**Case Study**

**Data Source**

To illustrate the methods, we used the MIMIC-III dataset. This is a freely available database that comprises data of 62722patients admitted to the ICU of the Beth Israel Deaconess Medical Center in Boston, Massachusetts between 2001 and 2012. The data consists of demographics, vital sign measurements made every hour, laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality.

**Data of Interest**

For the current case study, a few specific variables are of interested, based on the Sepsis-related Organ Failure Assessment (SOFA). The SOFA is a measure that is designed to describe a pattern of complications in critically ill patients that are related to organ failure and death. For computation of the SOFA score, the variables FiO2, PaO2, the number of platelets, the bilirubin level, the Glasgow coma score, MAP, creatinine level, and the urine output are of importance. Of specific interest are the ‘worst’ scores of these variables in the last 24 hours, i.e., the level in the last 24 hours that would yield the worst SOFA score.

**Data Extraction**

We extracted the data using the ‘ricu’ package in R. For each of the relevant variables, the worst value in 24 hours before discharge from the ICU was extracted, as well as a variable indicating whether they died or not. Explorations showed that 1830 patients had NA’s on each variable other than death. Since these variables will act as predictors of death, the data of these patients are not of use for the current research and were thus deleted. Table […] shows descriptives of the relevant continuous variables of the 60892 remaining patients. In addition, 6589 of the patients died and 54303 did not. The data presented 115 different missing data patterns.

**Table […].** *Descriptives of Worst Values in the Last 24 Hours of Predictors from MIMIC Dataset.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Predictor** | ***M*** | ***SD*** | ***NA’s*** | ***Complete Rate*** |
| FiO2 | 79.71 | 25.58 | 33162 | .46 |
| PaO2 | 91.55 | 384.86 | 48030 | .21 |
| Platelets | 190.04 | 102.59 | 1160 | .98 |
| Bilirubin | 3.23 | 5.58 | 20666 | .66 |
| Glasgow Coma Score (non-sedated) | 12.65 | 3.47 | 8537 | .86 |
| Glasgow Coma Score (total) | 9.31 | 5.03 | 31853 | .48 |
| MAP | 109.31 | 26.09 | 8459 | .86 |
| Creatinine | 1.81 | 1.90 | 7153 | .88 |
| Urine Output | 1360.93 | 1168.65 | 11502 | .81 |

115 patterns total

Table […] shows the results of the simulations. When considering FLR, the RMSE, Brier score, and C-index all indicate that single draw imputation performs the worst of the four methods, and the pattern submodel approach performs best. Regarding random forests, these measures also indicate single draw imputation to have the worst performance, and they indicate pattern submodels and surrogate splits to have a similarly high performance.

# with logistic model, the pattern submodel approach works best

# with rf, the pattern submodel and surrogate split methods both perform well

# imputation methods perform less good on the real-world data -> distributional assumptions/MNAR?

# within the imp methods, the random draw from the distribution clearly underperforms compared to regression imp and multiple imp

# these results do not show a clear superiority for multiple imputation (link to other research!!)

# overall, the method that performs best on real-world data is surrogate splitting

# this method also has advantages in explainability, and does not require the miss mech to be equal across datasets

# a small disadvantage is that the method is very slow to train, but with increasing computational power, that shouldn't be a problem

Table […]. Performance of Methods on MIMIC Data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | RMSE | Brier | C-index | CITL | Slope |
| FLR | JMI-CM | 0.308 | 0.095 | 0.803 | 0.142 | 0.192 |
| JMI-SD | 0.321 | 0.103 | 0.774 | 0.155 | 0.189 |
| JMI-MD | 0.312 | 0.098 | 0.791 | 0.155 | 0.187 |
| PS | 0.300 | 0.090 | 0.826 | 0.102 | 0.210 |
| RF | JMI-CM | 0.302 | 0.091 | 0.837 | 0.151 | 0.278 |
| JMI-SD | 0.313 | 0.098 | 0.825 | 0.183 | 0.254 |
| JMI-MD | 0.309 | 0.095 | 0.830 | 0.184 | 0.255 |
| PS | 0.286 | 0.082 | 0.862 | 0.096 | 0.290 |
| SS | 0.287 | 0.082 | 0.865 | 0.093 | 0.270 |

Legend – RMSE: root mean squared error; C-index: concordance-index; CITL: calibration-in-the-large; FLR: flexible logistic regression; RF: random forest; JMI-CM: conditional mean imputation; JMI-SD: single draw imputation; JMI-MD: multiple draw imputation; PS: pattern submodels; SS: surrogate splits.

**References**

Pollard, T.J., & Johnson, A.E.W. *The MIMIC-III Clinical Database* (2016) [http://dx.doi.org/10.13026/C2XW26](https://doi.org/10.13026/C2XW26)