# Detailed Documentation: Univariate Analysis of Factors Affecting Healing Duration in Polytrauma Patients

## Project Overview

This document provides comprehensive documentation of the univariate analysis performed to identify factors affecting healing duration in polytrauma patients. The analysis is part of the broader Polytrauma Analysis Project, which aims to understand the factors influencing recovery time and outcomes for patients with multiple traumatic injuries.

## Analysis Purpose and Scope

The univariate analysis was designed to examine how individual factors (without considering interactions) impact the healing duration of polytrauma patients. Specifically, the analysis focused on:

1. The impact of injuries to specific body parts
2. The relationship between the number of injured body parts and healing duration
3. The effect of demographic factors (age and gender) on healing time

The analysis serves as the foundation for subsequent multivariate analysis and potential predictive modeling of healing duration.

## Data Source and Preparation

### Source Data

* Primary dataset: Polytrauma\_Analysis\_Processed.xlsx (from step1 of the processing pipeline)
* The dataset contained 153 records representing 30 unique patients
* Each record represents a patient visit, with multiple visits per patient

### Patient-Level Data Preparation

* A patient-level dataset was created with one row per patient (n=30)
* Healing duration was defined as days from accident to last recorded visit
* For categorical fields (e.g., presence of injuries), a patient was marked as “Ja” if any visit showed “Ja”
* For each patient, information on injuries to different body parts was extracted
* Key statistics for healing duration: Min=182 days, Max=1233 days, Mean=579.4 days, Median=540.0 days

## Methodological Approach

### Statistical Methods

1. **Normality Assessment**:
   * Shapiro-Wilk test to check for normal distribution in subgroups
   * Used to determine appropriate statistical tests (parametric vs. non-parametric)
2. **Body Part Analysis**:
   * Comparison of healing duration between patients with vs. without each type of injury
   * T-tests for normally distributed data
   * Mann-Whitney U tests for non-normally distributed data or small sample sizes
   * Effect size calculation (Cohen’s d) to quantify the magnitude of differences
   * Visualization using violin plots with embedded box plots
3. **Injury Count Analysis**:
   * Correlation analysis (Pearson’s r) between injury count and healing duration
   * Categorization into severity groups (1-2, 3-4, 5+ injuries)
   * ANOVA to test for differences between severity categories
   * Visualization using scatter plots and box plots
4. **Demographic Analysis**:
   * Correlation analysis for age vs. healing duration
   * Age decade categorization and ANOVA comparison
   * Gender-based comparison (limited by data availability)
   * Visualization using scatter plots and box plots

### Significance Criteria

* Conventional statistical significance level: p < 0.05
* Near significance noted at: 0.05 ≤ p < 0.10
* Effect size interpretation:
  + Cohen’s d < 0.2: negligible
  + 0.2 ≤ d < 0.5: small
  + 0.5 ≤ d < 0.8: medium
  + d ≥ 0.8: large

## Results and Findings

### Body Part Injuries

#### Statistically Significant Factors

1. **Abdominal Injuries**:
   * 7 patients with injury, 23 without
   * Mann-Whitney U-Test: U=136.00, p=0.0049
   * Effect size: Cohen’s d = 1.48 (large)
   * Mean healing duration: With injury = 818.3 days, Without = 479.1 days
   * Mean difference: +339.2 days longer with abdominal injury
2. **Head Injuries**:
   * 15 patients with injury, 15 without
   * T-test: t=-2.83, p=0.0090
   * Effect size: Cohen’s d = 1.03 (large)
   * Mean healing duration: With injury = 694.0 days, Without = 444.9 days
   * Mean difference: +249.1 days longer with head injury

#### Nearly Significant Factors

1. **Spine Injuries**:
   * 17 patients with injury, 13 without
   * T-test: t=-1.99, p=0.0591
   * Effect size: Cohen’s d = 0.76 (medium)
   * Mean healing duration: With injury = 685.6 days, Without = 492.6 days
   * Mean difference: +193.0 days longer with spine injury
2. **Combined Arm Injuries**:
   * 19 patients with injury, 11 without
   * T-test: t=1.88, p=0.0706
   * Effect size: Cohen’s d = 0.65 (medium)
   * Mean healing duration: With injury = 625.3 days, Without = 455.7 days
   * Mean difference: +169.6 days longer with arm injury
3. **Right Arm Injuries**:
   * 8 patients with injury, 22 without
   * T-test: t=1.64, p=0.1333
   * Effect size: Cohen’s d = 0.79 (medium)
   * Mean healing duration: With injury = 722.9 days, Without = 519.4 days
   * Mean difference: +203.5 days longer with right arm injury

#### Non-Significant Factors (Small or Negligible Effects)

* Left Arm Injuries: d = 0.33 (small), p = 0.3721
* Thorax Injuries: d = 0.24 (small), p = 0.5393
* Pelvis Injuries: d = 0.17 (negligible), p = 0.4316
* Combined Leg Injuries: d = 0.15 (negligible), p = 0.6909
* Left Leg Injuries: d = 0.12 (negligible), p = 0.7537
* Right Leg Injuries: d = 0.01 (negligible), p = 0.9828
* Neck Injuries: Insufficient sample (n=1) for statistical testing

### Injury Count Analysis

* Correlation between injury count and healing duration: r = 0.0473, p = 0.8038 (not significant)
* Categorized analysis (ANOVA): F = 2.79, p = 0.0793 (approaching significance)
* Mean healing duration by category:
  + 1-2 injuries (n=9): 464.0 days
  + 3-4 injuries (n=10): 728.0 days
  + 5+ injuries (n=11): 538.7 days
* This suggests a non-linear relationship where moderate polytrauma may require longer healing times than either minor or very severe cases

### Age Analysis

* Overall correlation: r = -0.2829, p = 0.1298 (not significant)
* Age decade ANOVA: F = 3.74, p = 0.0102 (significant)
* Mean healing duration by decade:
  + 20s (n=6): 513.5 days
  + 30s (n=3): 690.7 days
  + 40s (n=5): 943.6 days (highest)
  + 50s (n=7): 450.9 days
  + 60s (n=4): 582.8 days
  + 70s (n=2): 399.0 days
  + 80s (n=2): 304.5 days (lowest)
* The results show a non-linear pattern with middle-aged patients (40s) having the longest healing durations

### Gender Analysis

* The log file noted insufficient gender data for meaningful analysis: “WARNING - Nicht genügend Geschlechtsdaten für Analyse vorhanden”

## Technical Implementation Details

### Software and Libraries

* Python for data processing and statistical analysis
* Key libraries:
  + pandas: Data manipulation and analysis
  + numpy: Numerical computing
  + scipy.stats: Statistical tests (t-test, Mann-Whitney U, ANOVA)
  + matplotlib and seaborn: Data visualization
  + logging: Comprehensive logging of process and results

### Implementation Structure

The analysis was implemented in the univariate\_analysis.py script with the following key functions:

1. **setup\_logging()**: Configured logging for the script
2. **load\_dataset()**: Loaded and validated the dataset
3. **prepare\_patient\_level\_data()**: Created a patient-level dataset for analysis
4. **test\_normality()**: Tested if data follows a normal distribution
5. **analyze\_body\_part\_impact()**: Analyzed impact of body part injuries on healing duration
6. **analyze\_injury\_count\_impact()**: Analyzed relationship between injury count and healing duration
7. **analyze\_demographic\_impact()**: Analyzed impact of demographic factors on healing duration
8. **export\_results\_to\_excel()**: Exported results to Excel format
9. **generate\_summary\_report()**: Generated a comprehensive Markdown report of findings

### Output Files and Locations

1. **Data Files**:
   * Patient-level data: data/output/step4/univariate\_analysis/patient\_level\_data.xlsx
   * Analysis results: data/output/step4/univariate\_analysis/univariate\_analysis\_results.xlsx
2. **Visualization Files** (all stored in plots/step4/univariate\_analysis/):
   * Body part visualizations (e.g., Kopf\_Heilungsdauer.png)
   * Injury count visualizations:
     + Verletzungsanzahl\_Heilungsdauer\_Scatter.png
     + Verletzungsanzahl\_Heilungsdauer\_Box.png
   * Age visualizations:
     + Alter\_Heilungsdauer.png
     + Altersdekade\_Heilungsdauer.png
3. **Documentation**:
   * Analysis report: data/output/step4/univariate\_analysis/univariate\_analysis\_report.md
   * Detailed log: logs/step4/univariate\_analysis.log

## Limitations and Considerations

### Statistical Limitations

1. **Sample Size**: The sample of 30 patients limits statistical power, especially for subgroup analyses
2. **Multiple Testing**: Tests across multiple body parts increase the risk of Type I errors (false positives)
3. **Uneven Subgroups**: Some injury types had few cases (e.g., neck injury n=1)
4. **Missing Data**: Gender analysis was not possible due to insufficient data

### Methodological Considerations

1. **Healing Duration Definition**: Time to last visit may not perfectly correspond to complete physiological recovery
2. **Univariate Approach**: Analysis does not account for interactions between factors
3. **Patient Selection**: Results may be influenced by the specific patient cohort and may not generalize to all polytrauma cases
4. **Temporal Aspects**: The analysis doesn’t account for potential changes in treatment approaches over time

## Recommended Next Steps

Based on the findings of this univariate analysis, the following next steps are recommended:

1. **Multivariate Analysis**:
   * Multiple regression modeling using significant factors (abdominal injuries, head injuries, spine injuries, age)
   * Consider including interaction terms, particularly between injury types
2. **Survival Analysis**:
   * Implement time-to-event analysis to better handle the temporal nature of healing
   * Consider different definitions of “recovery events” beyond the last visit
3. **Feature Prioritization**:
   * Focus on the most significant factors (abdominal and head injuries) in subsequent analyses
   * Investigate the age patterns more thoroughly to understand the non-linear relationship with healing duration
4. **Clinical Correlation**:
   * Review results with medical specialists to validate findings against clinical experience
   * Consider inclusion of quality-of-life or functional outcome measures alongside healing duration
5. **Data Enhancement**:
   * Consider additional variables that might explain the counterintuitive findings (e.g., treatment intensity by age group)
   * Address missing data fields, particularly gender information

## Conclusion

The univariate analysis has successfully identified several factors that significantly impact healing duration in polytrauma patients. Abdominal and head injuries emerged as the most significant factors, with large effect sizes and statistical significance. The non-linear relationships observed with both injury count and age suggest complex patterns that warrant further investigation through multivariate approaches. Despite the limited sample size, this analysis provides valuable insights to guide subsequent analytical steps and clinical decision-making for polytrauma patients.