PharmaSUG 2025 – Paper OS-182

Interactive Longitudinal Data Analysis and Visualization in Clinical Research Using R

María José Gómez Ramírez, Intego Group

ABSTRACT

Analyzing a longitudinal study with clinical data allows us to understand changes that have occurred over time in patient outcomes (such as biomarkers, the effects of treatments, and disease progression, among others). Graphical visualization of the data is crucial, as it enables decision-makers to interpret analytics in a visual format, making it easier to grasp complex concepts while identifying new patterns. Moreover, when you can interact dynamically with the application presenting the data, they can deepen their understanding of an event or uncover specific insights.

The use of R and its different free available packages makes it possible to create user-friendly data visualization applications that allow for diverse interactive forms. This paper will demonstrate an example of an application using R packages such as “*shiny”, “lme4”, “plotly”, “DT”* and others, to present an interactive visualization of a longitudinal data analysis using the Admiral’s publicly available dataset.

INTRODUCTION

In clinical research, a longitudinal study design permits the discovery of individual characteristics that can explain interindividual differences in changes in the study object’s outcomes (biomarkers, diseases, treatment effects, among others) over time (Fitzmaurice et al., 2011). Evaluating the results of a longitudinal analysis can provide important information for medical decision making and enable more effective interventions.

Various software solutions facilitate data analysis. Specifically, R is an open-source program with a wide range of packages that makes it a powerful tool for data analysis and visualization. The integration of this software with statistical models and machine learning techniques allows prediction of responses and identification of relevant data patterns. Additionally, incorporating interactive visualizations enhances exploration and understanding of results, making R a flexible platform for longitudinal analysis.

The shiny package in R allows the combination of various R packages for visualization in an HTML-style page, where you can access information interactively and in a user-friendly manner. This paper presents an example of a shiny application to visualize a longitudinal study dataset provided by the Admiral library.

APPLICATION

The application developed for this article was structured into modules. This modular approach makes it easier to update the R code when changes or improvements need to be made. In this case, four distinct modules are presented: Data, Statistics, Model, and Machine Learning, which are described below.

DATA

This serves as the homepage. It allows the uploading of .csv files and displays the data once uploaded. Additionally, it includes a column selection filter for the dataset. Figure 1 shows the initial display of the application, and Figure 2 shows the screen after selecting a database. The interface provides global filtering via a search bar outside the table and column-specific filtering options. These features are facilitated by the DT package in R, which specializes in handling tables in HTML format.

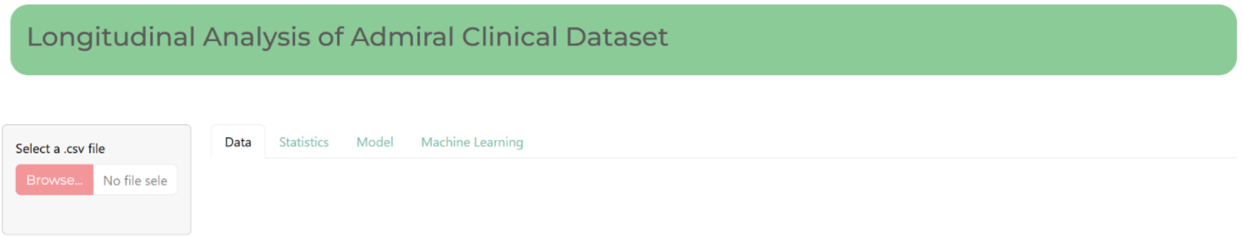
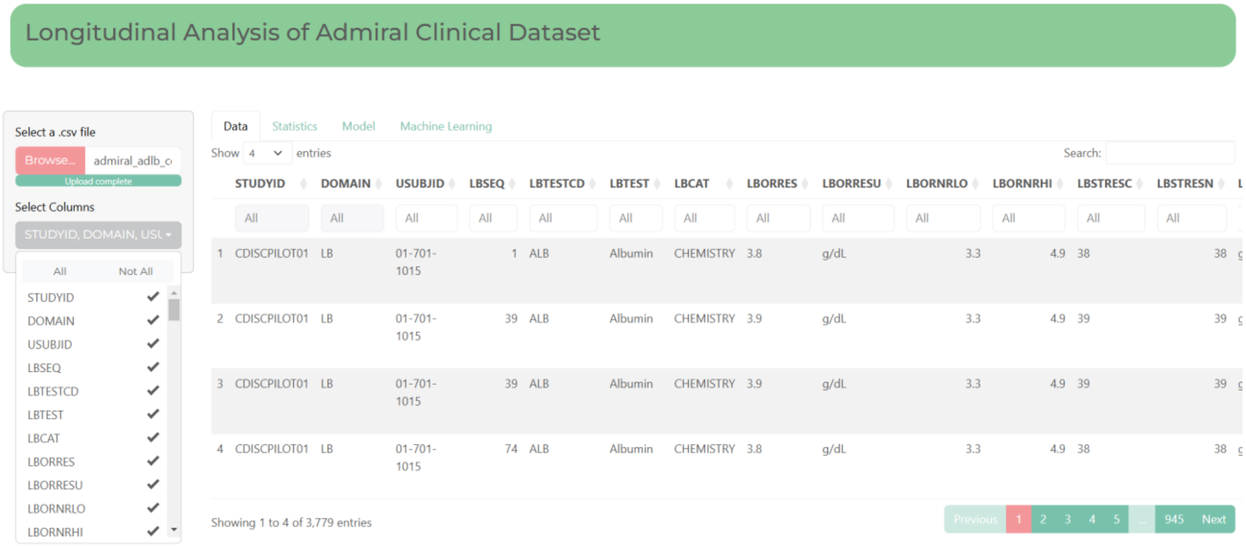


Figure 1. *Data tab*



	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBORRESU	LBORNRLO	LBORNRHI	LBSTRESC	LBSTRESN
	All	All	All	All	All	All	All	All	All	All	All	All	All
1	CDISCPILOT01	LB	01-701-1015	1	ALB	Albumin	CHEMISTRY	3.8	g/dL	3.3	4.9	38	38
2	CDISCPILOT01	LB	01-701-1015	39	ALB	Albumin	CHEMISTRY	3.9	g/dL	3.3	4.9	39	39
3	CDISCPILOT01	LB	01-701-1015	39	ALB	Albumin	CHEMISTRY	3.9	g/dL	3.3	4.9	39	39
4	CDISCPILOT01	LB	01-701-1015	74	ALB	Albumin	CHEMISTRY	3.8	g/dL	3.3	4.9	38	38

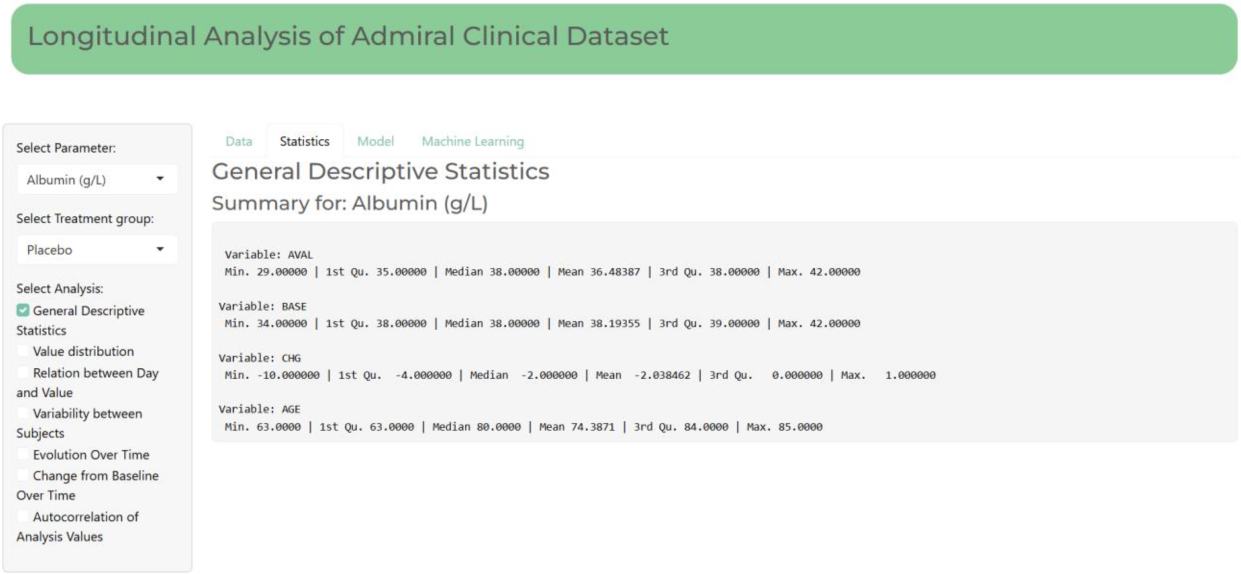
Showing 1 to 4 of 3,779 entries

Figure 2. *Table display*

STATISTICS

This module displays descriptive statistical graphs that differentiate based on the selected filter parameter and shows boxes for the selected variables in the left-side menu. By default, upon entering the tab, a summary of some numerical variables is displayed. The graphs are created using “*plotly”*, which provides interactive engagement, such as obtaining information based on mouse position, zooming in or out on sections of the graph, and more.

The left-side menu allows you to filter graphs by biomarker and treatment group. Figure 3 illustrates the initial display of the Statistics tab.



Longitudinal Analysis of Admiral Clinical Dataset

Select Parameter: Albumin (g/L)

Select Treatment group: Placebo

Select Analysis:
- ☑ General Descriptive Statistics
- ☐ Value distribution
- ☐ Relation between Day and Value
- ☐ Variability between Subjects
- ☐ Evolution Over Time
- ☐ Change from Baseline Over Time
- ☐ Autocorrelation of Analysis Values

General Descriptive Statistics

Summary for: Albumin (g/L)

```
 Variable: AVAL
 Min. 29.00000 | 1st Qu. 35.00000 | Median 38.00000 | Mean 36.48387 | 3rd Qu. 38.00000 | Max. 42.00000

Variable: BASE
 Min. 34.00000 | 1st Qu. 38.00000 | Median 38.00000 | Mean 38.19355 | 3rd Qu. 39.00000 | Max. 42.00000

Variable: CHG
 Min. -10.000000 | 1st Qu.  -4.000000 | Median  -2.000000 | Mean  -2.038462 | 3rd Qu.   0.000000 | Max.   1.000000

Variable: AGE
 Min. 63.0000 | 1st Qu. 63.0000 | Median 80.0000 | Mean 74.3871 | 3rd Qu. 84.0000 | Max. 85.0000
```

Figure 3. *Statistics tab and General Descriptive Statistics*

You can access the distribution of values by parameter, where the frequency of analysis values is visualized in a histogram. Additionally, a boxplot displays the distribution, outliers, and trends observed across different visits (Figure 4).

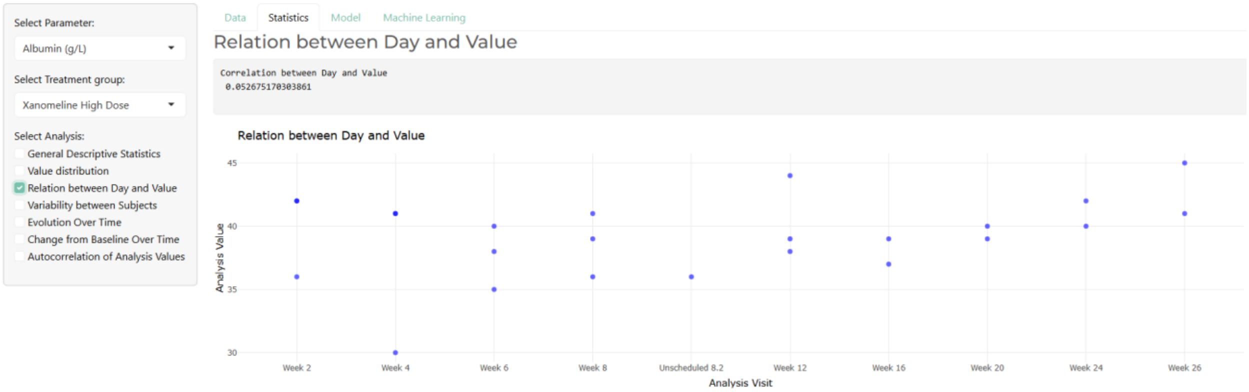
Select Parameter: Albumin (g/L)

Select Treatment group: Xanomeline High Dose

Select Analysis:
- ☐ General Descriptive Statistics
- ☑ Value distribution
- ☐ Relation between Day and Value
- ☐ Variability between Subjects
- ☐ Evolution Over Time
- ☐ Change from Baseline Over Time
- ☐ Autocorrelation of Analysis Values

Figure 4. *Value Distribution*

The relationship between the visit and the analysis value is illustrated through correlation metrics and scatter plots, helping identify potential patterns or trends over time (Figure 5).


Select Parameter: Albumin (g/L)

Select Treatment group: Xanomeline High Dose

Select Analysis:
- ☐ General Descriptive Statistics
- ☐ Value distribution
- ☑ Relation between Day and Value
- ☐ Variability between Subjects
- ☐ Evolution Over Time
- ☐ Change from Baseline Over Time
- ☐ Autocorrelation of Analysis Values

```
Correlation between Day and Value
 0.052675170303861
```

Figure 5. *Relation between Day and Value*

Inter-subject variability is analyzed using boxplots that compare the distribution of analysis values across different subjects, along with a standard deviation plot to assess variability within individual cases (Figure 6).


Select Parameter: Albumin (g/L)

Select Treatment group: Xanomeline High Dose

Select Analysis:
- ☐ General Descriptive Statistics
- ☐ Value distribution
- ☐ Relation between Day and Value
- ☑ Variability between Subjects
- ☐ Evolution Over Time
- ☐ Change from Baseline Over Time
- ☐ Autocorrelation of Analysis Values

Figure 6. *Variability between Subjects*

The evolution of values over time is displayed by showing the average values per visit with a LOESS smoothed trend line to visualize changes over time. Additionally, a line graph for each subject represents individual trajectories, and a boxplot illustrates distributional changes and variability in the parameter over time (Figure 7).

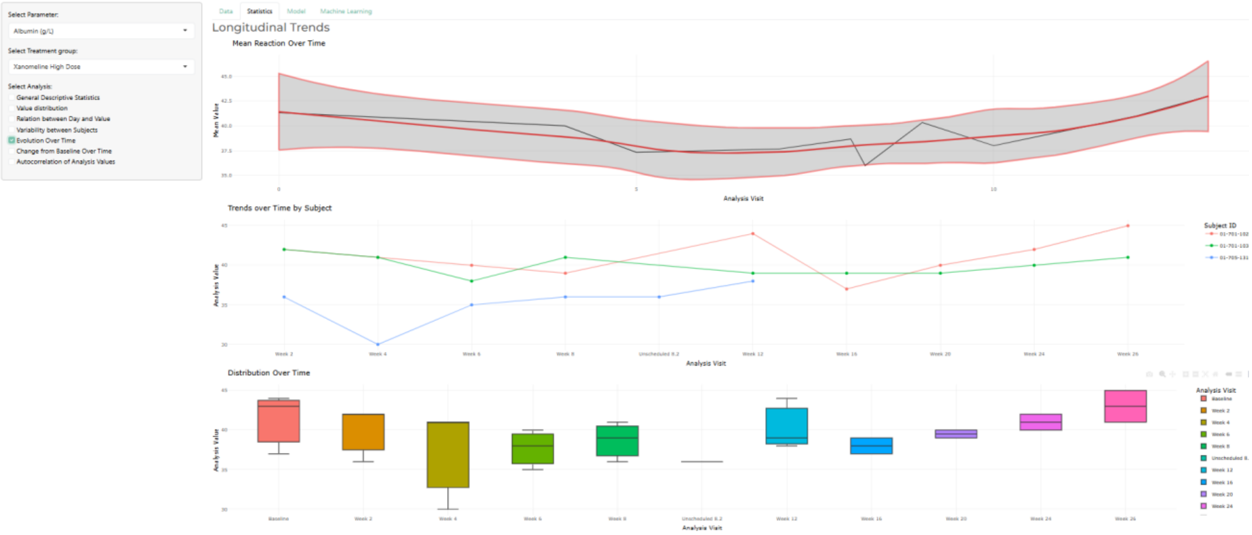

Select Parameter:
Albumin (g/L)
Select Treatment group:
Xanomeline High Dose
Select Analysis:
☐ General Descriptive Statistics
☐ Value distribution
☐ Relation between Day and Value
☐ Variability between Subjects
☑ Evolution Over Time
☐ Change from Baseline Over Time
☐ Autocorrelation of Analysis Values

Figure 7. *Evolution over Time*

Changes from baseline are analyzed using a line plot that displays the evolution of changes from baseline per visit, and a boxplot that visualizes the distribution of changes across different visits (Figure 8).

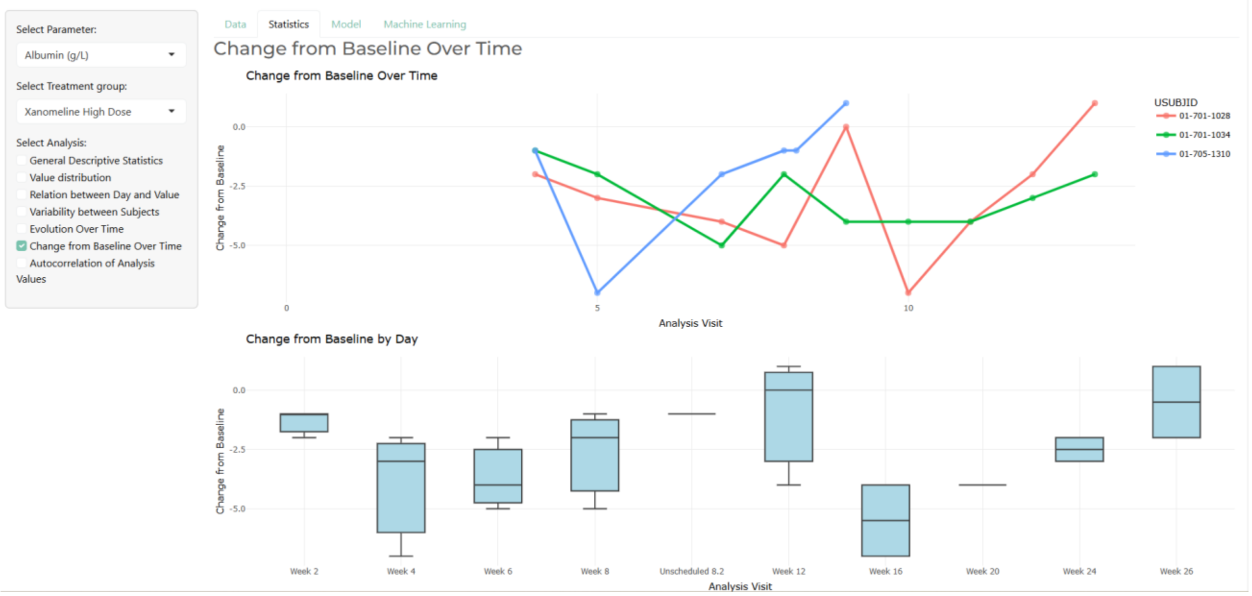
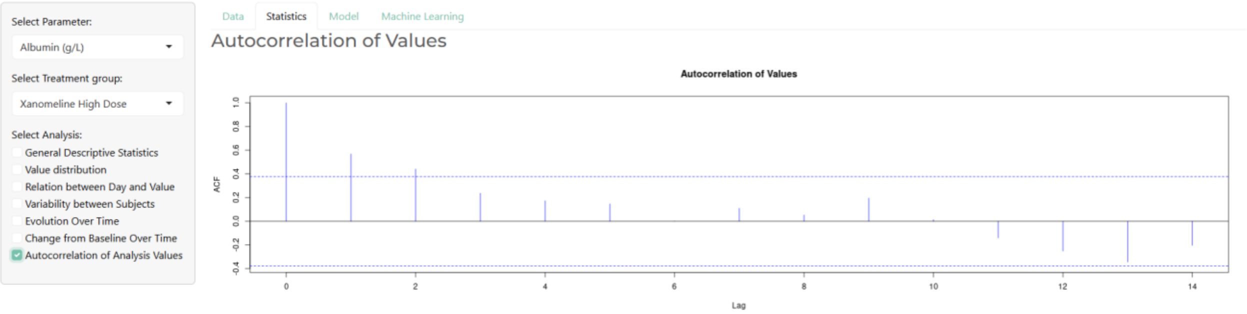

Select Parameter:
Albumin (g/L)
Select Treatment group:
Xanomeline High Dose
Select Analysis:
☐ General Descriptive Statistics
☐ Value distribution
☐ Relation between Day and Value
☐ Variability between Subjects
☐ Evolution Over Time
☑ Change from Baseline Over Time
☐ Autocorrelation of Analysis Values

Figure 8. *Change from Baseline over Time*

Finally, the autocorrelation of analysis values is presented, showing the correlation of a value with its previous measurements to evaluate whether there are persistent or delayed effects in the data. (Figure 9).



Select Parameter: Albumin (g/L)

Select Treatment group: Xanomeline High Dose

Select Analysis:

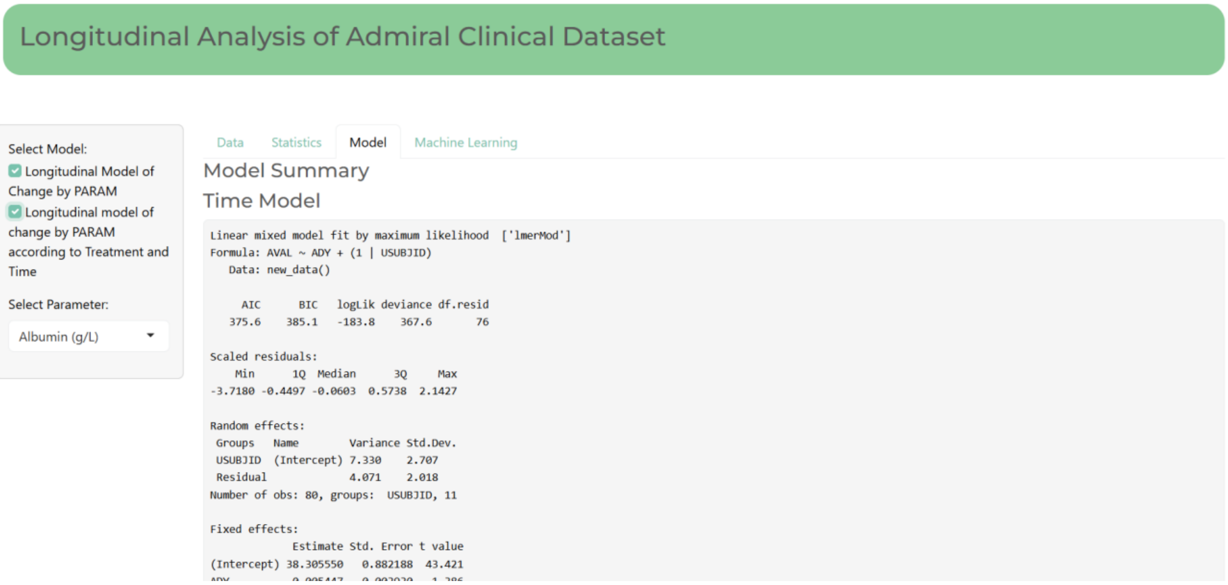
- ☐ General Descriptive Statistics
- ☐ Value distribution
- ☐ Relation between Day and Value
- ☐ Variability between Subjects
- ☐ Evolution Over Time
- ☐ Change from Baseline Over Time
- ☑ Autocorrelation of Analysis Values

Figure 9. *Autocorrelation of Analysis Value*

MODEL

This module allows to compare two linear mixed-effects models with different parameters, facilitated by the *"lmer"* function from the "*lme4*" package in R. Upon selecting a model, a summary, convergence diagnostics, and normality check graphs are displayed. If both models are selected, an additional comparison is made using the AIC and BIC methods.

These models assess differences in analysis values over time while considering intersubject variability and treatment differences over time. Figures 10 and 11 illustrate the model tab and results comparison.



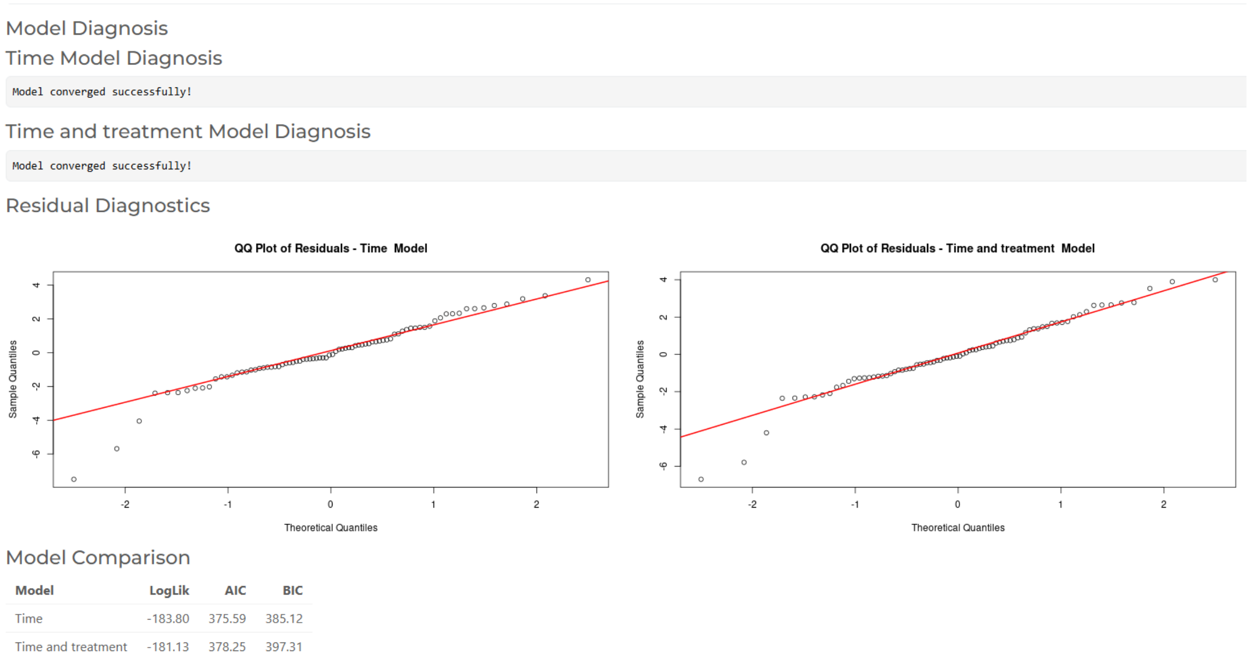
Select Model:

- ☑ Longitudinal Model of Change by PARAM
- ☑ Longitudinal model of change by PARAM according to Treatment and Time

Select Parameter: Albumin (g/L)

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: AVAL ~ ADY + (1 | USUBJID)
   Data: new_data()

     AIC      BIC   logLik deviance df.resid
   375.6    385.1   -183.8    367.6       76

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7180 -0.4497 -0.0603  0.5738  2.1427

Random effects:
 Groups   Name        Variance Std.Dev.
 USUBJID  (Intercept) 7.330    2.707
 Residual             4.071    2.018
Number of obs: 80, groups:  USUBJID, 11

Fixed effects:
             Estimate Std. Error t value
(Intercept) 38.305550   0.882188  43.421
```

Figure 10. *Model summary*



Model Diagnosis

Time Model Diagnosis

```
Model converged successfully!
```

Time and treatment Model Diagnosis

```
Model converged successfully!
```

Residual Diagnostics

Model Comparison

Model	LogLik	AIC	BIC
Time	-183.80	375.59	385.12
Time and treatment	-181.13	378.25	397.31

Figure 11. *Results of the model's comparison*

MACHINE LEARNING

This tab was created for exploring machine learning models. In this case, Random Forest and XGBoost are implemented to predict analysis values based on study day, subject, and treatment. The models display variable importance and Root Mean Squared Error (RMSE) for the selected parameter in the left-side menu. These models are powered by the "*randomForest*" and "*xgboost*" packages in R.

The main objective is to evaluate how well the model predicts analysis values, identify the most impactful variables, and detect potential patterns in treatment response. Figure 12 shows the screen layout of the machine learning results.

Longitudinal Analysis of Admiral Clinical Dataset

Select Parameter: Albumin (g/L)

Select ML Model: Random Forest

Model Performance

```
Root Mean Squared Error (RMSE)
 1.255
```

Feature Importance

Figure 12. *Machine learning tab*

CONCLUSIONS

Longitudinal data analysis enables the study of changes in various research variables over time. When applied to clinical data, it focuses on biomarkers, treatment effects, and other relevant aspects. Interactive visualization tools enhance interpretation by enabling more detailed, intuitive, and flexible analyses.

Using R as a platform provides a wide range of packages and functions for longitudinal analysis, including "*shiny*", "*plotly*", "*lme4*", "*DT*", and others. Combining these tools allows for the development of applications that enhance data exploration, significantly improving analysis and decision-making in medical research by presenting information in formats that are more interactive and easier to understand.

Furthermore, the modularization provided by shiny ensures that applications remain adaptable to evolving research needs, allowing for flexibility in updating data sources, modifying visualizations, and integrating new analytical methods. This adaptability is particularly valuable in clinical studies, where continuous monitoring and iterative analyses are necessary for assessing treatment efficacy and patient outcomes.

REFERENCES

1. Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). Applied Longitudinal Analysis. Wiley Series in Probability and Statistics. doi:10.1002/9781119513469
2. Shiny, https://cran.r-project.org/web/packages/shiny/shiny.pdf
3. plotly, https://cran.r-project.org/web/packages/plotly/plotly.pdf
4. lme4, https://cran.r-project.org/web/packages/lme4/lme4.pdf
5. DT, https://cran.r-project.org/web/packages/DT/DT.pdf
6. randomForest, https://cran.r-project.org/web/packages/randomForest/randomForest.pdf
7. xgboost, https://cran.r-project.org/web/packages/xgboost/xgboost.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

María José Gómez Ramírez

Intego Group

maria.gomez@intego-group.com