



Exploratory Data Analysis: 2022 NEMSIS Research Dataset

Aaron Fihn, 2023-10-30



Dataset Overview

- 51MM de-identified patient care reports (“PCRs”) from EMS activations in 2022
- Collected by U.S. states/territories
- 42 text files, 175 GB
- Released for research requests by National Emergency Medical Services Information System (NEMSIS), an NHTSA program
- Vague column names (“eDisposition_12”) explained in online data dictionary

“Information documented by the EMS clinician about the patient encounter. It identifies the type of injury, scene of injury, medications administered, procedures performed, and more.”

<https://nemsis.org/wp-content/uploads/2022/11/NEMSIS-Guide-to-Data-Sets-1.pdf>



Context: Project Goals

- Given a **cardiac arrest**, we wish to:
 - know the effect of urbanity/rurality on patient outcomes
 - build a model for survivability based on available predictors
- NEMSIS gives [explicit criteria](#) for identifying cardiac arrests
- Filtering to these gives 296,731 records



Data Limitations

- No patient identifiers (records are event-based)
- No geographic identifiers
- Convenience sample
 - States have different criteria for data submission
- Missing data
 - *"In most cases, NEMSIS data are not missing at random and analyses, therefore, are subject to bias if missing data are ignored."*
- Data quality
 - *"Checked for completeness, logical consistency, and proper formatting"*
 - Problem data files are flagged for state review
 - If not corrected, **these files are included anyway**

https://nemsis.org/wp-content/uploads/2023/04/2022-NEMSIS-RDS-340-User-Manual_v1-FINAL.pdf



Data Structure

```
'PcrKey'~|~'eDispatch_01'~|~'eDispatch_02'~|~'eArrest_14'~|~'eArrest_01'~|~'eArrest_02'~|~'eArrest_05'~|~'eArrest_07'~|~'eArrest_11'~|~'eArrest_16'~|~'eArrest_18'~|~
26669835 ~|~2301069 ~|~2302007 ~|~Not Recorded ~|~3001001 ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~4212033 ~|~
31570619 ~|~2301045 ~|~2302007 ~|~Not Applicable ~|~3001001 ~|~7701001 ~|~7701001 ~|~7701001 ~|~7701001 ~|~7701001 ~|~7701001 ~|~4212033 ~|~
45315246 ~|~2301071 ~|~2302001 ~|~Not Recorded ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~4212033 ~|~
45457991 ~|~2301025 ~|~2302007 ~|~Not Applicable ~|~3001001 ~|~7701001 ~|~7701001 ~|~7701001 ~|~7701001 ~|~7701001 ~|~7701001 ~|~4212033 ~|~
64915249 ~|~2301013 ~|~2302007 ~|~Not Recorded ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~7701003 ~|~4212025 ~|~
```

- Delimited text files with “PcrKey” as the row identifier on all tables
- Mostly categorical data
- Some files can list a PcrKey zero, once, or many times
- Largest file is “Pub_PCRevents.txt”
 - Each PcrKey appears at most once
 - Holds all fields that have at most one entry per PcrKey
 - Primary research focus



Relevant Columns

- **PcrKey:** Uniquely identifies a patient care report across files
- Filters PCRs to “cardiac arrest” case definition:
 - **eArrest_01:** Cardiac Arrest (categorical, Yes/No)
 - **eArrest_02:** Cardiac Arrest Etiology (categorical, Cardiac/Non-Cardiac)
 - **eSituation_09:** Primary Symptom (ICD-10-CM)
 - **eSituation_10:** Other Associated Symptoms (ICD-10-CM)
 - **eSituation_11:** Provider’s Primary Impression (ICD-10-CM)
 - **eSituation_12:** Provider’s Secondary Impression (ICD-10-CM)



Relevant Columns (cont.)

- Critical features:
 - **Urbanicity**: Urban/Suburban/Rural/Wilderness)
- Probably important features:
 - **eArrest_05**: CPR Care Provided Prior to EMS Arrival (categorical, Yes/No)
 - **eDisposition_23**: Hospital Capability (General Hospital, Trauma Center, etc)
 - **ePatient_15**: Age (integer, 1-120)
 - **ePatient_16**: Age Units (days/hours/minutes/months/years)
- Outcome-related columns:
 - **eArrest_18**: End of EMS Cardiac Arrest Event (Expired/ROSC/Ongoing)
 - **eOutcome_01**: Emergency Department Disposition (Discharged/Deceased/Admitted...)



Next Problems

- Which column do we take as the “source of truth” for patient outcomes?
- How do we perform imputation method sensitivity testing?



Mandatory Disclosures

- “All Information derived from the NEMSIS National EMS Database shall remain the full property of The National Highway Traffic Safety Administration and shall be so noted in educational material, website presentations, and publications.”
- “National Highway Traffic Safety Administration (NHTSA), National Emergency Medical Services Information System (NEMSIS). The content reproduced from the NEMSIS Database remains the property of the National Highway Traffic Safety Administration (NHTSA). The National Highway Traffic Safety Administration is not responsible for any claims arising from works based on the original Data, Text, Tables, or Figures.”