Supplement to Example Manuscript Template for a Data Analysis Project

# 1. Overview

This supplement has a few exploratory figures and discussion of additional methods that can be used for disproportionality analysis.

# 2. Code and file information

To reproduce this study: First run the processing code file. This processed the raw data in the ‘data’ subfolder and saves it as an rds file. Then, run the exploratory data (eda) code file. This file produces summary tables and figures. The figures are available in this supplement. Then, run the analysis file. This conducts a disproportionality analysis and runs a random forest model to assess serious vs non-serious cases based on age, sex, and ICI.

# 3. Additional Method Details

Three machine learning models were fit - logistic regression, decision tree, and random forest. We found that the models were not able to distinguish adequately between the serious and non-serious reactions. This most probably occurs due to the class imbalance - 305 serious, 60 non-serious cases.

# 4. Additional results

[Figure 1](#fig-result) shows the number of reactions by drug, along with the seriousness of the reaction.

here() starts at C:/Pooja/Pooja/Spring 2025/EPID-BIOS-8060E-MADA/Gokhale-MADA-Project

|  |
| --- |
| Figure 1: Bar chart of outcome of immune checkpoint inhibitors |

We can see that a large proportion of the reactions are serious.

[Figure 2](#fig-result1) shows the bar chart by type of organ system involved in the reaction.

|  |
| --- |
| Figure 2: Bar chart of types of reactions |

The most common reactions are dermatological, followed by gastrointestinal.

[Figure 3](#fig-result2) shows the ROC curve for the logistic regression model.

|  |
| --- |
| Figure 3: ROC Curve - Logistic Regression |

We see that the curve covers about 60% of the area.

[Figure 4](#fig-result3) shows the ROC curve for the decision tree model.

|  |
| --- |
| Figure 4: ROC Curve - Decision Tree |

The curve is a staright diagonal line (covering about 50% of the area). Thus, this model is not able to distinguish between serious and non-serious cases better than at random.

[Figure 5](#fig-result4) shows the ROC curve for the random forest model.

|  |
| --- |
| Figure 5: ROC Curve - Random forest |

We see that the curve covers about 80% of the area.

# 5. Discussion

Overall, the random forest model performs better. However, all three models are not able to distinguish between serious and non-serious reactions, possibly due to the class imbalance.