BIG DATA in the context of Pharmacovigilance

ML. Kürzinger
Pharmacoepidemiologist
Global pharmacovigilance and Epidemiology
Sanofi R&D

SANOFI Research & Development

Paris BD 2016 - Télécom ParisTech, 24th March 2016

AGENDA

1. **Social media = New sources of data for pharmacovigilance**
2. **Big data and pharmacovigilance: potential for web-based data mining**
    1. Examples of ongoing initiatives across different data sources
        1. Social media and WEB RADR
        2. Query logs and Microsoft
        3. Patients forums and Kappa Santé Detec't
3. **Conclusion**

Definitions

1. Pharmacovigilance

Pharmacovigilance (PV) is defined as the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem.

2. Signal

A ‘signal’ consists of reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously.

UPCOMING NEW PHARMACOVIGILANCE DATA SOURCES

FULLY ESTABLISHED

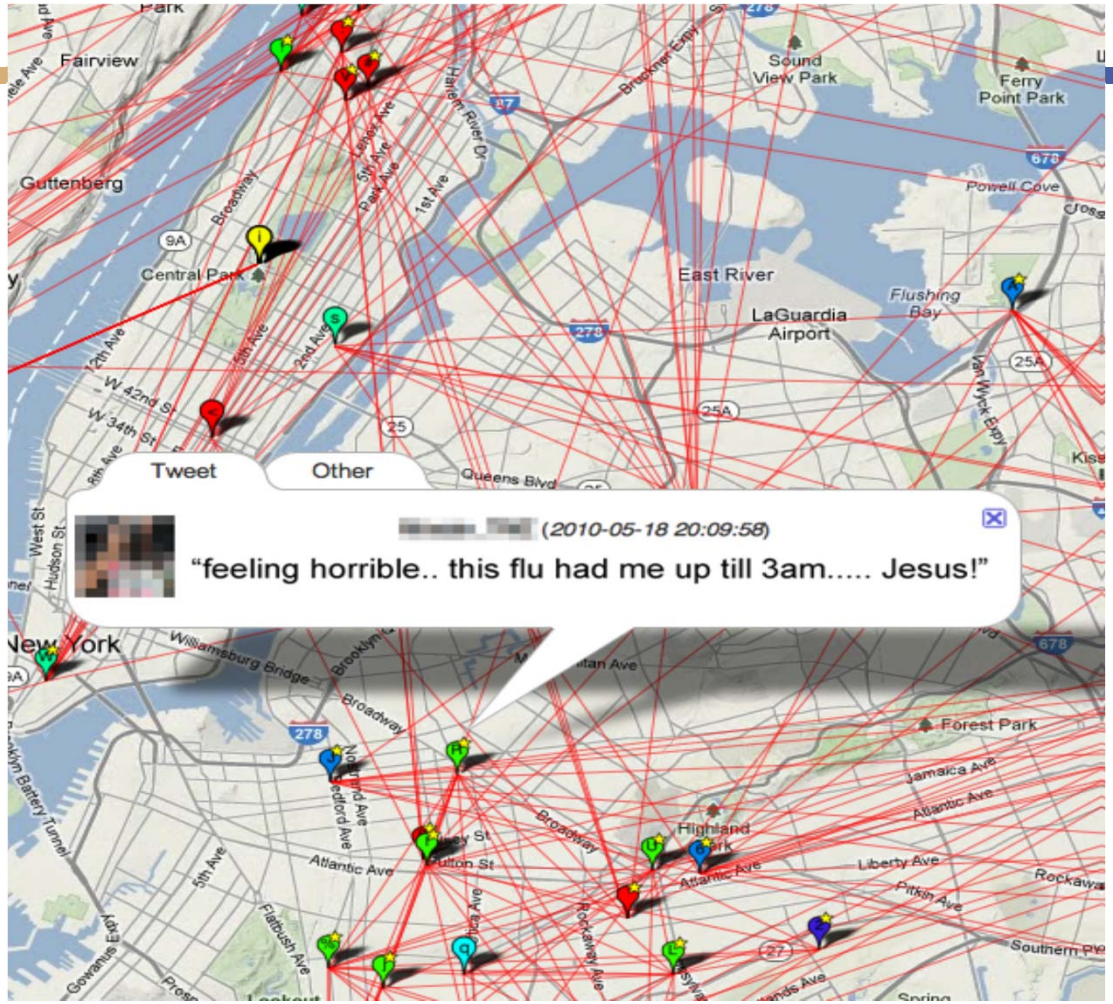
- Patients, health care professionals, pharmacists
- Electronic medical records
- Claims databases
- Spontaneous reporting system

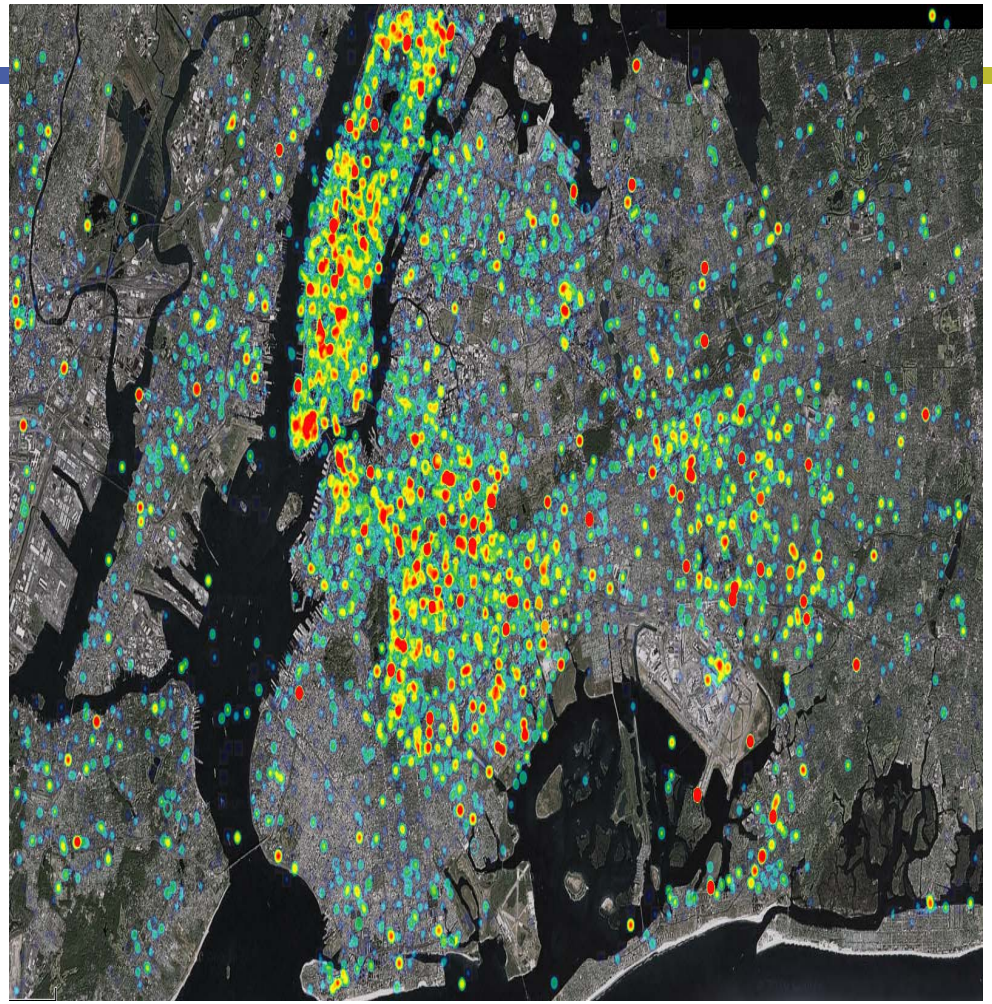
plus

UNDER DEVELOPMENT

- Web-based, Internet search (e.g., Google, Bing)
- Social media (e.g., Facebook, Twitter)
- Patient Forums (e.g. PatientsLikeMe, Doctissimo)

TWITTER AND FLU IN NYC





New York City, Twitter friends:

Texting flu (+ specific drug) could mean a signal for that drug

New York City, heat map of Twitter users:

The redder the dot means the larger the number of reports

Source: Sadilek A, Kautz H, Silenzio V. Modeling Spread of Disease from Social Interactions.

http://www.cs.rochester.edu/~sadilek/publications/Sadilek-Kautz-Silenzio_Modeling-Spread-of-Disease-from-Social-Interactions_ICWSM-12.pdf

NOT ALWAYS SUCCESSFUL!

JOHN ANGEL LLOYD/UPP NEWSCO M

The latest US influenza season is more severe and has caused more deaths than usual.

EPIDEMIOLOGY

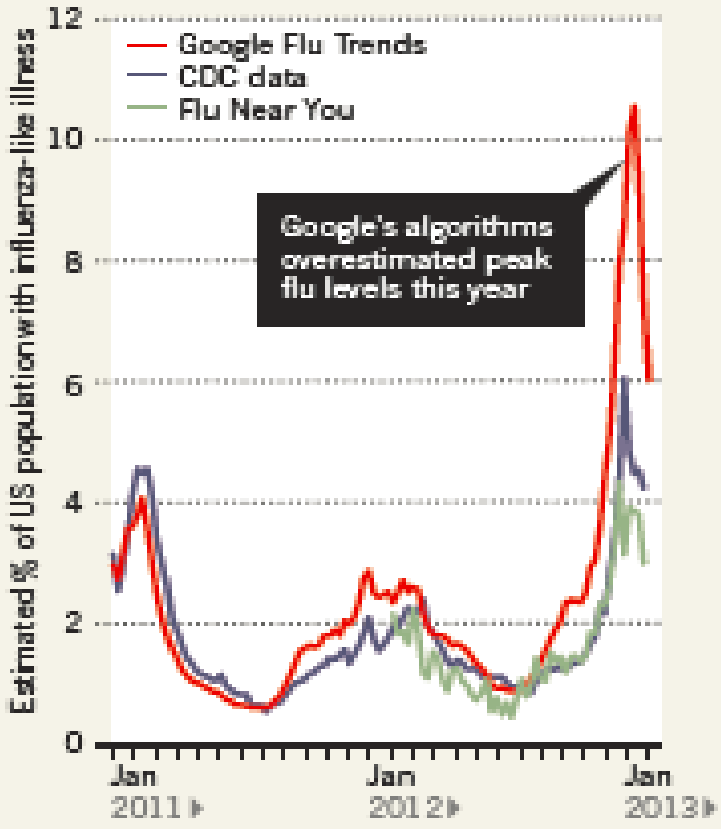
When Google got flu wrong

US outbreak foxes a leading web-based method for tracking seasonal flu.

14 FEBRUARY 2013 | VOL 494 | NATURE | 155

FEVER PEAKS

A comparison of three different methods of measuring the proportion of the US population with an influenza-like illness.



Challenges

- “When Google got flu wrong” (Nature, 14 February 2013)
  - Drastically overestimated peak flu level in 2012
  - Due to widespread media coverage which may have triggered many flu-related searches by people who were not ill
  - Constant adaptation and recalibration are needed

HUGE VARIETY OF SOURCES AND VOLUME OF INFORMATION

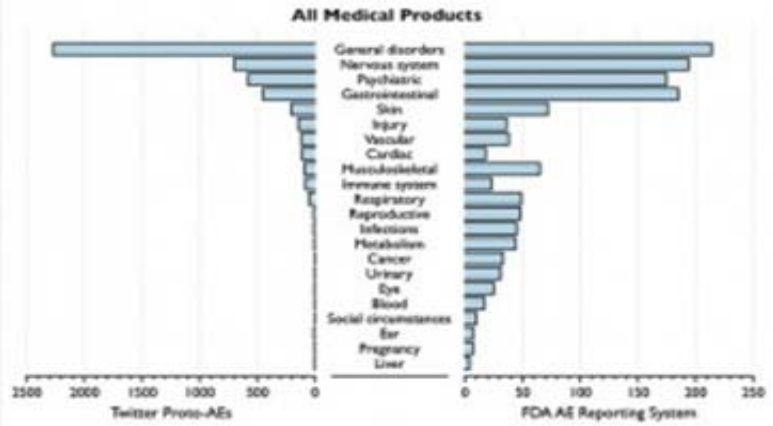
June 2015: FDA Partners With Networking Forum To Gather Adverse Event Data Directly From Patients

FDA taps PatientsLikeMe to test the waters of social media adverse event reporting

By: Jonah Comstock | Jun 15, 2015

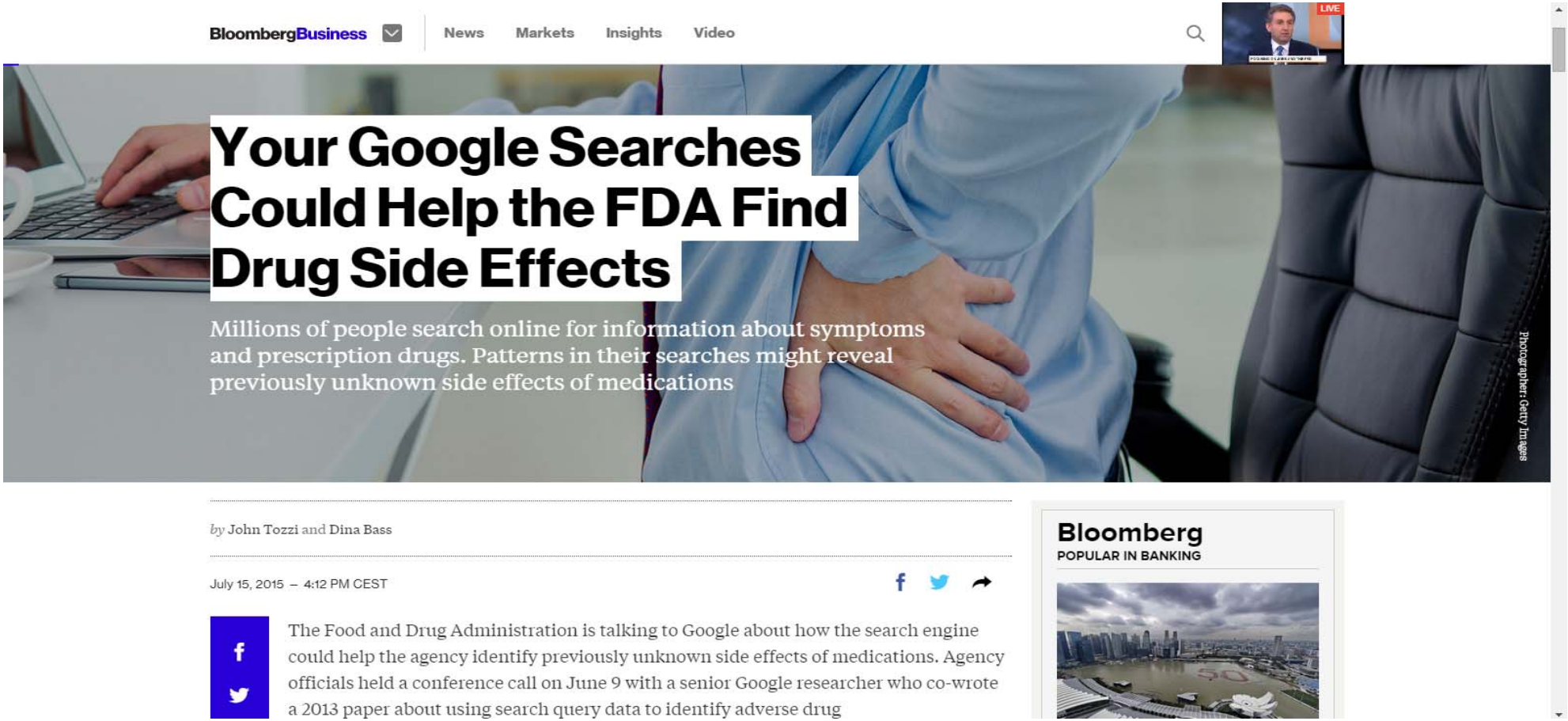
Tags: adverse event reporting | crowdsourced patient data | FDA | PatientsLikeMe | social media adverse event reporting | US FDA |

Online patient community **PatientsLikeMe** has found another partner for its massive repository of patient-generated data on health conditions and treatments, but it's not another pharma company or retail pharmacy. PatientsLikeMe has announced a research partnership with the FDA: The agency will assess the platform's feasibility as a way to generate adverse event reports, which the FDA uses to regulate drugs after their release into the market.



A study last year found that many people reported adverse drug events on Twitter.

July 2015: FDA Talking To Google About Using Data Mining To Identify Unknown Drug Side Effects



BloombergBusiness | News | Markets | Insights | Video

Your Google Searches Could Help the FDA Find Drug Side Effects

Millions of people search online for information about symptoms and prescription drugs. Patterns in their searches might reveal previously unknown side effects of medications

Photographer: Getty Images

by John Tozzi and Dina Bass

July 15, 2015 — 4:12 PM CEST

The Food and Drug Administration is talking to Google about how the search engine could help the agency identify previously unknown side effects of medications. Agency officials held a conference call on June 9 with a senior Google researcher who co-wrote a 2013 paper about using search query data to identify adverse drug

Bloomberg
POPULAR IN BANKING

NEW PHARMACOVIGILANCE DATA SOURCES

- More and more patients discuss online
- Traditional adverse reporting systems a slow to adapt
- Regulation is changing (FDA, EMA)
  - MAHs should regularly screen internet or digital media for potential reports of suspected adverse reaction (Module VI, GPV, EMA)

What is the role/advantages of Social Media in PV?

- **Real time** => early signal detection
- **Massive** scale (millions of messages) => detect unknown signals
- **Patient** insights (voice from the patient directly)

Questions

- “What methods should be used?
- What data sources (what type of web-media)?
  - Query logs
  - Facebook, Twitter
  - Forums
- How good is web-based Pharmacovigilance?
  - How reliable – compared to other sources
  - How valid – compared to “gold standards”

WEB RADR (IMI PROJECT)

WB2B ANALYTICS

http://web-radr.eu/

Innovative Medicines Initiative

SANOFI Research & Development

efpia European Federation of Pharmaceutical Industries and Associations

European Commission

WEB-RADR - Recognising Adverse Drug Reactions

- Public private partnership between the European Commission and European Federation of Pharmaceutical Industries and Associations
- Consortium of organisations including European medicines regulators, academics and the pharmaceutical industry
- 3 year project to develop new ways of gathering information on suspected adverse drug reactions (ADRs)
  - to develop a mobile app for healthcare professionals and the public to report suspected ADRs to national EU regulators.
  - to investigate the potential for publicly available social media data for identifying potential drug safety issues

WP2B ANALYTICS – DATA SOURCES AND METHODS

Social media data

from Jan 2010 Twitter

from Jun 2012 Facebook

Predefined list of drugs

AERS

VIGIBASE

Spontaneous reporting system

(time-indexed reference)

ANALYTICS

Signal detection

PRR

IC025

Assessment of performance

PPV sensitivity

Novelty value

Timing metrics

WEB BASED SIGNAL DETECTION PROJECT USING QUERY LOGS

Collaboration with Microsoft

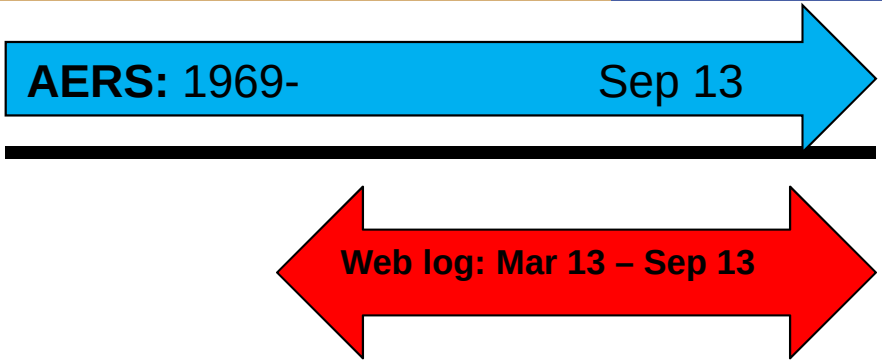
CHALLENGES AND OBJECTIVES

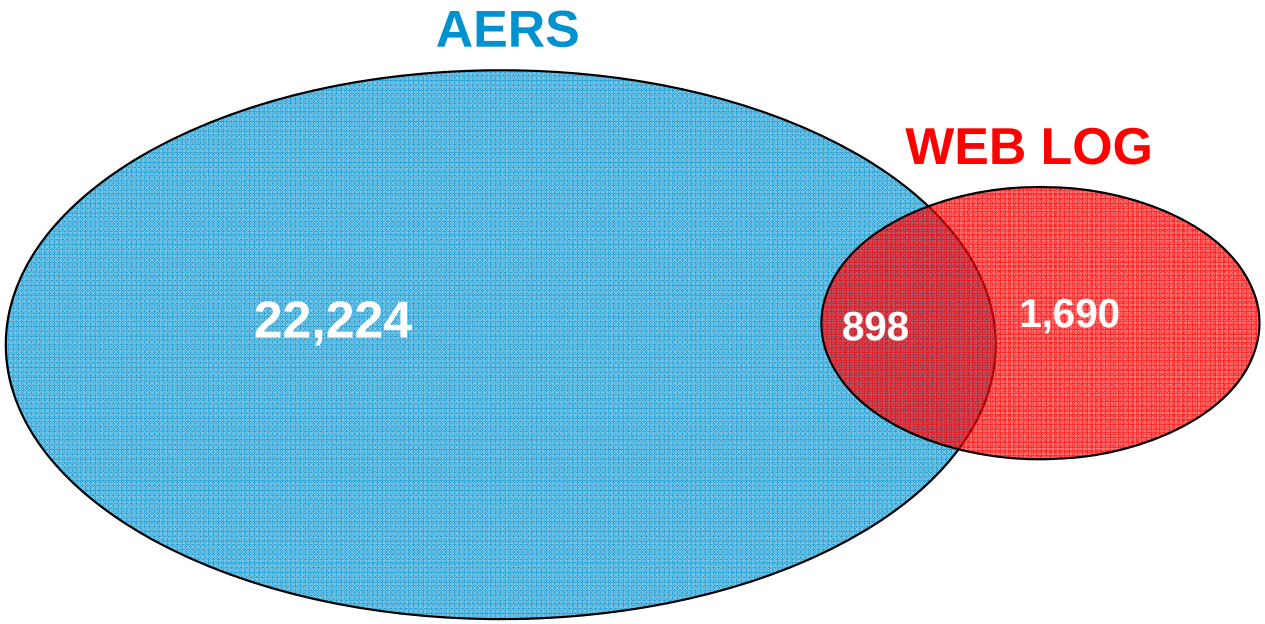
- What methods should be used?
  - To develop and evaluate different methods
- How good is web-based Pharmacovigilance?
  - To estimate the reliability/validity of those methods using different “gold standards”

DATA SOURCES

- Web Log database: Query logs from Microsoft Bing search engines
  - Over 55 million users with at least 1 query
  - Pre-dominantly US internet users (very small proportion non-US)
- FDA AERS database (“gold standard”)
  - Over 9 million reports (since 1969)
  - Over 70% US reports
  - Routinely utilized by GPE since 2001
- Target of 10 marketed drugs
  - From different therapeutic areas, recently marketed or under the market for many years

TIME PERIOD AND DRUG-EVENT PAIRS COUNT





Results: PQR Sensitivity & Specificity (%)

Based on 898 drug-event pairs

FDA AERS	Query log	Sensitivity	Specificity	PPV	NPV
EB05 ≥ 2	PQR ≥ 1	54.17	56.12	6.52	95.59
EBGM ≥ 2	PQR ≥ 1	47.06	55.84	10.03	90.98
EBGM ≥ 4	PQR ≥ 1	81.82	56.03	2.26	99.60
N≥3 and PRR≥2 and PRR_CHISQ≥4	PQR ≥ 1	47.41	56.01	13.78	87.78

NEXT STEPS

- Web log data create too much “noise”, not true signal, “false positive”
- Relies on web-based search – not true diagnosis
- Sensitive to increase in media coverage resulting in increased search
- Prone to changes in people’s search behavior
- No true denominator – could easily underestimate or overestimate peak
- Needs continuous updates on modeling

=> New methods need to be developed for web-based signal detection

WEB BASED SIGNAL DETECTION PROJECT
USING PATIENT FORUMS

Collaboration with Kappa Santé

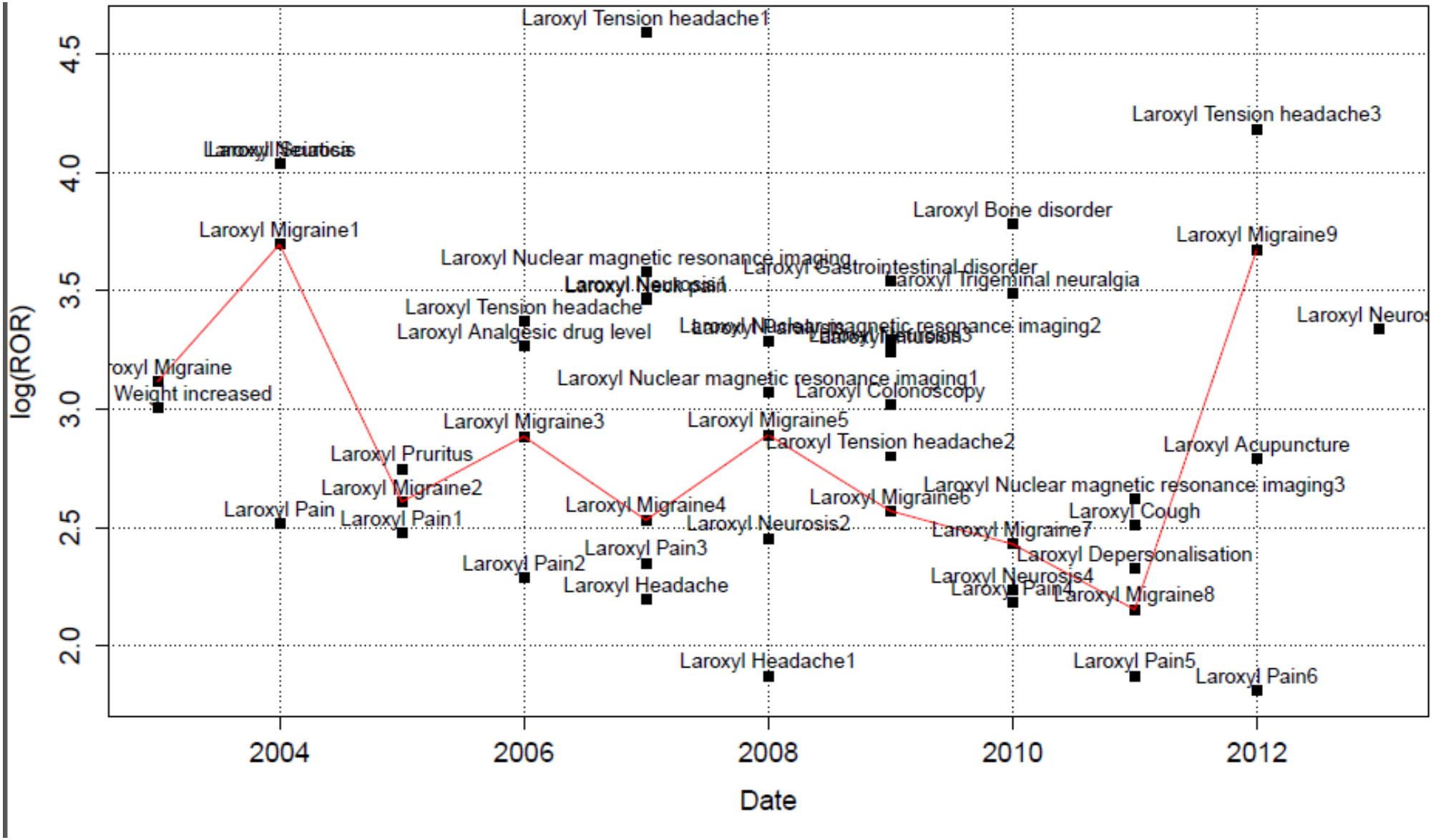
CHALLENGES AND OBJECTIVES

- How to leverage web-based data to early signal detection?
- What are the best methods for web-based signal detection?
- How to measure whether or not the goals have been reached (indicators)?
  - Performance indicators
    - number of new signals detected while undetected by traditional methods,
    - delay between web-based proto-signal and traditional signal

DATA SOURCES

- Patients forums
  - 17,703,218 messages processed over the past decade
  - Data mining techniques
    - Web-crawler
    - Data pre-processing
    - Data processing
      - Annotation including classification (ATC and MEDDRA)
      - Relevance
- FDA AERS database (“gold standard”)
  - Over 9 million reports (since 1969)
  - Over 70% US reports
  - Routinely utilized by GPE since 2001

EXPECTED RESULTS:
TEMPORAL ANALYSIS OF DETECTED SIGNALS



CONCLUSION
BIG DATA ARE ALREADY IN PHARMACOVIGILANCE

- Valuable knowledge can be extracted from social media which has a large volume of timely user generated content
- Data mining pathways being implemented in different sources
- Performance of web-based signal detection being assessed
- Social media guidance being prepared by Health Authorities

Thank you!

Merci!

Gracias!

Danke!

謝謝！

ありがとう！

METHODS USED

Web based query log

Query for the event	Query for the drug?	
	No	Yes
Before Day 0	a	b
After Day 0	c	d
	a+c=N1	b+d=N2

Query Log Reactions Score (**QLRS**)

Proportional query ratio (**PQR**)

PQR = (d/N2)/(c/N1)

FDA AERS

Reported AEs	Event of interest	All other events	Total
Drug of interest	a	b	a+b = M1
All other drugs	c	d	c+d = M2
	a+c = N1	b+d = N2	N

Proportional Reporting Ratio

PRR = (a/M1) / (c/M2)

Empirical Bayes Geometric Mean (**EBGM**)

SOME RECENT PUBLICATIONS

- Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, Upadhaya T, Gonzalez G. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform. 2015 Apr;54:202-12.
- Yang M, Kiang M, Shang W. Filtering big data from social media--Building an early warning system for adverse drug reactions. J Biomed Inform. 2015 Apr;54:230-40
- Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, Dasgupta N. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. Drug Saf. 2014 May;37(5):343-50. Erratum in: Drug Saf. 2014 Jul;37(7):555
- https://webradr.files.wordpress.com/2014/11/web-radr-poster.pdf