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

<table>
<thead>
<tr>
<th>Hyperglycemia Signs and Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>appetite increase</td>
</tr>
<tr>
<td>blurred vision</td>
</tr>
<tr>
<td>blurry vision</td>
</tr>
<tr>
<td>breathing difficulty</td>
</tr>
<tr>
<td>breathing trouble</td>
</tr>
<tr>
<td>breathless</td>
</tr>
<tr>
<td>breathlessness</td>
</tr>
<tr>
<td>coma</td>
</tr>
<tr>
<td>confused</td>
</tr>
<tr>
<td>confusion</td>
</tr>
<tr>
<td>decreased libido</td>
</tr>
<tr>
<td>decreased sex drive</td>
</tr>
<tr>
<td>decreased sexual desire</td>
</tr>
<tr>
<td>dehydrated</td>
</tr>
<tr>
<td>dehydration</td>
</tr>
<tr>
<td>difficulty breathing</td>
</tr>
<tr>
<td>dizziness</td>
</tr>
<tr>
<td>dizzy</td>
</tr>
<tr>
<td>drowsiness</td>
</tr>
<tr>
<td>Drowsy</td>
</tr>
<tr>
<td>dry mouth</td>
</tr>
<tr>
<td>dry skin</td>
</tr>
<tr>
<td>erectile dysfunction</td>
</tr>
<tr>
<td>fatigue</td>
</tr>
<tr>
<td>fatigued</td>
</tr>
<tr>
<td>feet tingling</td>
</tr>
<tr>
<td>frequent urinating</td>
</tr>
</tbody>
</table>
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

A large set of health markers
Column People keep leaping to their deaths from iconic Pasadena bridge. How do we stop them?

Suicide

Cancer

Memory

Diabetes

Home / Diseases Conditions
8 reasons for unexplained bruises on your body
By Mita Majumdar | Updated: Wed, November 25, 2015 10:59 am
Do you bruise in places that you don't even remember bumping or hurtting, and the bruise doesn't go away for months together? Read more about what could cause it.

I can't find my shopping list.
When will you be picking me up?
I will be there at 1 pm. Already told you an hour ago. Did you forget?
I saw your shopping list in the kitchen next to the fridge.

Timeline Photos
have been thought of. F11. This is called “Acanthosis Nigricans”, and is a symptom of INSULIN RESISTANCE, and a sign that you are diabetic (or soon to be), or you are Obese, or have a Thyroid or Pituitary Disorder. So don't go giving your kids chemical burns and mental trauma. Take them to the doctor and cut out those sweets and carbs! #memoryLane
www.the413challenge.com
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
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 Participating.
Building Lexicon to Detect Health Synonyms on Screens

Reddit Corpus

Pre-trained word2vec embeddings

Cosine similarity > 0.5 between seed words and other words

Drug lexicon for screening
"zoloft", "paroxetine", "wellbutrin", "effexor", "prozac", "wellbutrin", "SSRI"

Disease lexicon for screening
"suicidal thoughts", "anxiety", "deep depression", "manic", "anxiety attacks", "self-harm"

Antidepressants
"sertraline", "fluoxetine", "citalopram"

Drug seed words

Disease seed words

Depression
"sad" "depressed" "depression" "suicidal"
Overall Exploratory Workflow

1. Extract text from individual Screenome via OCR
2. Filter stopwords and non-words from extracted text
3. Scan for drug disease hits using lexicons
4. Identify drug and/or disease related device interactions
Overall Exploratory Workflow

1. Extract text from individual Screenome via OCR
2. Filter stopwords and non-words from extracted text
3. Scan for drug disease hits using lexicons
4. Identify drug and/or disease related device interactions
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

“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

~25 days (in seconds)
“diabetes” words found

“diabetes meds” words found

~25 days (in seconds)
Individual #2

~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
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
Cluster Analysis of 3-Day Screenomes $N = 30$