**Multivariate Analysis of Factors Affecting Healing Duration in Polytrauma Patients**

**Project Overview**

This document provides comprehensive documentation of the multivariate analysis performed to identify factors affecting healing duration in polytrauma patients. The analysis builds upon the univariate findings and explores how multiple factors interact to influence recovery time. Both regression models and survival analysis approaches were implemented to provide complementary perspectives on the data.

**1. Purpose and Scope**

The multivariate analysis aimed to:

* Identify which factors remain significant when controlling for others
* Quantify the combined effect of multiple injury types
* Develop predictive models for healing duration
* Explore time-to-event patterns through survival analysis

**2. Data Source and Preparation**

**2.1 Source Data**

* Patient-level dataset created during the univariate analysis phase
* 30 unique patients with information on injury types, demographics, and healing duration
* Structure: One row per patient (rather than per visit)

**2.2 Key Variables**

* **Dependent variable**: Heilungsdauer (healing duration in days)
* **Independent variables**:
  + Body part injuries (Kopf, Abdomen, Wirbelsaeule, Arm, etc.)
  + Demographics (Alter/Age)
  + Visit information (Anzahl\_Besuche)

**2.3 Variable Definition and Transformation**

* **Healing Duration**: Days from accident to last recorded visit
* **Injury Variables**: Binary indicators (1/0 or Ja/Nein) for each body part
* **Age**: Patient age at time of accident
* **Age\_Squared**: Square of age (to capture potential non-linear effects)

**3. Methodology**

**3.1 Analysis Approaches**

Two complementary analytical approaches were implemented:

1. **Multiple Linear Regression**:
   * Models with increasing complexity
   * Focus on prediction and quantification of effects
2. **Survival Analysis**:
   * Kaplan-Meier curves for visualization
   * Cox proportional hazards modeling for hazard ratios

**3.2 Regression Models**

Three progressively complex regression models were built:

1. **Model 1**: Included only the two factors found significant in univariate analysis
2. Heilungsdauer ~ Abdomen + Kopf
3. **Model 2**: Added nearly significant factors from univariate analysis
4. Heilungsdauer ~ Abdomen + Kopf + Wirbelsaeule + Arm
5. **Model 3**: Added age with quadratic term to capture non-linear effects
6. Heilungsdauer ~ Abdomen + Kopf + Wirbelsaeule + Arm + Alter + Age\_Squared

**3.3 Survival Analysis Framework**

* **Event Definition**: "Recovery" defined as reaching the last visit
* **Time Variable**: Heilungsdauer (days from accident to last visit)
* **Event Indicator**: All set to 1 (all patients reached their last visit)
* **Kaplan-Meier Estimation**: For visual comparison of injury groups
* **Cox Proportional Hazards Model**: To quantify the effect of factors on "recovery" rate

**3.4 Assumptions**

1. **Regression Assumptions**:
   * Linear relationship between predictors and outcome
   * Independence of observations
   * Homoscedasticity (equal variance of errors)
   * Normal distribution of residuals
2. **Survival Analysis Assumptions**:
   * Proportional hazards (effect of variables remains constant over time)
   * Independence of observations
   * Last visit represents completion of recovery process
3. **General Assumptions**:
   * Binary injury variables adequately capture injury presence/absence
   * The 30 patients represent a meaningful sample of polytrauma cases
   * There are no systematic biases in follow-up patterns

**4. Results**

**4.1 Regression Models**

**Model 1: Basic Model with Significant Univariate Factors**

* **R²**: 0.419 (41.9% of variance explained)
* **Adjusted R²**: 0.376
* **Significant Predictors**:
  + Abdomen: +285.51 days (p=0.0054)
  + Kopf: -191.96 days (p=0.0233)

**Model 2: Extended Model with Additional Factors**

* **R²**: 0.545 (54.5% of variance explained)
* **Adjusted R²**: 0.472
* **Significant Predictors**:
  + Abdomen: +339.71 days (p=0.0010)
  + Wirbelsaeule: -186.75 days (p=0.0383)
  + Kopf: -110.06 days (p=0.1805) - no longer significant

**Model 3: Full Model with Age Effects**

* **R²**: 0.553 (55.3% of variance explained)
* **Adjusted R²**: 0.436
* **Significant Predictors**:
  + Abdomen: +321.55 days (p=0.0041)
  + Wirbelsaeule: -167.80 days (p=0.0858) - borderline significance

**4.2 Survival Analysis Results**

**Kaplan-Meier Analysis**

* The visualization showed clear separation between patients with and without abdominal injuries
* Patients with abdominal injuries had consistently longer times to "recovery"
* Spine injury curves showed faster times to "recovery" compared to patients without spine injuries

**Cox Proportional Hazards Model**

* **Concordance Index**: 0.738 (good discrimination)
* **Log-Likelihood Ratio Test p-value**: 0.0003 (highly significant model)
* **Significant Factors**:
  + Abdomen: HR=0.16 (95% CI: 0.05-0.52, p=0.0025)
    - Interpretation: 84% lower "recovery" rate (longer healing)
  + Wirbelsaeule: HR=3.68 (95% CI: 1.03-13.14, p=0.0448)
    - Interpretation: 268% higher "recovery" rate (shorter healing)

**4.3 Model Comparison and Selection**

* Model 2 had the highest adjusted R² (0.472), suggesting the best balance between explanatory power and parsimony
* The concordance index of the Cox model (0.738) indicated good predictive performance
* The findings from regression and survival analysis showed consistent patterns

**5. Key Findings and Interpretation**

**5.1 Confirmed Effect of Abdominal Injuries**

* **Finding**: Abdominal injuries significantly extend healing duration by approximately 322-340 days
* **Consistency**: This effect remained robust across all models and analytical approaches
* **Significance**: The effect was highly significant (p<0.01) in all models
* **HR in Cox Model**: 0.16, indicating substantially reduced "recovery" rate

**5.2 Counterintuitive Findings for Spine Injuries**

* **Finding**: Spine injuries were associated with *shorter* healing durations by approximately 167-187 days
* **Emergence**: This effect became significant only in the multivariate models
* **HR in Cox Model**: 3.68, indicating faster "recovery" rate
* **Interpretation**: Potential explanations include:
  1. More intensive or specialized treatment protocols for spine injuries
  2. Confounding factors not captured in the dataset
  3. Different follow-up patterns for spine injury patients

**5.3 Declining Significance of Head Injuries**

* **Finding**: Head injuries lost statistical significance when accounting for other factors
* **Pattern**: The effect diminished from Model 1 (-192 days, p=0.023) to Model 2 (-110 days, p=0.181)
* **Interpretation**: Head injury effects may be partially confounded with other injury types

**5.4 Age-Related Effects**

* **Finding**: Neither linear nor quadratic age terms reached statistical significance
* **Observation**: Non-significant positive linear term (2.43 days/year) and negative quadratic term (-0.04)
* **Interpretation**: Age effects observed in univariate analysis may be explained by injury patterns

**6. Visualizations Created**

1. **Regression Coefficients Comparison**: Bar chart showing coefficient magnitude and significance across models
2. **Actual vs. Predicted Healing Duration**: Scatter plot with R²=0.553 showing model fit
3. **Kaplan-Meier Curves by Injury Type**:
   * Overall survival curve
   * Curves for Abdomen (Ja/Nein)
   * Curves for Kopf (Ja/Nein)
   * Curves for Wirbelsaeule (Ja/Nein)
   * Curves for Arm (Ja/Nein)
4. **Cox Model Hazard Ratios**: Forest plot with 95% confidence intervals

**7. Limitations**

1. **Sample Size**: Limited to 30 patients, constraining statistical power
2. **Definition of Healing Duration**: Based on last visit rather than clinical recovery
3. **Multicollinearity**: Potential correlation between injury types
4. **Missing Variables**: Other factors not measured may influence healing duration
5. **Generalizability**: May not represent all polytrauma populations

**8. Completion Status in Analysis Plan**

The following steps from the analysis outline have been completed:

**Completed Steps:**

1. **Definition of Healing Duration Metric**: Successfully implemented using days from accident to last visit
2. **Univariate Analysis for Body Part Injuries**: Completed in previous phase
3. **Demographic Factor Analysis**: Age and its non-linear effects analyzed
4. **Multiple Regression Models**: Successfully implemented with increasing complexity
5. **Survival Analysis**: Implemented Kaplan-Meier curves and Cox proportional hazards model

**Steps Not Yet Completed:**

1. **Time Interval Analysis**: Analyzing problems identified in specific time periods
2. **Time-Based Analysis**: Detailed treatment timeline analysis
3. **Critical Injury Impact Analysis**: Further detailed analysis of specific injury types
4. **Professional Reintegration Analysis**: Analysis of work status outcomes

**9. Next Steps**

Based on the completed work and remaining items in the analysis plan, the next steps should be:

**9.1 Time-Based Analysis**

* Analyze relationship between first visit timing and overall healing duration
* Investigate if early intervention correlates with shorter healing
* Calculate average time between visits for each patient and test correlation with healing duration

**9.2 Professional Reintegration Analysis**

* Analyze healing duration for patients with different work status outcomes
* Compare healing duration between cases marked as "Arbeitsfaehig" vs. "Arbeitsunfähig" in final records
* Analyze if vocational rehabilitation interventions correlate with healing time

**9.3 Development of a Prediction Tool**

* Create a user-friendly tool (e.g., spreadsheet or web application)
* Implement the best predictive model (Model 2)
* Allow input of patient characteristics to estimate healing duration

**10. Technical Implementation**

The analysis was implemented in Python using the following libraries:

* pandas: For data manipulation
* statsmodels: For regression modeling
* lifelines: For survival analysis
* matplotlib/seaborn: For visualizations

**11. Conclusion**

The multivariate analysis has provided valuable insights into the factors affecting healing duration in polytrauma patients, explaining approximately 55% of the variance. Abdominal injuries emerged as the most consistent predictor of prolonged healing, while spine injuries showed an unexpected association with shorter healing durations. The findings have important implications for patient care, resource allocation, and rehabilitation planning.

The next phase of analysis should focus on temporal aspects of recovery and professional outcomes to further enhance our understanding of the polytrauma recovery process.