
TARRAGONA DATATHON 2018
Saturday 10th and Sunday 11th November

Team-8 Study

Members

1. David Riaño, Universitat Rovira i Virgili, Data Scientist, david.riano@urv.cat
2. Alejandro Rodríguez, UCI-HJXXIII, Physician, ahr1161@yahoo.es
3. Laura Claverias, Hospital de Tortosa Verge de la Cinta, Physician, lauraclaverias@gmail.com
4. Jordi Guix, Independent consultant, Data-Base Expert, jordiguix@gmail.com
5. Mohammed Sayed, Universitat Rovira i Virgili, Data Scientist, mgamal.sayed921@gmail.com
6. Luis Felipe Reyes, University of Texas, Physician, reyeslf@utexas.edu
7. Georgina Ballvé, Universitat Rovira i Virgili, Data Scientist, ginabs.dev@gmail.com

Medical Question and Challenge

- Question:

Is there a difference in the evolution of community-acquired pneumonia patients when they receive a macrolide-based treatment or not?

- Challenge:

Determine whether the above question can be answered by means of the analysis of the data contained in the MIMIC-III database.

Problem Approach: Methodology

1. Analysis of the problem (2h)
   Literature review; Identification of relevant data in MIMIC-III
2. Data Extraction (8h)
   SQL queries: incremental approach
3. Data Analysis
   Classification of admissions in Macrolide-based and not Macrolide-based; All patients analysis and SOFA 4-5 (severe) analysis
4. Conclusions of the Study
5. Future actions

Analysis of the Problem (1): Literature Review

OBJECTIVES

- Determine the Cohort
  - Definition and Codification of CAP
- Determine the relevant data
  - Identify signs, risk factors, comorbidities, etc. that are relevant to CAP

RESULT

- Two CAP patients (ICD9 codes)
  - Prim Diag: *480.0-483.99, 485.X-487.X* (Pneumonia codes)
  - Prim Diag: *518.81*(Resp. Failure) or *0.38.XX* (sepsis) + Sec Diag: Pneumonia Codes
- CAP related signs (literature)
- CAP related risk factors (literature)
- CAP related procedures and ATB (literature)

Analysis of the Problem (2): Parameters in MIMIC-III

OBJECTIVES

- Identify in MIMIC-III table *d_item* the available data that are relevant to the analysis
- Note the tables contained these parameters

RESULTS

- Involved Tables (Metavision): *CHARTEVENTS, DATETIMEEVENTS, INPUTEVENTS_MV, PROCEDUREEVENTS_MV*
- Patient descriptors (num param)
- Treatment descriptors (num param)
- Evolution descriptors (num param):

Data Extraction (Q1): Cohort

OBJECTIVES

- Isolate the patients that are under the definition of *CAP and older than 17*
- List the admission-IDs of these cases
- Determine the N (num of CAP episodes)

RESULTS

- *N = 1749 patients*

QUERY

```
WITH edad AS (
  SELECT adm.HADM_ID , pat.GENDER, DATE_DIFF(DATE(adm.ADMITTIME),
DATE(pat.dob), YEAR) AS age, pat.SUBJECT_ID
  FROM `physionet-data.mimiciii_clinical.patients` AS pat INNER JOIN
`physionet-data.mimiciii_clinical.admissions` AS adm
  ON pat.SUBJECT_ID = adm.SUBJECT_ID
),
pneumonia AS (
  SELECT * FROM edad GROUP BY HADM_ID, GENDER, age, SUBJECT_ID
  HAVING ( age > 17 )
)
SELECT diag.HADM_ID, diag.SUBJECT_ID, pneumonia.age, pneumonia.gender,
diag.ICD9_CODE, diag.SEQ_NUM
 FROM `physionet-data.mimiciii_clinical.diagnoses_icd` AS diag
 INNER JOIN pneumonia
 ON diag.HADM_ID = pneumonia.HADM_ID
 WHERE (SEQ_NUM = 1)
   AND (REGEXP_CONTAINS(ICD9_CODE, "^48[0123567]"))
 GROUP BY diag. HADM_ID, diag.SUBJECT_ID, pneumonia.age,
pneumonia.gender, diag.ICD9_CODE, diag.SEQ_NUM
```

Data Extraction (Q2): ATB treatment type

OBJECTIVES

- Determine which patients receive which ATB in the first stage of treatment

RESULTS

- *N = 747 (ATB in the first time treatment)*
- *Macrolides 143 patients*
- *NO macrolides 604 patients*

QUERY

```
SELECT
        hadm_id ,
        subject_id ,
        itemid,
        starttime,
        amount,
        amountuom,
        patientweight
FROM `physionet-data.mimiciii_clinical.inputevents_mv`
WHERE HADM_ID =     TARGET_PATIENTS
```

Data Extraction (Q3): Patient Illness Severity

OBJECTIVES

- Determine type of patient according to SOFA

RESULTS

- *SOFA value (with marcrolides) = 3.57*
- *SOFA value (without marcrolides) = 3.61*
- *SOFA 4-5 (with marcrolides) = 63*
- *SOFA 4-5 (without marcrolides) = 275*

QUERY

```
SELECT
        subject_id,
        hadm_id,
        icustay_id,
        SOFA
FROM `physionet-data.mimiciii_derived.sofa`
WHERE HADM_ID = TARGET_PATIENTS
```

Conclusions

- **Comunity-Acquired Pneumonia**
- **Macrolide-Based vs. non macrolide-Based Treatment**

- **We found no statistically significant differences on mortality considering all patients**
- **We didn’t find differences even after adjusting for disease severity (SOFA 4-5)**

Future Actions

- Better analysis of the patient parameters
- Clustering of patients in order to determine those groups that have a better evolution for Macrolide-based and or non macrolide-Based treatments
- More accurate application of ML methods

THANKS TO

Datathon Organizers at UCI-HJ23

MIT folks for their support during the meeting