

The Feature Definition Spreadsheet for NEMESIS-Data Based Research and Beyond

The Need for a New Approach to NEMESIS-Data Based Research

This paper introduces a novel tool, the Feature Definition Spreadsheet, designed to directly address the significant challenges researchers face when striving for transparency and reproducibility in NEMESIS-based research, particularly in the critical area of cardiac arrest patient sub-grouping.

While simply accessing NEMESIS data is often cited as a hurdle, the complexities extend far beyond data acquisition. The lack of standardized feature definitions and the ambiguity surrounding code interpretation create major roadblocks to reproducing research findings. This is especially concerning in cardiac arrest research, where precise patient sub-grouping is paramount for drawing meaningful conclusions about the efficacy of treatments and ultimately, patient outcomes.

Introducing the Feature Definition Spreadsheet

Our proposed solution, the Feature Definition Spreadsheet, tackles these challenges. This multi-sheet spreadsheet acts as a centralized repository documenting every step involved in defining and implementing patient sub-groups (features) for analysis.

Sheet-by-Sheet Insight:

Several sheets are interrelated, here is a description of what they do and how they interact:

- "Key" Sheet: This sheet serves as the roadmap, providing clear definitions for the content and organization of all other sheets within the spreadsheet.
- "Features_Sheet" Sheet: Here, each patient sub-group (feature) is identified as a row or group of rows. This sheet is the nucleus of the spreadsheet and houses these columns:
 - Confirmation Authority: Identifies the team member/entity vouching for the definition's accuracy.
 - Characteristics Required for Reproducibility: Outlines the essential elements for others to replicate the feature.
 - Feature Description: A concise explanation of the feature's meaning and purpose.
 - Entry Restricted Column: Flags features with restricted data, linking to the "Validation_Sheet" for clarity.
 - Notes: Offers additional context and explanations for specific codes used in the feature definition.
 - Table in NEMESIS PDF: Specifies the NEMESIS table(s) containing the source data.
 - Column Name: Lists the specific column(s) used for feature creation.
 - Primary/Secondary Data Source ID: Documents any external resources used in defining the feature.
 - Include/Exclude: Specifies which codes within a column are included or excluded in the feature definition.
 - Characteristic by Code: Links to detailed explanations of each code used in defining the feature.

- "Data_Sources_Sheet" Sheet: This sheet delves into the origin of each feature, providing a link to the document that gives the most complete information about the fields and codes within that we could find.
- "Code_Sheet" Sheet: This sheet provides a comprehensive code repository which holds codes grouped by feature linked to from the "features_sheet." Proposed and confirmed features are often defined by a group of codes which don't fit well on a single summary line in the features_sheets, this sheet gives them their own space without losing the clear connection to features involved.
- "Validation_Sheet" Sheet: Contains lists of valid data entries by column for describing and defining features. Different researchers may abbreviate capitalize things differently, this sheet creates a dictionary of possible entries to use when defining features in the "features_sheet."

Benefits: Transparency, Collaboration, and Reproducibility

The Feature Definition Spreadsheet offers a robust framework that promotes:

- Enhanced Transparency: The current status of each attribute is prominently displayed giving the user a roadmap for which features are ready to use and which require attention to define.
- Improved Collaboration: The spreadsheet acts as a shared knowledge base, fostering seamless communication among researchers.
- Simplified Reproducibility: By providing a clear, standardized approach to feature definition, the spreadsheet empowers other researchers to accurately replicate analyses and validate findings.

Conclusion: A Path to Reproducible Research Using NEMSIS Public Data and Beyond

The Feature Definition Spreadsheet is not just a tool; it represents a fundamental shift towards more transparent and reproducible research practices. By combining robust data access procedures with this meticulous approach to feature definition, we can significantly enhance the reliability and impact of NEMSIS-data based research, particularly in the critical field of cardiac arrest.