

Social Media Mining for Pharmacovigilance: challenges and opportunities

Case-control studies from Twitter???

Health Language Processing Lab – Penn IBI

Graciela Gonzalez-Hernandez, PhD

email: gragon@pennmedicine.upenn.edu



@gracielagon

Funded by NLM/NIH Grant 5R01LM011176

ASCPT 2018 #Science Session – Orlando, Florida



Perelman
School of Medicine
UNIVERSITY of PENNSYLVANIA

Social media as an “online health report”?

- ◆ **26% of internet users actively discuss health information. Of that group ...¹**
 - 30% *changed behavior* as a result
 - 42% discussed *current* medical conditions
- ◆ **“Extrapolating” this to Twitter...^{2,3}**
 - Given 317 million active monthly users (Q3 2015): **about 24 million** would change their health behavior
 - Given 350,000 tweets/minute: **about 38,220 tweets / minute** about their current medical conditions



¹<http://www.pewinternet.org/fact-sheets/health-fact-sheet/>

²<http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>

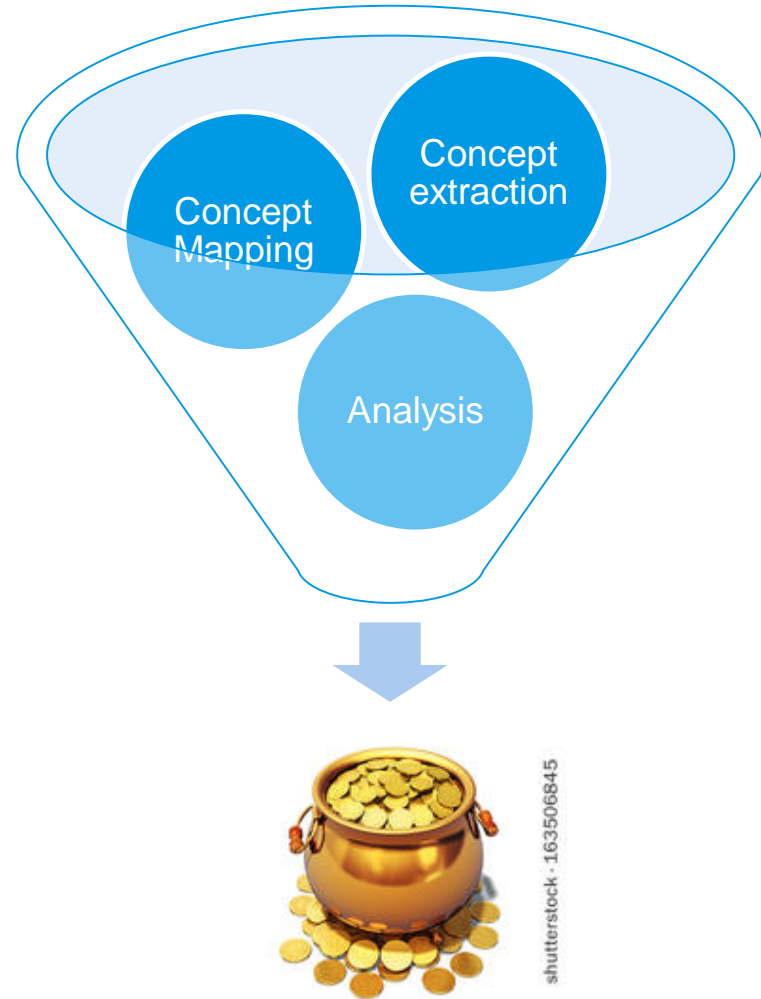
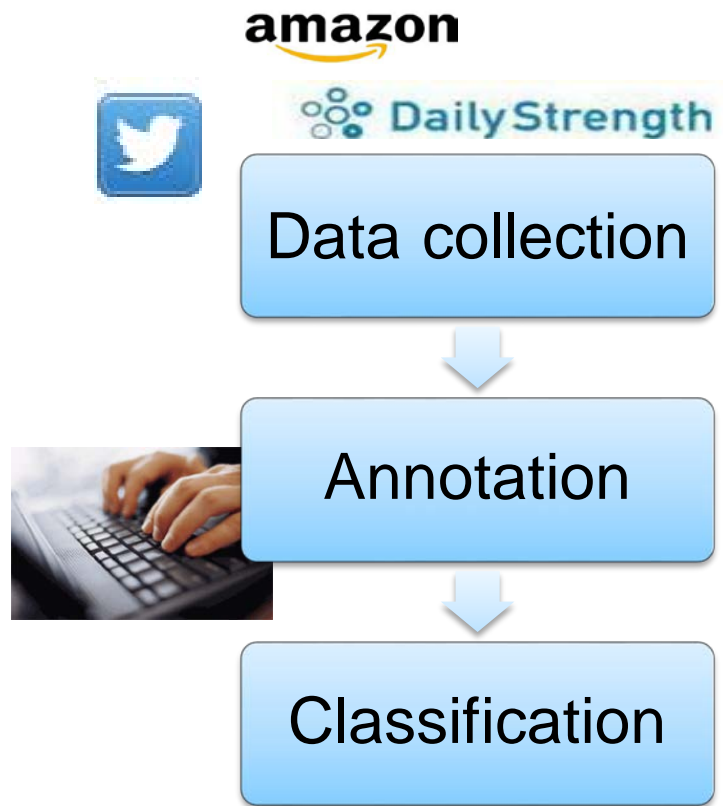
³<http://www.internetlivestats.com/twitter-statistics/>

SM data for pharmacovigilance studies

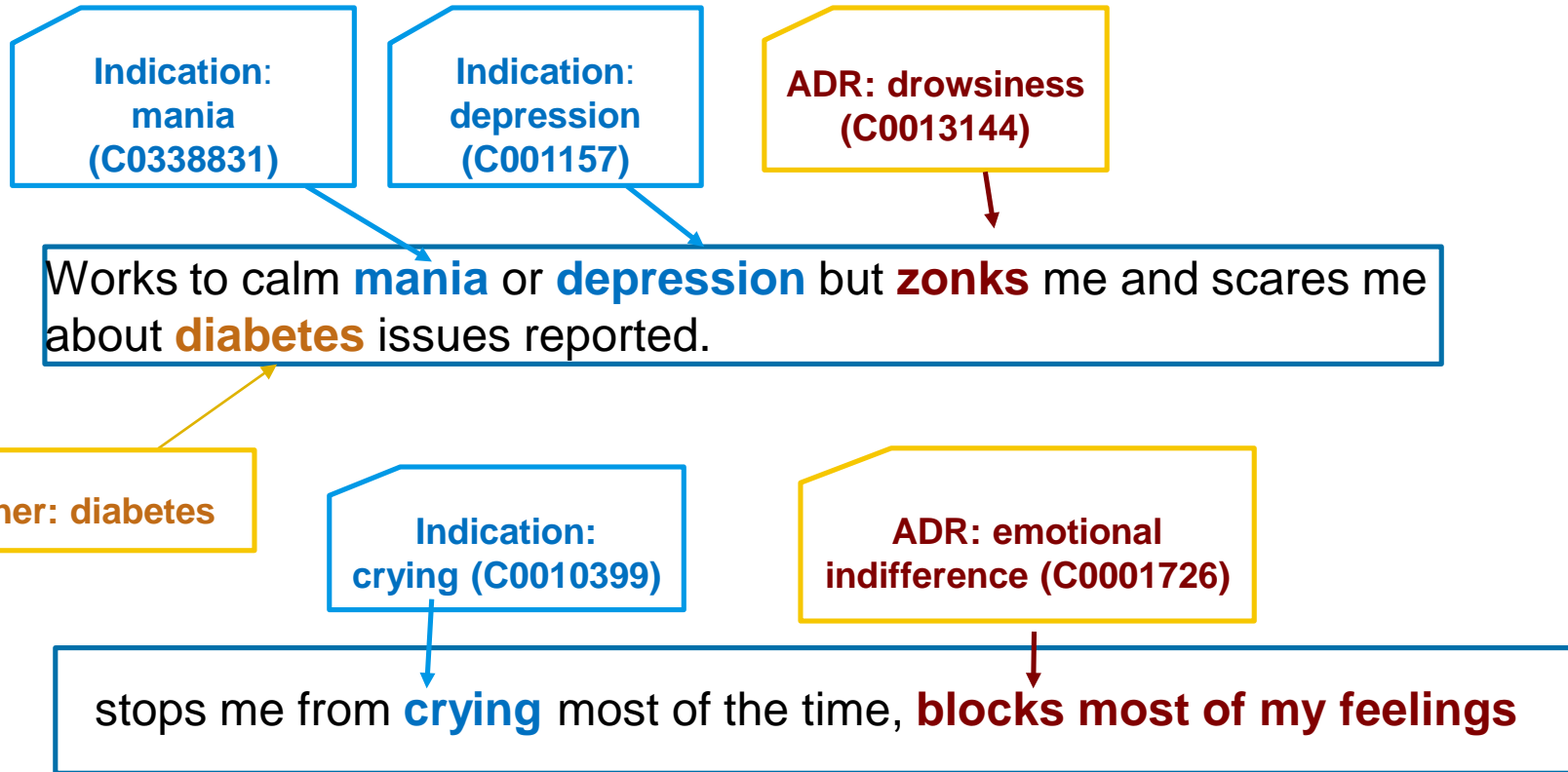
- ◆ **Patient reporting brings different perspective** (34 studies): more detail, info on severity and impact of ADRs in daily life. (PMID 27558545).
- ◆ **Abundant adverse event reports in SM** (29 studies): a higher frequency of adverse events are found in SM, particularly for 'mild' adverse events. (PMID 26271492).



Social Media Mining pipeline



Adverse event annotation example



Concept extraction: finding adverse events

- ◆ Given a set of postings, need to find a specific concepts automatically
- ◆ An algorithm is trained to find (extract) mentions of AEs and other concepts
- ◆ Our approach: conditional random fields and deep learning clusters to capture descriptive expressions beyond what lexicon-based approaches could¹

b) *I felt awful, it made my stomach hurt_{ADR} with bad heartburn_{ADR} too, horrid taste in my mouth_{ADR} tho it does tend to clear up the infection_{Indication}.*

¹ Nikfarjam *et al.* Pharmacovigilance from social media.. cluster features. **JAMIA**. 2015.

Unsupervised learned clusters

Cluster	Topic	Examples of clustered words
c ₁	Drug	abilify, adderall, ambien, ativan, aspirin, citalopram, effexor, paxil, ...
c ₂	Signs/Symptoms	hangover, headache, rash, hive, ...
c ₃	Signs/Symptoms	anxiety, depression, disorder, ocd, mania, stabilizer, ...
c ₄	Drug dosage	1000mg, 100mg, .10, 10mg, 600mg, 0.25, .05, ...
c ₅	Treatment	anti-depressant, antidepressant, drug, med, medication, medicine, treat, ...
c ₆	Family member	brother, dad, daughter, father, husband, mom, mother, son, wife, ...
c ₇	Date	1992, 2011, 2012, 23rd, 8th, april, aug, august, december, ...

Classification task examples

- ◆ Adverse event mention classification¹
- ◆ Drug abuse classification²
- ◆ Medication intake mention classification³
- ◆ Pregnancy announcement classification⁴

1 Sarker and Gonzalez. Portable automatic text classification. **J Biomed Inform.** 2015.

2 Sarker *et al.* Social media mining for toxicovigilance. **Drug Saf.** 2016.

3 Klein *et al.* Detecting Personal Medication Intake in Twitter. **ACL BioNLP Workshop 2017**

4 Sarker *et al.* Discovering cohorts of pregnant women .. [J Med Internet Res.](#) 2018

Case-control study with SM data?

- ◆ **Select cohort of pregnant women from SM¹**
 - About 120 thousand, 700 million tweets
- ◆ **Within that, find cases of interest**
 - *“Women who gave birth to a child with a birth defect and whose public tweets include tweets during pregnancy”*
- ◆ **Annotate (100% of the data found)**
- ◆ **Find matching (control) subjects**
 - *“Women pregnant around the same time, for whom there is no evidence that their child was born with a birth defect”*

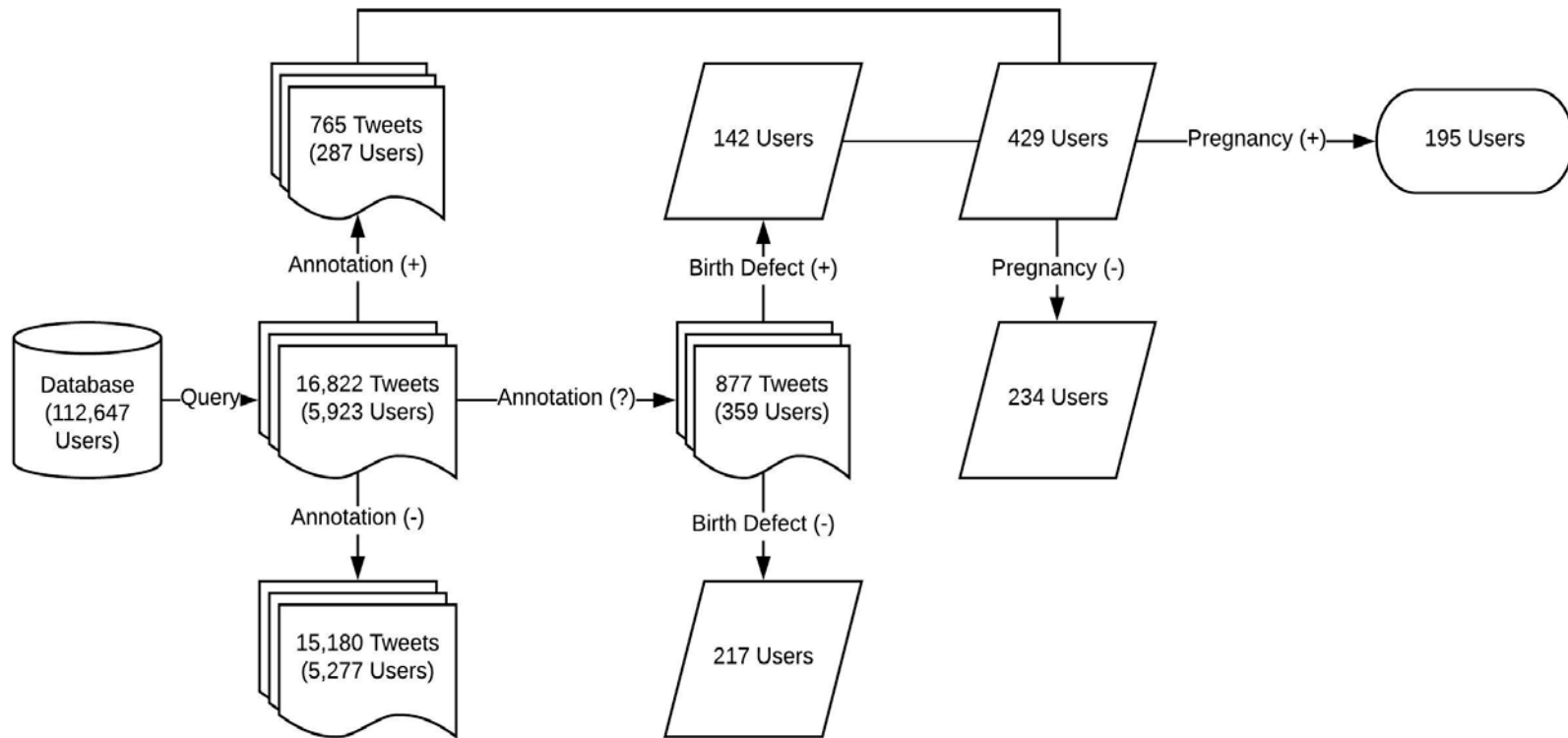
1. Sarker *et al*/ Discovering cohorts of pregnant women .. [J Med Internet Res](#). 2018

From Twitter, “I am 12 weeks pregnant”

The image shows a vertical scroll of five tweets from Twitter. Each tweet is partially obscured by a blue rectangular redaction box. The tweets are as follows:

- Tweet 1:** Profile picture of a woman in a white coat. Text: "Today, **I am** officially **12 weeks pregnant!** Here's my first personal blog post in two years... [instagram.com/p/BgoHF_-leBC/](https://www.instagram.com/p/BgoHF_-leBC/)"
- Tweet 2:** Profile picture with "I ❤️ MCB" text. Text: "**I am 12.5 weeks pregnant** and suffering terrible morning sickness all day - any recommendations on what I could take to settle it? I've tried everything :([#help](#)"
- Tweet 3:** Profile picture of a woman. Text: "I have a feeling **I am 12 weeks pregnant** because of how bloated my belly is, I can't wait to get a blood test to find out what's going on, I was supposed to have an ultrasound but didn't have it yet this month 🤰"
- Tweet 4:** Profile picture of a woman. Text: "Fast forward to this year and now here **I am** sitting down watching this video currently **12 weeks pregnant**. Thanks for helping me smile Mark :) you have the most beautiful heart and never change you big goof :)"
- Tweet 5:** Profile picture of a man and woman. Text: "my son is 15 months and my wife is **12 weeks pregnant**. when **I am** home it's funny dealing with his high energy and tantrums"
- Tweet 6:** Profile picture with "CLASSIC105" logo. Text: "'**I am 12 Weeks Pregnant!**,' Janet Mbugua Reveals She Is Expecting Baby Number Two [classic105.com/i-am-12-weeks-...](https://www.classic105.com/i-am-12-weeks-...)"

Finding cases – birth defects cohort



Birth defects data from Social Media

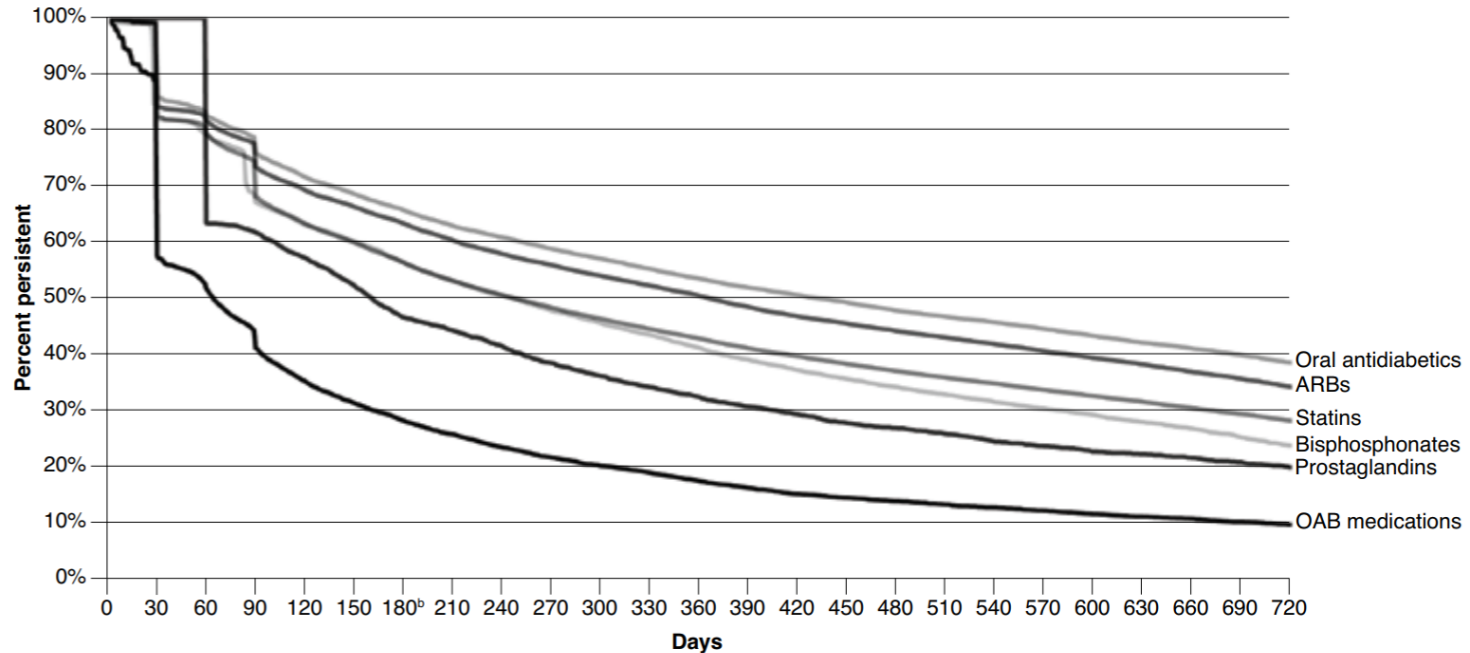
	Cases (n=197)	Controls (n=196)	OR or t-test [95% CI]	P-value
Age				
Median Age (IQR)	23 (20 to 28)	21 (19 to 23)	2 (1 to 3)	0.0001
Mean Age (range)	25 (17 to 42)	22 (16 to 37)	2.52 (1.38 to 3.66)	<0.0001
Women <30 years	80% (134/168)	91% (129/141)	0.37 (0.17 to 0.77)	0.004
Women <35 years	93% (156/168)	98% (138/141)	0.28 (0.05 to 1.08)	0.04
Missing data on age	14% (28/196)	28% (55/196)	0.43 (0.25 to 0.73)	0.0008
Race/Ethnicity				
Caucasian	73% (120/164)	55% (102/184)	2.19 (1.36 to 3.54)	chi ² = 23.69, d.f. = 5 P < 0.001
Black	13% (22/164)	27% (51/184)	0.40 (0.22 to 0.72)	
Hispanic	9% (14/164)	12% (21/184)	0.72 (0.33 to 1.56)	
Asian	2% (4/164)	3% (5/184)	0.90 (0.17 to 4.24)	
Other (Islander, Native American/Indian, Multiracial/Mixed)	2% (4/164)	2% (5/184)	0.90 (0.17 to 4.24)	
Missing data on race	16% (32/196)	6% (12/196)	0.99 (1.44 to 6.58)	

Adherence/persistence studies from SM

- ◆ **Social media may be particularly useful for identifying sources of intolerability that lead to non-adherence/non-persistence**
- ◆ **These are often not reported by physicians or patients through standard means because are considered “mild”, “not serious” or are unexpected**
- ◆ **Significant problem, given that, on average:**
 - 30% of treated patients have a beneficial response
 - 30% do not respond
 - 10% have only side effects
 - 35%-70% are non-adherent / non-persistent, often due to side-effects or perceived/real non-response

6-month persistence rate

FIGURE 2 Time to Discontinuation^a of 6 Chronic Therapy Classes, Allowing for 60-Day Treatment Gap



Copyright© 2009, Academy of Managed Care Pharmacy. All rights reserved.

- prostaglandin analogs 47%
- statins 56%
- bisphosphonates 56%
- oral antidiabetics 66%
- angiotensin II receptor blocker 63%
- overactive bladder medications 28%

“I stopped taking” & “made me”

Y'all I literally went a whole week without 1 suicidal thought. And that's without needing to take my antidepressants. I'm so happy. You have no idea how proud I am of myself. Like I'm literally crying while typing myself. #8daysdownmanymoreto come

If anyone's wondering which I doubt, the reason I stopped taking my antidepressants was because it messed with my appetite and made me feel extra drowsy and just emotionally numb. I constantly felt like a zombie, so I figured I see how I felt without them.

And well without them while in a better state of mind. I've gone without them before but those were my darker days aka like 2 weeks ago lol

Ladies, anyone on Trigestrel? Hows the side effects? #GirlTalkZA

I used to be on it. It made me nauseous, gave me headaches, made me bloated & made my period pains worse. I stopped taking it.

It's a cheaper version of Triphasil. After that, I promised not to use cheap contraceptives again. The expensive ones have less side effects

Social media data: challenges

◆ **Incompleteness:**

- Not all health conditions may be revealed
- Complete data about individual cases may be difficult to obtain: taking drug X, but dosage, frequency, length of treatment, may be missing
- Participants from the cohort may dropout at higher rates

◆ **Accessibility:**

- Data from social media is not easily collected: API limitations
- Not easily processed once collected: challenging to process using automatic methods
- Data collection methods may have to be changed frequently over time

◆ **Authenticity:**

- Bots – a large portion of social media is generated by bots, making it harder to mine reliable data
- *Automatic* processing of postings is often misleading: for example postings mentioning a drug might not necessarily mean intake.

Thank you!



gr
TV
HI
ht

