



Data Analytics Strategy

value of the portfolio

Gustavo B. Cruz & Jakob Olbrich

CONTENTS

Overview	02
Download Database	03
Clean Data	04
Step 1: Solution	05
Step 2: Solution	06
Discussion	07

Overview

To address the interests of the doctors from Germany regarding the possible adverse effects of drugs used to treat neurological pain, particularly comparing Tramal and Lyrica, the analysis will proceed as follows:

- We Analyzed the dataset to determine the ten most common adverse effects associated with Tramal medication;
- Subsequently, compared the adverse effects of Tramal with those of Lyrica to assess their similarity;
- Finally, a discuss the findings obtained from these studies and their implications.

Download Database

The database was obtained from the FAERS webpage for the year 2019, this data is publicly accessible to everyone. I opted to download the data in ASCII format.

2019		
October - December 2019 <i>posted on 5-Feb-2020</i>	ASCII (ZIP - 60MB)	XML (ZIP - 113MB)
July - September 2019 <i>posted on 7-Nov-2019</i>	ASCII (ZIP - 62MB)	XML (ZIP - 118MB)
April - June 2019 <i>posted on 1-Aug-2019</i>	ASCII (ZIP - 62MB)	XML (ZIP - 118MB)
January - March 2019 <i>posted on 8-May-2019</i>	ASCII (ZIP - 56MB)	XML (ZIP - 103MB)



Once downloaded, the data was imported into RStudio using R script.

Data	
DRUG19Q1	826599 obs. of 20 variables
DRUG19Q2	920704 obs. of 20 variables
DRUG19Q3	943128 obs. of 20 variables
DRUG19Q4	843516 obs. of 20 variables
REAC19Q1	638358 obs. of 4 variables
REAC19Q2	718607 obs. of 4 variables
REAC19Q3	755582 obs. of 4 variables
REAC19Q4	679416 obs. of 4 variables

Clean Data

After importing this data into RStudio, it was imperative to organize it. Initially, only the 'primaryid' and 'drugname' columns were extracted from the dataset. Eventually, they were merged based on the 'primaryid' column, resulting in a dataset comprising approximately 12 million rows and three columns.

```
dnames12 <- rbind(dnames1,dnames2)
head(dnames12, n=5)
dnames34 <- rbind(dnames3,dnames4)
dnames <- rbind(dnames12, dnames34)
rnames12 <- rbind(rnames1,rnames2)
head(rnames12, n=5)
rnames34 <- rbind(rnames3,rnames4)
rnames <- rbind(rnames12, rnames34)
```



```
R | Global Environment
df 12179537 obs. of 3 variables
df <- merge(dnames, rnames, by = "primaryid", all = TRUE)
```

All = True was used to fill up possible missing values with NaN

Subsequently, it was essential to filter the dataset for the studied medications, Tramal and Lyrica. This cleaning process ensured that the data was prepared for analysis.

```
tramal = filter(df, drugname == "Tramal")
lyrica = filter(df, drugname == "LYRICA")
```



```
R | Global Environment
lyrica 82278 obs. of 3 variables
 $ primaryid: chr "1001090113" "1001090113" "1001090113" "1001090113" ...
 $ drugname : chr "LYRICA" "LYRICA" "LYRICA" "LYRICA" ...
 $ pt       : chr "Periarthritis" "Osteoporosis" "Hyperuricaemia" "Diarrhoea" ...

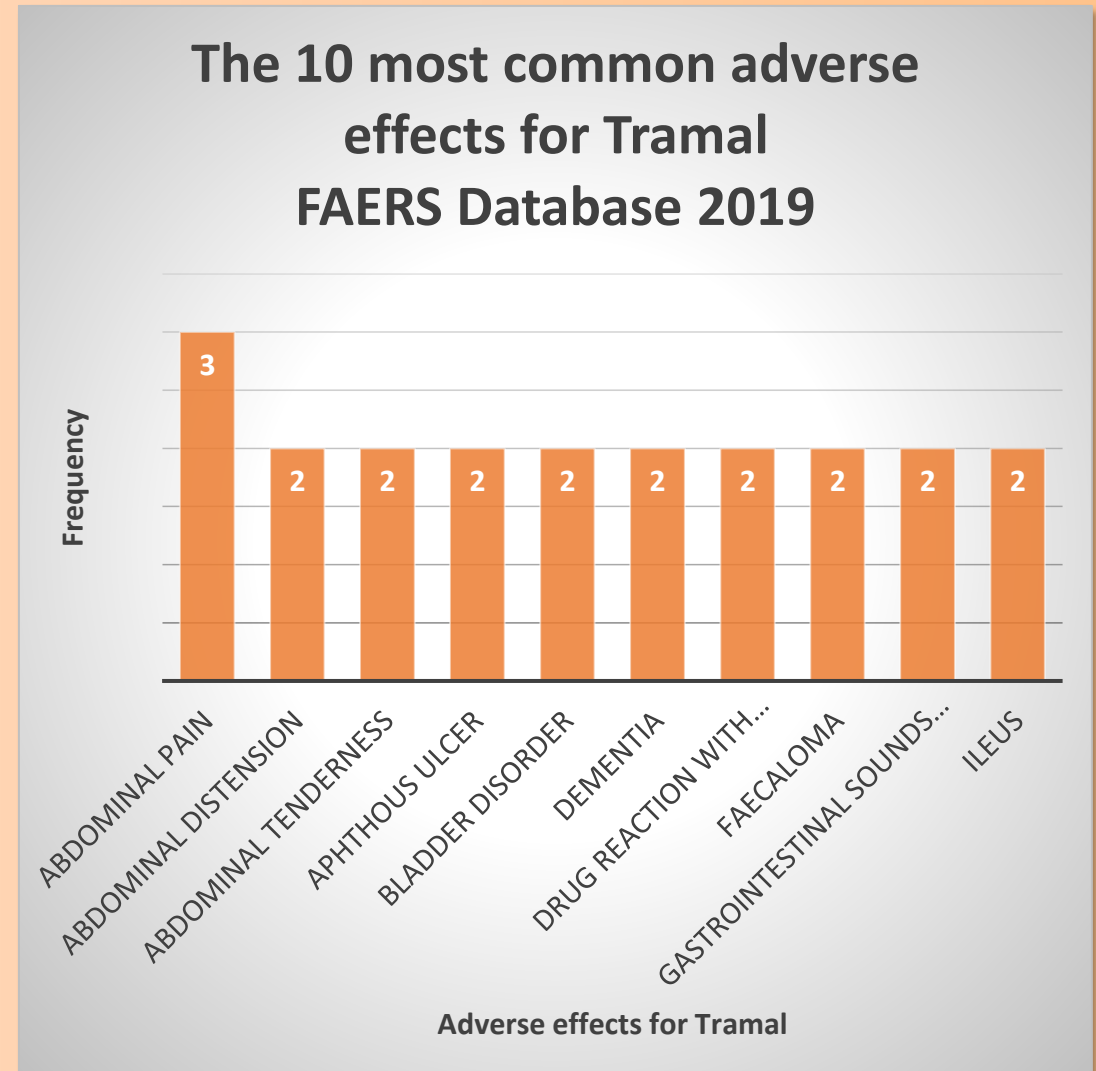
tramal 74 obs. of 3 variables
 $ primaryid: chr "151315613" "1519267810" "1519267810" "1519267810" ...
 $ drugname : chr "Tramal" "Tramal" "Tramal" "Tramal" ...
 $ pt       : chr NA "Faecaloma" "Ileus" "Subileus" ...
```

Step 1: Solution

To determine the most common adverse effects associated with Tramal medication, it was essential to compile them into a table and arrange them in descending order:

```
freq_tramal <- sort(table(tramal$pt), decreasing = TRUE)
freq_tramal_percent <- sort(100*prop.table(table(tramal$pt)),
                           decreasing = TRUE)
freq_tramal[1:10]
freq_tramal_percent[1:10]
```

Following this analysis, a bar graph was generated, illustrating that Abdominal pain emerged as the most prevalent adverse effect for these medications, occurring three times out of 74 instances, accounting for 5.26% of the total. The remaining adverse effects manifested with the same frequency, occurring twice each (3.51%).



Step 2: solution

To compare whether Tramal and Lyrica medications share similar adverse effects, it was imperative to intersect them. Subsequently, the number of rows in the new dataframe was counted, and the results are detailed in the table provided right here:

```
tra_lyr <- intersect(tramal$pt, lyrica$pt)
length(tra_lyr) #33-1 dados semelhantes (1 is NaN)
|
tra_lyr
write.csv(tra_lyr, file = "C:/Users/Startklar/Desktop/GetJob/Documents/PwC_Business/task5")
```



Abdominal distension
Abdominal pain
Abdominal tenderness
Acute generalised exanthematous
pustulosis
Anaemia
Aphthous ulcer
Bladder disorder
Cholangitis
Death
Delirium
Dementia
Drug reaction with eosinophilia and
systemic symptoms
Faecaloma
Gastric ulcer
Gastrointestinal sounds abnormal
Hepatitis
Ileus
Influenza
Maternal exposure during pregnancy
Maternal exposure during pregnancy
Migraine
Nausea
Neurotoxicity
Oesophageal varices haemorrhage
Pancreatitis
Product dose omission
Psychomotor hyperactivity
Pyrexia
Respiratory tract infection
Skin discolouration
Subileus
Uveitis
White blood cell count increased

Discussion

It's intriguing to compare the medications Tramal and Lyrica. When we analyze the most commons adverse effects associated with Lyrica, we observe that they are precisely the same as those for Tramal. Remarkably, in a dataset comprising 82,878 rows, the proportions remain identical as well.

Most 10 common adverse effects for Tramal and Lyrica	Frequency (%)
Abdominal pain	5.26%
Abdominal distension	3.51%
Abdominal tendernes	3.51%
Aphthous ulcer	3.51%
Bladder disorder	3.51%
Dementia	3.51%
Drug reaction with eosinophilia and systemic symptoms	3.51%
Faecaloma	3.51%
Gastrointestinal sounds abnormal	3.51%
Ileus	3.51%


I uploaded the Rscript that I used in this case into a cloud service, you can find this file [here](#).




Thank you for your attention!

Gustavo B. Cruz & Jakob Olbrich

CONTACT INFORMATION

 +41 76 541 56 35

 gboaventuracruz@gmail.com

 [LinkedIn](#)

 <https://www.pwc.ch/>