Mining digital life for precision predictions and early detection

Thomas Robinson, MD, MPH

School of Medicine, Departments of Pediatrics and Medicine

Byron Reeves, PhD

School of Humanities & Sciences, Department of Communication and Graduate School of Education

Russ Altman, MD, PhD & Adam Lavertu

School of Medicine, Departments of Bioengineering, Genetics, Medicine and Biomedical Data Science

Nilam Ram, PhD

Department of Human Development & Family Studies and Psychology, Pennsylvania State University

Previous work: Can we detect drug interactions in internet search logs?

- Yes
- For two drugs we associated with hyperglycemia, we looked at BING search logs for mention of 50 hyperglycemia-related words.
- We compared
  - Baseline search for hyperglycemia words
  - Search with one drug + hyperglycemia
  - Search with second drug + hyperglycemia
  - Search for both drugs + hyperglycemia

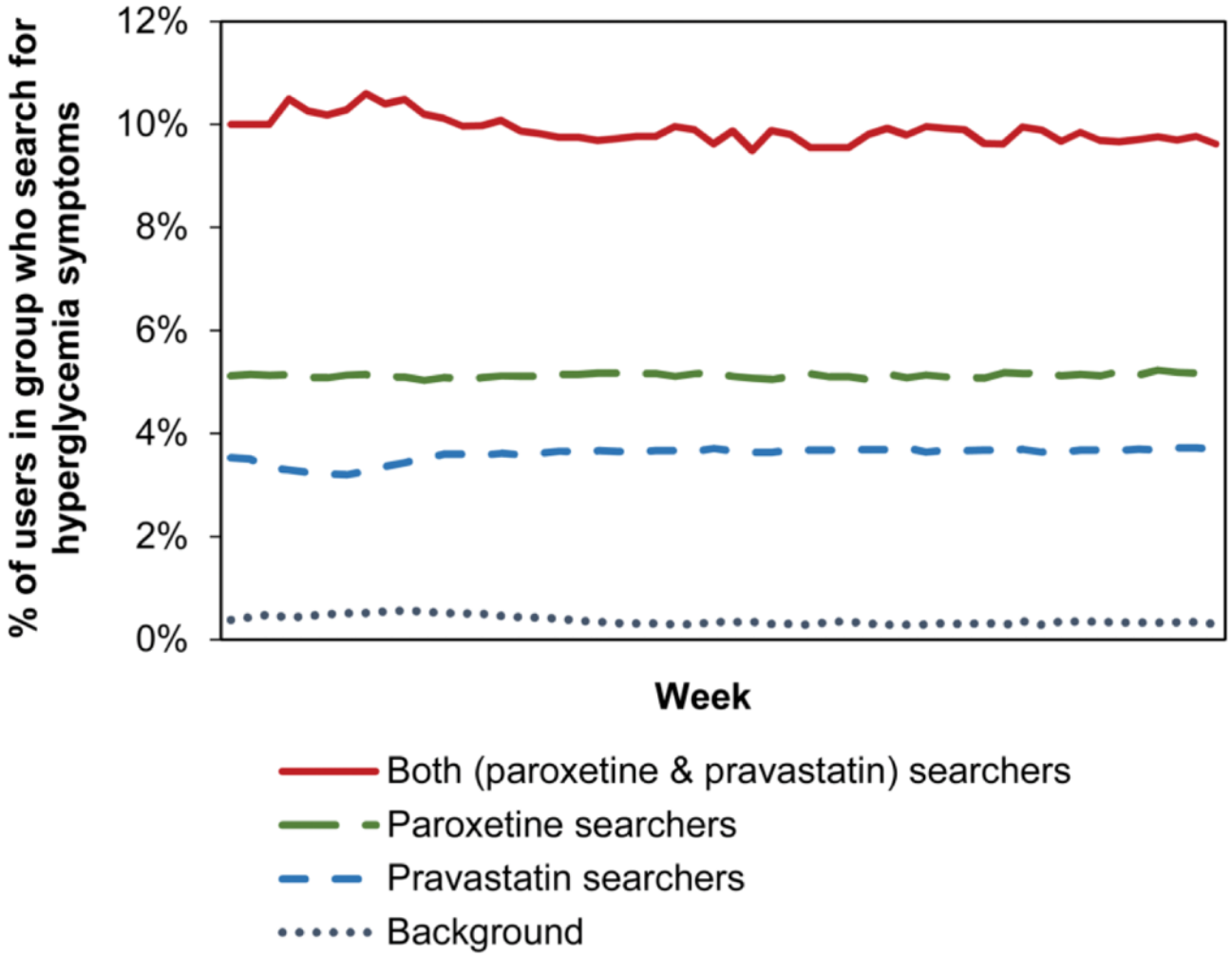
White et al, J Am Med Inform Assoc. 2013 May-Jun; 20(3): 404–408.

50“hyperglycemia” words/phrases

appetite increase
blurred vision
blurry vision
breathing difficulty
breathing trouble
breathless
breathlessness
coma
confused
confusion
decreased libido
decreased sex drive
decreased sexual desire
dehydrated
dehydration
difficulty breathing
dizziness
dizzy
drowsiness
Drowsy
dry mouth
dry skin
erectile dysfunction
fatigue
fatigued
feet tingling
frequent urinating
frequent urination
heel tingling
hunger
hungry
impotence
impotent
increased appetite
increased urination
itchy skin
labored breathing
light headed
lightheaded
light-headed
lightheadedness
loss in weight
loss of weight
low sex drive
polydypsia
polyphagia
polyuria
poor healing
poor wound healing
short of breath shortness of breath
skin tingling
sleepiness
sleepy
slow healing
slow wound healing
thirst
thirstiness
thirsty
tingling feet
tingling heel
tingling skin
tired
tiredness
trouble breathing
unconscious
unconsciousness
urinating
urination
Xerostomia

Web searches? Patient search for pravastatin & paroxetine and DM-related words more frequently.

White et al, J Am Med Inform Assoc. 2013 May 1;20(3):404-8



A large set of health markers

Biology

Genome
Epigenome
Transcriptome
Proteome
Metabolome
Microbiome

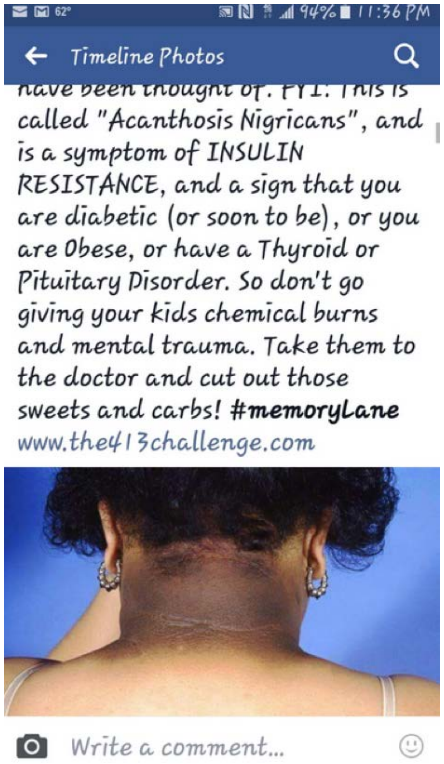
Screenome

Behavior-Environment

Suicide

Cancer

Memory



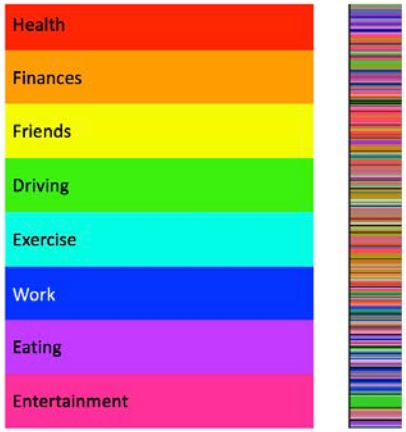
Diabetes

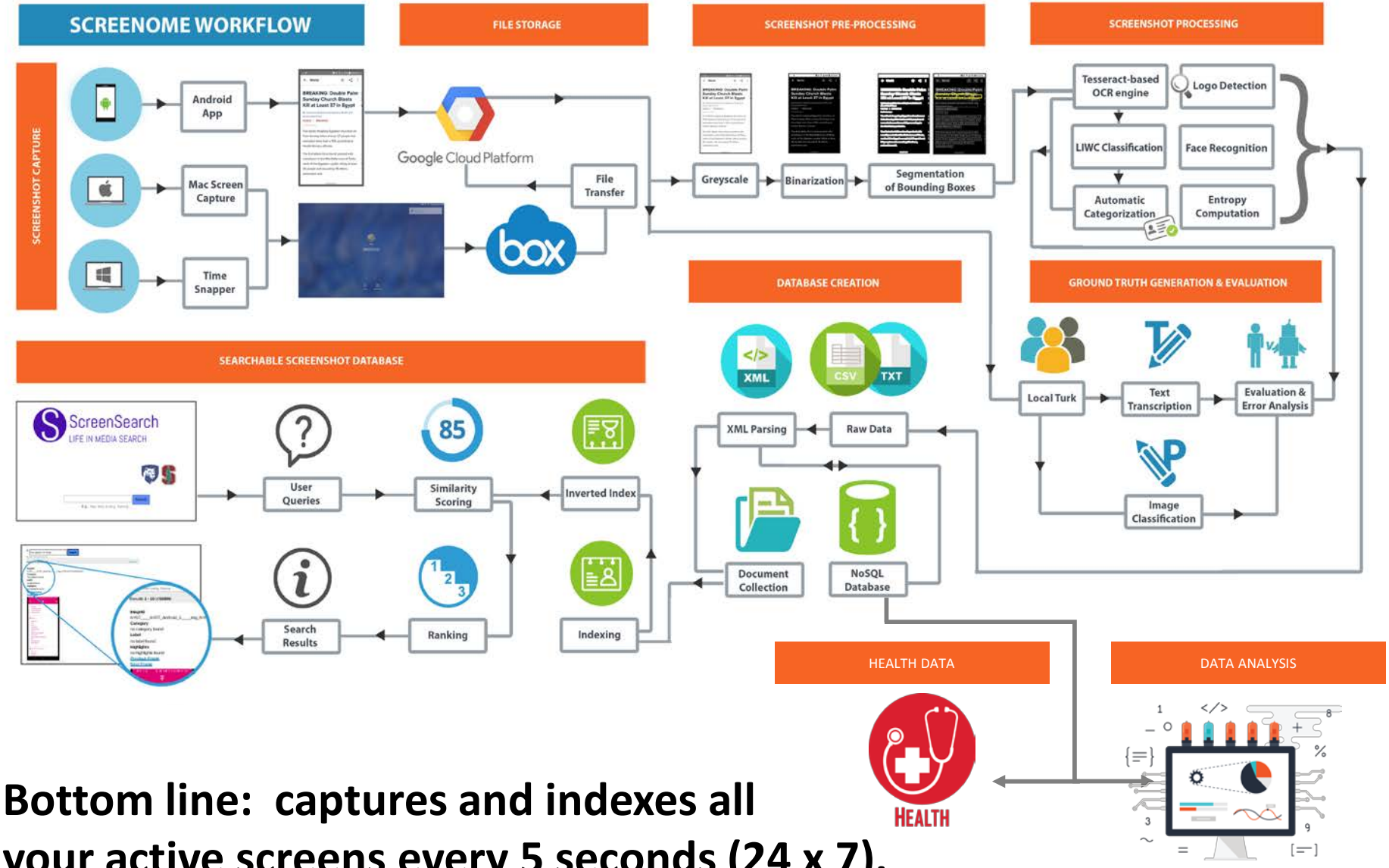
Hypothesis

- Individual use of digital devices provides unique information about signs and symptoms for early prediction, prevention and detection of risk factors and disease

Why **screens**?

- Captures **broad spectrum** of life
  - Fulfillment of the digital promise
- Increasing **generational relevance**
  - The new iGen
- Captures change at the new **speed of life**
  - Quick changes between radically different content
- Captures specific **attention**
  - What people are actually looking at
- **Threads** rather than buckets **of experience**
  - Examine sequence, context, and interdependence
- **Passive** data collection
  - No other devices, sensors, chargers, things to carry or wear

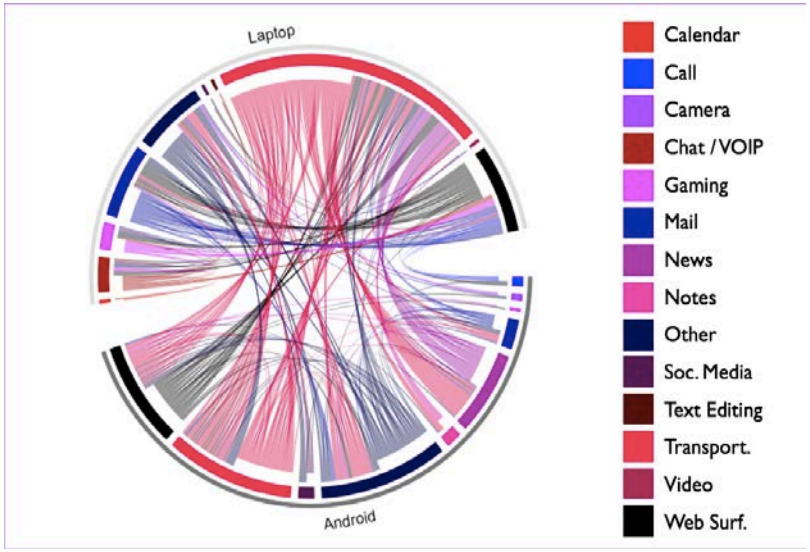


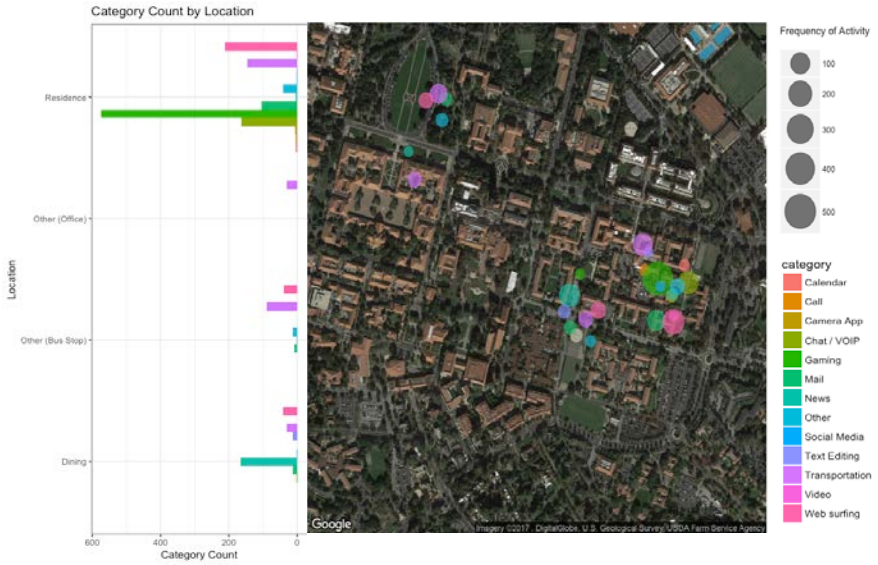


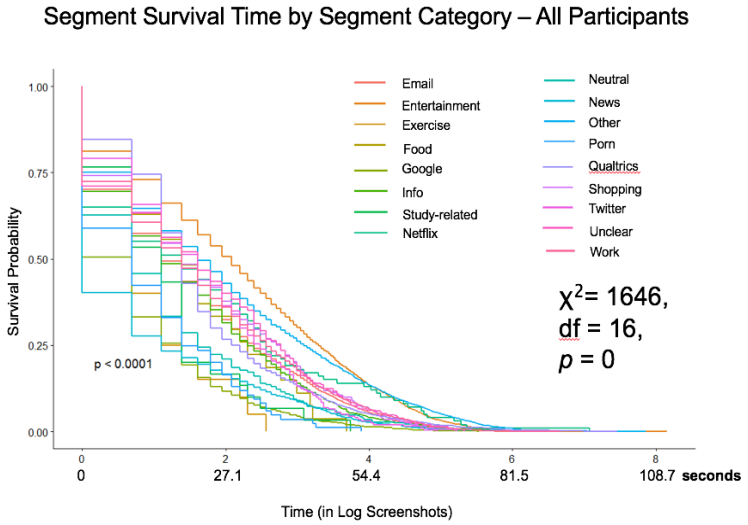
Bottom line: captures and indexes all your active screens every 5 seconds (24 x 7).

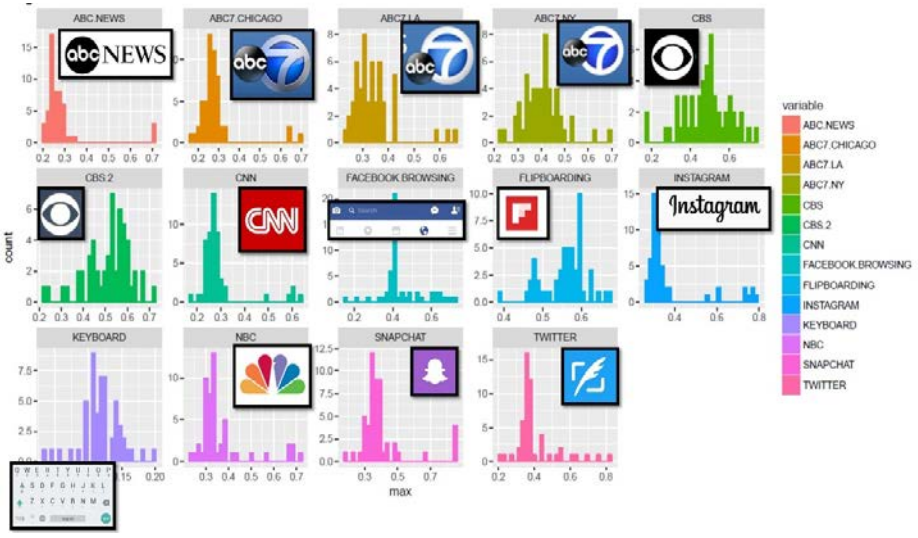
Security, trust, privacy, recruiting

- Genome vs. **screenome**
  - Changing attitudes about personal data
- **Risk reward** ratio
  - Big ask but big reward
- **Human subjects**
  - IRB approvals and discussions
- Data **security**
  - Encryption, secure servers, de-identification, strict access limits, offline storage
- **Recruiting**
  - Paid volunteers aged 18-45
  - Digitally active
  - ~60% acceptance to date


Laptop
Android
Calendar
Call
Camera
Chat / VOIP
Gaming
Mail
News
Notes
Other
Soc. Media
Text Editing
Transport.
Video
Web Surf.


Category Count by Location
Residence
Other (Office)
Other (Bus Stop)
Dining
Location
600
400
200
0
Category Count
Google
Frequency of Activity
100
200
300
400
500
category
Calendar
Call
Camera App
Chat / VOIP
Gaming
Mail
News
Other
Social Media
Text Editing
Transportation
Video
Web surfing


Segment Survival Time by Segment Category – All Participants
Survival Probability
1.00
0.75
0.50
0.25
0.00
p < 0.0001
Email
Entertainment
Exercise
Food
Google
Info
Study-related
Netflix
Neutral
News
Other
Porn
Qualtrics
Shopping
Twitter
Unclear
Work
χ2= 1646,
df = 16,
p = 0
0
27.1
54.4
81.5
108.7 seconds
Time (in Log Screenshots)


ABC.NEWS
ABC7.CHICAGO
ABC7.LA
ABC7.NY
CBS
CBS.2
CNN
FACEBOOK.BROWSING
FLIPBOARDING
INSTAGRAM
KEYBOARD
NBC
SNAPCHAT
TWITTER
count
max
variable
ABC.NEWS
ABC7.CHICAGO
ABC7.LA
ABC7.NY
CBS
CBS.2
CNN
FACEBOOK.BROWSING
FLIPBOARDING
INSTAGRAM
KEYBOARD
NBC
SNAPCHAT
TWITTER

Informatics challenges adapting **Screenome** to **health**

- Create **health-specific infrastructure**
  - Ensure participant privacy/security
  - Identify health terminologies and ontologies (bioontology.org)
  - Map non-expert terminologies to controlled vocabularies
  - Map screen behaviors to health concepts (e.g. typing speed ~ motor skill)
- **Exploratory** analysis of screenome data
  - Cluster within-individual data, across-individual data
  - Characterize modal behaviors, persistence of features over time
  - Associate clusters with health concepts/behaviors/diseases
- Focused **hypothesis-driven** analyses
  - Identify & validate screenome-derived risk factors for disease
  - Identify & validate screenome for disease severity and treatment

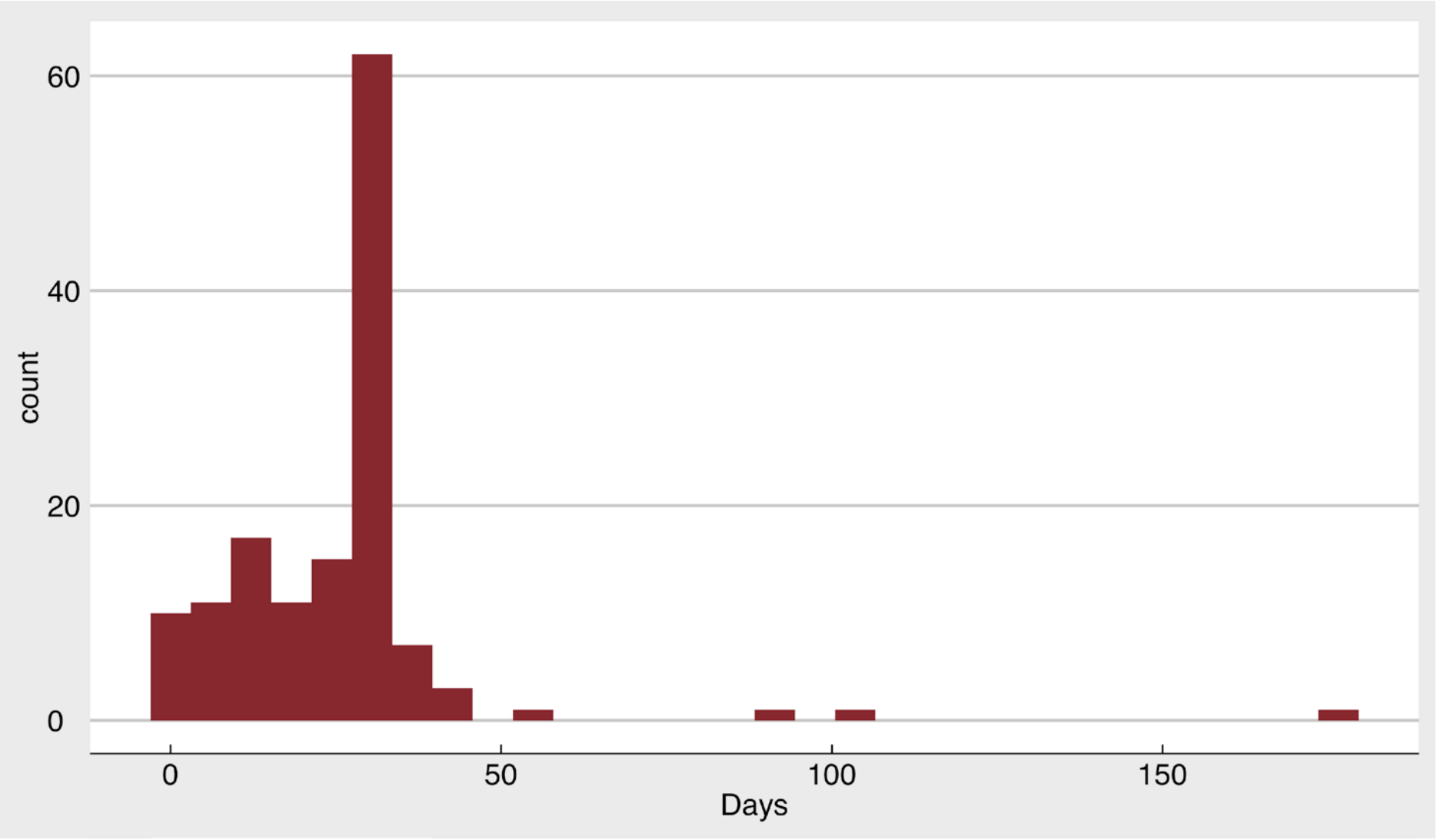
Potential Promise of Screenome

- **Integrate** with other **diagnostics**
- Improve **predictability** of diagnosis
- **Not** dependent on any one commercial product
- Include **previously unobserved** features of life

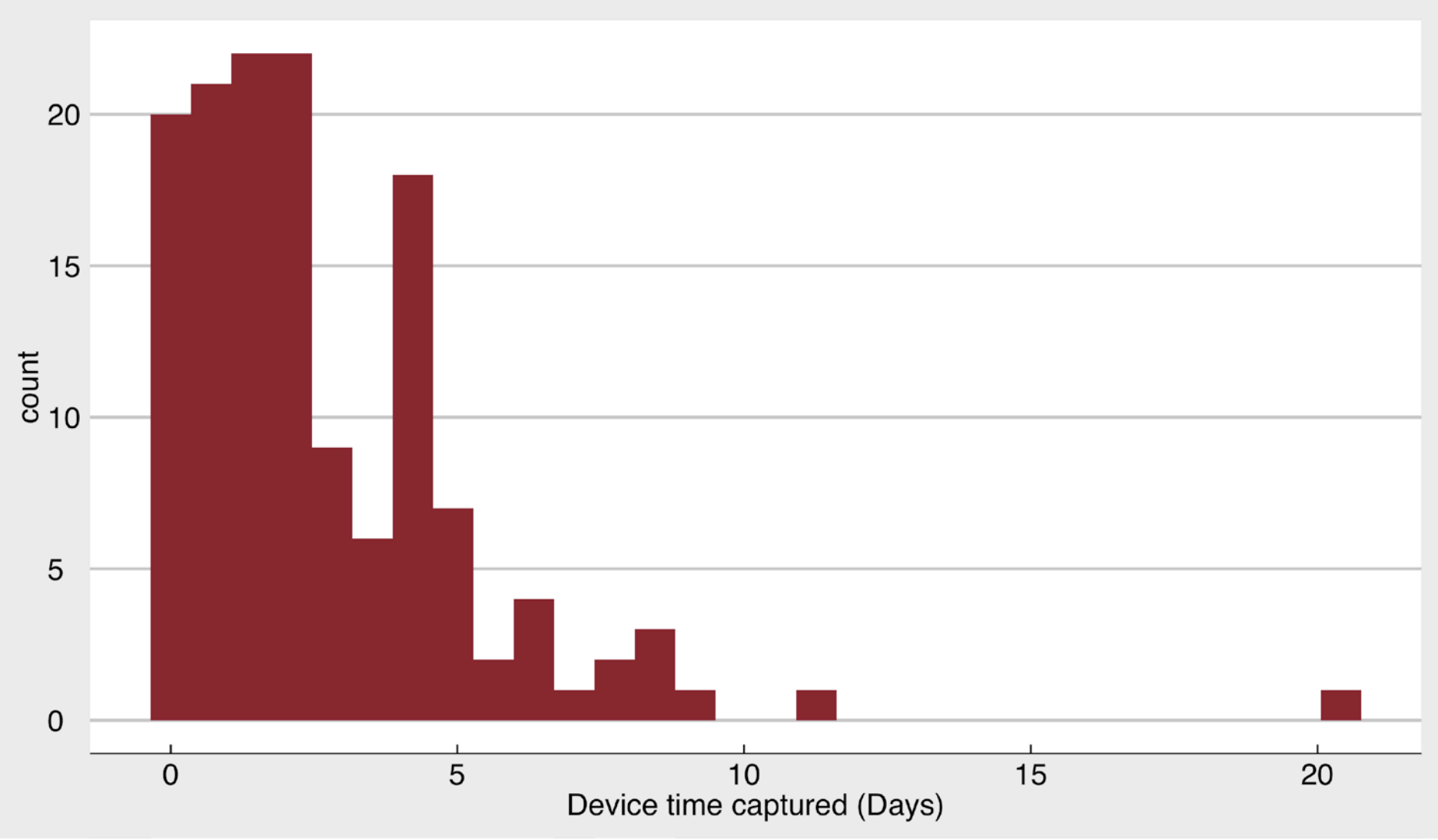
Initial Exploratory Health/Drug Analysis (Adam Lavertu)

- 140 individuals
- LA, Chicago, NY, Stanford
- Screen capture every 5 seconds when device is active
- 6,541,755 screen captures so far

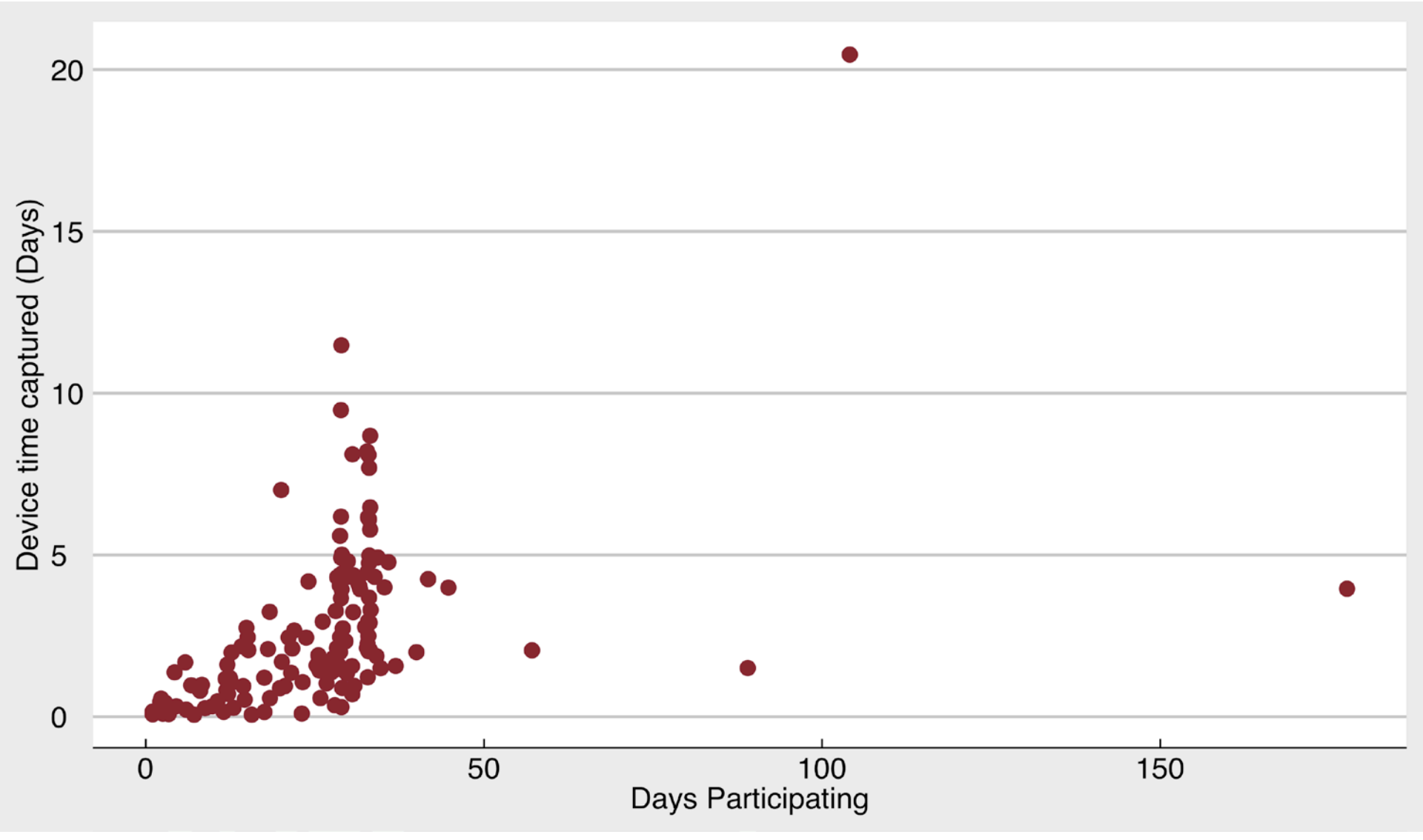
of Days Participating



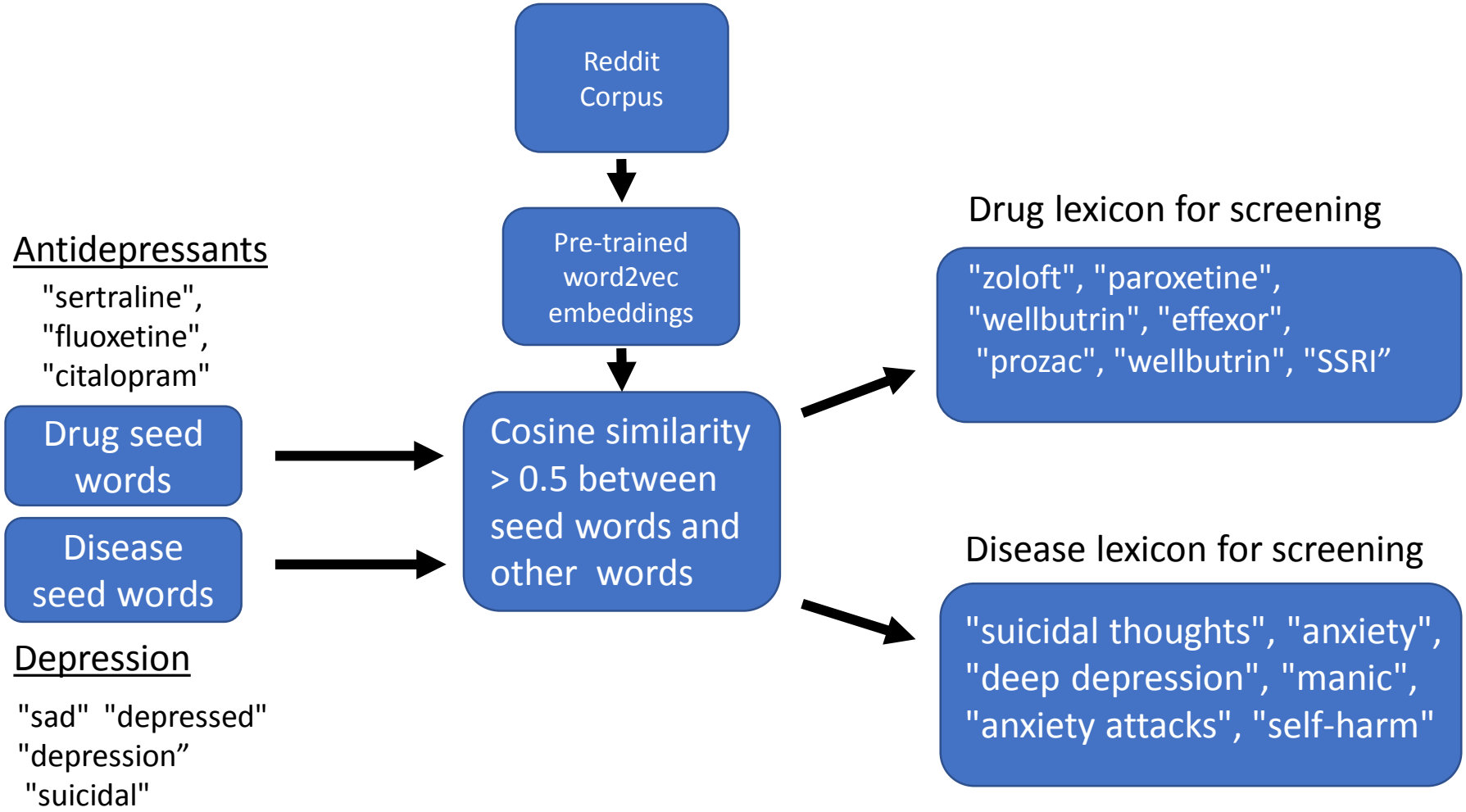
Total Device time captured



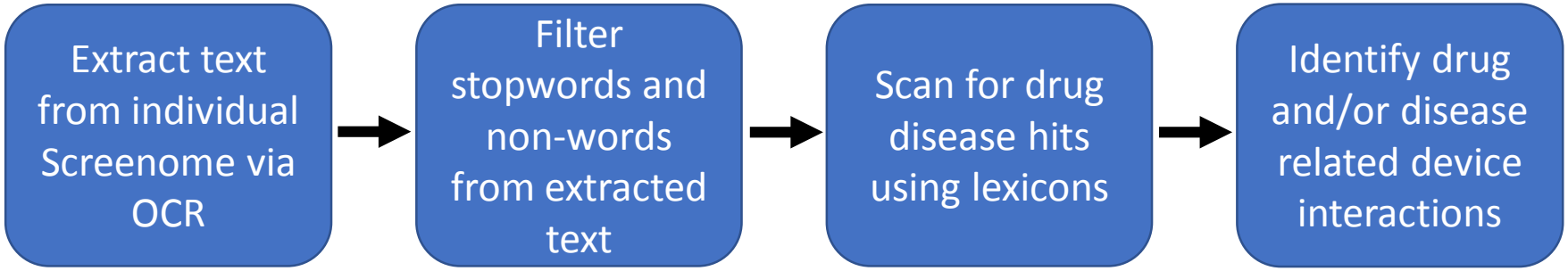
Time captured vs. Days of Partic.



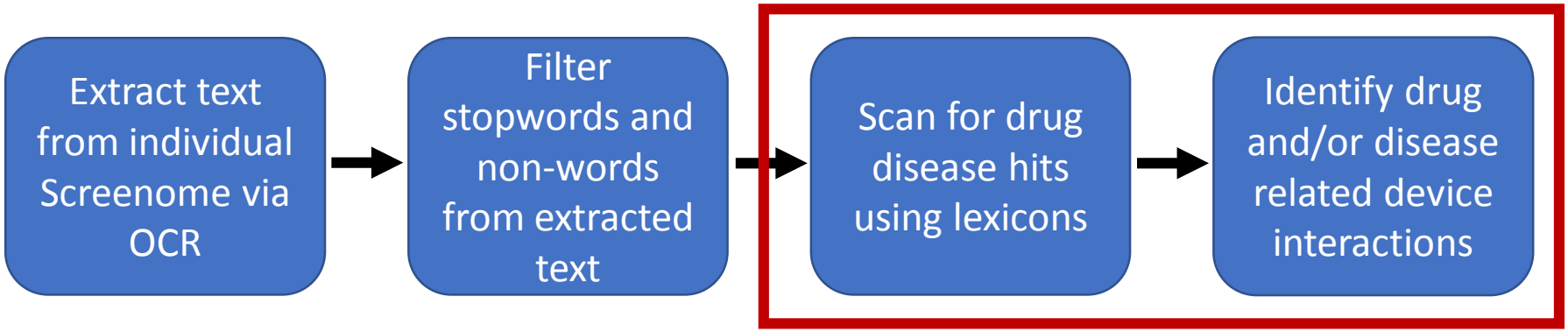
Building Lexicon to Detect Health Synonyms on Screens



Overall Exploratory Workflow

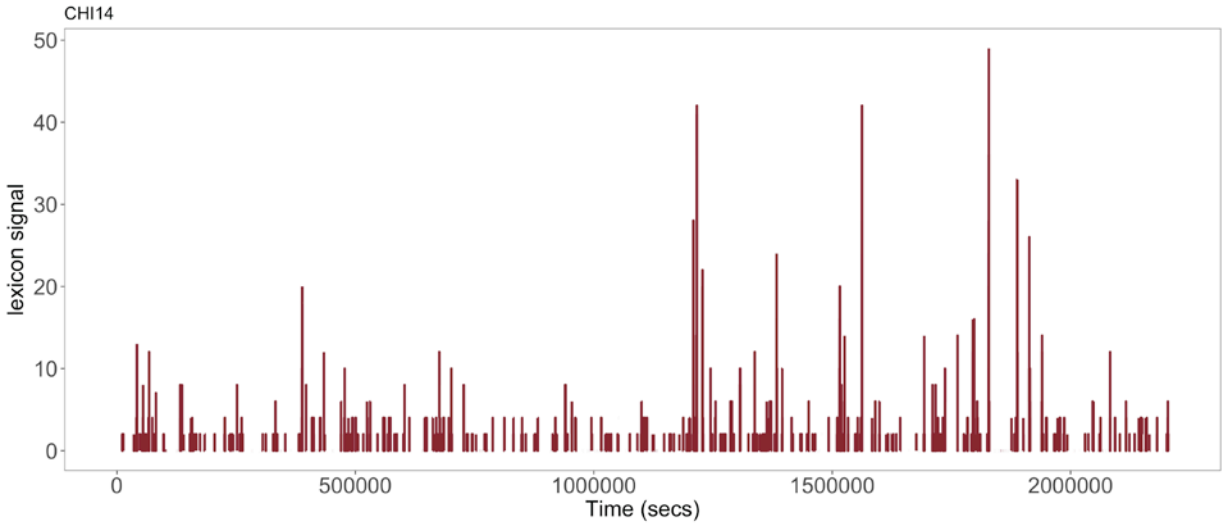


Overall Exploratory Workflow

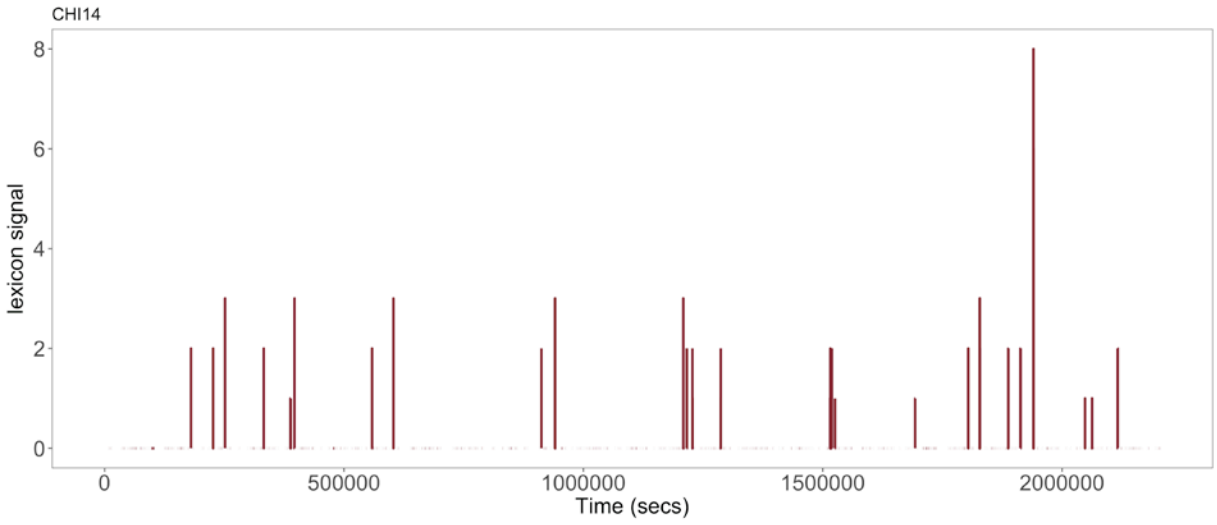


Individual #1

“pain” words found



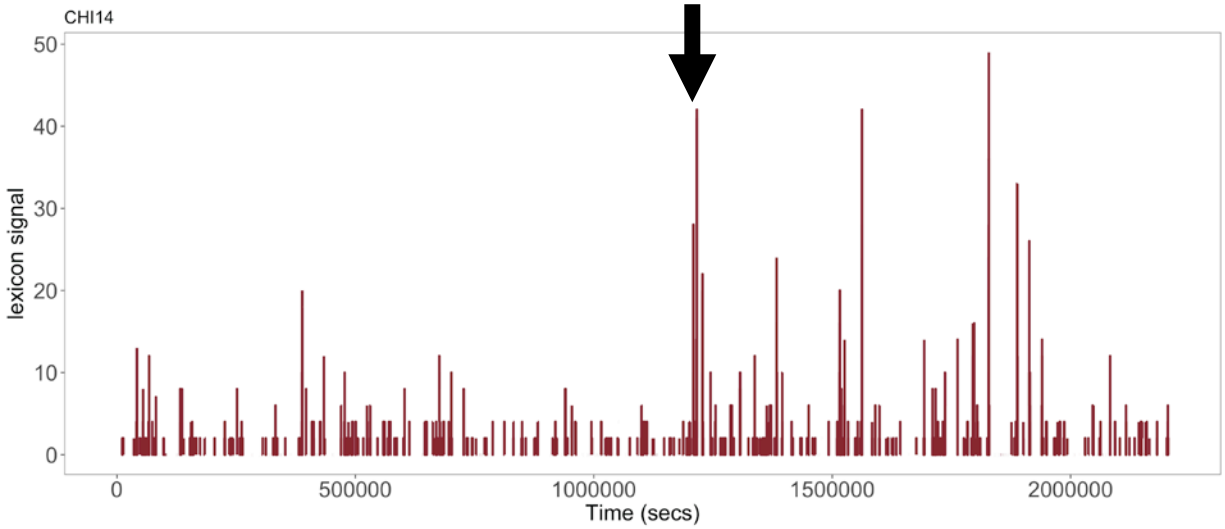
“NSAID” words found



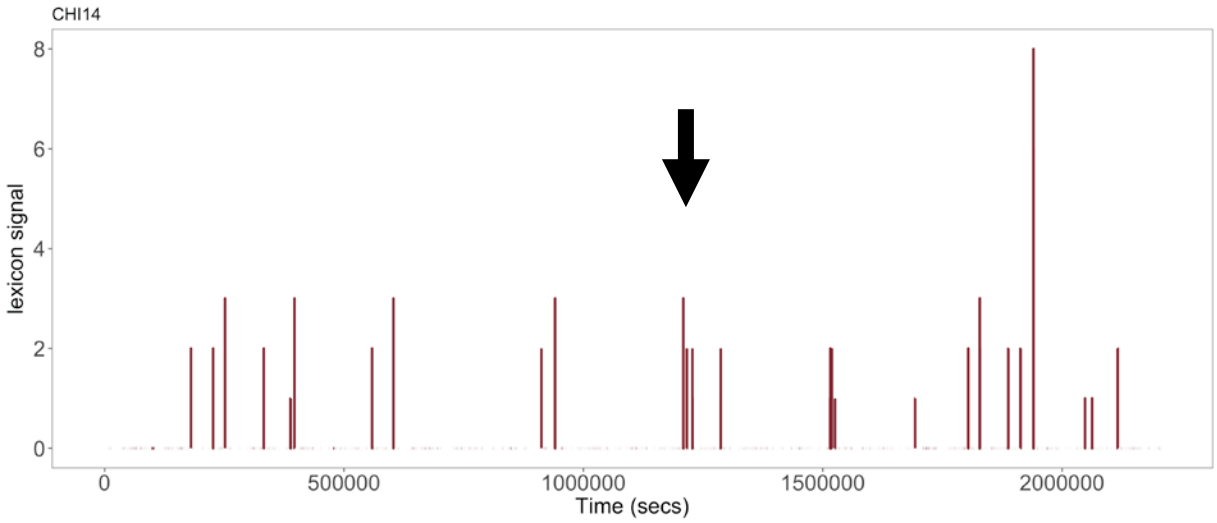
~25 days (in seconds)

Individual #1

“pain” words found



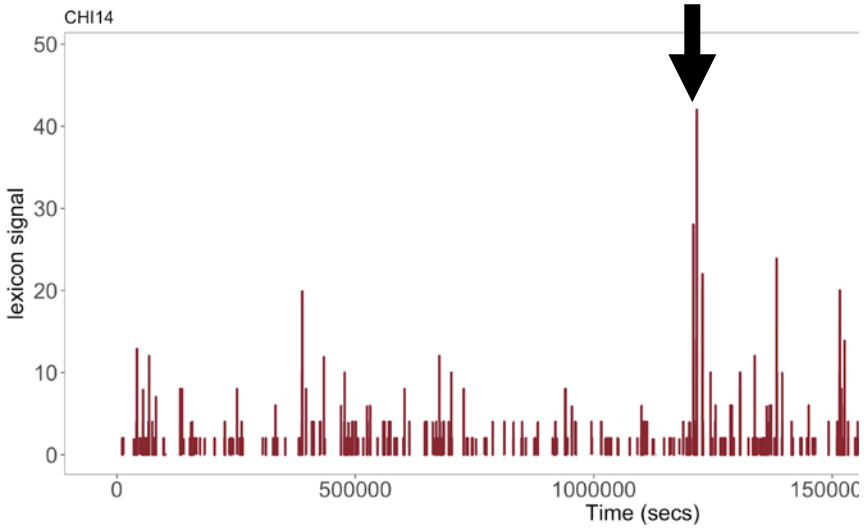
“NSAID” words found



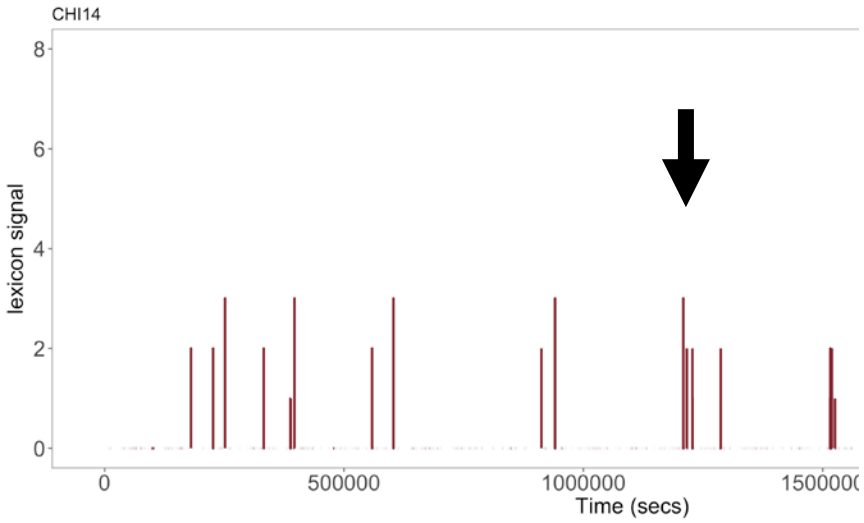
~25 days (in seconds)

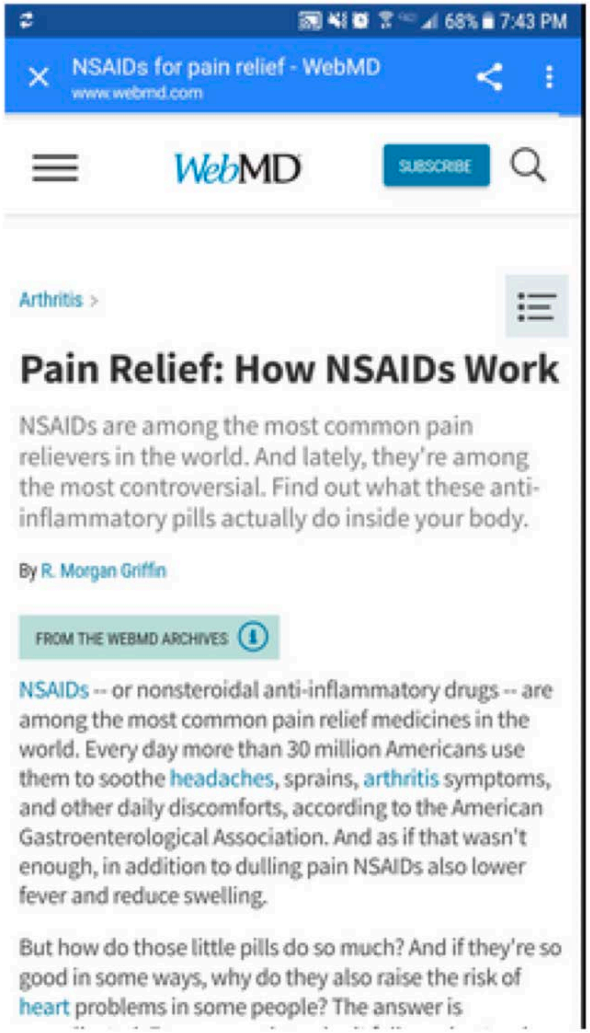
Individual #1

“pain” words found



“NSAID” words found

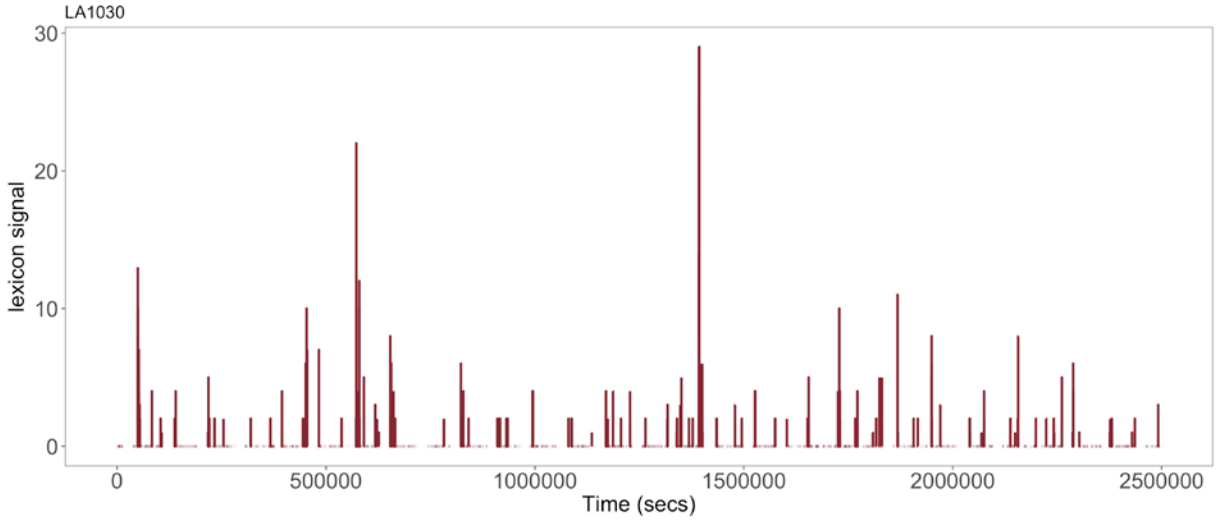




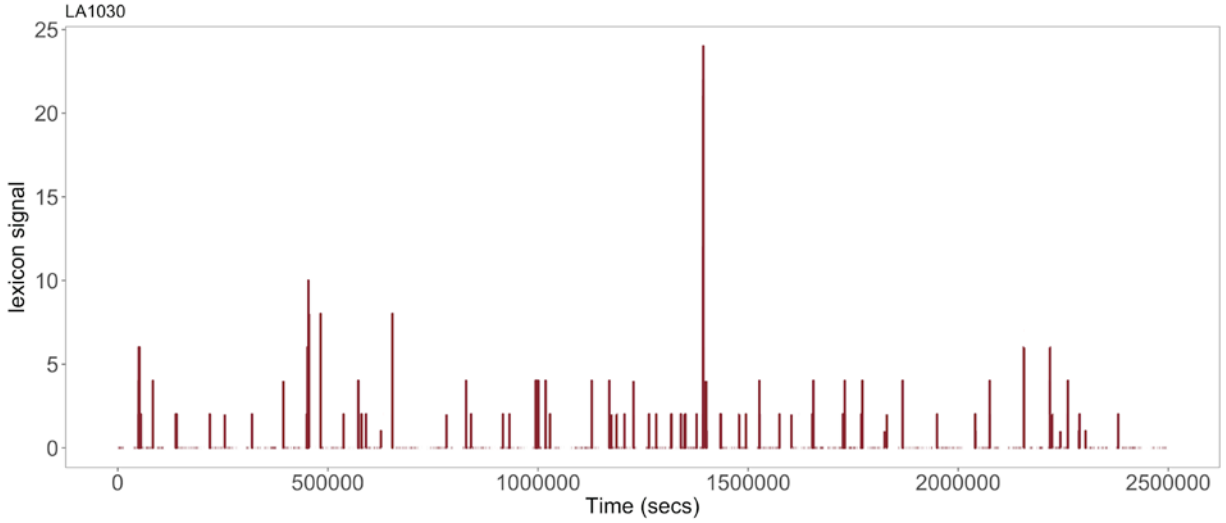
~25 days (in seconds)

Individual #2

“diabetes” words found



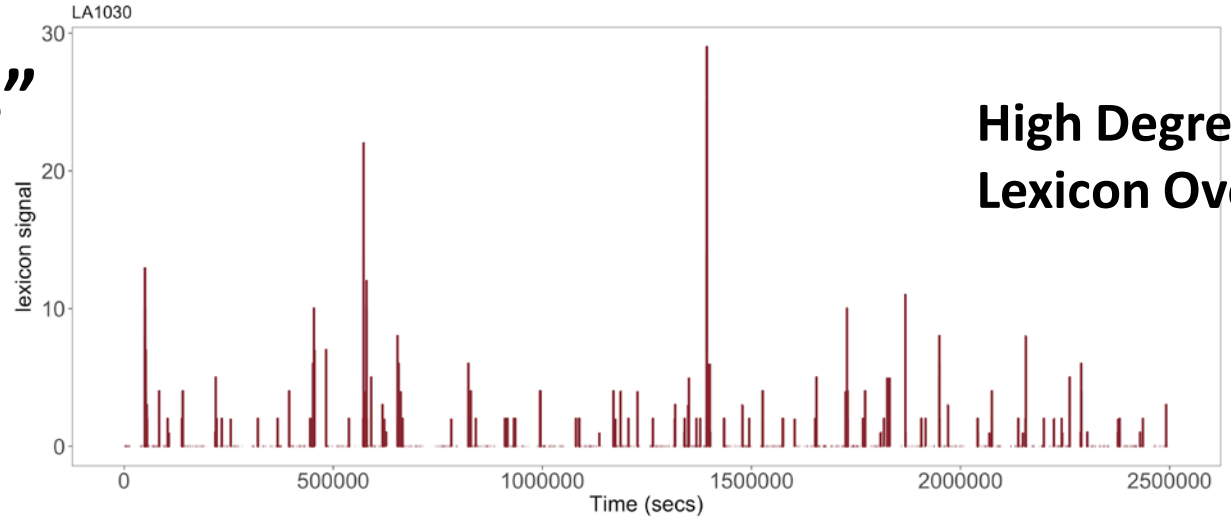
“diabetes meds” words found



~25 days (in seconds)

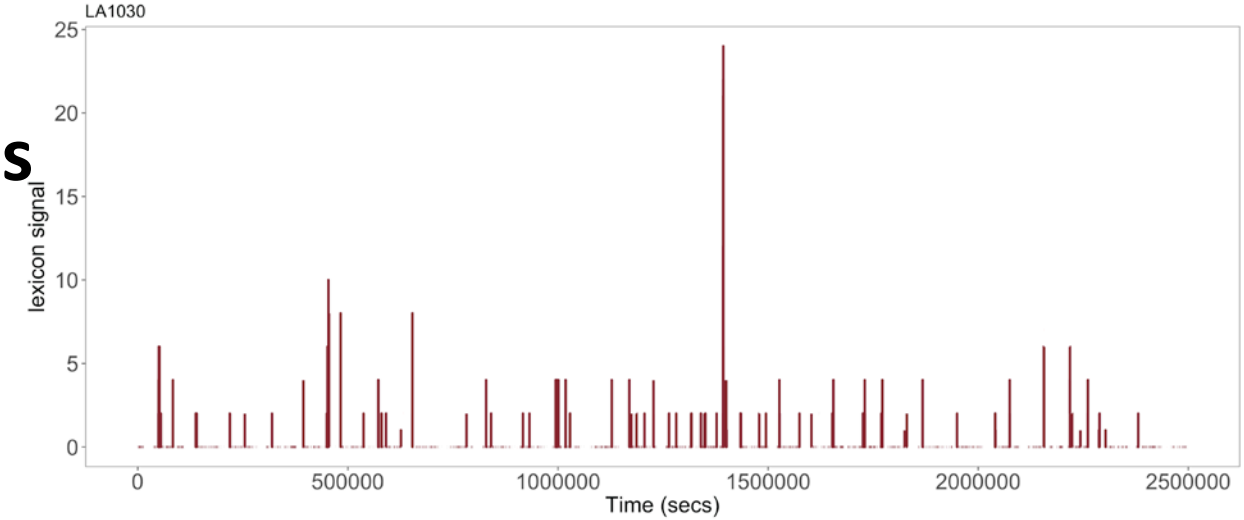
Individual #2

“diabetes” words found



High Degree of Lexicon Overlap

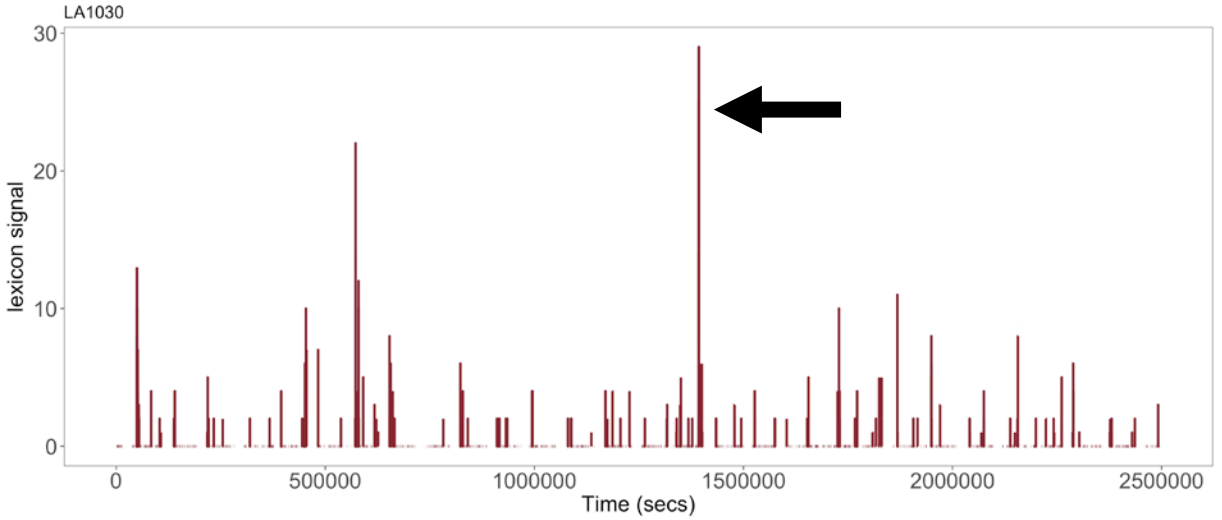
“diabetes meds” words found



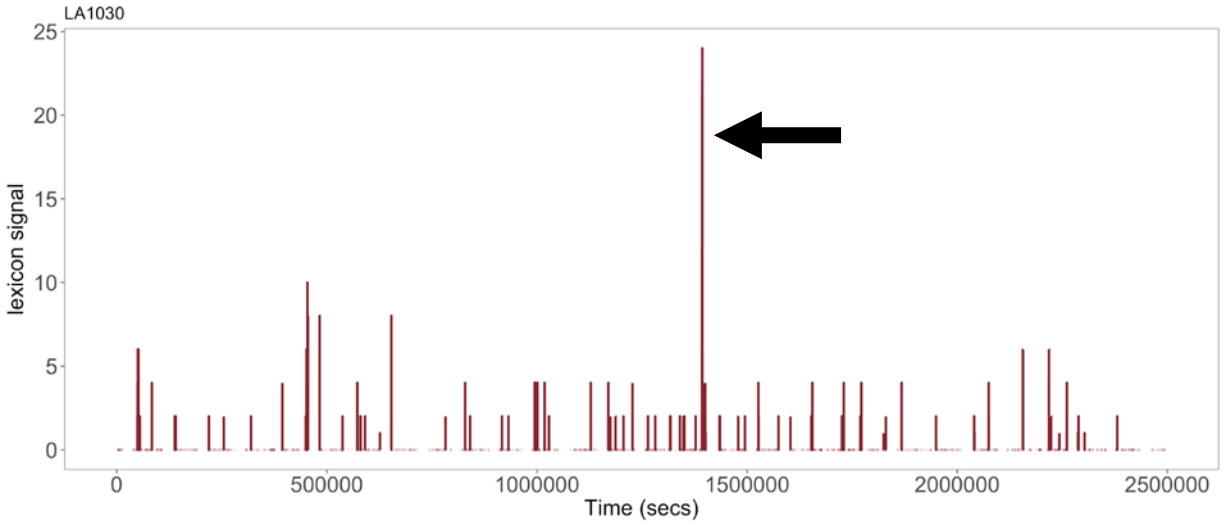
~25 days (in seconds)

Individual #2

“diabetes” words found



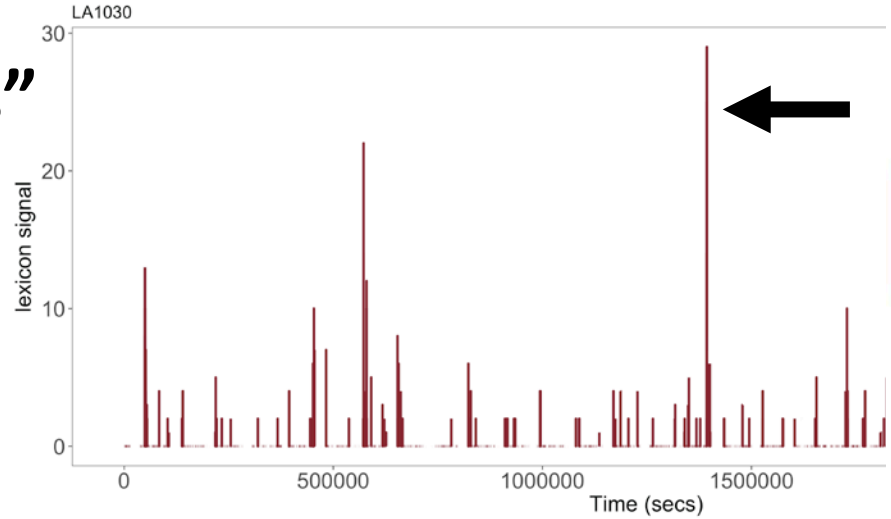
“diabetes meds” words found



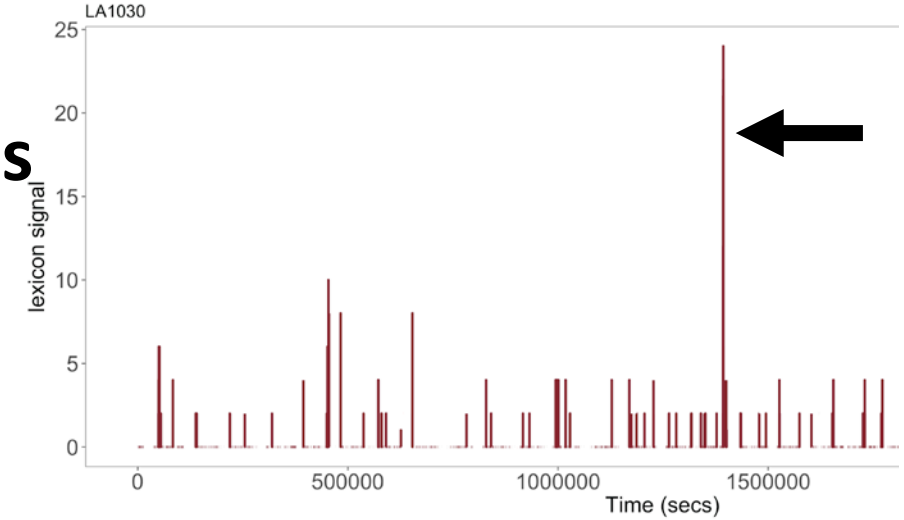
~25 days (in seconds)

Individual #2

“diabetes” words found



“diabetes meds” words found

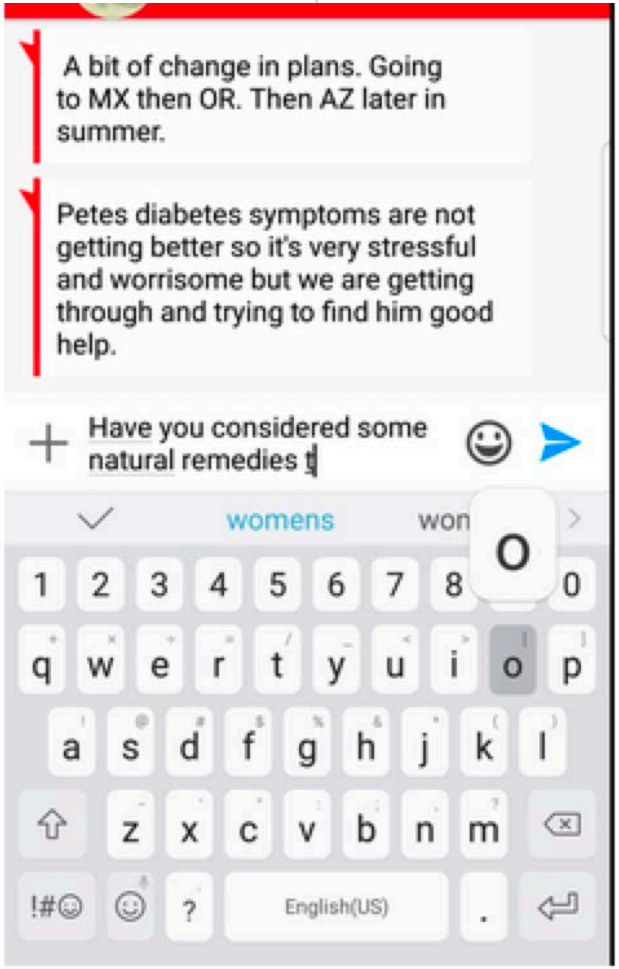


~25 days (in seconds)

Individual #2

“diabetes” words found

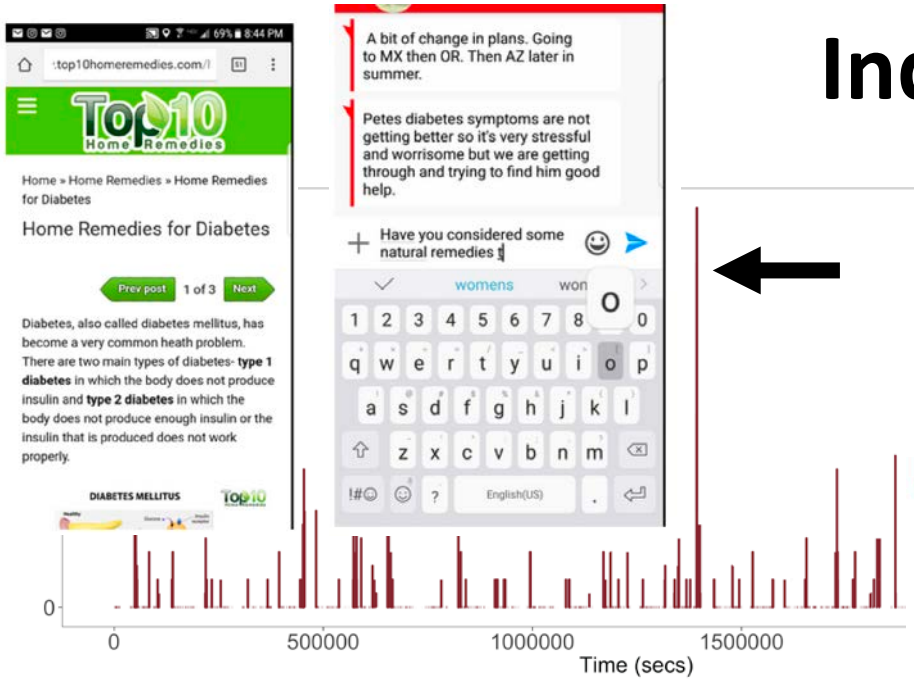
“diabetes meds” words found



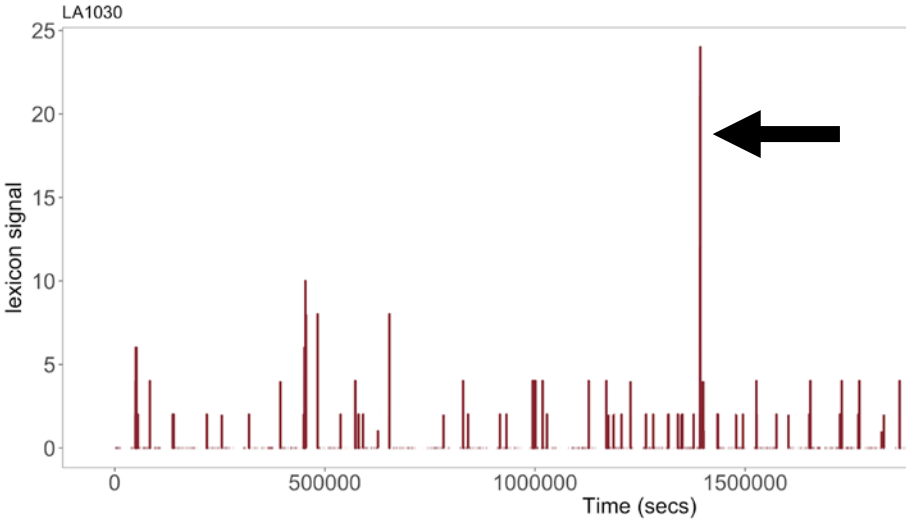
~25 days (in seconds)

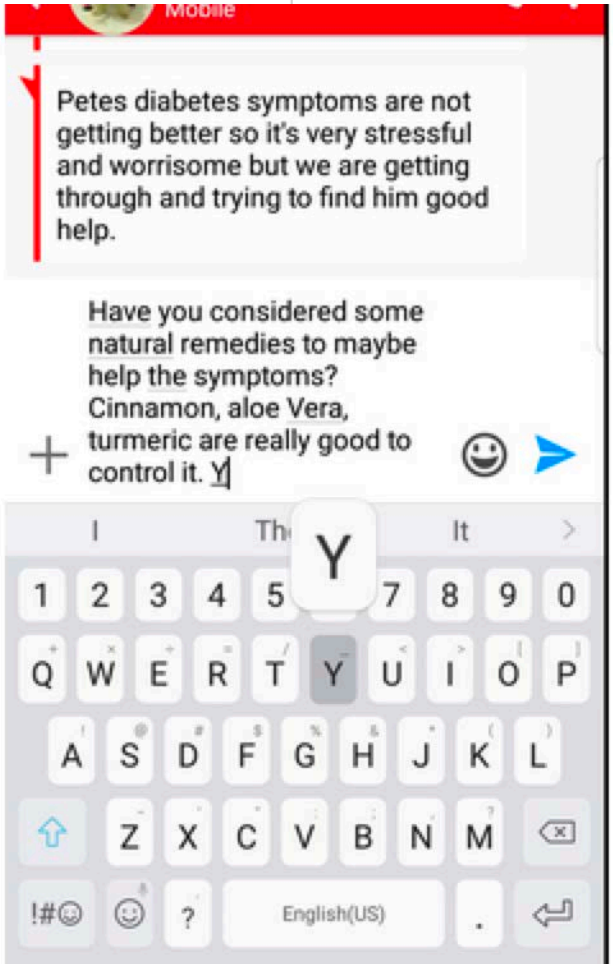
Individual #2

Diabetes Lexicon



Diabetes Drugs Lexicon



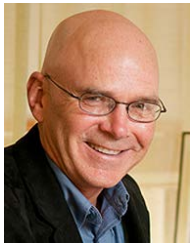


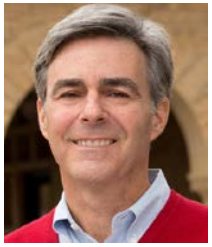
~25 days (in seconds)

Analytical Challenges

- Pop-Up Advertisements (sometimes relevant, sometimes false positives)
- Who is the patient vs. caregiver/relative
- Other languages
- Many many more…

Thanks! russ.altman@stanford.edu


Byron Reeves


Tom Robinson


Russ Altman


Nilam Ram

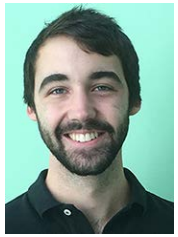

Lee Giles


Jennifer Pan


Jim Cummings


MJ Cho


Katie Roehrick

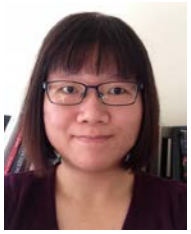

Dan Muise


Yingdan Lu

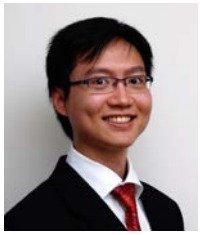

Anu Gagneja

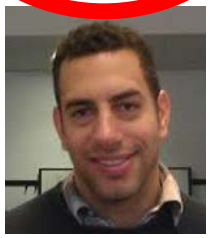

Leo Yeykelis


Agnese Chiatti


Xiao Yang


Mimi Brinberg





Growing **literature** on **social media** and health…

- Focused on **single media** (Facebook, Twitter, patient forums, internet search)
- **Not integrated** across activities
- **Difficult** to understand **context**
  - Difficult to untangle statistical bias
  - Very low temporal resolution

Web-scale pharmacovigilance: listening to signals from the crowd FREE

Ryen W White ✉, Nicholas P Tatonetti, Nigam H Shah, Russ B Altman, Eric Horvitz

Journal of the American Medical Informatics Association, Volume 20, Issue 3, 1 May 2013, Pages 404–408, https://doi.org/10.1136/amiajnl-2012-001482

Published: 07 March 2013 **Article history** ▾

JAMIA

ELSEVIER

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Exploring online communication about cigarette smoking among Twitter users who self-identify as having schizophrenia

Yulin Hswen[a,b,*], John A. Naslund[c], Pooja Chandrashekar[d], Robert Siegel[e], John S. Brownstein[b,f,g], Jared B. Hawkins[b,f,g]

ELSEVIER CrossMark

Surgery for Obesity and Related Diseases 13 (2017) 1369–1375

SURGERY FOR OBESITY AND RELATED DISEASES

Original article

Examination of bariatric surgery Facebook support groups: a content analysis

Afton M. Koball, Ph.D., A.B.P.P.[a,*], Dylan J. Jester, B.S.[b], Sarah E. Domoff, Ph.D.[c,d], Kara J. Kallies, M.S.[b], Karen B. Grothe, Ph.D., A.B.P.P.[e], Shanu N. Kothari, M.D., F.A.C.S., F.A.S.M.B.S.[f]

Published on 21.01.11 in Vol 13, No 1 (2011): Jan-Mar

This paper is in the following e-collection/theme issue:

Medicine 2.0: Social Media, Open, Participatory, Collaborative Medicine

Journal of Medical Internet Research

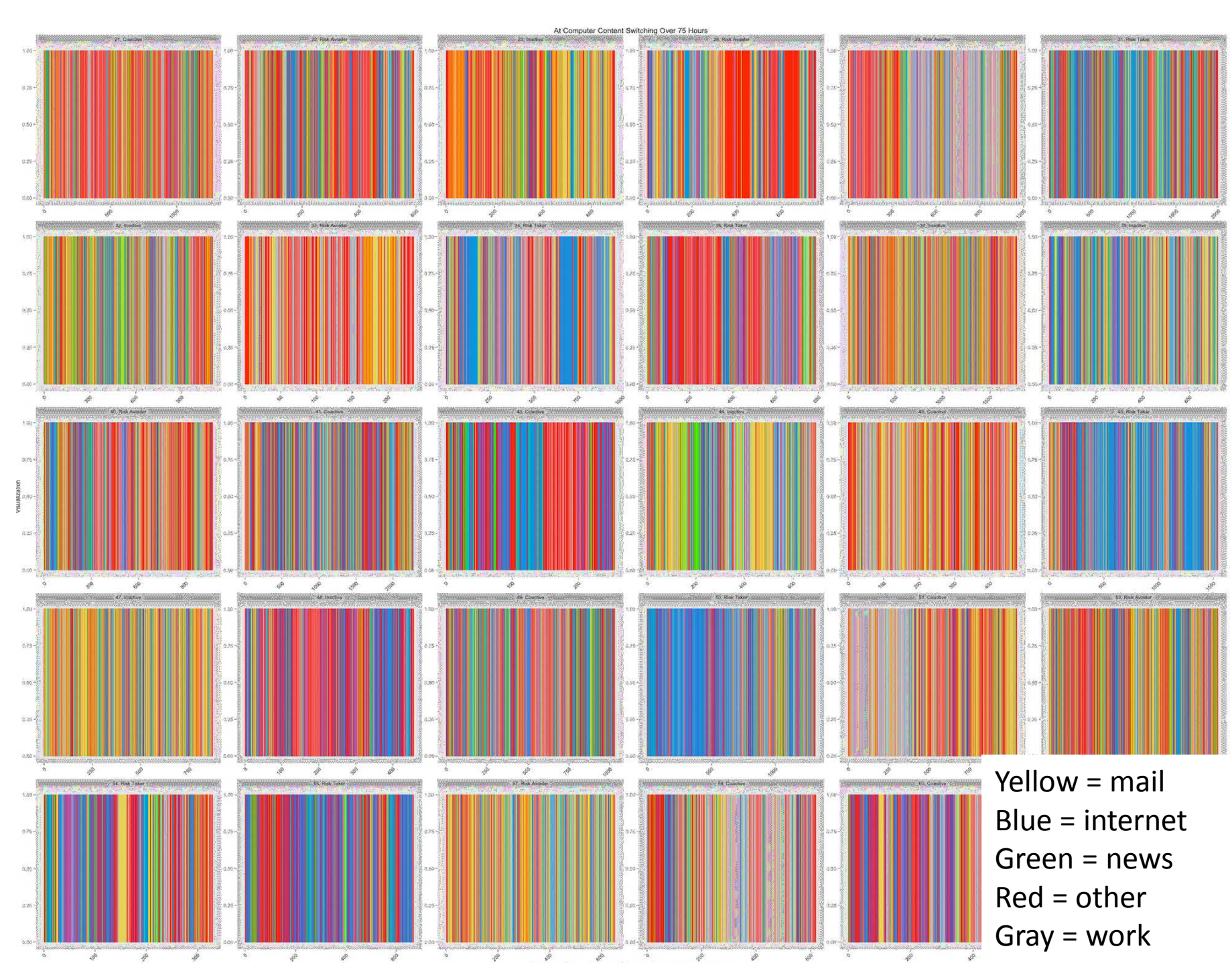
patientslikeme

Article | Cited By (50) | Tweetations (14) | Metrics

Original Paper

Patient-reported Outcomes as a Source of Evidence in Off-Label Prescribing: Analysis of Data From PatientsLikeMe

Jeana Frost[1], PhD; Sally Okun[2], RN; Timothy Vaughan[2], PhD; James Heywood[2], BS; Paul Wicks[2], PhD


At Computer Content Switching Over 75 Hours
Visualization
Content Category
email
entertainment
news
other
work
Yellow = mail
Blue = internet
Green = news
Red = other
Gray = work

Cluster Analysis of 3-Day Screenomes *N* = 30