**Polytrauma Injury Category Analysis Documentation**

**1. Overview**

This documentation outlines the injury category distribution analysis performed on polytrauma patient data. The analysis examines the prevalence of nine main injury categories across 30 unique patient cases. Unlike previous analyses that focused on subcategories within each main category, this analysis provides a cross-category comparison to understand the overall injury landscape.

**2. Methodology**

**2.1 Data Source**

The analysis used the processed dataset from Polytrauma\_Analysis\_Processed.xlsx, which contained 153 records representing 30 unique patients with multiple visits. The data was loaded with proper typing (Schadennummer as string) and processed to ensure consistent field names.

**2.2 Category Aggregation**

A key methodological decision was to merge separate limb categories:

* "Arm links" and "Arm rechts" were merged into a single "Arm" category
* "Bein links" and "Bein rechts" were merged into a single "Bein" category

This aggregation was implemented using the following logic:

# Merge Arm links and Arm rechts into a single Arm category

df['Arm'] = df[['Arm links', 'Arm rechts']].apply(

lambda row: 'Ja' if 'Ja' in row.values else 'Nein', axis=1

)

# Merge Bein links and Bein rechts into a single Bein category

df['Bein'] = df[['Bein links', 'Bein rechts']].apply(

lambda row: 'Ja' if 'Ja' in row.values else 'Nein', axis=1

)

**2.3 Category Structure**

The analysis used the following nine main categories:

1. **Körperteil** (Body Parts): Kopf, Hals, Thorax, Abdomen, Arm, Wirbelsaeule, Bein, Becken
2. **Somatisch** (Physical Conditions): Funktionsstoerung, Schmerz, Komplikationen
3. **Personenbezogen** (Personal Factors): Psychische Probleme/Compliance, Entschädigungsbegehren, Migrationshintergrund, Suchtverhalten, Zusätzliche Erkrankungen
4. **Taetigkeit** (Employment Status): Arbeitsunfähig, Wiedereingliederung, Arbeitsfaehig, BU/EU, Altersrentner, Ehrenamt, Zuverdienst
5. **Umwelt** (Environmental Factors): Beziehungsprobleme, Soziale Isolation, Mobilitaetsprobleme, Wohnsituatuation, Finazielle Probleme
6. **Med RM** (Medical Case Management): Arzt-Vorstellung, Arzt-Wechsel, Organisation ambulante Therapie, Organisation medizinische Reha, Wietere Krankenhausaufenthalte, Psychotherapie, Organisation Pflege
7. **Soziales RM** (Social Case Management): Lohnersatzleistungen, Arbeitslosenunterstuetzung, Antrag auf Sozialleistungen, Einleitung Begutachtung
8. **Technisches RM** (Technical Case Management): Hilfmittelversorgung, Mobilitätshilfe, Bauliche Anpassung, Arbetsplatzanpassung
9. **Berufliches RM** (Vocational Case Management): Leistungen zur Teilhabe am Arbeitsleben, Integration/berufliche Neurientierung allgemeiner Arbeitsmarkt, Wiedereingliederung geförderter Arbeitsmarkt, BEM

**2.4 Analysis Assumptions**

1. **Case Definition**: Each unique Schadennummer represents a separate patient case.
2. **Positive Case Definition**: A case is marked as positive for a category if any subcategory has "Ja" across any visit.
3. **Missing Values**: Missing values in categorical columns were treated as "Nein" during processing.
4. **Aggregation Logic**: For merged categories (Arm and Bein), a case is considered positive if either the left or right side (or both) is marked as "Ja".
5. **Percentages Calculation**: All percentages are calculated based on the total number of unique cases (n=30).

**3. Implementation**

The analysis was implemented using Python with the following key components:

**3.1 Key Functions**

The implementation is now split across two complementary scripts:

**injury\_category\_distribution.py (Category-level Analysis):**

1. **setup\_logging()**: Configures logging for the script
2. **load\_dataset()**: Loads and validates the dataset
3. **create\_merged\_categories()**: Creates unified Arm and Bein categories
4. **analyze\_categories()**: Performs the core analysis of category prevalence
5. **visualize\_category\_comparison()**: Creates various visualizations comparing categories
6. **create\_heatmap()**: Generates a heatmap of top subcategories
7. **export\_results()**: Exports results to Excel and JSON
8. **create\_summary\_report()**: Creates a markdown summary report

**injury\_subcategory\_distribution.py (Subcategory-level Analysis):**

1. **setup\_logging()**: Configures logging for the script
2. **load\_dataset()**: Loads and validates the dataset
3. **analyze\_category\_distribution()**: Analyzes distribution of subcategories within each category
4. **plot\_category\_distribution()**: Creates visualizations for individual categories
5. **validate\_categories()**: Performs validation checks on category statistics
6. **generate\_summary\_report()**: Creates a comprehensive markdown report

**3.2 Visualizations Created**

The combined scripts produce multiple visualizations:

**Category-level Visualizations:**

1. **Category Comparison Bar Chart**: Vertical bar chart showing percentages for all categories
2. **Horizontal Category Comparison**: Horizontal bar chart for easier comparison
3. **Category Distribution Pie Chart**: Shows the relative distribution of categories
4. **Category Radar Chart**: Displays coverage across all domains on a circular axis
5. **Top Subcategories Heatmap**: Shows the most prevalent subcategories across main categories

**Subcategory-level Visualizations:**

1. **Individual Category Histograms**: Separate histograms for each category showing subcategory distribution
2. **Category Percentage Histogram**: Overview of all category percentages
3. **Category Distribution Radar**: Alternative view of category prevalence

**3.3 Output Files**

The analysis generates several output files:

1. **Excel Reports**:
   * injury\_category\_analysis.xlsx with category-level analysis
   * injury\_subcategory\_statistics.xlsx with detailed subcategory statistics
2. **JSON Data**:
   * injury\_category\_analysis.json with structured category-level results
   * category\_statistics.json with detailed subcategory-level results
3. **Markdown Reports**:
   * injury\_category\_analysis\_report.md with key category-level findings
   * category\_distribution\_report.md with detailed subcategory analysis
4. **Visualizations**: Multiple PNG image files with different chart types

**4. Results**

**4.1 Overall Category Distribution**

| **Category** | **Positive Cases** | **Percentage** |
| --- | --- | --- |
| Körperteil | 30 | 100.0% |
| Taetigkeit | 30 | 100.0% |
| Med RM | 30 | 100.0% |
| Somatisch | 26 | 86.7% |
| Technisches RM | 21 | 70.0% |
| Personenbezogen | 18 | 60.0% |
| Umwelt | 15 | 50.0% |
| Berufliches RM | 15 | 50.0% |
| Soziales RM | 10 | 33.3% |

**4.2 Top Subcategories by Category**

**Körperteil (Body Parts)**

* Arm: 63.3% (19 cases)
* Thorax: 60.0% (18 cases)
* Wirbelsaeule: 56.7% (17 cases)

**Somatisch (Physical Conditions)**

* Funktionsstoerung: 83.3% (25 cases)
* Schmerz: 53.3% (16 cases)
* Komplikationen: 36.7% (11 cases)

**Med RM (Medical Case Management)**

* Organisation ambulante Therapie: 96.7% (29 cases)
* Arzt-Vorstellung: 80.0% (24 cases)
* Organisation medizinische Reha: 63.3% (19 cases)

**Taetigkeit (Employment Status)**

* Arbeitsunfähig: 70.0% (21 cases)
* Arbeitsfaehig: 46.7% (14 cases)
* Wiedereingliederung: 43.3% (13 cases)

**Technisches RM (Technical Case Management)**

* Hilfmittelversorgung: 63.3% (19 cases)
* Mobilitätshilfe: 40.0% (12 cases)
* Bauliche Anpassung / Arbetsplatzanpassung: 10.0% (3 cases each)

**5. Visualization Descriptions and Interpretations**

**5.1 Category Comparison Bar Chart (Image 1)**

**Description:** This vertical bar chart displays the percentage of cases affected by each main category, sorted in descending order from highest to lowest prevalence. Each bar is annotated with both the percentage and absolute number of cases (e.g., "100.0% (30/30)"). The chart title "Vergleich der Hauptverletzungskategorien" (Comparison of Main Injury Categories) and all labels are in German.

**Interpretation:**

* Three categories (Körperteil, Taetigkeit, Med RM) show 100% prevalence, indicating that all patients had some form of body part injury, activity limitation, and medical case management.
* Somatisch (86.7%) and Technisches RM (70.0%) also show high prevalence, indicating that most patients experience physical symptoms and require technical assistance.
* Soziales RM has the lowest prevalence (33.3%), suggesting that social welfare interventions are required for only a third of patients.

**Usage:** This visualization provides a quick overview of which aspects of rehabilitation are most commonly needed following polytrauma, helping to prioritize resources and interventions at a program level.

**5.2 Horizontal Category Comparison (Image 2)**

**Description:** This horizontal bar chart presents the same data as the vertical chart but in a horizontal format, which may be easier to read for comparing values across categories. The chart has the title "Verletzungskategorien (sortiert nach Prävalenz)" (Injury Categories (Sorted by Prevalence)) and all labels are in German.

**Interpretation:**

* The horizontal layout makes it easier to compare the exact percentages across categories.
* The clear sorting from highest to lowest prevalence helps identify the most and least common categories.
* The consistent labeling of both percentages and case counts provides context for the relative frequencies.

**Usage:** This visualization is particularly useful for presentations and reports where direct category comparisons are needed.

**5.3 Category Distribution Pie Chart (Image 3)**

**Description:** This pie chart displays the relative distribution of categories, focusing on the top five categories with the highest prevalence and grouping the rest as "Andere Kategorien" (Other Categories). The chart title "Verteilung der Verletzungskategorien" (Distribution of Injury Categories) and all labels are in German.

**Interpretation:**

* The three 100% prevalence categories (Körperteil, Taetigkeit, Med RM) each represent 15.4% of the total distribution.
* Somatisch represents 13.3% of the distribution.
* Technisches RM accounts for 10.8% of the distribution.
* The remaining categories combined make up 29.7% of the distribution, labeled as "Andere Kategorien" in German.

**Usage:** This visualization helps understand the proportional representation of different categories within the overall rehabilitation landscape.

**5.4 Category Radar Chart (Image 4)**

**Description:** The radar chart (also called spider plot) displays the percentage of cases for each main category around a circular axis, with 0% at the center and 100% at the outer edge. Categories are arranged radially, and connected values form a polygon. The chart title "Abdeckung über Verletzungskategorien" (Coverage Across Injury Categories) and all labels are in German.

**Interpretation:**

* The chart shows the relative prevalence of each category compared to others.
* The Körperteil, Med RM, and Taetigkeit categories extend to the outer edge (100%), forming a triangular pattern.
* Somatisch extends to about 87%, while Technisches RM reaches 70%.
* Soziales RM creates the shortest spoke, indicating it has the lowest prevalence.

**Usage:** The radar chart allows for quick visual comparison across all categories simultaneously and highlights the multidimensional nature of polytrauma rehabilitation. The area of the polygon represents the overall burden of injury and intervention needs across all categories.

**5.5 Top Subcategories Heatmap (Image 5)**

**Description:** This heatmap visualizes the top three subcategories within each main category, with color intensity representing the prevalence percentage. The matrix layout allows for comparisons across both categories and subcategories. The title "Top-Unterkategorien nach Hauptkategorien" (Top Subcategories Across Main Categories) and all labels are in German.

**Interpretation:**

* Organisation ambulante Therapie is the most prevalent subcategory overall (96.7%).
* Funktionsstörung (83.3%) and Arzt-Vorstellung (80.0%) are also highly prevalent.
* Many subcategories have zero prevalence, indicated by the dark purple cells.
* The heatmap reveals clusters of high-prevalence subcategories that may be related.

**Usage:** This visualization helps identify specific intervention points that are most commonly needed across different rehabilitation domains, which can inform clinical protocols and resource allocation.

**6. Key Findings and Implications**

**6.1 Prevalence Patterns**

1. **Universal Categories**: Three categories (Körperteil, Taetigkeit, Med RM) show 100% prevalence, indicating core aspects of polytrauma rehabilitation.
2. **High Prevalence Interventions**:
   * Organisation ambulante Therapie (96.7%) is nearly universal
   * Funktionsstoerung (83.3%) affects most patients
   * Arzt-Vorstellung (80.0%) is needed for the vast majority
3. **Moderate Prevalence Issues**:
   * Technical assistance (70.0%) is needed in most cases
   * Personal factors (60.0%) affect the majority of patients
   * Environmental factors (50.0%) affect half of all patients
4. **Lower Prevalence Issues**:
   * Social case management (33.3%) is needed in only a third of cases
   * Certain subcategories (e.g., Ehrenamt, Zuverdienst, Suchtverhalten) show zero prevalence

**6.2 Extremity Injury Patterns**

The merged analysis revealed:

1. **Upper Extremity**: Arm injuries (combined) are among the most common body parts affected (63.3%, 19 cases)
2. **Lower Extremity**: Leg injuries (combined) affect more than half of patients (53.3%, 16 cases)
3. **Bilateral Consideration**: The merged approach provides a more comprehensive view of extremity involvement by considering overall limb injury regardless of laterality

**6.3 Implications for Rehabilitation**

1. **Integrated Care Model**: The 100% prevalence of body part injuries, activity limitations, and medical management needs confirms the necessity of multidisciplinary approaches.
2. **Resource Allocation**: The varying prevalence rates can guide resource allocation, with more resources directed toward high-prevalence categories and subcategories.
3. **Ambulatory Care Focus**: The extremely high prevalence of ambulatory therapy organization (96.7%) suggests that outpatient services are a critical component of polytrauma rehabilitation.
4. **Functional Prioritization**: With functional disorders affecting 83.3% of patients, functional rehabilitation should be a primary focus.
5. **Technical Support**: The high rate of assistive device needs (63.3% requiring mobility aids) indicates the importance of technical support in rehabilitation.

**7. Limitations and Considerations**

1. **Sample Size**: The analysis is based on only 30 unique patient cases, which may limit generalizability.
2. **Temporal Dynamics**: The analysis does not account for changes in categories over time during the rehabilitation process.
3. **Data Completeness**: Some fields had missing data that was interpreted as "Nein" which may underestimate actual prevalence.
4. **Category Interdependence**: Relationships between categories were not analyzed but may be important for comprehensive rehabilitation planning.
5. **Merged Categories Limitation**: While merging arm and leg categories provides a clearer overall picture, it loses information about laterality which may be clinically important in some contexts.

**8. Technical Implementation Details**

**8.1 Environment Setup**

The script uses environment variables loaded from a .env file:

* DATASET: Path to the processed dataset
* OUTPUT\_FOLDER: Directory for output files
* LOG\_FOLDER: Directory for log files
* GRAPHS\_FOLDER: Directory for visualization images

**8.2 Dependencies**

The implementation relies on several Python libraries:

* pandas: For data manipulation and analysis
* matplotlib and seaborn: For visualization creation
* numpy: For numerical operations
* logging: For comprehensive logging
* json: For exporting results in JSON format
* pathlib: For path handling

**8.3 Folder Structure**

The script uses a structured folder approach:

step3\_output\_folder = os.path.join(output\_folder, "step3")

step3\_log\_folder = os.path.join(log\_folder, "step3")

step3\_plots\_folder = os.path.join(graphs\_folder, "step3", "injury\_category\_analysis")

**9. Conclusion**

The injury category analysis provides a comprehensive view of the prevalence and distribution of different injury types and rehabilitation needs across polytrauma patients. The findings highlight the universal need for body part injury treatment, activity support, and medical case management, while revealing varying prevalence rates for other categories like technical assistance, personal factors, and social support.

The visualizations offer multiple perspectives on the data, from direct category comparisons to proportional distributions and domain coverage. The analysis results can inform rehabilitation planning, resource allocation, and clinical protocol development for polytrauma patients.

The merged approach to extremities (combining left and right limbs) provides a streamlined view of overall limb involvement while maintaining the ability to examine specific body regions affected.

**10. Future Directions**

Potential next steps for analysis include:

1. **Correlation Analysis**: Examining relationships between categories (e.g., do certain body part injuries correlate with specific functional disorders?)
2. **Temporal Analysis**: Analyzing how category prevalence changes over time during the rehabilitation process
3. **Outcome Prediction**: Investigating how category patterns relate to rehabilitation outcomes
4. **Subgroup Analysis**: Comparing category patterns across different demographic groups or injury severity levels
5. **Intervention Effectiveness**: Assessing which interventions are most effective for specific category combinations

**11. Technical Notes on Plot Labels**

It's important to note that all visualizations now use consistent German labeling:

* All plot titles are in German (e.g., "Vergleich der Hauptverletzungskategorien" instead of "Comparison of Main Injury Categories")
* All category labels remain in German (Körperteil, Taetigkeit, Med RM, etc.)
* The label "Andere Kategorien" in the pie chart is also in German (previously "Other Categories")
* All axis labels, percentages, and annotations are in German

This ensures a consistent presentation throughout the application, standardizing on German for all user-facing elements.

**12. Script Architecture**

The analysis is now split into two separate scripts:

1. **injury\_category\_distribution.py**: Focuses on analyzing and visualizing the main injury categories with cross-category comparisons
2. **injury\_subcategory\_distribution.py**: Focuses on detailed analysis of subcategories within each main category

This separation of concerns allows for more focused analysis and clearer presentation of results at different levels of detail.

*This documentation was created as part of the Polytrauma Analysis Project, March 2025.*