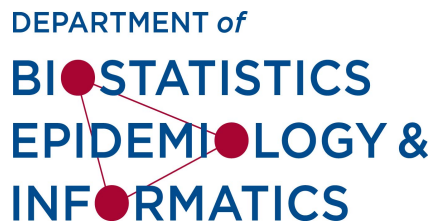




AI & ML for Digital Epidemiology of Birth Defects using Twitter Data

Graciela Gonzalez-Hernandez, Ph.D.
Health Language Processing Center
Twitter: @gracielagon
AI & Health
January 27, 2020

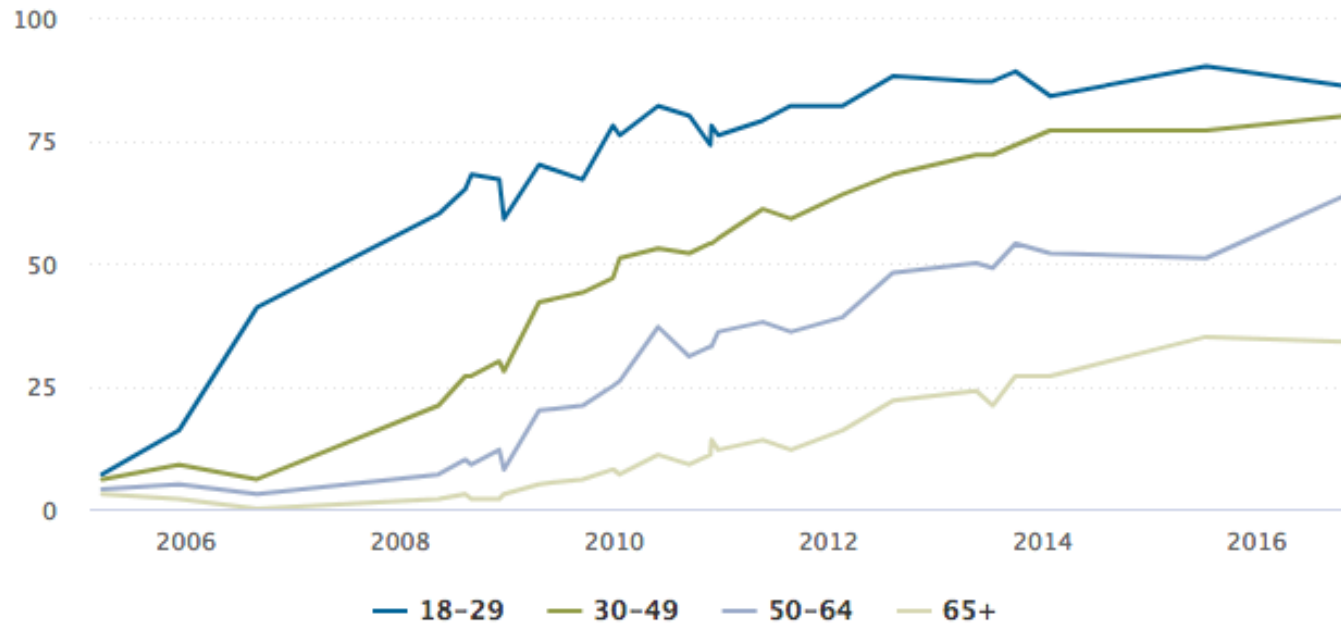


Motivation

- Birth defects are the leading cause of infant mortality in the U.S.
- Etiology of the majority of birth defects remains largely unknown.
- Methods for observing pregnancies with birth defects are limited.
 - Clinical trials
 - Animal models
 - Registries

Who uses social media in the US?

% of U.S. adults who use at least one social media site, by age



Source: Surveys conducted 2005-2016.

PEW RESEARCH CENTER

Our central question

- Can social media data fill the gap?
- Can it be used as a complementary resource for observational studies of birth defects?

Twitter data for health research

- Pros

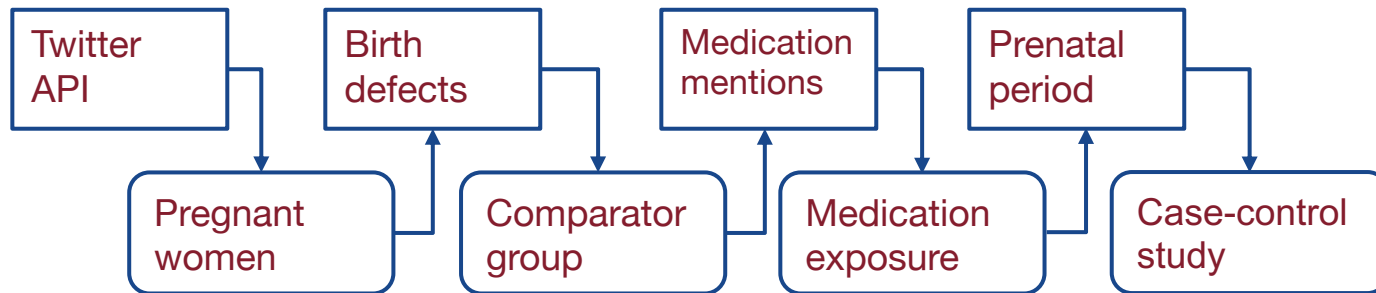
- Large volumes of user-generated text
- Tweets are publicly available to view
- User timeline allows observations over time
- Twitter provides an application programming interface (API)

- Cons

- Unstructured, free text
- Mostly “noise”
- Informal language and misspellings
- End-to-end processing pipelines are difficult to construct

What have we accomplished?

- Collected a cohort* of 500,000 pregnant women over 3 years.
- Twitter data for a case-control study of birth defects** that scales***



Pregnancy	2016-02-01
	2016-02-11	I woke up today with a sore throat and stiff neck and feel even worse now. #Advil #bed #ihateviruses
	2016-07-11	It's time to announce that I'm having a baby and it's going to be a boy!!!!
	2016-11-07	Well today's my due date and the baby's not here. #babyboy #imready
	2016-11-13	After 6 crazy days I'm finally announcing the birth of our son!
	2016-12-12	I hadn't heard of Hirschsprungs Disease until 5 weeks ago. Now I feel like an expert. Amazing how life can change overnight.

*Sarker A et al. Discovering cohorts of pregnant women... Journal of Medical Internet Research 2017;19(10):e361.

**Golder S et al. Pharmacoepidemiologic evaluation of birth defects... Drug Safety 2019;42(3):389-400.

*** Klein AZ et al. Towards scaling Twitter for digital epidemiology of birth defects. npj Digital Medicine 2019;2:96.

Determining the prenatal period

- Due date

2016-06-04 Hello 3rd trimester! **12 weeks until my due date.** Can't wait to meet you! → 2015-11-21 - 2016-08-27

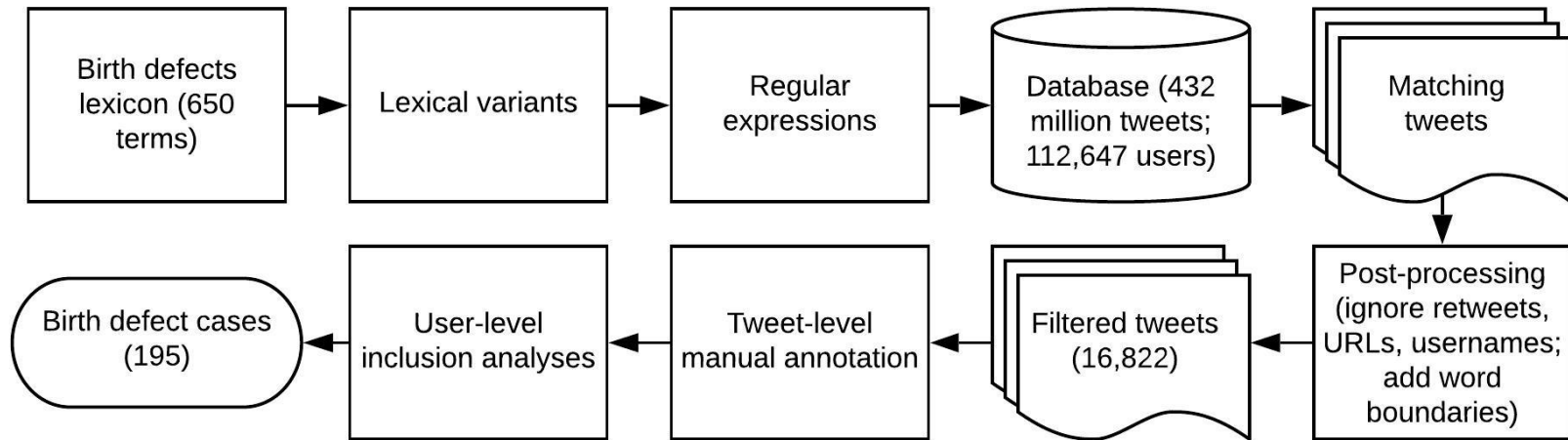
- Gestational age

2015-12-14 I'm **20 weeks and 3 days pregnant.** The baby is the size of a banana! → 2015-07-24 - 2016-04-29

- Date of birth

2015-04-26 **My baby was born 1 MONTH ago** today. Time is already flying :(→ 2014-06-19 - 2015-03-26

Birth defect cases



```
r'cleft(\b(the|this|that|these|those|a|an|
|his|her|of|and|(w/|with|w)|their|between
|in|on|under|both)\b|\W)*palate'
```

```
cleft palate
cleft on her palate
cleft-palate
#CleftPalateAwareness
```

```
My little miracle, we are so blessed to have you #hlhs
```

+

```
Scheduling an appointment for [name] to confirm craniosynostosis.
```

?

```
Can't watch this cleft palette infomercial, I can't wake up depressed
```


-

*Source: Klein AZ, Sarker A, Cai H, Weissenbacher D, Gonzalez-Hernandez G. Social media mining for birth defects research: a rule-based, bootstrapping approach to collecting data for rare health-related events on Twitter. *Journal of Biomedical Informatics* 2018;87:67-78.

Medication mention and exposure

- Generating misspellings of medications (e.g., Klonopin)*

Semantic



klonopin ativan
 xanax clonazepam
klonipin lorazepam
 trazodone valium
klonapin clonopin
 vistaril ambien celexa
 remeron lexapro
 diazepam seroquel
 buspar xanex vistarel

Lexical

klonopim klonipin klonepin
 kolnopin clonopin klomopin
 klonapins klonoin klopopin
 klonodine klonopan klonopon
 klonapin klonopn klanopin
 Klenopin clonopine klonopins
 klonopen klonipine klolopin
 Klonopine klonipins klonopin

- Detecting when tweets mention medication intake**

- maybe i'm tired because i had 2 **tramadol** my bk is sore ... #scoliosis
- I'm going to cry it's so painful 🙄gonna take **codeine** this morning

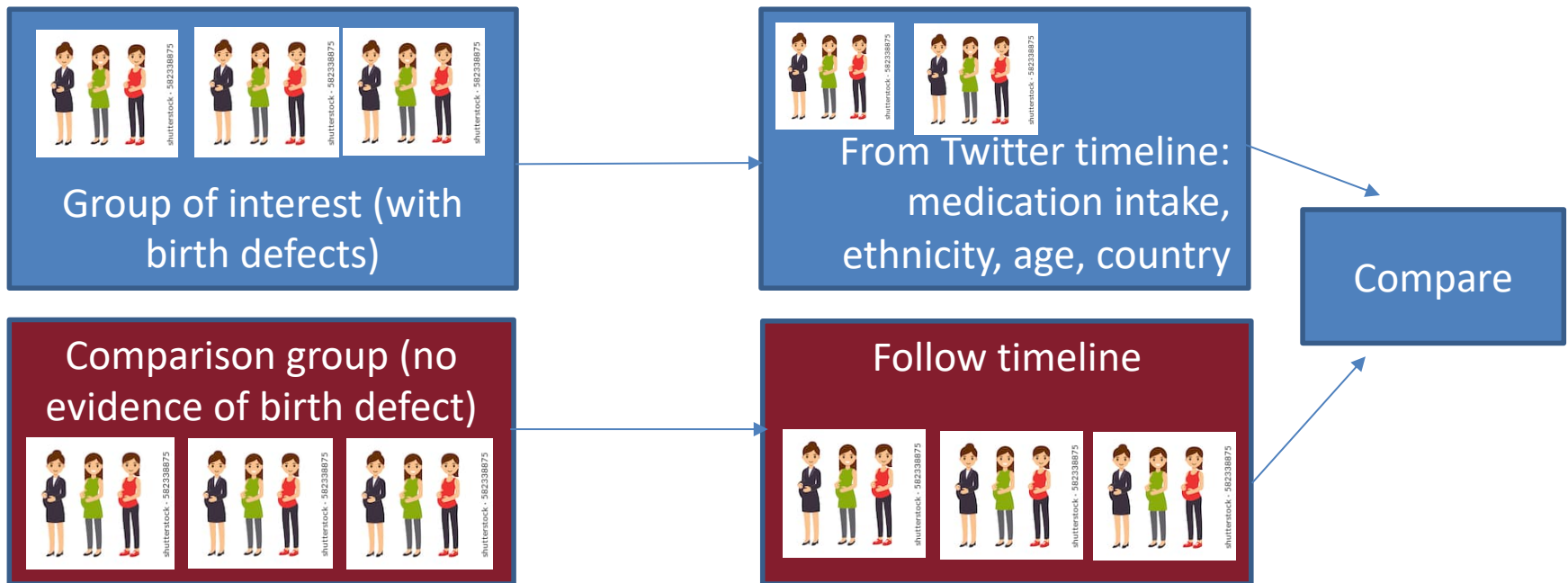
- Using Deep Neural Networks to find medications mentions ***

*Sarker A, Gonzalez-Hernandez G. An unsupervised and customizable misspelling generator for mining noisy health-related text sources. *Journal of Biomedical Informatics* 2018;88:98-107.

**Klein AZ, Sarker A, Rouhizadeh M, O'Connor K, Gonzalez G. Detecting personal medication intake in Twitter: an annotated corpus and baseline classification system. In: *Proceedings of the BioNLP 2017 Workshop*; 2017:136-142.

***Weissenbacher D, Sarker A, Klein A, O'Connor K, Magge A, Gonzalez-Hernandez G. Deep neural networks ensemble for detecting medication mentions in tweets. *Journal of the American Medical Informatics Association* 2019;26(12):1618-1626.

Setup of a Twitter nested case-control study



- Reports of medication exposure during pregnancy were statistically significantly greater among the cases than the controls.
- Limitation: not enough cases for studying specific birth defects.

Improving our comparator group (controls)

User	Tweet	Outcome	Date	Tweet +/-	User +/-
A	17 days till my due date, so sad its going to be over but cant wait to see my baby's beautiful face 🥰💙	Term	2015-09-21	+	-
	last month i was in the most horrible pain delivering my baby. 1 month later i have a happy & healthy 9lb 7oz baby 🌍💙	NB	2015-10-28	-	
B	I'm 38 weeks pregnant! Wow! I've never been this pregnant 😭😭	Term	2016-10-11	+	-
	Today is #WorldPrematurityDay. My 36 weeker was born weighing 6lbs5ozs and is now a healthy 18 month old!	NB	2015-11-17	-	
C	Today's my due date, but life clearly had other plans #5weeksold today!	Term	2016-04-01	-	-
	My sister-in-law had her baby! Born January 22 at 8:44pm - 8lbs. 4oz. We really appreciate our first year anniversary gift!	NB	2010-01-23	-	
D	mine still fits over my 39+ week belly with some room to spare. (today is my due date).	Term	2016-05-31	+	-
	My baby was born on Saturday 8/2 @ 2:45pm. 7lbs. 3oz. 20 in.	NB	2014-08-05	+	
E	I'm 2 weeks away from my due date. I've gotta keep my blood pressure normal 😊	Term	2016-09-17	+	+
	Meet my darling. Born at 17:15, weighing 2.8kg. Mom is in loads of pain	NB	2016-10-12	+	

Source: Klein AZ, Gebreyesus A, Gonzalez-Hernandez G. Automatically identifying comparator groups on Twitter for digital epidemiology of pregnancy outcomes. AMIA Joint Summits Translational Science Proceedings 2020. Preprint at <https://arxiv.org/abs/1908.06015>.

Towards scaling Twitter-based studies

- Annotated 22,999 tweets to train and evaluate Multinomial Naïve Bayes, SVM, and LSTM classifiers for birth defects.
- Challenge: class imbalance associated with automatically detecting rare events in practice (90% of tweets merely mention birth defects).
- Various under-sampling and over-sampling experiments to address class imbalance problem.

	F(+)	F(?)	F(-)
NB	0.54	0.44	0.94
SVM	0.65	0.51	0.96
LSTM	0.60	0.35	0.96

- Deployed classifiers on 20,457 unlabeled tweets, identifying 542 additional users for potential inclusion as cases.

Thank you! Questions?



Contact: gragon@upenn.edu

Twitter: @gracielagon

Upcoming workshop: #SMM4H at COLING

HLP Center @ UPenn:

<https://healthlanguageprocessing.org>

