intervention causes an increase in physiological stress. The data's period of collection (during the COVID-19 pandemic) and the chronic stress during the time may lead to maladaptive neuroplasticity, where the brain's stress circuits become hypersensitive, reinforcing stress response over time (Radley et al., 2015). Increased cortisol could be due to initial physiological stress from spinal adjustments, rather than worsening stress. This rise may reflect the body's "wear and tear" during early therapeutic realignment, before long-term benefits can appear. Such short-term elevations in cortisol in early intervention phases may indicate a transitional stress adaptation, rather than worsening stress.

While the study provides initial insights into the relationship between chiropractic intervention and stress outcomes, these limitations highlight the need for more

comprehensive, longitudinal, and controlled research to draw stronger and more generalizable conclusions.

Future work should focus on extending the duration of the study, as it is essential to assess whether the changes in cortisol and DASS scores are sustained or evolve over time. A longer timeline could also help detect delayed physiological responses to chiropractic care that may not manifest within a 12-week window, especially in populations with chronic stress or dysregulated stress systems. Increasing the sample size and including participants from a broader range of ethnic backgrounds and health statuses would allow for deeper investigation of age-related and demographic differences. This would also enhance the statistical power and generalizability of the findings.

Future studies should also consider incorporating additional objective stress-related biomarkers such as heart rate variability, sleep quality measures, and galvanic skin response, which would provide a more holistic view of stress regulation and its modulation through chiropractic care. Collecting the stress measures data more frequently (e.g., weekly) would provide greater temporal resolution and allow researchers to track short-term fluctuations (for e.g. stress response after every intervention session). Additionally, standardising the timing and method of subluxation scoring could reduce noise and improve model accuracy.

At last, given that the data collection period coincided with the COVID-19 pandemic, which introduced widespread and chronic stress, future research should

include measures to account for external stressors. This may involve using questionnaires to quantify daily life stressors or controlling for known environmental stress factors.

This study found improvements in self-reported psychological stress (DASS scores) but an increase in physiological stress markers (cortisol) post-intervention. Predictive models showed that baseline biological markers were stronger predictors than treatment factors, suggesting stress response may be biologically driven. Future work should explore non-linear models, longer follow-up, and include broader stress indicators for clearer insights.

Section 9: References

Ahuja, R., & Banga, A. (2019). Mental stress detection in university students using machine learning algorithms. *Procedia Computer Science, 152*, 349–353. https://doi.org/10.1016/j.procs.2019.05.007

Brockman, S. (2007). The role of chiropractic manipulation in promoting an individual's perception of psychological well-being. *Clinical Chiropractic*, *10*(1), 8-23. https://doi.org/10.1016/j.clch.2006.08.004

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy*, *35*(1), 79-89. https://doi.org/10.1016/s0005-7967(96)00068-x

Ding, C., Zhang, Y., & Ding, T. (2023). A systematic hybrid machine learning approach for stress prediction. *PeerJ. Computer science*, *9*, e1154. https://doi.org/10.7717/peerj-cs.1154

Ferguson, K. K., McElrath, T. F., Chen, Y.-H., Mukherjee, B., & Meeker, J. D. (2015). Urinary Phthalate Metabolites and Biomarkers of Oxidative Stress in Pregnant Women: A Repeated Measures Analysis. Environmental Health Perspectives, 123(3), 210-216. https://doi.org/doi:10.1289/ehp.1307996

Haas, M., Groupp, E., & Kraemer, D. F. (2004). Dose-response for chiropractic care of chronic low back pain. *The Spine Journal*, *4*(5), 574-583. https://doi.org/10.1016/j.spinee.2004.02.008

Hadi, W., El-Khalili, N., AlNashashibi, M., Issa, G., & AlBanna, A. A. (2019). *Application of data mining algorithms for improving stress prediction of automobile drivers: A case study in Jordan*. *Computers in Biology and Medicine, 114*, 103474. https://doi.org/10.1016/j.compbiomed.2019.103474

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, *34*(2), 163-171. https://doi.org/10.1016/j.psyneuen.2008.10.026

Herr, R. M., Almer, C., Loerbroks, A., Barrech, A., Elfantel, I., Siegrist, J., Gündel, H., Angerer, P., & Li, J. (2018). Associations of work stress with hair cortisol concentrations – initial findings from a prospective study. *Psychoneuroendocrinology*, *89*, 134-137. https://doi.org/10.1016/j.psyneuen.2018.01.011

Kaptchuk, T. J., & Eisenberg, D. M. (1998). Chiropractic: Origins, Controversies, and Contributions. *Archives of Internal Medicine*, *158*(20), 2215-2224. https://doi.org/10.1001/archinte.158.20.2215

Mentis, A.-F. A., Makris, G., Tzachanis, D., Fradelos, E. C., Christodoulou, N., Chrousos, G. P., & Darviri, C. (2024). Machine learning applications for stress and post-

traumatic stress disorder detection and prediction. *Molecular Psychiatry, 29*(5), 1882–1894. https://doi.org/10.1038/s41380-023-02047-6

Radley, J., Morilak, D., Viau, V., & Campeau, S.. (2015). Chronic stress and brain plasticity: Mechanisms underlying adaptive and maladaptive changes and implications for stress-related CNS disorders. *Neuroscience & Biobehavioral Reviews*, *58*, 79–91. https://doi.org/10.1016/j.neubiorev.2015.06.018

Rubinstein, S. M., Leboeuf-Yde, C., Knol, D. L., De Koekkoek, T. E., Pfeifle, C. E., & Van Tulder, M. W. (2007). The Benefits Outweigh the Risks for Patients Undergoing Chiropractic Care for Neck Pain: A Prospective, Multicenter, Cohort Study. *Journal of Manipulative and Physiological Therapeutics*, *30*(6), 408-418. https://doi.org/10.1016/j.jmpt.2007.04.013

Sinclair, S. J., Siefert, C. J., Slavin-Mulford, J. M., Stein, M. B., Renna, M., & Blais, M. A. (2012). Psychometric Evaluation and Normative Data for the Depression, Anxiety, and Stress Scales-21 (DASS-21) in a Nonclinical Sample of U.S. Adults. *Evaluation & the Health Professions*, *35*(3), 259-279. https://doi.org/10.1177/0163278711424282

Steudte, S., Stalder, T., Dettenborn, L., Klumbies, E., Foley, P., Beesdo-Baum, K., & Kirschbaum, C. (2011). Decreased hair cortisol concentrations in generalised anxiety disorder. *Psychiatry Research*, *186*(2-3), 310-314. https://doi.org/10.1016/j.psychres.2010.09.002

Takai, N., Yamaguchi, M., Aragaki, T., Eto, K., Uchihashi, K., & Nishikawa, Y. (2004). Effect of psychological stress on the salivary cortisol and amylase levels in healthy young adults. *Archives of Oral Biology*, *49*(12), 963-968. https://doi.org/10.1016/j.archoralbio.2004.06.007

Tuchin, P. J. (1998). The effect of chiropractic spinal manipulative therapy on salivary cortisol levels. *Australas Chiropr Osteopathy*, *7*(2), 86-92. https://pubmed.ncbi.nlm.nih.gov/17987159/

Zhang, J., Dean, D., Nosco, D., Strathopulos, D., & Floros, M. (2006). Effect of Chiropractic Care on Heart Rate Variability and Pain in a Multisite Clinical Study. *Journal of Manipulative and Physiological Therapeutics*, *29*(4), 267-274. https://doi.org/10.1016/j.jmpt.2006.03.010

Mixed Linear Model Regression Results

Model:	MixedLM	Dependent Variable:	BloodCortisol_Long
No. Observations:	244	Method:	REML
No. Groups:	122	Scale:	0.0976
Min. group size:	2	Log-Likelihood:	-237.9358
Max. group size:	2	Converged:	Yes
Mean group size:	2.0		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	0.983	0.487	2.020	0.043	0.029	1.936
Timepoint[T.PostIntervention]	0.175	0.041	4.278	0.000	0.095	0.255
C(ActivePassive)[T.Passive]	0.321	0.224	1.429	0.153	-0.119	0.760
DASSBaseline_log	0.204	0.061	3.360	0.001	0.085	0.323
Mid_Intervention_Score	-0.013	0.018	-0.716	0.474	-0.049	0.023
Last_Intervention_Score	-0.031	0.016	-1.942	0.052	-0.062	0.000
SalivaBaseline_log	1.320	0.079	16.680	0.000	1.165	1.475
Group Var	0.746					
Group x Timepoint[T.PostIntervention] Cov	-0.079					
Timepoint[T.PostIntervention] Var	0.009					

Figure 8: *Regression Results for log-transformed Blood Cortisol*

Section 10: Appendix

Appendix A: Regression Results and Residuals for Linear Mixed-Effects Model (Blood Cortisol)

Figure 8 is the summary of results for the Linear Mixed Effects Model for Blood Cortisol (with random intercepts and random slopes for Timepoint).

The Histograms and the Q-Q plots satisfy the normality assumptions; however, the scatter plots show some bias despite log-transforming the stress variables (Figure 9). The residuals are not perfectly met and indicate heteroscedasticity.

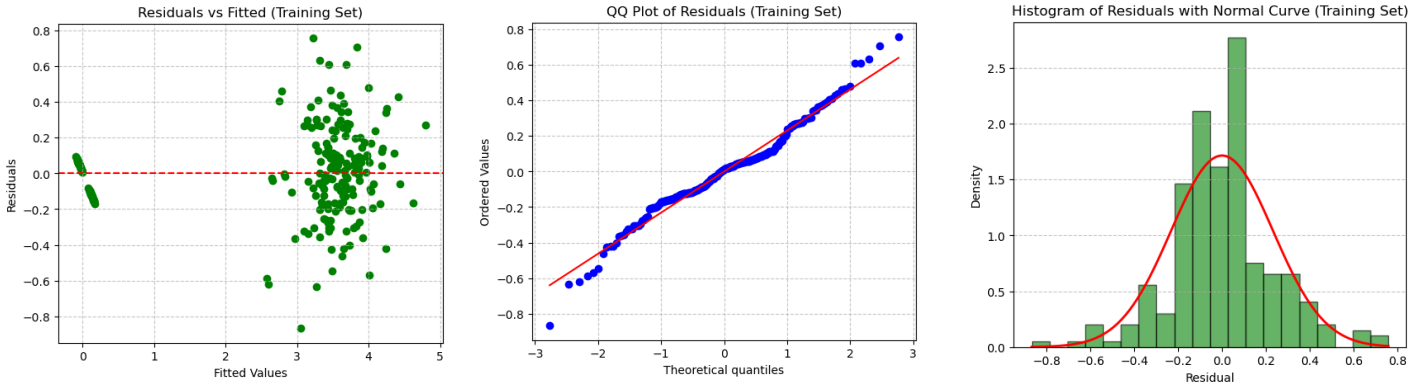

Figure 9: *Residuals for the LMM Model (Blood Cortisol)*

Mixed Linear Model Regression Results

Model:	MixedLM	Dependent Variable:	SalivaCortisol_Long
No. Observations:	244	Method:	REML
No. Groups:	122	Scale:	0.2062
Min. group size:	2	Log-Likelihood:	-245.7964
Max. group size:	2	Converged:	Yes
Mean group size:	2.0		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	-0.447	0.319	-1.399	0.162	-1.073	0.179
Timepoint[T.PostIntervention]	0.573	0.067	8.510	0.000	0.441	0.706
C(ActivePassive)[T.Passive]	0.148	0.142	1.045	0.296	-0.130	0.427
Mid_Intervention_Score	0.025	0.012	2.104	0.035	0.002	0.048
Last_Intervention_Score	-0.017	0.010	-1.630	0.103	-0.037	0.003
BloodBaseline_log	0.647	0.033	19.877	0.000	0.584	0.711
Group Var	0.215					
Group x Timepoint[T.PostIntervention] Cov	-0.068					
Timepoint[T.PostIntervention] Var	0.142					

Figure 10: *Regression Results for log-transformed Saliva Cortisol*

Figure 11: *Residuals for the LMM Model (Saliva Cortisol ADULTS)*

Appendix B: Regression Results and Residuals for Linear Mixed-Effects Model (Saliva Cortisol ADULTS)

Figure 10 is the summary of results for the Linear Mixed Effects Model for Saliva Cortisol in Adults (with random intercepts and random slopes for Timepoint).

The Histograms and the Q-Q plots satisfy the normality assumptions; however, the scatter plots show some bias despite log-transforming the stress variables (Figure 11). The residuals are not perfectly met, and indicate heteroscedasticity.

Mixed Linear Model Regression Results

Model:	MixedLM	Dependent Variable:	Dass_Long
No. Observations:	366	Method:	REML
No. Groups:	122	Scale:	0.7018
Min. group size:	3	Log-Likelihood:	-518.0865
Max. group size:	3	Converged:	Yes
Mean group size:	3.0		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	1.291	0.483	2.671	0.008	0.344	2.239
Timepoint[T.PostIntervention]	-0.440	0.107	-4.105	0.000	-0.651	-0.230
Timepoint[T.Week6]	-0.107	0.107	-0.993	0.321	-0.317	0.104
C(ActivePassive)[T.Passive]	-1.551	0.617	-2.515	0.012	-2.760	-0.342
AnxietyBaseline_log	0.768	0.072	10.741	0.000	0.628	0.908
Mid_Intervention_Score	-0.006	0.016	-0.358	0.721	-0.038	0.026
Last_Intervention_Score	-0.022	0.019	-1.174	0.240	-0.058	0.015
Last_Intervention_Score:C(ActivePassive)[T.Passive]	0.071	0.029	2.458	0.014	0.014	0.127
Group Var	0.293	0.706				
Group x Last_Intervention_Score Cov	-0.003	0.041				
Last_Intervention_Score Var	0.000	0.002				

Figure 12: *Regression Results for log-transformed DASS Scores (With Random Slopes)*

Mixed Linear Model Regression Results

Model:	MixedLM	Dependent Variable:	Dass_Long
No. Observations:	366	Method:	REML
No. Groups:	122	Scale:	0.7018
Min. group size:	3	Log-Likelihood:	-518.2908
Max. group size:	3	Converged:	Yes
Mean group size:	3.0		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	1.317	0.482	2.733	0.006	0.373	2.262
Timepoint[T.PostIntervention]	-0.440	0.107	-4.105	0.000	-0.651	-0.230
Timepoint[T.Week6]	-0.107	0.107	-0.993	0.321	-0.317	0.104
C(ActivePassive)[T.Passive]	-1.544	0.605	-2.552	0.011	-2.729	-0.358
AnxietyBaseline_log	0.765	0.072	10.684	0.000	0.625	0.905
Mid_Intervention_Score	-0.006	0.016	-0.394	0.693	-0.038	0.025
Last_Intervention_Score	-0.023	0.019	-1.210	0.226	-0.059	0.014
Last_Intervention_Score:C(ActivePassive)[T.Passive]	0.071	0.028	2.517	0.012	0.016	0.126
Group Var	0.307	0.103				

Figure 13: *Regression Results for log-transformed DASS Scores (Without Random Slopes)*

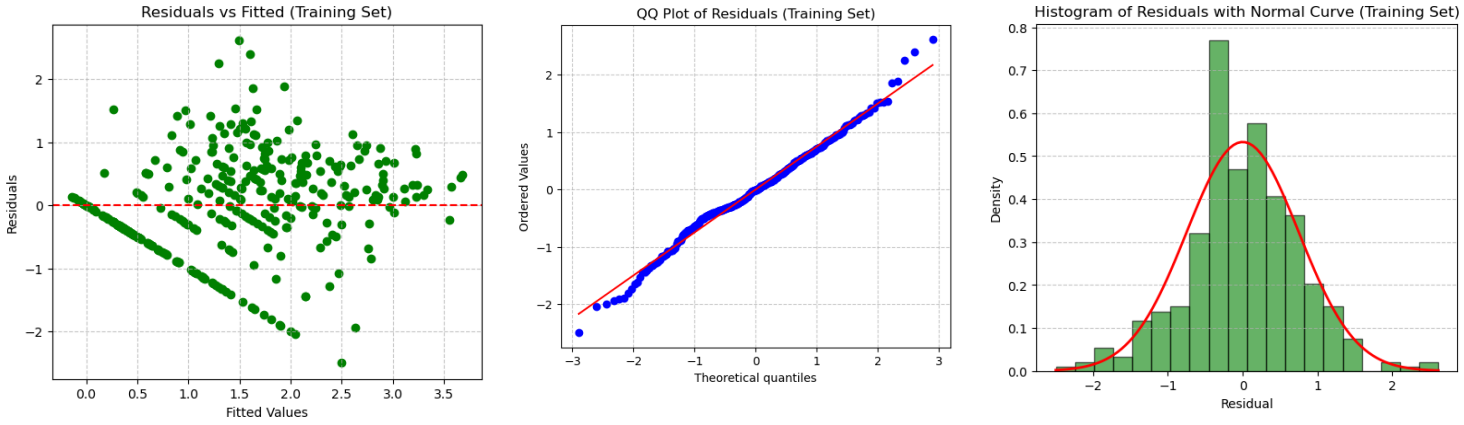

Figure 14: *Residuals for the LMM Model (DASS Scores)*

Appendix C: Regression Results and Residuals for Linear Mixed-Effects Model (DASS Scores)

Figure 12 is the summary of results for the Linear Mixed Effects Model for DASS Scores (with random intercepts and random slopes for Timepoint). However, the variance for random slopes for 'Last Intervention Score' is 0, indicating a lack of variation within participants. Since it does not improve the model performance, a simpler model was retained (Figure 13).

The Histograms and the Q-Q plots satisfy the normality assumptions; however, the scatter plots show some bias despite log-transforming the stress variables (Figure 13). The residuals are not perfectly met, and indicate heteroscedasticity.

Mixed Linear Model Regression Results

Model:	MixedLM	Dependent Variable:	SalivaCortisol_Long
No. Observations:	160	Method:	REML
No. Groups:	80	Scale:	14.6774
Min. group size:	2	Log-Likelihood:	-510.3473
Max. group size:	2	Converged:	Yes
Mean group size:	2.0		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	11.055	4.428	2.497	0.013	2.377	19.734
Timepoint[T.PostIntervention]	10.191	0.963	10.583	0.000	8.304	12.078
C(ActivePassive)[T.Passive]	-5.479	6.075	-0.902	0.367	-17.385	6.427
Hair_Baseline	-0.004	0.006	-0.597	0.550	-0.016	0.008
Mid_Intervention_Score	-0.078	0.190	-0.413	0.679	-0.450	0.293
C(ActivePassive)[T.Passive]:Mid_Intervention_Score	-0.077	0.245	-0.316	0.752	-0.557	0.402
Last_Intervention_Score	-0.080	0.208	-0.386	0.699	-0.487	0.327
C(ActivePassive)[T.Passive]:Last_Intervention_Score	0.358	0.266	1.344	0.179	-0.164	0.880
Group Var	2.900					
Group x Timepoint[T.PostIntervention] Cov	1.902					
Timepoint[T.PostIntervention] Var	44.823					

Figure 15: *Regression Results for log-transformed Saliva Cortisol (CHILDREN)*

Figure 16: *Residuals for the LMM Model (Saliva Cortisol CHILDREN)*

Appendix D: Regression Results and Residuals for Linear Mixed-Effects Model (Saliva Cortisol CHILDREN)

Figure 15 is a summary of regression results for the Linear Mixed Effects Model for Saliva Cortisol in Children (with random intercepts and random slopes for Timepoint). The

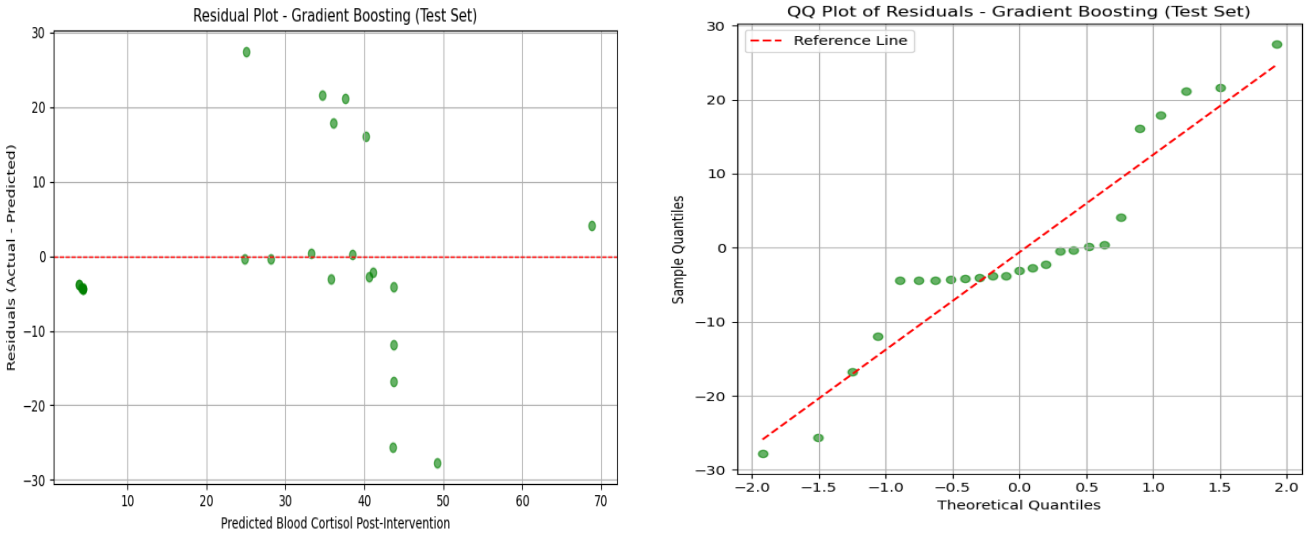

Figure 17: *Residuals for Gradient Boosting Model (Blood Cortisol)*

residuals indicate clear bias and violate the assumptions of homoscedasticity, as the scatter plot shows clear patterns.

Appendix E: Residuals for Gradient Boosting Model (Post-Intervention Blood Cortisol)

Residuals indicate some heteroscedasticity (Figure 17), but the spread of residuals is reasonable, and the model generally makes unbiased predictions.

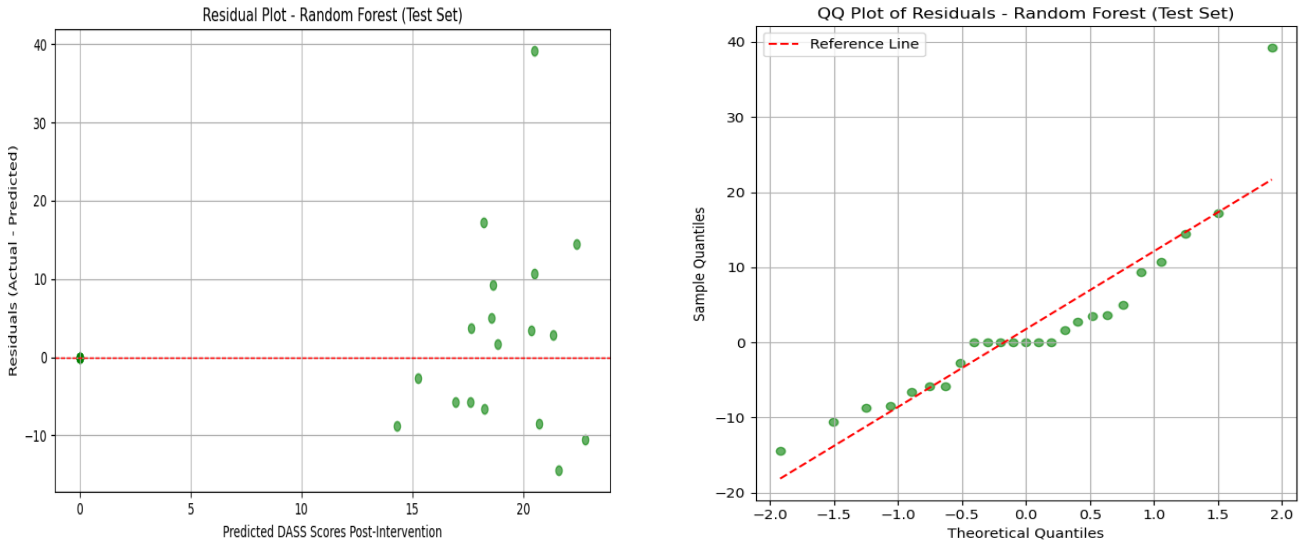

Figure 18: *Residuals for Random Forest Model (Saliva Cortisol ADULTS)*

Appendix F: Residuals for Random Forest Model (Post-Intervention Saliva Cortisol ADULTS)

The residuals indicate heteroscedasticity, and there is obvious bias, as seen in Figure 18. The model struggles to predict higher levels of saliva cortisol. The points are scattered to the right, violating the assumption of homoscedasticity.

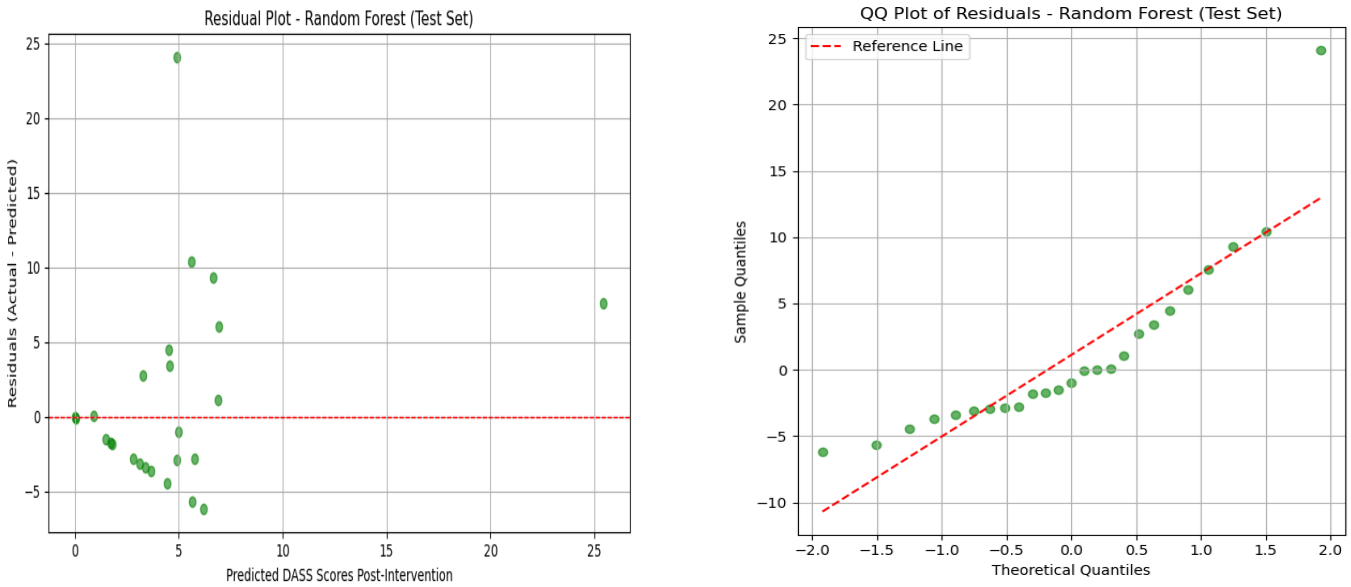

Figure 19: *Residuals for Random Forest Model (DASS Scores)*

Appendix G: Residuals for Random Forest Model (Post-Intervention DASS Scores)

The residuals indicate heteroscedasticity, and there is obvious bias, as seen in Figure 19. The model struggles to predict lower levels of DASS scores. The points are scattered to the left, violating the assumption of homoscedasticity.