June 3, 2016

Overview of Social Listening Methods

Carrie Pierce, Director, Client Delivery

Epidemico, Inc.

Cardiac Safety Research Consortium • FDA White Oak Campus

epidemico

Perspective

Digital Disease Detection — Harnessing the Web for Public Health Surveillance

John S. Brownstein, Ph.D., Clark C. Freifeld, B.S., and Lawrence C. Madoff, M.D.

Share: f ✔ g+ in +

Article   **References**   **Citing Articles (73)**

The Internet has become a critical medium for clinicians, public health practitioners, and laypeople seeking health information. Data about diseases and outbreaks are disseminated not only through online announcements by government agencies but also through informal channels, ranging from press reports to blogs to chat rooms to analyses of Web searches (see Digital Resources for
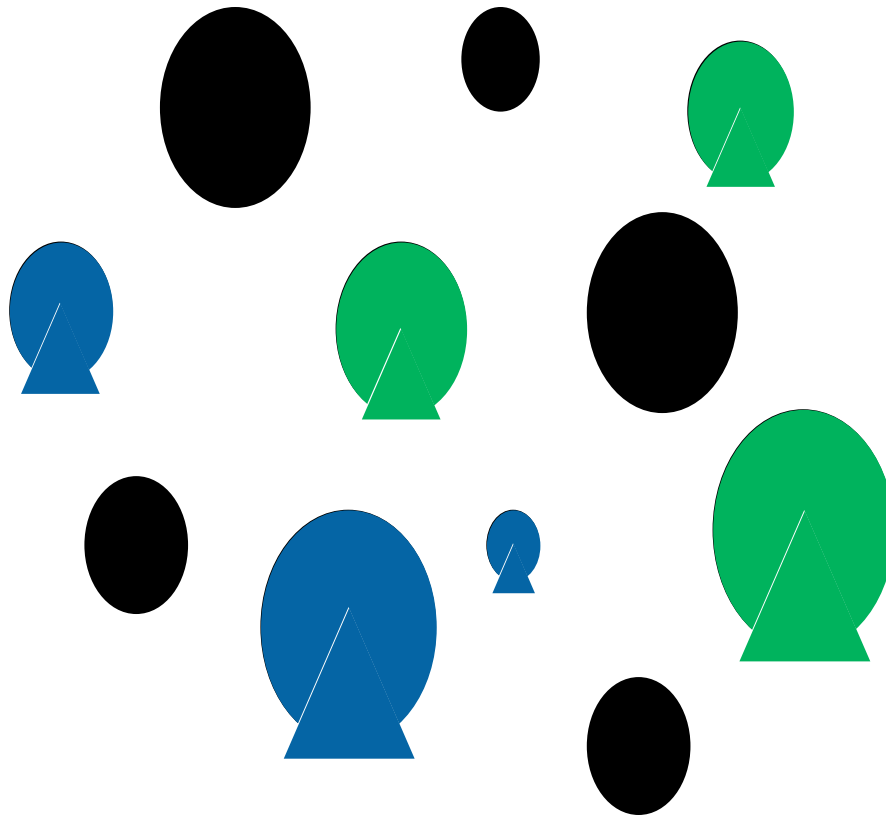
Machine Learning
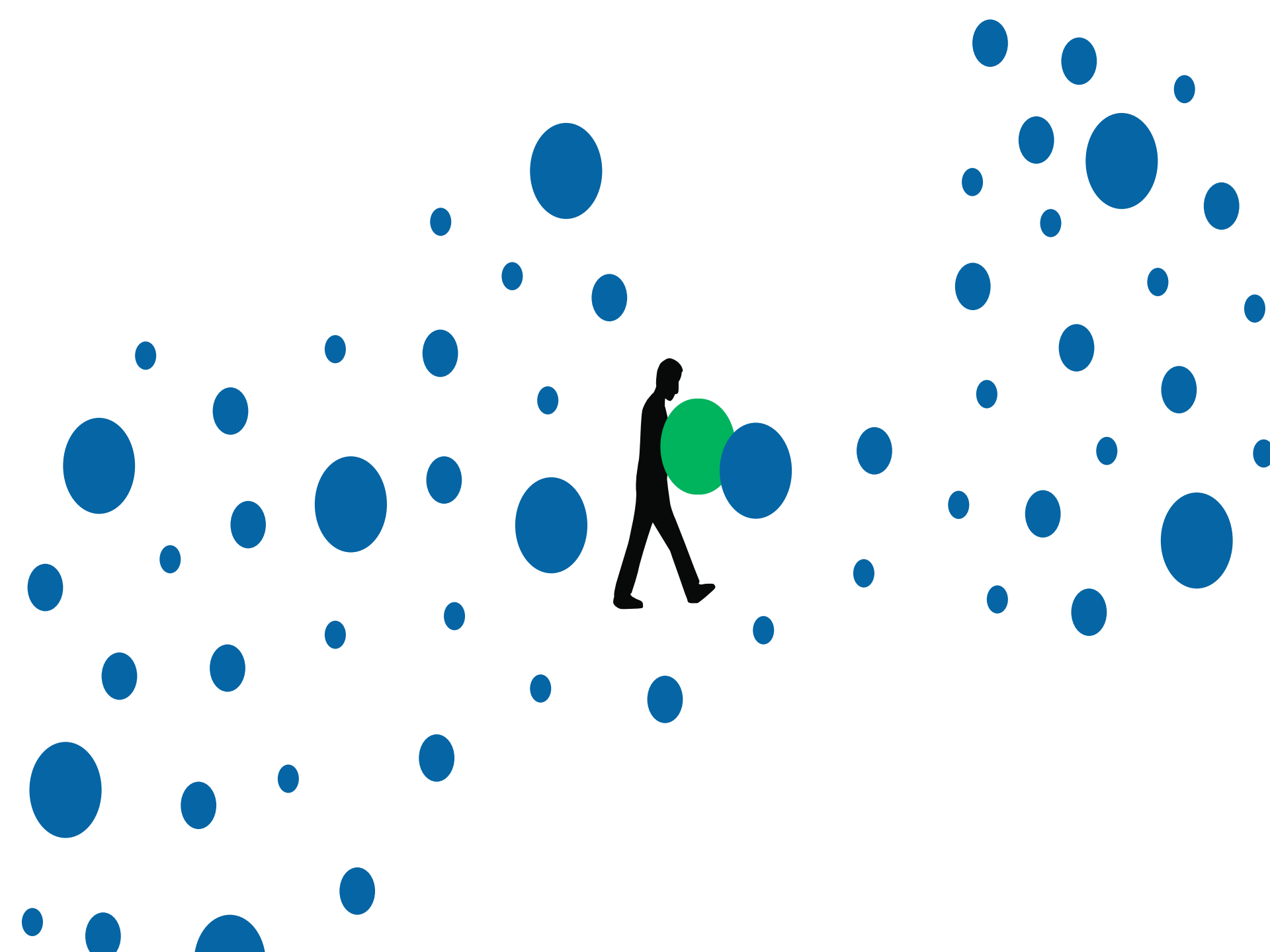# vs.
Natural Language Processing

Natural Language Processing

Extract the parts of important elements.
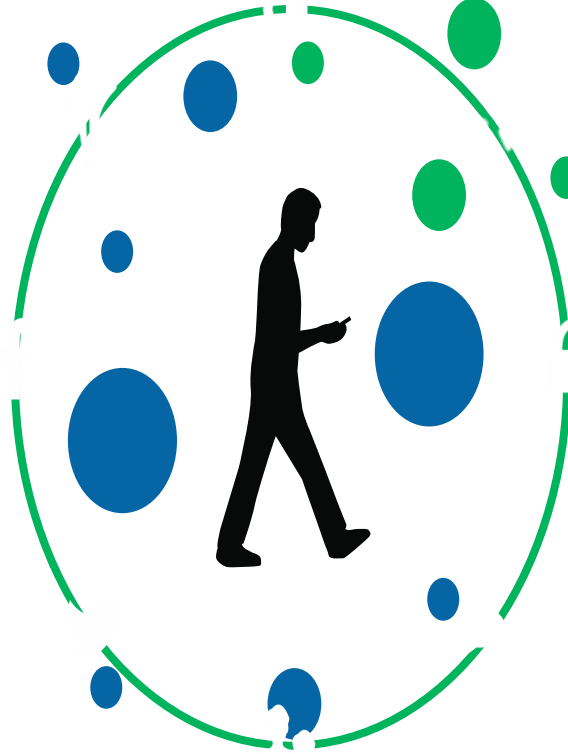
Data Processing

Acquire

Collect unstructured data from social media APIs, third-party authorized resellers, and automated scraping.

Process

Data are passed through a series of apps, emerging as meaningful bits of information.

| geo-tag | de-identify | detect benefits | detect adverse events | detect sentiment | language translation | statistics | consolidate multiples |

Export

Relevant data are passed to another series of apps in preparation for human interpretation and analysis.

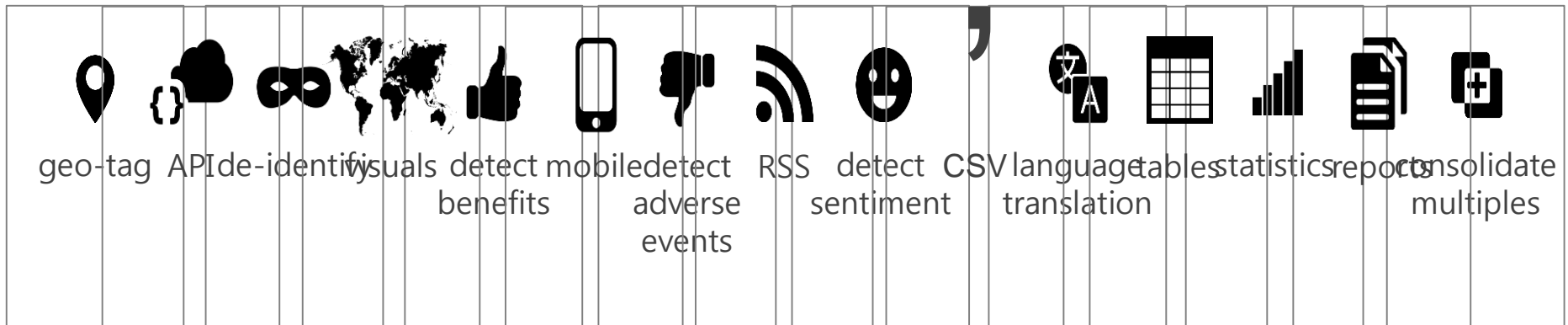| geo-tag | API | de-identify | visuals | detect benefits | mobile | detect adverse events | RSS | detect sentiment | CSV | language translation | tables | statistics | reports | consolidate multiples |

Export

Relevant data are passed to another
series of apps in preparation for
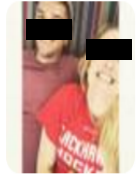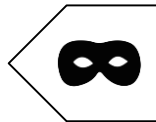human interpretation and analysis.

visuals

mobile

CSV

tables

reports

"Proto-AE"

Albuterol has me feeling all sorts of dizzy and weak this morning 😣💦 🙂

6:21 AM - 17 ... 5
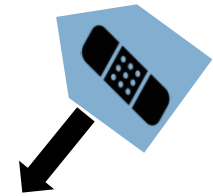
Asthenia
MedDRA: 10003549

Dizziness
MedDRA: 10013573

Follow

visual snow dont want to eat cross vision lost my appetite

apetite surpressed googley eyed no appetitey doublevision

blind appetite is nonexistent blurry feeling hu

anor visión doble Spelling Typos nt eat

uldn't vision change blurry

st appetite lost th Emoticons ages doub

car te

Invented sin hambre g cou

cross eyed seeing weird colour stomach small changes in vis

lack of apetite never want to eat seeing weird color never h

visual snow   dont want to eat   cross vision   lost my appetite

apetite surpressed   googley eyed   no appetitey   doublevision

blind   appetite is nonexistent   blurry vision 😞   miss feeling hu

anorexic   visión doble   #notevenhungry   blindness   cant eat

uldn't see   making me eat like a mouse vision change   blurry

st appetite   lost their eyesight   didn't get hungry seeing doub

can't eat   killed my apetite   googly eyed   lost teh appetite

Visual impairment
MedDRA 10047571

Decreased appetite
MedDRA 10061428

Visual impairment
SNOMED 397540003

Loss of appetite
SNOMED 79890006

Classifier Performance

Albuterol has me feeling all sorts of dizzy and weak this morning

1. Tokenization

albuterol
albuterol has
albuterol has me
has me
has me feeling
feeling
feeling all
feeling all sorts
sorts
sorts of dizzy
dizzy
dizzy and weak
weak
weak this
weak this morning
morning
morning😓
💧

2. Score calculation

For each token "w"

$$b(w) = \frac{(\text{number of positives containing w})}{(\text{total number of positives})}$$

$$g(w) = \frac{(\text{number of negatives containing w})}{(\text{total number of negatives})}$$

$$p(w) = \frac{b(w)}{b(w) + g(w)}$$
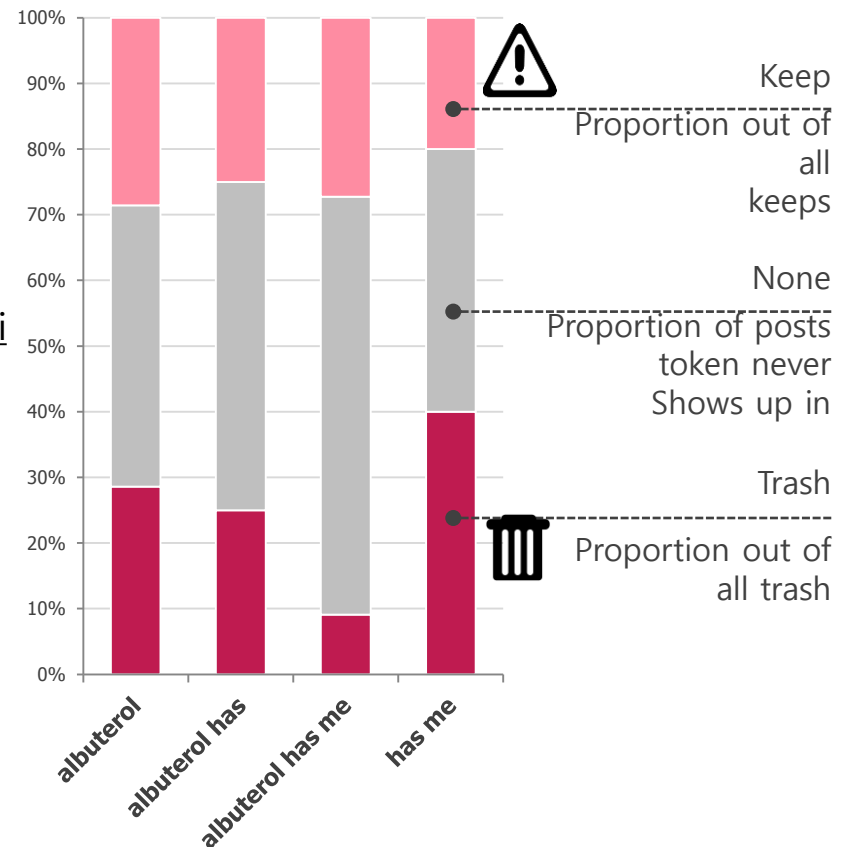
Keep
Proportion out of all keeps

None
Proportion of posts token never Shows up in

Trash
Proportion out of all trash

Classifier Performance

Albuterol has me feeling all sorts of dizzy and weak this morning 😓

0.8208

PRO

INDICATOR SCORE

CON
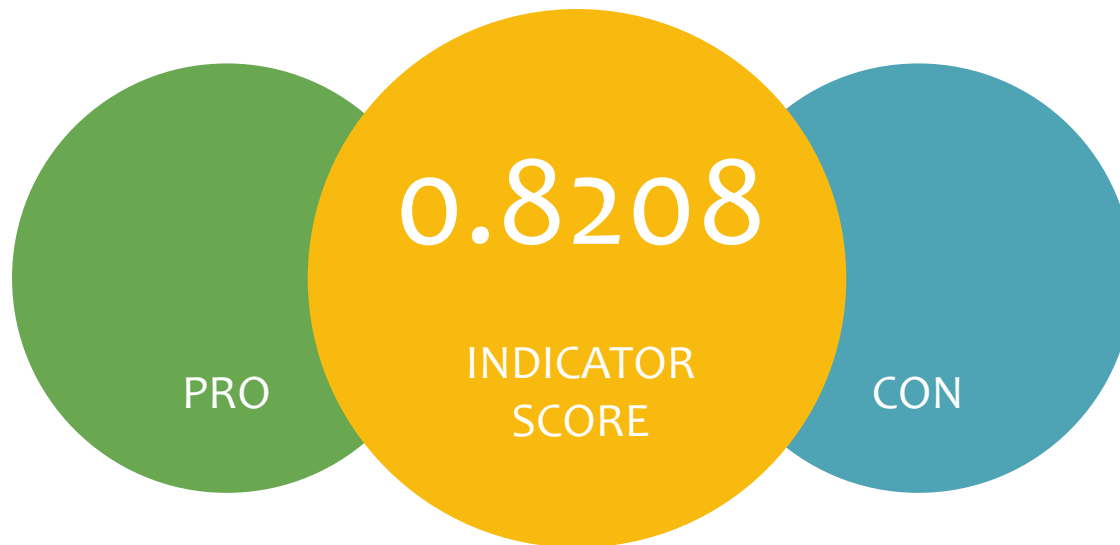
Positive Predictive Value or Precision
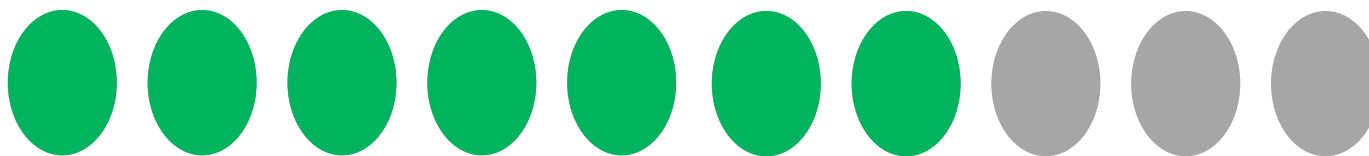
7 OF 10 posts contain adverse event information (0.68).

Can increase to 100% with manual curation (may vary by product).

If the algorithm says it is a Proto-AE, then 7 out of 10 times is actually is.

Sensitivity or Recall
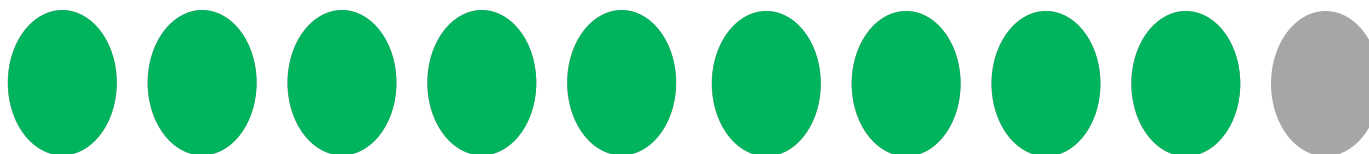
Automated tools identify **9** **OF** **10** adverse events

across all products, all time, all data sources (0.88).

The algorithm correctly identifies 9 out of 10 Proto-AEs from the pool of everything.

June 3, 2016

Thank you!

demo.medwatcher.org
@epidemico
carrie@epidemico.com

epidemico

Back Up Slides

Proto-AEs
per week

Legend:
- Nonspecific reaction (Tecfidera)
- Flushing (Tecfidera)
- Abdominal pain (Tecfidera)
- Pain (Tecfidera)
- Drug ineffective (Tecfidera)
- Fatigue (Tecfidera)
- Pruritus (Tecfidera)
- Nausea (Tecfidera)
- Malaise (Tecfidera)
- Therapy change (Tecfidera)

▲ 1/10 ▼

Y-axis: Proto-AEs (0, 5, 10, 15, 20, 25, 30, 35)
X-axis: Jan '14, Jul '14

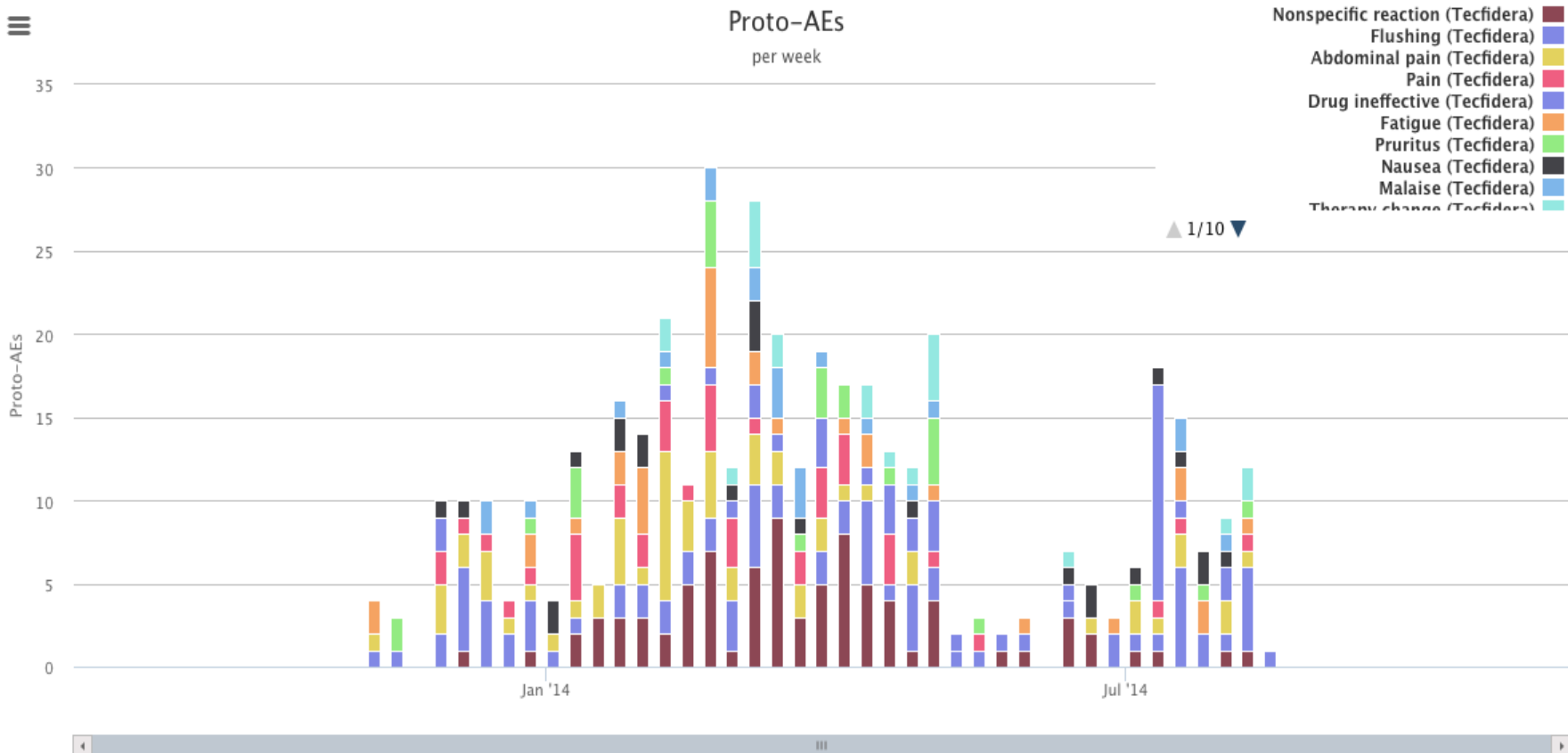| Twitter | Facebook | Forums | ⬇ Export | ☁ Contextualize | ⬇ Hide |

Post		Date	Indicator	Actions
Search post texts	or set filters »			
✔ Haven't checked in for a bit. Tecfidera is still causing stomach to not play nice. But hopeful this side effect will go away soon.		8/18/2014	0.7571	
✔ 🌐 A week off Tecfidera and this weird shaky nerve thing happens with my hand...definitely weird. #MsSucks		8/5/2014	0.73364	
✔ 📍 To my #msfamily that takes Tecfidera, do any of y'all experience itching taking this medicine?		7/31/2014	0.76234	

DIGITAL DRUG SAFETY SURVEILLANCE

Digital drug safety surveillance: monitoring pharmaceutical products in twitter. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, Dasgupta N. *Drug Safety*. 2014 May;37(5):343-50. doi: 10.1007/s40264-014-0155-x. PMID: 24777653
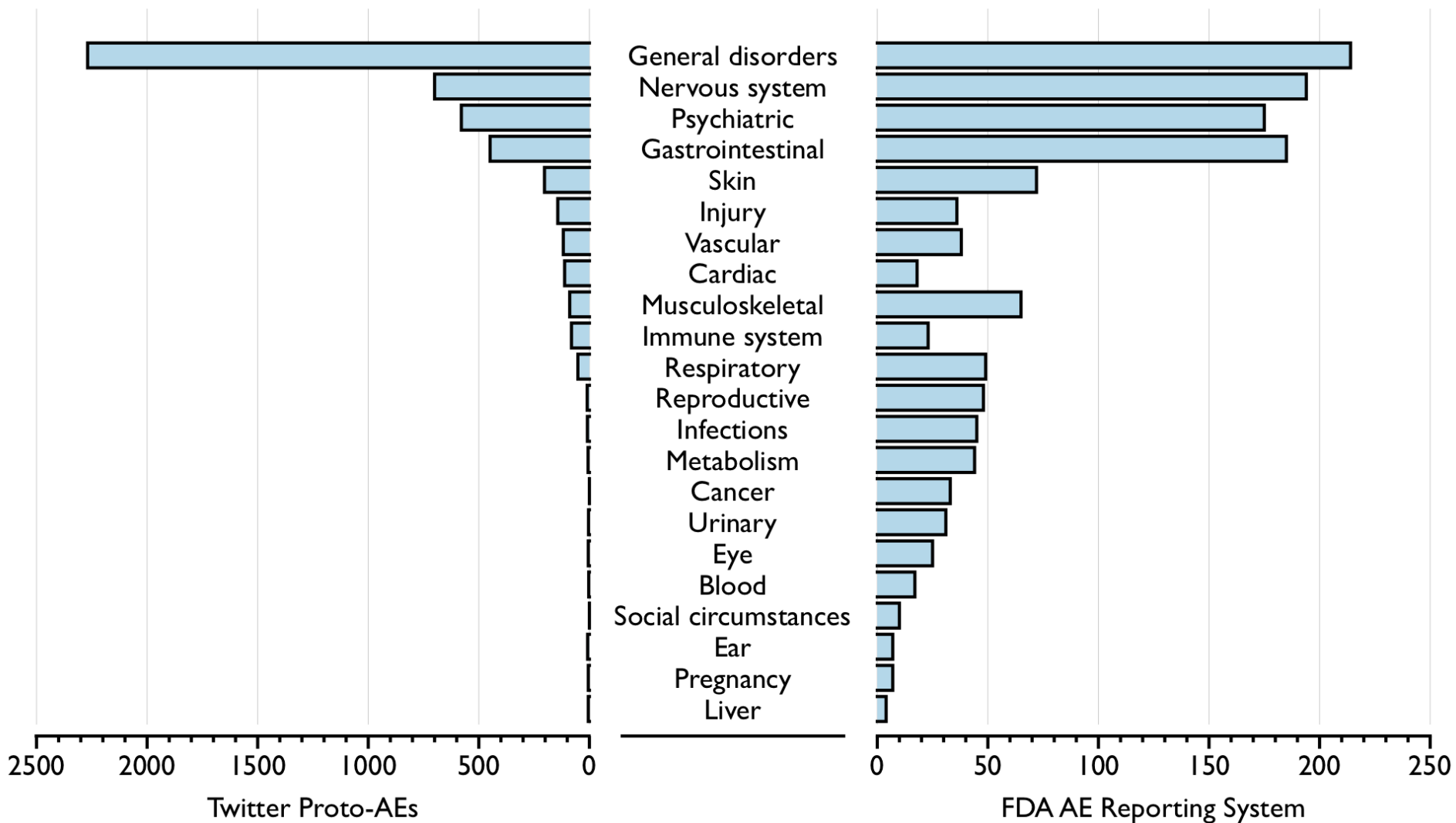
- Co-authored with FDA

- Evaluate similarity between Twitter posts mentioning AEs and FAERS spontaneous data

- Collected and classified 6.9MM Twitter posts mentioning 23 products

- Served as validation study for system and patients' voice

- Groundwork (for) research in this area

All Medical Products

Category	Twitter Proto-AEs	FDA AE Reporting System
General disorders	~2250	~215
Nervous system	~700	~195
Psychiatric	~600	~175
Gastrointestinal	~450	~185
Skin	~200	~70
Injury	~150	~35
Vascular	~110	~38
Cardiac	~110	~18
Musculoskeletal	~90	~65
Immune system	~80	~23
Respiratory	~50	~48
Reproductive		~47
Infections		~44
Metabolism		~44
Cancer		~33
Urinary		~31
Eye		~25
Blood		~17
Social circumstances		~10
Ear		~6
Pregnancy		~7
Liver		~4

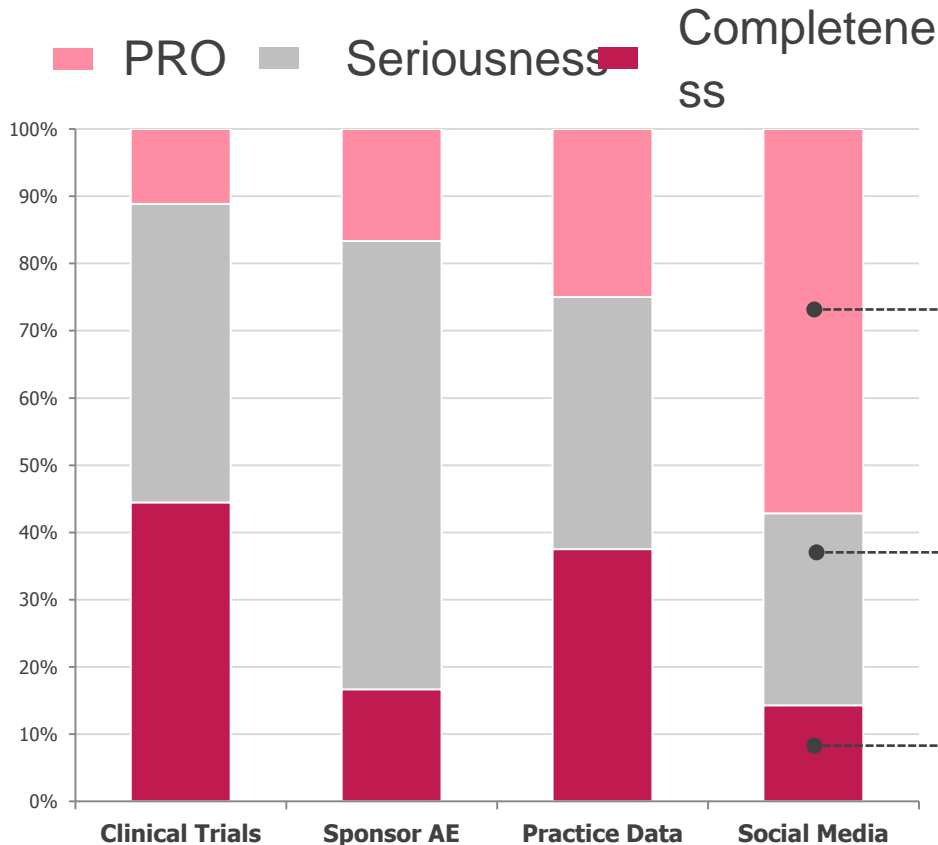Twitter Proto-AEs

FDA AE Reporting System

Rank-order concordance observed at the MedDRA SOC level between Twitter and FAERS data

WHAT WE KNOW

Each source of product safety information provides different perspectives on patient experience

Hypothetical data for illustrative purposes only.

Legend: PRO, Seriousness, Completeness

Chart y-axis: 0% to 100%
Categories: Clinical Trials, Sponsor AE, Practice Data, Social Media

Three Dimensions of Bias

1. Patient-Reported Outcomes
Social media can elucidate what patients experience most frequently.

2. Seriousness
Traditional PV focus on most serious of events may be less central in social media.

3. Completeness
Social media posts may by less complete.
Conversely, contain less sensitive identifying information than electronic health

MWS TOTALS 2016

63,000,000

Posts collected

Public posts from Facebook, Twitter & patient forums, back to 2012

380,000

Manually classified

Curator inter-rater reliability 0.88

1,700,000

Proto-AEs

~2.8% of all product mentions

Legend:
- All Posts
- Proto-AEs