insight into value

Erwan LE COVEC
Data Scientist
erwan.lecovec@keyrus.com

INNOVATIVE CONTRACT
RESEARCH ORGANIZATION

ADVERSE DRUG REACTIONS ON SOCIAL MEDIA: BIAS AND LIMITATION

DMB

KEYRUS
biopharma

AGENDA

Context

- Context
- Previous paper
- Goal

Method

- Data sources
- Languages
- Metadata

Results & Discussion

- Results
- Discussion

KEYRUS
biopharma

AGENDA

Context

- Context
- Previous paper
- Goal

Method

- Data sources
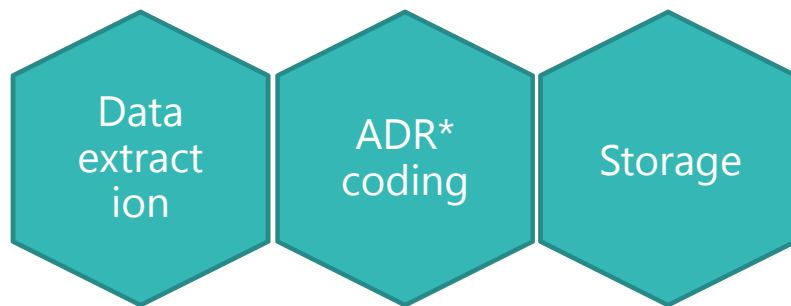- Languages
- Metadata

Results & Discussion

- Results
- Discussion

CONTEXT – PREVIOUS PAPER

"Patient-generated Health Data (Social Media) is a Potential Source for ADR Reporting"

Goal

Data extraction

ADR* coding

Storage

*Adverse Drug Reaction

Sources

Twitter®, a micro-blogging social network

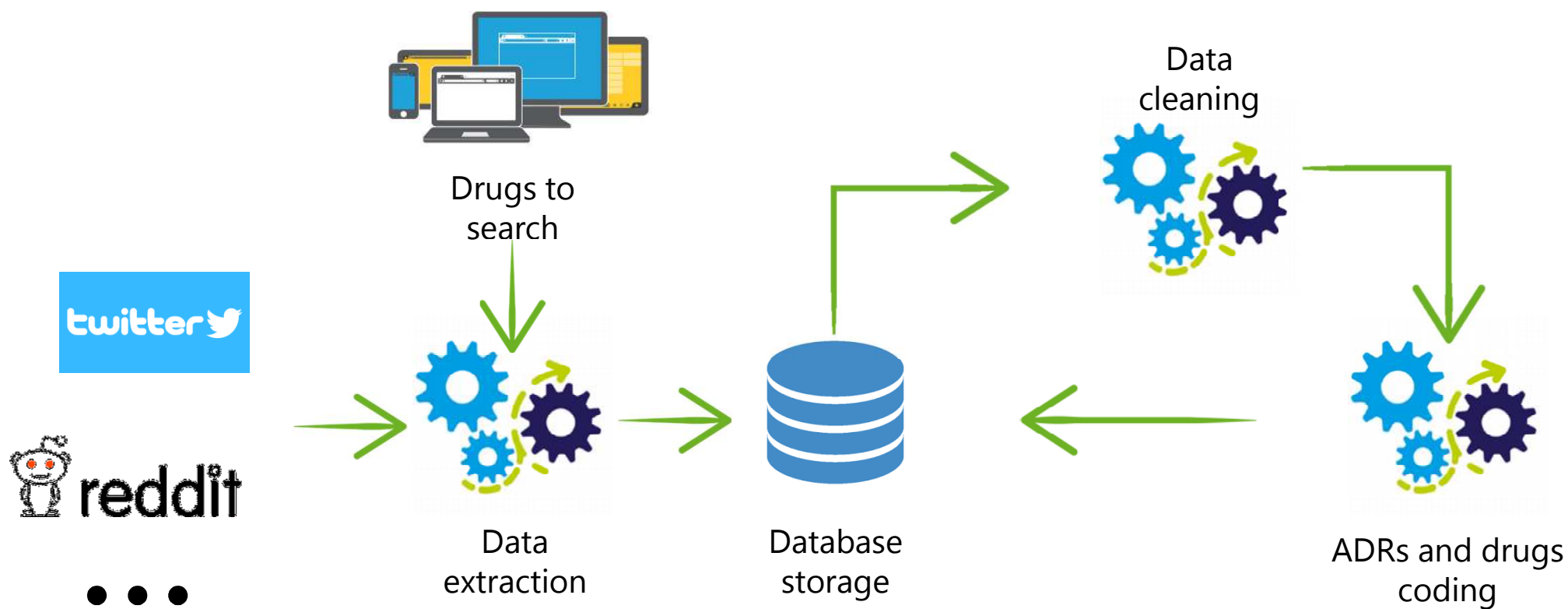*Reddit*®, a widely used forum

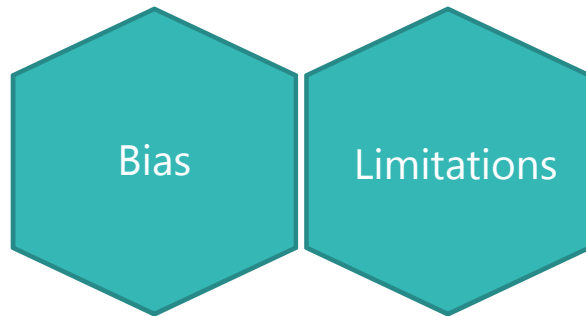**ADR detection possible but subject to limitation:**

1) type of sources   2) languages   3)method

PROCESS FLOW

Drugs to search

twitter

reddit

● ● ●

Data extraction

Database storage

Data cleaning

ADRs and drugs coding

CONTEXT - GOAL

Find & *Estimate*

Bias   Limitations

Of doing ADRs detection on social media

By *collected data* using:

+ Source: classic social media and health related websites

+ Additional language support: using more than just English

Better understand extracted data
Prepare machine learning methods

AGENDA

Context

- Context
- Previous paper
- Goal

Method

- Data sources
- Languages
- Metadata

Results & Discussion

- Results
- Discussion

KEYRUS
biopharma

DATA SOURCES 1/2

- **_Twitter®_**, a micro-blogging social network

- **_Reddit®_**, a widely used forum

- **Carenity®**, an online patient community for patient and caregiver

Are all data sources equal ?

KEYRUS
biopharma

DATA SOURCES 2/2

Twitter

- 12th *most used* website in the *world*

- Used to send short messages (280 characters)

- Used to *share* news, pictures and *everyday life*

reddit

- 9th *most used* website in the *world*

- Used for online discussion

- Contains various *sections* from video games to *drugs usage*

carenity

- *Health-related* forum

- Used to share disease experience with other patient

- Contains **1,000+ pathologies**
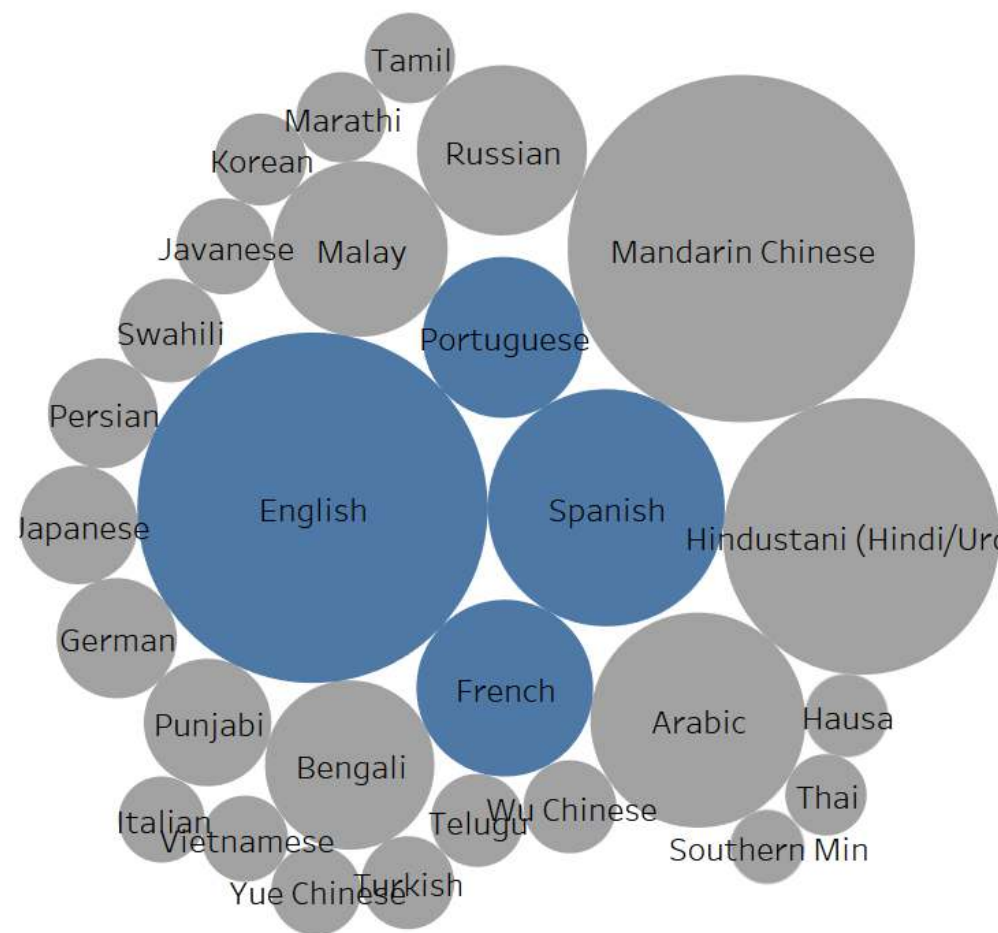
LANGUAGES

Languages used for extraction:

English	French
Spanish	Portuguese

32% of the *world*'s population

Language influence on ADRs detection ?

KEYRUS
biopharma

METADATA

Metadata: data that provides information about other data.

Carenity®:

Created by Chameleon Design
from Noun Project

Gender

Created by Creative Mahira
from Noun Project

Age

Reddit®:

- *None available* for anonymization purpose

Twitter®:

- *Not trustworthy* and difficult to exploit

How can we use metadata for ADRs detection?

AGENDA

Context

- Context
- Previous paper
- Goal

Method

- Data sources
- Languages
- Metadata

Results & Discussion

- Results
- Discussion

KEYRUS
biopharma

RESULTS

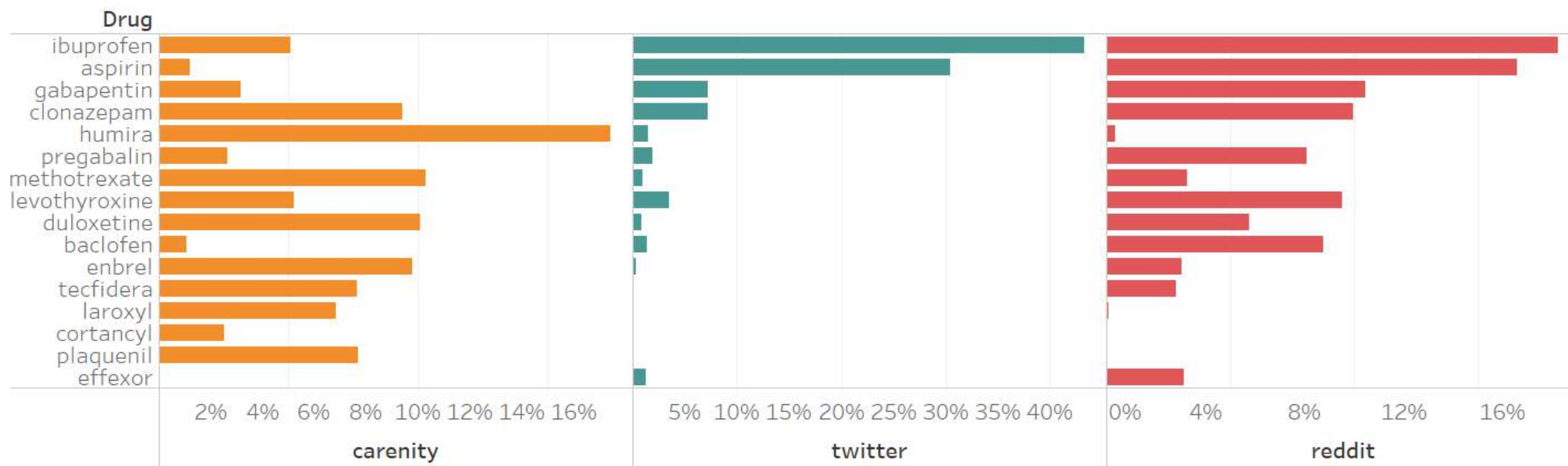|	Twitter	Reddit	Carenity
Period	1 year	2007 - 2018	2011 - 2018
# records	352,285	163,426	12,637

Disclaimer:

The data contained in this paper was obtained for demonstration purposes only using the techniques presented. Anyone analyzing data from social media sites, either public or membership based, should investigate the source terms and conditions relevant to data reuse and obtain any necessary permissions if required.

Detection ~40% of records with potential ADR*

*Potential ADR = ADR and other medical events

KEYRUS
biopharma

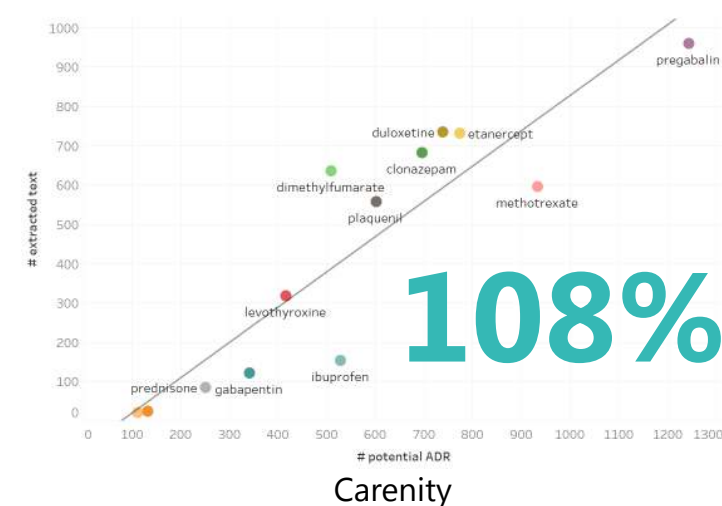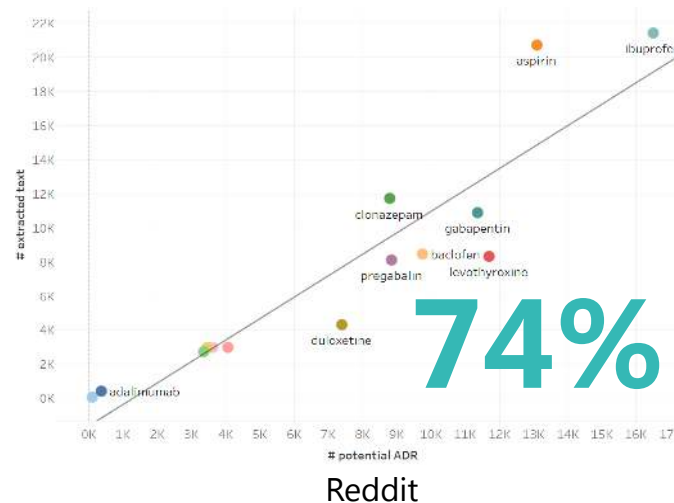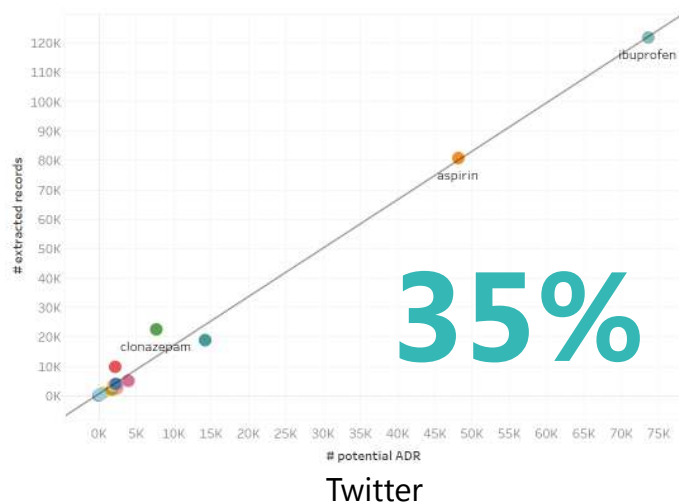# DRUG REPRESENTATION IN TWITTER® VS REDDIT® VS CARENITY®

- *Not similar* data *repartition*
- *Aspirin* and *Ibuprofen* over *representation* on Twitter
- *Better* drugs representation on *forums*

All data sources are not equal

KEYRUS
biopharma

RECORDS AND POTENTIAL ADR RATIO
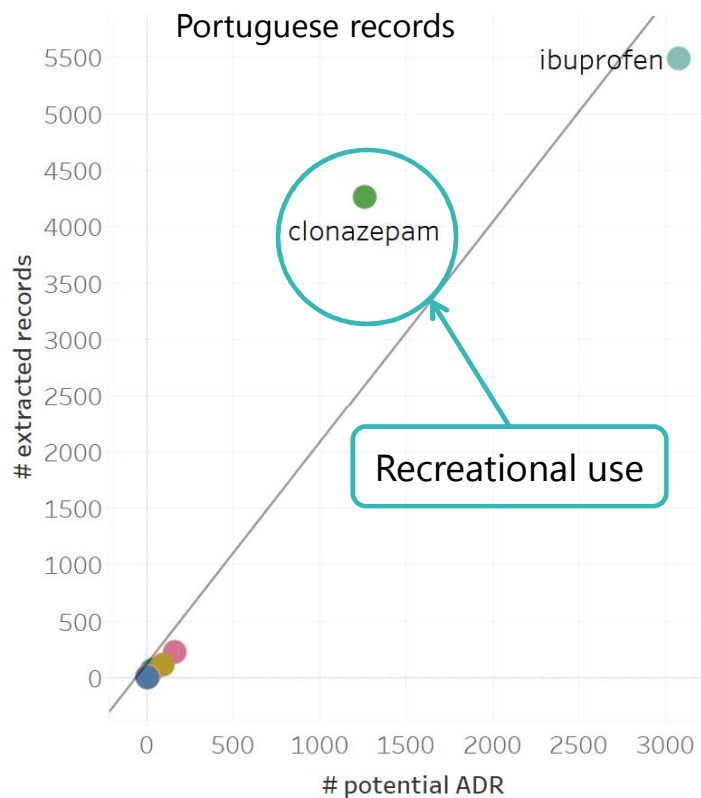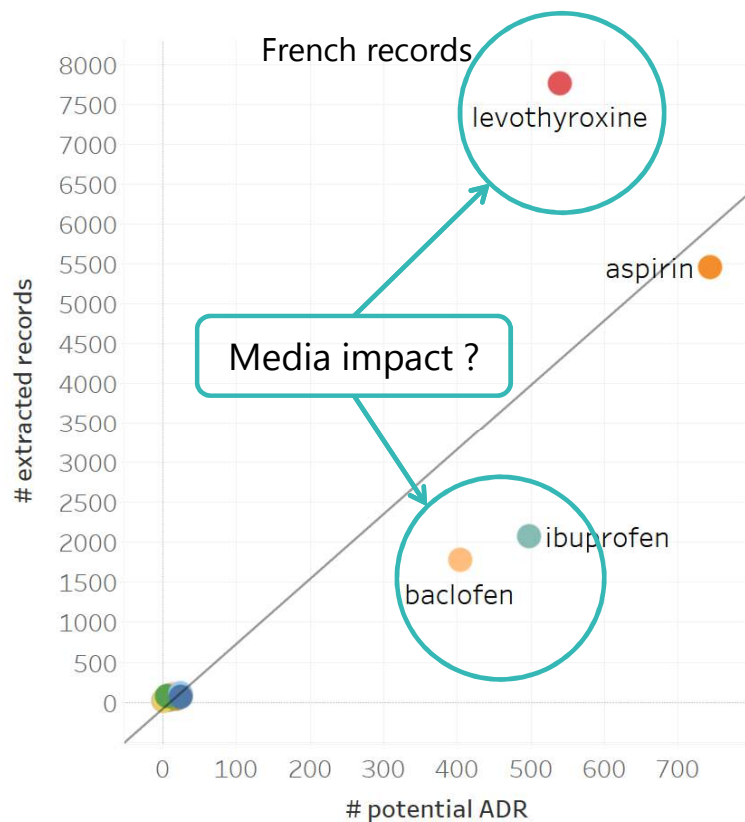

35%

Twitter

74%

Reddit

108%

Carenity

- *Ratio* very *different* depending on *source*

- *Forums* = *higher* ratio

- Twitter *noisier*

Forums carry more valuable information ?

POTENTIAL ADR RATIO ON TWITTER® (FRENCH & PORTUGUESE RECORDS)

French records

levothyroxine

aspirin

Media impact ?

ibuprofen

baclofen

extracted records

potential ADR

Portuguese records

ibuprofen

clonazepam

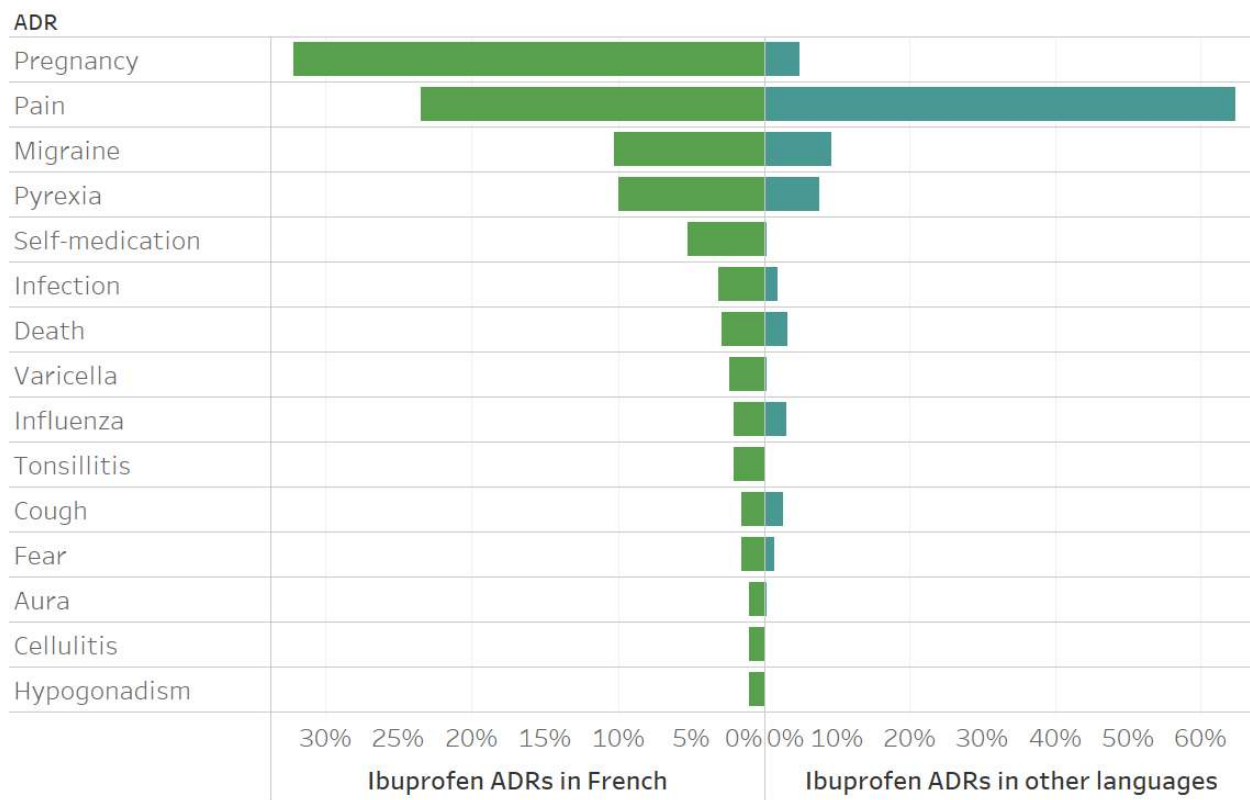Recreational use

extracted records

potential ADR

Low ratio (potential ADR/records) *in French*: ~5 to 35%

- Language complexity

- French market oriented

Low ratio (potential ADR/records) *in Portuguese*: ~25 to 40%

Languages difference

KEYRUS biopharma

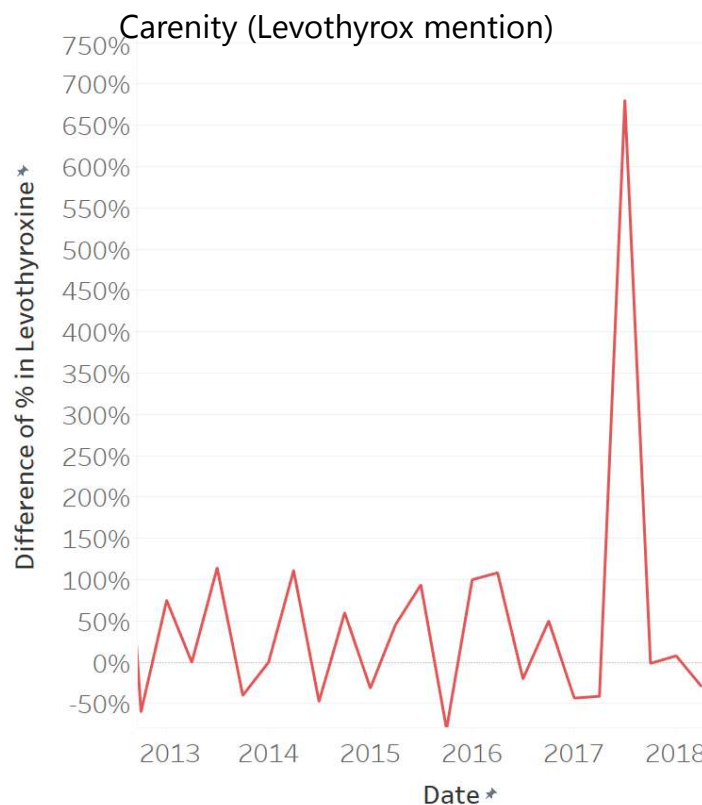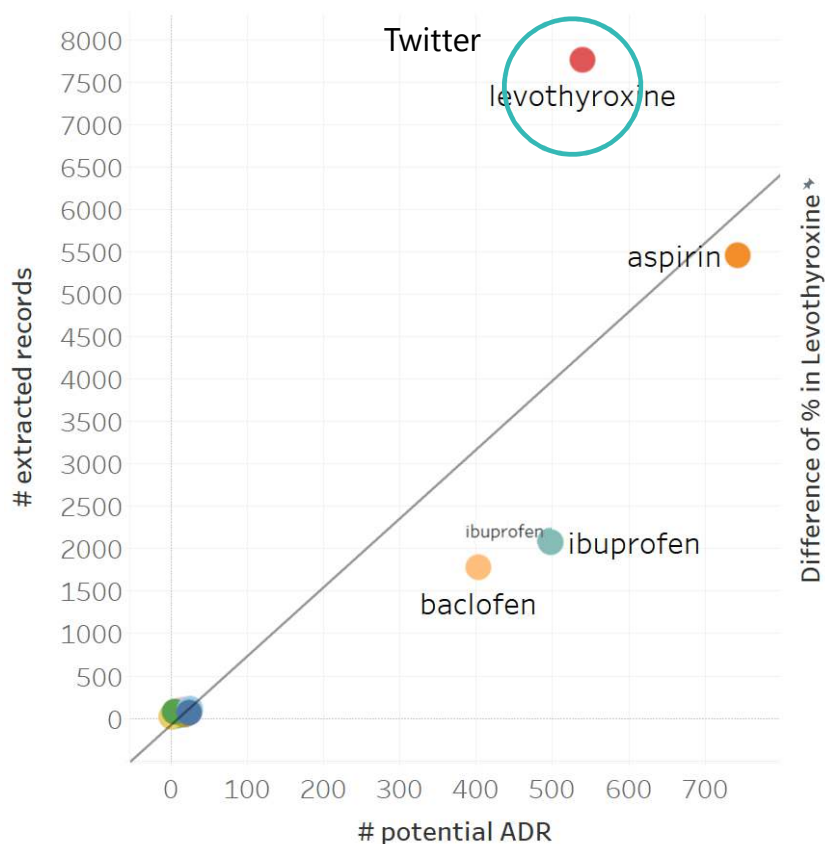# ADRS DETECTED FOR IBUPROFEN: FRENCH VS OTHER LANGUAGES IN TWITTER

Detected ADR are not similar:

- French first potential ADR: Pregnancy

- First potential ADR for other languages: Pain

→ Cultural and not grammatical characteristic

Country specific

MEDIA IMPACT IN FRENCH

- Twitter: *lots of Levothyroxine* records

- Carenity: *high peak* of levothyroxine mention in *2017*
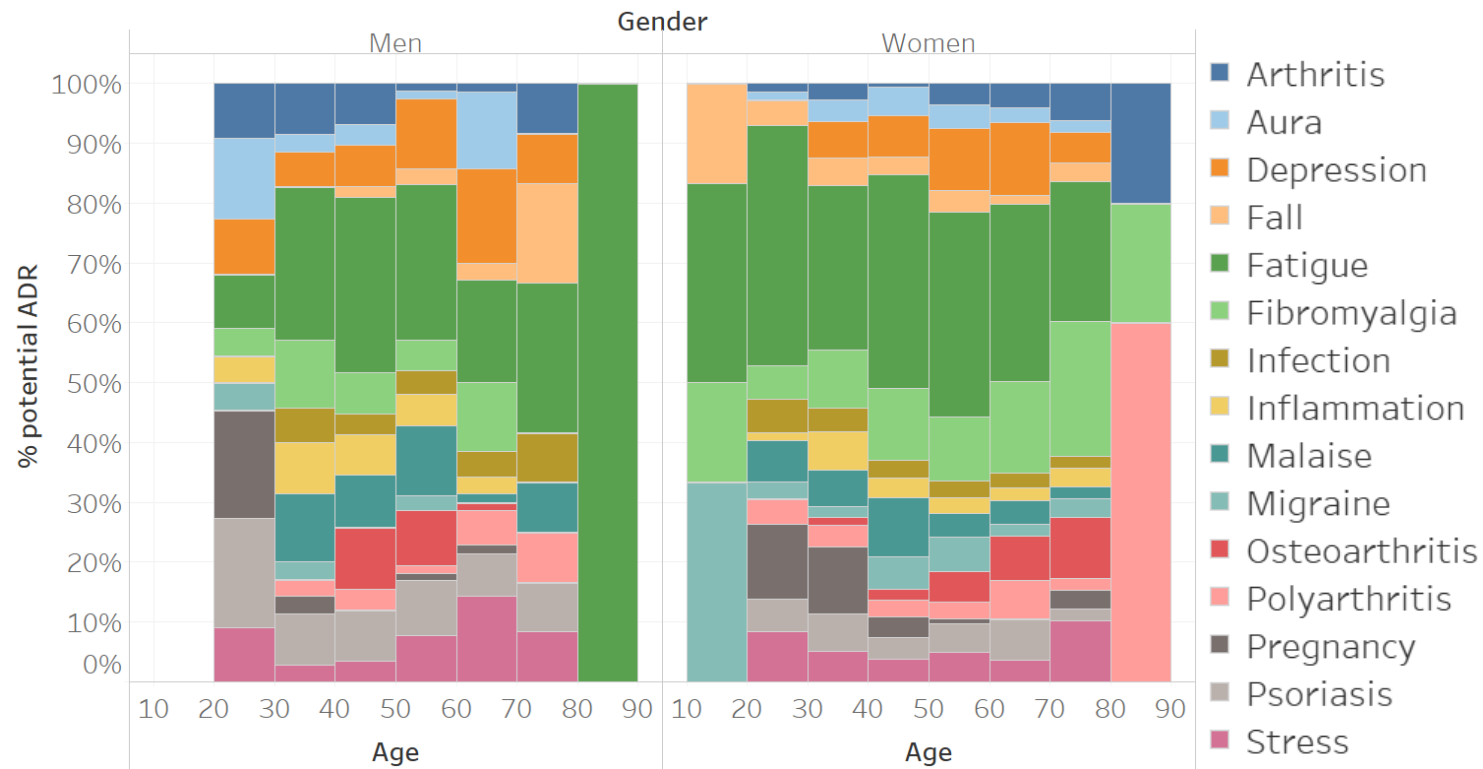
→ *Media impact*

Levothyrox : l'Agence du médicament reconnaît un défaut d'information aux patients

5 000 signalements d'effets inhabituels ont été enregistrés fin août. Parmi les 3 millions de patients traités, certains se reportent sur des solutions prévues pour les enfants.

LE MONDE | 01.09.2017 à 11h35 • Mis à jour le 02.09.2017 à 06h39 | Par Pascale Santi

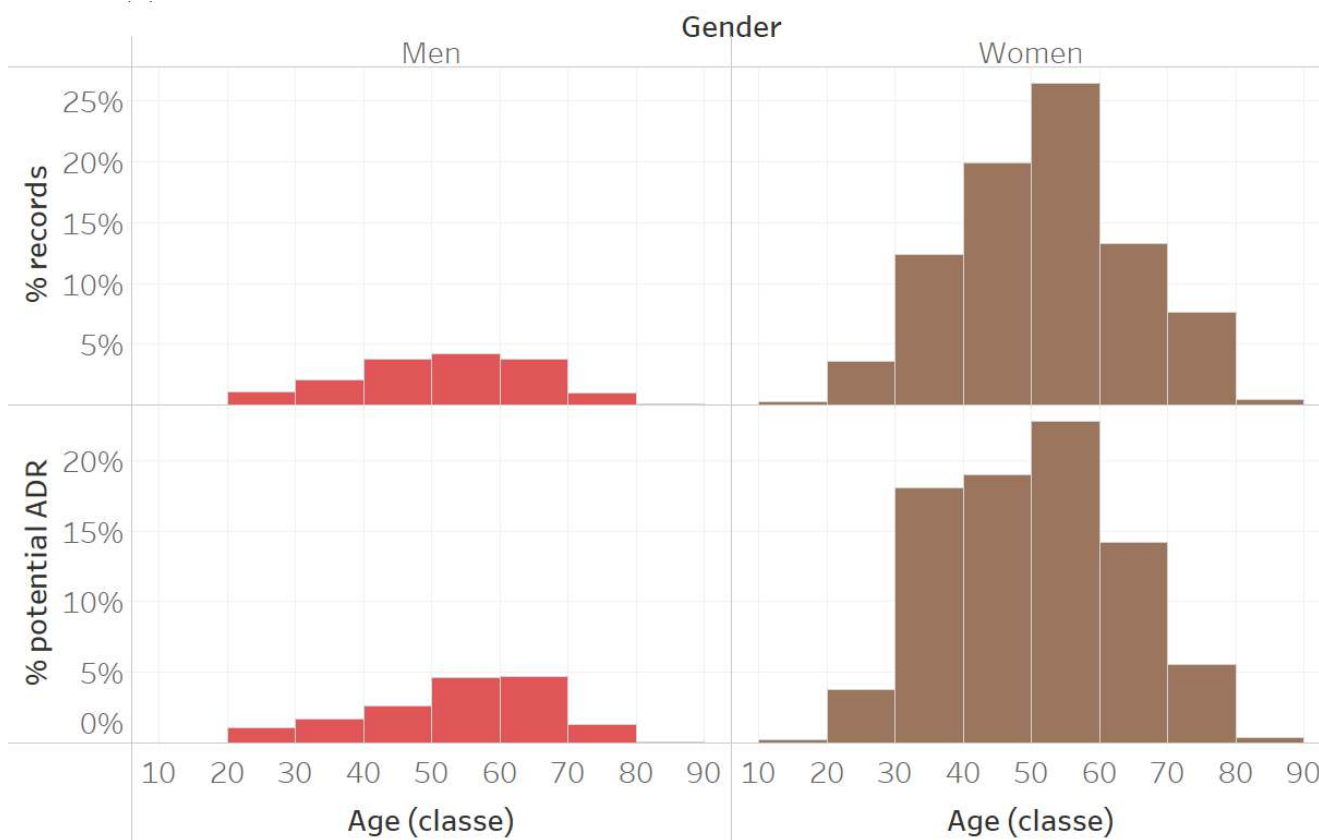## All sources subject to media impact

POTENTIAL ADRS DETECTION BY AGE AND GENDER ON CARENITY

Close but *not similar* distribution between *men* and *women*

ADRs and pathologies not similar between gender

KEYRUS
biopharma

AGE & GENDER REPARTITION ON CARENITY

- **Consistent** age distribution

- **Women**: **80%** of the records

- Women **~35yo** mention **more ADRs**

Women more willingly to share their experiences ?

KEYRUS
biopharma

REAL ADR VS CONDITION (SUBSET FOR METHOTREXATE)

	Reddit® En	Twitter® En	Carenity® En	Carenity® Fr
# detected ADR	318	318	318	318
# real ADR % of the total	31 *10%*	17 *5%*	54 *17%*	23 *7%*
# ADR missed % of missed	6 *16%*	4 *19%*	4 *7%*	32 *58%*

- *Carenity* and *Reddit* in *English*: most *accurate* sources

- *Twitter*: *not* a lot of *useful* information

- *Carenity* in *french*: *lot* of ADR *missed*

  → Language barrier ❓

Forums seems more valuable, especially health-related ones

KEYRUS
biopharma

CONCLUSION

Objectives

Detect bias and limitation by:
- sources number
- number of languages

Next steps

method
false positives
false negatives

Results

Sources quality
Languages specification
Gender influence

PHARMACOEPIDEMIOLOGY PERSPECTIVES

Pharmacoepidemiology, the study in *real conditions* and on *large populations* of:

Effectiveness

Use

Risk of drugs

More suitable than Pharmacovigilance

QUESTIONS

Erwan Le Covec

Lise Radoszycki

Stéphane Chollet

**Keyrus
Biopharma**

KEYRUS
biopharma