Advanced Biostatistics Methods

Investigating the Impact of Various Treatments on the Recovery of Multiple Sclerosis

PROJECT BY: AYODEJI OLUWATOBILOBA EJIGBO

PANTHERID: 002703371

04/24/2024

INTRODUCTION

An idiopathic inflammatory illness of the central nervous system is multiple sclerosis. Axonal degeneration that follows demyelination is a pathogenic feature of it. This type of illness typically strikes young individuals, with twice as many women as men affected. According to PA (2004), common presenting symptoms include numbness, weakness, blurred vision, loss of balance, dizziness, urgency in the bladder, exhaustion, and depression.

Some individuals with the illness risk becoming unable to work or losing their capacity to walk independently. Others might need to experience protracted remissions without experiencing any new symptoms. Regretfully, there's no known treatment for multiple sclerosis currently. Treatments, however, can alleviate discomfort, manage symptoms, and hasten recovery from episodes.

This study aims to investigate two different treatment options for multiple sclerosis. Azathioprine (AZ), an immunosuppressive drug used to avoid rejection, is the initial course of treatment. It can be administered intravenously or consumed orally. The combination of methylprednisolone (MP) and AZ is the second therapy. A corticosteroid drug called MP reduces inflammation and suppresses the immune system. AFCR, a marker of immunological responses, was recorded following each treatment. A lower AFCR indicates a strengthening immune system. Put differently, this suggests the efficacy of the treatment.

This project will conduct a longitudinal analysis to investigate the effects of AZ and MP treatment on multiple sclerosis disease. The study questions are: (1) Do the effects of AZ and MP treatment differ from one another? and (2) Do the effects of prior treatments differ from one another? SAS 9.4 is the program utilized in this research.

METHODOLOGY

Data description

There are 150 observations in the dataset used for this study that are afflicted with the illness. Seventy-five of them (group 1) received AZ treatment, and the remaining seventy-five (group 2) received AZ plus MP treatment. The autoimmune measure, AFCR, was measured at the beginning of treatment (baseline, time 0) and then at 3, 6, 9, 12, 15, and 18 months. The information was obtained from the website.

- **Response Variable**: AFCR: measure of autoimmunity, the lower the better.
- **Predictor Variables:**
    1. **Time (in months)**: Baseline, 3, 6, 9, 12, 15 and 18 Group: Treatment group, 1= AZ, 2 = AZ+MP
    2. **Indicator**: Prior history on previous treatment with either of the study agents (0=no, 1=yes)
    3. **Age (in years):** age at baseline

Research Question

1. Do the immune system responses to AZ treatment alone and AZ plus MP treatment differ?
2. Does the immune system respond differently to past medical treatment?

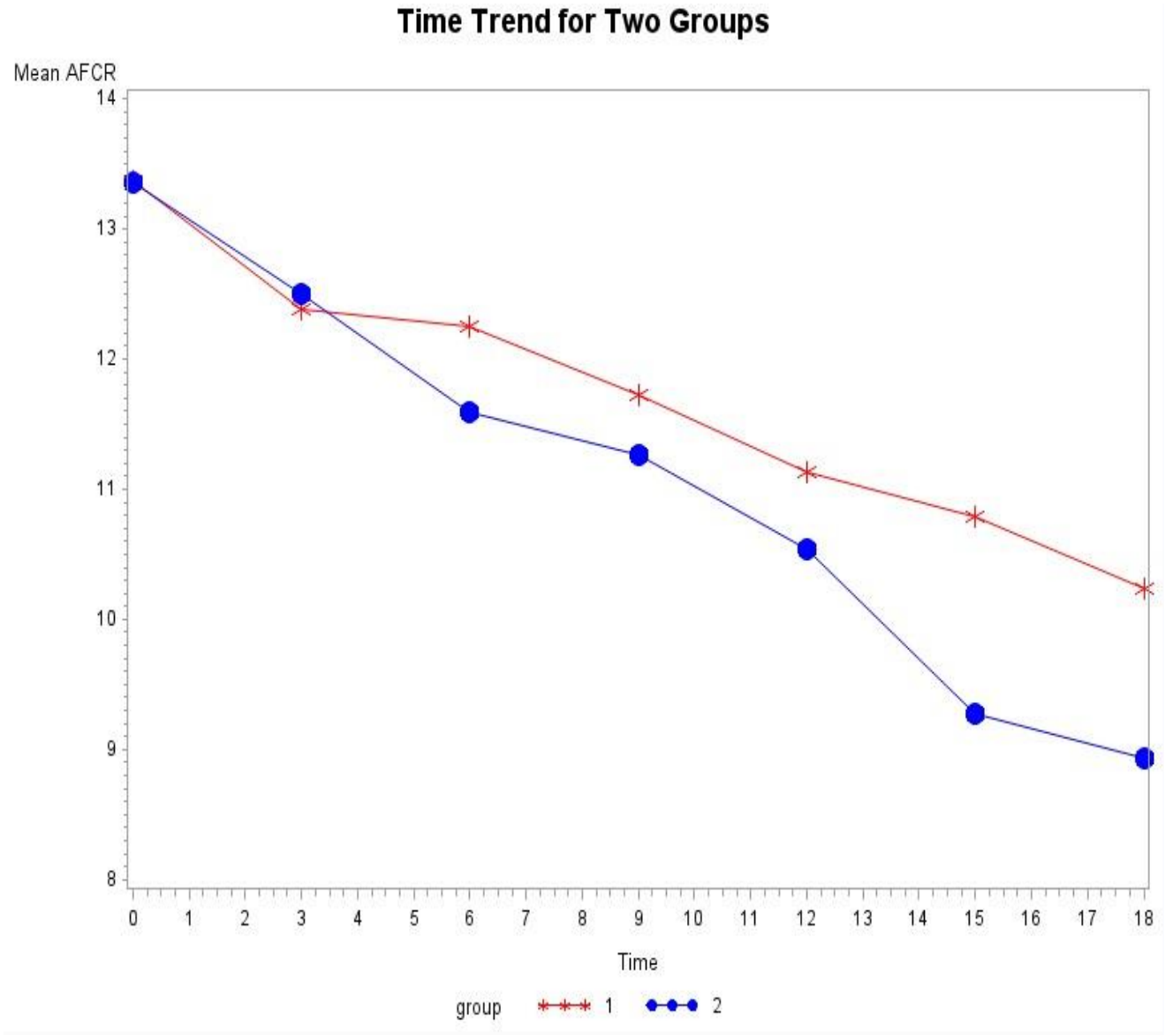
Descriptive Analysis

A few descriptive statistics of the AFCR for each time point are displayed for the two therapy groups in Table 1. Over time, both groups' mean AFCRs decline. The baseline mean AFCRs of the two groups are relatively similar, but the AZ+MP group has a lower mean AFCR at month eighteen. The mean AFCR of the AZ only group is lower than that of the AZ+MP group only at the third month.

Table 1: Descriptive Analysis for two groups at each time occasion

Analysis Variable : AFCR					
group	time	N Obs	Mean	Std Dev	Variance
1	0	59	13.3694915	2.3053352	5.3145704
	3	58	12.3827586	1.8197877	3.3116273
	6	62	12.2532258	1.9848374	3.9395796
	9	65	11.7276923	2.8025941	7.8545337
	12	62	11.1370968	2.4748727	6.1249947
	15	64	10.7890625	2.1612840	4.6711483
	18	59	10.2355932	2.4373362	5.9406078
2	0	64	13.3609375	2.3668757	5.6021007
	3	58	12.5000000	2.3141062	5.3550877
	6	62	11.5983871	2.4135201	5.8250793
	9	58	11.2603448	1.9181442	3.6792771
	12	63	10.5396825	2.0355343	4.1433999
	15	53	9.2698113	1.7845043	3.1844557
	18	60	8.9333333	2.3438102	5.4934463

The trend of the two groups' mean AFCR with respect to time is displayed in the time plot. The figure indicates that group 1's mean AFCR is higher than group 2's mean AFCR. Also, both groups' AFCR means fluctuate with time.



Variance-Covariance Matrix Selection

The following model will be applied in order to choose which variance-covariance matrix will be used in longitudinal analysis.

$$AFCR = \beta_0 + \beta_1 \times group + \beta_2 \times time1 + \beta_3 \times time2 + \beta_4 \times time3 + \beta_5 \times time4 + \beta_6 \times time5$$
$$+ \beta_7 \times time6 + \beta_8 time7 + \beta_9 \times indicator + \beta_{10} \times group \times time + \beta_{11} \times indicator \times time + \varepsilon$$

Where:

- $\beta_0 = intercept$,
- β_1 to $\beta_{11} = Coefficients$,

time1 = 1 when baseline, else = 0,

time2= 1 when at 3 months, else = 0,

time3 = 1 when at 6 months, else = 0,

time4 = 1 when at 9 months, else = 0,

time5 = 1 when at 12 months, else = 0,

time6 = 1 when at 15 months, else = 0,

time7 = 1 when at 18 months, else 0.

The likelihood-ratio test serves as the basis for selecting the variance-covariance matrix. This study takes into consideration five different types of matrixes:

- Compound symmetry.
- Heterogeneous compound symmetry.
- Unstructured covariance.
- Heterogeneous autoregressive.
- Autoregressive.

Each matrix type will be evaluated against a non-structured variance-covariance matrix to calculate the G2 statistic using the formula $2(\hat{l_{full}} - \hat{l_{reduced}})$. Table 3 presents the outcomes of selecting the variance-covariance matrix. All models are compared to the unstructured covariance model. The heterogeneous compound symmetry covariance model shows the lowest G^2 value at 13.7 with 7 degrees of freedom ($p < .0001$), suggesting it may adequately fit the data. Additionally, this model has the second lowest AIC. Therefore, for this project, we will employ the heterogeneous compound symmetry covariance model. Compound symmetry covariance model.

Table 3 Variance-Covariance Matrix Selection				
Model	***-2 Log-Likelihood***	***G_2***	***AIC***	***BIC***
(1) Unstructured Covariance	***3472.4***	***NA***	***3570.4***	***3717.9***
(2)Compound Symmetry	***3943.7***	***471.3, df=1(p<.0001)***	***3539.7***	***3609.0***
(3) Heterogeneous Compound Symmetry	***3486.1***	***13.7, df=7(p<.0001)***	***3544.1***	***3631.4***
(4) Autoregressive	***3587.0***	***114.6, df=1(p<.0001)***	***3633.0***	***3702.2***
(5) Heterogeneous Autoregressive	***3577.7***	***105.3, df=7 (p<.0001)***	***3635.7***	***3723.0***

Model Selection for Covariate

For model selection in this project, we will use the backward method. Table 4 provides the -2 log likelihood, G^2, AIC, and for each model, with comparisons made against Model 1. According to the table, the second model, which incorporates the interaction between time and treatment groups, reports the smallest G^2 = 3.1 (df = 1*, p-value <.0001*) and the lowest AIC at 3624.1. Due to its minimal AIC value, the second model will be selected as our preferred model for this analysis.

Model 1:

$$AFCR = \beta_0 + \beta_1 \times group + \beta_2 \times time + \beta_3 \times indicator + \beta_4 \times group \times time + \beta_5 indicator \times time + \varepsilon$$

Model 2:

$$AFCR = \beta_0 + \beta_1 \times group + \beta_2 \times time + \beta_3 \times indicator + \beta_4 \times group \times time + \varepsilon$$

Model 3:

$$AFCR = \beta_0 + \beta_1 \times group + \beta_2 \times time + \beta_3 \times group \times time + \varepsilon$$

Model 4:

$$AFCR = \beta_0 + \beta_1 \times group + \beta_2 \times time + \varepsilon$$

Model 5:

$$AFCR = \beta_0 + \beta_1 \times group + \beta_2 \times time + \beta_3 \times indicator + \beta_5 indicator \times time + \varepsilon$$

Model 6:

$$AFCR = \beta_0 + \beta_1 \times group + \beta_2 \times time + \beta_3 \times indicator + \varepsilon$$

Table 4 Model Selection for Covariate

Model	***-2 Log-Likelihood***	***G_2***	***AIC***	***BIC***
1	***3587.0***	***NA***	***3633.0***	***3702.2***
2	***3590.1***	***3.1, df=1(p<.0001)***	***3624.1***	***3675.2***
3	***3610.2***	***23.2,df=2(p<.0001)***	***3642.2***	***3690.4***
4	***3628.0***	***42.0,df=3(p<.0001)***	***3648.0***	***3678.1***
5	***3605.3***	***18.3,df=1(p<.0001)***	***3639.3***	***3690.5***
6	*3608.7*	*21.7,df=2(p<.0001)*	*3630.7*	*3663.8*

Model Selection for Random Effects

To evaluate subject-specific effects, we will perform model selection for random effects by exploring three linear mixed-effects models: Model A with a random intercept, Model B with a random slope, and Model C with both random intercept and slope. We will calculate G^2 by comparing the first two models (Model A and Model B) with Model C, which includes both random intercept and slope. According to the results shown in Table 5, Model A not only has the lowest G^2, but also the lowest AIC. Thus, we will proceed with Model A for our analysis.

Model A:

$$AFCR_{ij} = \beta_0 + \beta_1 \times group + \beta_2 \times time_{ij} + \beta_3 \times indicator + \beta_4 \times group \times time_{ij} + b_{1i} + \varepsilon_{ij}$$

Model B:

$$AFCR_{ij} = \beta_0 + \beta_1 \times group + \beta_2 \times time_{ij} + \beta_3 \times indicator + \beta_4 \times group \times time_{ij} + b_{2i} \times time_{ij} + \varepsilon_{ij}$$

Model C:

$$AFCR_{ij} = \beta_0 + \beta_1 \times group + \beta_2 \times time_{ij} + \beta_3 \times indicator + \beta_4 \times group \times time_{ij} + b_{1i} + b_{2i} \times time_{ij} + \varepsilon_{ij}$$

where $AFCR_{ij}$ = j[th] response on the i[th] subject.

Table 5 Model Selection for Random Effects

Model	*-2 Log-Likelihood*	*G^2(Compare with C)*	*AIC*	*BIC*
A	*3505.3*	*-107.8, df=1(p>0.1)*	*3521.3*	*3545.4*
B	*3627.5*	*14.4,df=1(p<.0001)*	*3643.5*	*3667.6*
C	*3613.1*	*NA*	*3627.1*	*3648.2*

Model Selection for Quadratic Terms

It appears that your sentence was cut off. Could you please provide more details or complete your sentence regarding the quadratic model (Model Q)? This will help me provide a more accurate and helpful response.

$$AFCR_{ij} = \beta_0 + \beta_1 \times group + \beta_2 \times time_{ij} + \beta_3 \times indicator + \beta_4 \times group \times time_{ij} + \beta_5 \times time_{ij} \times time_{ij} + \beta_6 \times indicator \times time_{ij} \times time_{ij} + b_{1i} + \varepsilon_{ij}$$

Table 6 Model Selection for Quadratic Terms				
Model	-2 Log Likelihood	G^2	AIC	BIC
A	3505.3	-0.2, df=1(*p=0.87*)	3521.3	3545.4
Q	3505.1	NA	3525.1	3555.2

Table 6 shows the difference between Model A and Model Q. We use the Model Q as the base Model. Model A has smaller AIC and the p-value for G^2 is greater than 0.05. That means the liner model is adequate to fit the data.

To sum up the model selection process, Model B is the final model. The following analysis will be based on Model A, which is

$$AFCR_{ij} = \beta_0 + \beta_1 \times group + \beta_2 \times time_{ij} + \beta_3 \times indicator + \beta_4 \times group \times time_{ij} + b_{1i} + \varepsilon_{ij}$$

RESULTS

In the methodology section, we opted for the heterogeneous compound symmetry covariance model. Table 7 presents the results, showing the estimated covariance and correlation matrices under this model. Specifically, the estimated variance of the random intercept is 2.0081, indicating variability in baseline measurements among subjects. Additionally, the estimated variance of within-subject measurement errors is 2.7557, reflecting the variability in observations within each subject over time.

Table 7: Estimated Covariance and Correlation Matrices

Estimated G Matrix			
Row	Effect	id	Col1
1	Intercept	1	2.0081

Estimated G Correlation Matrix			
Row	Effect	id	Col1
1	Intercept	1	1.0000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Variance	id	2.0081
AR(1)	id	0
Residual		2.7557

Table 8 The final model for AFCR (Air Flow Capacity Rate) based on the coefficients presented in Table 8 is given by the following equation:

$$AFCR = 12.4386 + 0.2649 \times group - 0.1543 \times time + 1.0048 \times indicator - 0.09129 \times group \times time$$

- From the estimated final model, we can say when there is no previous treatment, at baseline the mean AFCR for AZ group is 12.4386.
- The mean slope is -0.1543, meaning the mean AFCR for AZ group with no previous treatment decreases 0.1543 every three months.
- The mean intercept of treatment group is 12.7035, indicating at baseline the AZ+MP group with no previous treatment experience has a mean AFCR which is 12.7035.
- The estimated different effect from two groups is 0.2649 (p = 0.3967). If we control the treatment groups, increasing 1 unit of indicator will increase the baseline mean AFCR by 1.0048 (*p-value <.001*).
- The group variable is not significant (*p-value = 0.3967*) on mean AFCR, but the effect of different treatment on the change of mean AFCR over time is significant (*p-value <.0001*).

- Previous treatment experience is significant on the mean AFCR (*p-value <.001*). However, from table 9 we can see, the previous treatment experience does not have significant effect on the change of mean AFCR over time (*p-value = 0.6284*).

Table 8: Output of Final Model

Solution for Fixed Effects						
Effect	group	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		12.4386	0.2824	147	44.04	<.0001
group	2	0.2649	0.3124	695	0.85	0.3967
group	1	0	.	.	.	.
time		-0.1543	0.01370	695	-11.26	<.0001
indicator		1.0048	0.2667	695	3.77	0.0002
time*group	2	-0.09129	0.01933	695	-4.72	<.0001
time*group	1	0	.	.	.	.

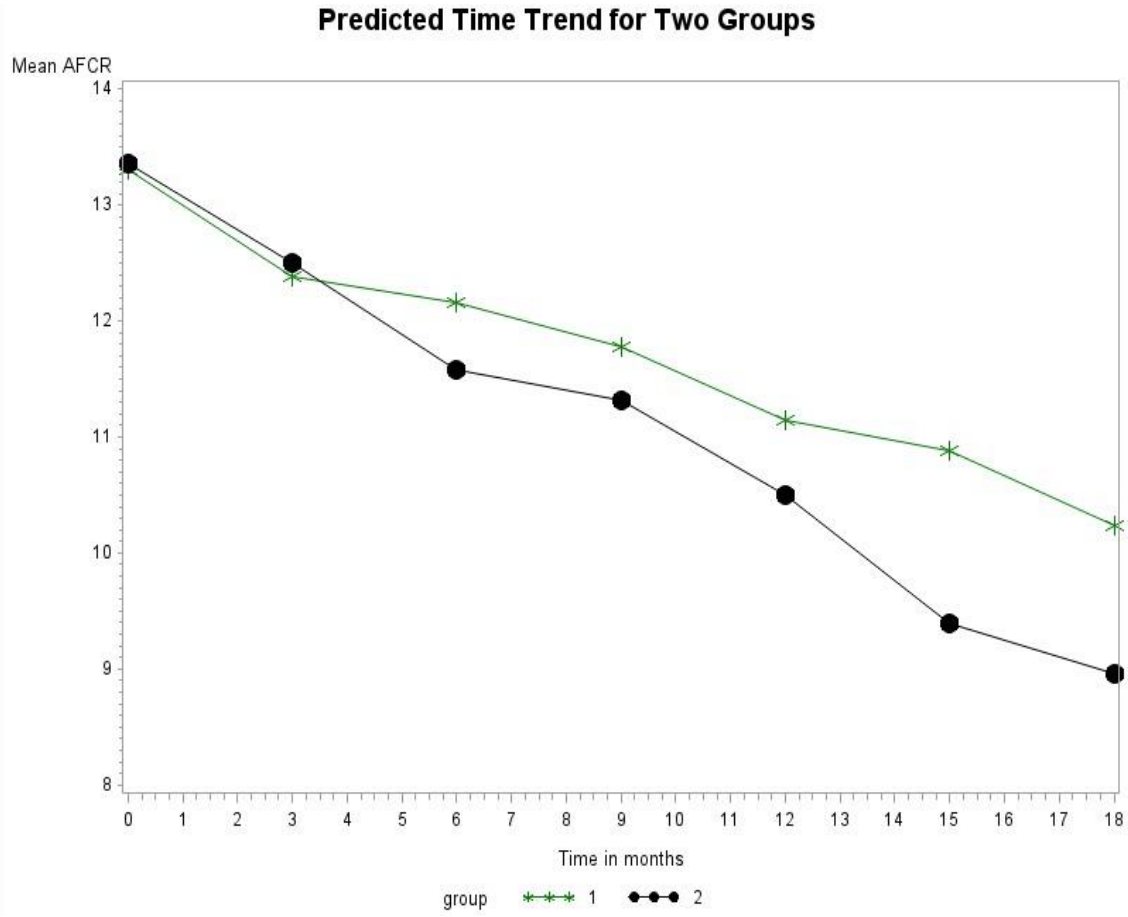
Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
group	1	695	0.72	0.72	0.3964	0.3967
time	1	695	427.97	427.97	<.0001	<.0001
indicator	1	695	14.19	14.19	0.0002	0.0002
time*group	1	695	22.30	22.30	<.0001	<.0001

Table 9: Output of Full Model

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
group	1	679	4.70	4.70	0.0302	0.0305
time	6	679	182.81	30.47	<.0001	<.0001
indicator	1	679	14.41	14.41	0.0001	0.0002
group*time	6	679	25.12	4.19	0.0003	0.0004
indicator*time	6	679	4.36	0.73	0.6284	0.6286

The time plot illustrates the trend of the average Air Flow Capacity Rate (AFCR) for two groups over time based on predicted data. The plot reveals that the average AFCR for Group 1 is higher than that for Group

2. Additionally, the average AFCR for both groups varies over time. The plot created using predicted data closely resembles the plot using the original data.



COUNCLUSION

Based on the analysis results, the initial AFCR (Air Flow Capacity Rate) is lower in the AZ treatment group compared to the AZ + MP (Methylprednisolone) group. Although the difference in treatment effects on AFCR is not statistically significant, long-term treatment with AZ + MP has been observed to decrease AFCR, suggesting a potential recovery of the immune system with prolonged treatment.

Previous experience with treatment significantly influences AFCR levels, although it does not significantly impact the change in AFCR over time.

While we conclude that AZ + MP treatment can reduce AFCR over time based on the current data, additional follow-up data is necessary to confirm these findings. Particularly, data from censoring events are crucial as they can provide more insights into the effectiveness of the treatment. Moreover, increasing the number of observations could lead to a more robust model, enhancing the precision and reliability of our conclusions.

REFERENCE

Calabresi PA. Diagnosis and management of multiple sclerosis. Am Fam Physician, 2004, 70(10).

Appendix:

SAS CODE:

```
/* Read data from a text file */
data afcr;
   infile 'C:\Users\ay\afcr.txt' dsd missover;
   informat id 1. time 2. AFCR 4.1 group 1. indicator 1. age 2.;
   input id $ 1-3 time 6-7 AFCR 10-13 group $ 17-18 indicator $ 21-22 age 24-25;
run;

/* Sort data by group and time */
proc sort data=afcr;
   by group time;
run;

/* Perform descriptive analysis */
title 'Descriptive Analysis';
proc means data=afcr mean std var;
   var AFCR;
   class group time;
   output out=data2 mean=mean_data;
run;

/* Means separated by group and time */
proc means data=afcr mean std var;
   var AFCR;
   by group time;
   output out=data2(drop=_TYPE_ _FREQ_) mean=mean_data;
run;

/* Reset graphics options and set axis labels */
goptions reset=all;
axis1 label=('Mean AFCR');
axis2 label=('Time');

/* Plot the time trend for the two gender groups */
title 'Time Trend for Two Groups';
proc gplot data=data2;
```

```
   plot mean_data*time=group / haxis=axis2 vaxis=axis1;
   symbol1 v=star c=red h=2 i=join;
   symbol2 v=dot c=blue h=2 i=join;
run;

/* Create binary indicators for each time */
data afcr2;
   set afcr;
   array times [7] time1-time7;
   do i = 1 to 7;
      if time=(i-1)*3 then times[i]=1;
      else times[i]=0;
   end;
   drop i;
run;

/* Unstructured Covariance */
title 'Unstructured Covariance';
proc mixed data=afcr method=ml;
   class id group(ref='1') time(ref='0');
   model AFCR=group indicator time group*time indicator*time / solution CHISQ outpm=pred;
   repeated time / type=UN subject=id R RCORR;
run;

/* Compound Symmetry Covariance */
title 'Compound Symmetry Covariance';
proc mixed data=afcr method=ml;
   class id group(ref='1') time(ref='0');
   model AFCR=group time indicator group*time indicator*time / solution CHISQ outpm=pred;
   repeated time / type=CS subject=id R RCORR;
run;

/* Heterogeneous Compound Symmetry Covariance */
title 'Heterogeneous Compound Symmetry Covariance';
proc mixed data=afcr method=ml;
   class id group(ref='1') time(ref='0');
   model AFCR=group time indicator group*time indicator*time / solution CHISQ outpm=pred;
   repeated time / type=CSH subject=id R RCORR;
run;

/* Autoregressive Covariance */
title 'Autoregressive Covariance';
proc mixed data=afcr method=ml;
   class id group(ref='1') time(ref='0');
```

```
   model AFCR=group time indicator group*time indicator*time / solution CHISQ outpm=pred;
   repeated time / type=AR(1) subject=id R RCORR;
run;

/* Heterogeneous Autoregressive Covariance */
title 'Heterogeneous Autoregressive Covariance';
proc mixed data=afcr method=ml;
   class id group(ref='1') time(ref='0');
   model AFCR=group time indicator group*time indicator*time / solution CHISQ outpm=pred;
   repeated time / type=ARH(1) subject=id R RCORR;
run;

/* Model Selection */
/* Model specifications are abbreviated for clarity, you can include other predictors if needed */

title 'Model 1';
/* Code for Model 1 - same as above, just an example */
proc mixed data=afcr method=ml;
   /* ... */
run;

/* Repeat for Models 2 through 6 with the appropriate model specifications */

/* Model Selection for Random Effects */
/* Model A: Random Intercept */
title 'Model A: Random Intercept';
proc mixed data=afcr method=ml;
   class id group(ref='1');
   model AFCR=group time indicator group*time / S CHISQ outpm=pred;
   random intercept / type=AR(1) subject=id g gcorr S;
run;

/* Model B: Random Slope */
title 'Model B: Random Slope';
proc mixed data=afcr method=ml;
   /* ... similar structure */
run;

/* Model C: Random Intercept and Slope */
title 'Model C: Random Intercept and Slope';
proc mixed data=afcr method=ml;
   /* ... similar structure */
run;
```

```
/* Model Selection for Quadratic Terms */
title 'Quadratic Terms';
proc mixed data=afcr method=ml;
   /* ... */
run;

/* Observe the Difference Between Observed Data and Predicted Data */
title 'Observed vs Predicted Comparison';
proc mixed data=afcr method=ml;
   /* ... model specification similar to Model A */
run;

proc means data=pred mean std var;
   var pred;
   by group time;
   output out=AFCR_pred mean=mean_AFCR_pred;
run;

goptions reset=all;
axis1 label=('Mean AFCR');
axis2 label=('Time in months');

symbol1 v=star c=green h=2 i=join;
symbol2 v=dot c=black h=2 i=join;

title 'Predicted Time Trend for Two Groups';
proc gplot data=AFCR_pred;
   plot mean_AFCR_pred*time=group / haxis=axis2 vaxis=axis1;
run;

/* Full Model */
title 'FULL MODEL';
proc mixed data=afcr method=ml;
   /* ... model specification similar to Model A */
run;
```