How online research opens new doors for computational psychiatry

Claire Gillan, PhD
Assistant Professor of Psychology and MQ Fellow
Trinity College Dublin
www.gillanlab.com
@clairegillanTCD
Outline

• Why take psychiatry research online?

• Online methods
  o Crowdsourcing
  o Smartphones
  o Scraping
  o Proxies for physiology

• Characteristics of online samples
  o Are they representative?
  o Are mental health data valid?
  o Are the data of acceptable quality?
  o Are findings relevant for diagnosed patients?

Spoiler: yes.
Why take psychiatry research online?
The problem: DSM disorders are the "ground-truth" for research
The problem: DSM disorders are the “ground-truth” for research
The problem: DSM disorders are the “ground-truth” for research

- Patients are heterogeneous within-disorder
The problem: DSM disorders are the “ground-truth” for research

- Patients are heterogeneous within-disorder
- Patients are similar across-disorder
The solution: **redraw the lines**

Symptom Dimension 1
Symptom Dimension 2
Symptom Dimension 3

**Treatment** Dimensions?  **Brain** Dimensions?  **Genetic** Dimensions?
**Cognitive** Dimensions?  **Environmental** Dimensions?
To do this well, we need big samples.

People living in Dublin: 500,000
People using the Internet: 3,200,000,000
Why you should think about going online

Lab-based
N=17
OCD

Lab-based
N=9
OCD
CBT longitudinal

Online w/phone
N=200
OCD/GAD
(40% longitudinal)

Crowdsourcing
N=1400
General population
Why you should think about going online

• Sample size (>3.2 billion people use the internet)

• Access select populations (geography, age, race, clinical condition, socio-economic status, etc.)

• Speed (hundreds/thousands of subjects per day)

• Low cost (but it depends)

• Anonymous participation (sensitive populations)

• Reproducibility/standardization (same test, different lab)

• Exploratory vs. confirmatory research
Crowdsourcing Platforms
Crowdsourcing Platforms

1. Internal Amazon Problems:
   What is the title of this book?

There are things that humans can do better than computers
Crowdsourcing Platforms

1. Internal Amazon Problems:
   What is the title of this book?

2. Internal Market Research
   Which cover do you prefer?

There are things that humans can do better than computers
Crowdsourcing Platforms

1. Internal Amazon Problems: *What is the title of this book?*

2. Internal Market Research *Which cover do you prefer?*

3. Opens to everyone: *Transcribing audio files, writing descriptions of complex scenes, answering surveys, ranking products and full-scale psychology experiments*

There are things that humans can do better than computers
Crowdsourcing Platforms

Conduct psychology research with sophisticated, interactive cognitive tests

• Recruit from this **global workforce** at an unprecedented rate. e.g. N>1400 in 10 days

• Reproduces classic psychology effects

Crump et al., 2013
Impairments in Goal-Directed Actions Predict Treatment Response to Cognitive-Behavioral Therapy in Social Anxiety Disorder

Gail A. Alvares, Bernard W. Balleine, Adam J. Guastella*  
Brain & Mind Research Institute, The University of Sydney, Sydney, New South Wales, Australia

Gillan et al., American Journal of Psychiatry, 2011
Are goal-directed deficits characteristic of a trans-diagnostic dimension?

Symptom Dimension 1

Symptom Dimension 2

Symptom Dimension 3

Obsessive Compulsive Disorder

Social Anxiety Disorder

Gillan et al., eLife, 2016
Are goal-directed deficits characteristic of a trans-diagnostic dimension?

But... these scores reflect overlapping constructs
Factor Analysis

Inter-correlation of 209 individual self-report questionnaire items

Gillan et al., eLife, 2016
Factor 2: “Compulsive Behaviour and Intrusive Thought”

- Binge eating
- Alcohol dependence
- Repetitive checking
- Disturbing thoughts
- Impulse buying
- Cannot control thoughts

Gillan et al., eLife, 2016
The factor structure can be independently reproduced

Rouault, Seow*, et al., Biological Psychiatry, 2018

The association with goal-directed planning replicates

Patzelt et al., Biological Psychiatry, 2019
Can we use this method to study **metacognition** in anxious-depression?
Is confidence expressed transdiagnostically?

Rouault*, Seow*, Gillan & Fleming, Biological Psychiatry, 2018
Using confidence to unpack the role of metacognition in anxious-depression

Metacognitive bias
over/under-confidence

Confidence

Metacognitive sensitivity
“correct about being correct”

**Metacognitive Sensitivity in anxious-depression**

- Apathy
- Social Anxiety
- Generalised Anxiety
- Impulsivity
- Depression
- Alcoholism
- Schizotypy
- OCD
- Eating Disorders

- ‘Anxious-Depression’
- ‘Compulsive Behavior and Intrusive Thought’
- ‘Social Withdrawal’

Rouault*, Seow*, Gillan & Fleming, Biological Psychiatry, 2018
Smartphones
Smartphones for Large-Scale Citizen Science

The model

- Gamified versions of cognitive tests
- Play as much as you like

What you get

- Really large samples 1000s
- Repeated within-subject, longitudinal assessment
Smartphones for Science

4.3 million downloads!

Hugo Spiers et al., at UCL
Spatial navigation differences across country and sex

But this difference is partially explained by the Gender Gap in each country (and GDP)

Coutrot et al., Current Biology, 2018
Big Data samples can rapidly generate norms.

Spatial navigation performs better than test of episodic memory

Coughlan et al., PNAS 2019
Scraping
Scared? You should be.

What do your thumbs say about you?

Study Design
• Acquired consent
• Psychometric test scores (e.g. IQ)
• Survey information (e.g. democrat?)
• Data from profile, ~170 Facebook Likes per person

Kosinski, Stillwell and Graepel, PNAS, 2013
Scraping social media

This information is in many cases publically available.

When not, it is available for purchase from Facebook

Band formerly called “Burn the Priest”

<table>
<thead>
<tr>
<th>Satisfied</th>
<th>Dissatisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satan</td>
<td>Auschwitz</td>
</tr>
<tr>
<td>Lucifer</td>
<td>Negan</td>
</tr>
<tr>
<td>Jesus</td>
<td>Voldemort</td>
</tr>
<tr>
<td>Bible</td>
<td>Voldemort</td>
</tr>
<tr>
<td>God</td>
<td>Voldemort</td>
</tr>
<tr>
<td>Satan</td>
<td>Auschwitz</td>
</tr>
<tr>
<td>Lucifer</td>
<td>Negan</td>
</tr>
<tr>
<td>Jesus</td>
<td>Voldemort</td>
</tr>
<tr>
<td>Bible</td>
<td>Voldemort</td>
</tr>
<tr>
<td>God</td>
<td>Voldemort</td>
</tr>
</tbody>
</table>

Kosinski, Stillwell and Graepel, PNAS, 2013
Is there a wrong way to do this?
DEEP NEURAL NETWORKS CAN DETECT SEXUAL ORIENTATION FROM FACES

Figure 1. Graphical illustration of the outcome produced by Face++. Panel A illustrates facial landmarks (colored dots, n=83) and facial frame (blue box). Panel B illustrates pitch, roll, and yaw parameters that describe the head's orientation in space. The location of the face in the image, outlines of its elements, and the head’s orientation were extracted using a widely used face-detection software: Face++. Figure 1 shows the output of Face++ in a graphical format. The colored dots (Panel A) indicate the location of the facial landmarks outlining the contour and elements of the face. Additionally, Face++ provided the estimates of the head’s yaw, pitch, and roll (Panel B).

Based on the Face++ results, we removed images containing multiple faces, partially hidden faces (i.e., with one or more landmarks missing), and overly small faces (i.e., where Face++ can be accessed at http://www.faceplusplus.com).

Scraping without consent

- With 1 image per person, AUC = .81
- With 5 images, AUC > .90
- Algorithm outperformed humans (N=1000)

Gender atypicality predicts gay and lesbian status

Yang & Kosinski, J Personality and Social Psychology, 2018
GLAAD and HRC call on Stanford University & responsible media to debunk dangerous & flawed report claiming to identify LGBTQ people through facial recognition technology.

"participants" (the people whose images were used) were not consented.

Why Stanford Researchers Tried to Create a ‘Gaydar’ Machine

The New York Times

THE A.I. “GAYDAR” STUDY AND THE REAL DANGERS OF BIG DATA

The New Yorker
Biology vs Cultural Norms

Do algorithms reveal sexual orientation or just expose our stereotypes?

Margaret Mitchell
Blaise Aguera y Arcas
Alex Todorov
Scraping Social Media as an alternative to Ecological Momentary Assessment
Can we approximate this using Twitter?
Scraping Social Media for Ecological Momentary Assessment in Psychiatry

Advertising exclusively through twitter

- soliciting mental health ‘influencer’ or organisation retweets

~5 minute survey

Within the past year, have you had an episode of major depression? *

Yes

Start Date *

02/04/2019

End Date *

03/06/2019
Scraping Social Media for Ecological Momentary Assessment in Psychiatry

Sentiment Analysis

- **LIWC** is a dictionary of ~12,000 words with 90 different output variables
  - linguistic characteristics (e.g. articles and pronouns)
  - psychological constructs (e.g. sadness and positive emotions)
  - general text information (e.g. punctuation and word count)

- **VADER** was built specifically for use with social media text
  - Emojis
  - Punctuation (!!!)
  - Capitalization (SAD)
  - degree modifiers (really, totally, very, etc)
Scrapping Social Media for Ecological Momentary Assessment in Psychiatry

De Choudhury et al., 2013
Tweet timings as a proxy for sleep

example participant w/ self-report depression

Time of Day

Tweet Percentage

0 5 10 15 20

0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10 0.11 0.12

Sean Kelley

sleep

De Choudhury et al., 2013
Proxies for Physiology: Eye-Tracking
Proxies for Physiology: Eye-Tracking

Proof of principle: What video features attract the most attention in babies?

Method
- Subjects sat on parents lap
- Initial video used to determine eligibility
- Then if eligible, babies were asked to watch different videos.
- Once complete, blind raters coded videos for looking time by ‘video features’

Key Findings:
- faces, camera zooms and rhyming and singing increased infant attention.

Cusack et al., 2017, Journal of Experimental Child Psychology
Proxies for Physiology: Eye-Tracking

**Proof of principle:**
*What video features attract the most attention in babies?*

**Method**
- Subjects sat on parents lap
- Initial video used to determine eligibility
- Then if eligible, babies were asked to watch different videos.
- Once complete, blind raters coded videos for looking time by ‘video features’

**Key Findings:**
- faces, camera zooms and rhyming and singing increased increased infant attention.

_Cusack et al., 2017, Journal of Experimental Child Psychology_
Proxies for Physiology: Phone Sensor Data

- GPS receiver
- Microphone
- Camera
- Pedometer
- Heart rate
- Text messages
- Social networking
- Typing speed
- Word-use complexity
- Time using phone
- ...

Bipolar Symptoms

- High or Low Mood
- Racing Thoughts or Psychomotor Retardation
- Excessive or Reduced Activity
- Increased or Reduced Sleep
- Level of Social Engagement

Smartphone Measures

- Self-Report
- Smartphone Microphone
- Accelerometer
- Geo-Location
- Social Media Patterns
- Light Sensor
- Smartphone Use

Matthews et al., 2016
Outline

• Why take psychiatry research online?

• Online methods
  o Crowdsourcing
  o Smartphones
  o Scraping
  o Proxies for physiology

• Characteristics of online samples
  o Are they representative?
  o Are mental health data valid?
  o Are the data of acceptable quality?
  o Are findings relevant for diagnosed patients?

Spoiler: yes.
Characteristics of Online Samples
Are Internet based samples representative?

### Race

<table>
<thead>
<tr>
<th>Race</th>
<th>Internet-Based</th>
<th>US census</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>90%</td>
<td>70%</td>
</tr>
<tr>
<td>Black</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>Asian</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>American Indian / Alaskan Native</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>Native Hawaiian / Pacific Islander</td>
<td>1%</td>
<td>1%</td>
</tr>
</tbody>
</table>

M-Turk is **more racially diverse** than regular online sample and college samples

Buhrmester at al., 2012

---

1. Gillan et al., under review
Are Internet based samples representative?

Age

Generic Internet-based (USA)

OCD/GAD patients (N=285)

Gillan et al., under review
Are there differences in mental health?

- The prevalence of depression among participants on Mechanical Turk was **consistent with prevalence in the general population**
  
  *Shapiro et al., 2013; Kessler, Chiu, Demler, & Walters, 2005*

- But… **social anxiety is up to 7x more prevalent on M-Turk.**
  
  *Shapiro et al., 2013, Gillan et al., 2016; Seow & Gillan, in prep; Hunter et al., BiorXiv*
Are people who they say they are?

**Overwhelmingly, YES!**

- Excellent **test-retest reliability** for self-report depression, $r=.87$ (Shapiro et al., 2013)
- We can **reproduce associations with mental health** found in in-person samples
- **Few get caught** out with trap questions (e.g. <1%, Gillan et al., eLife 2016)
- A more elaborate **malingering questionnaire caught just 10/530** (Shapiro et al., 2013)
- You can confirm location from IP address (but think about GDPR)
- Mechanical Turk allows you to select for certain demographics (age, gender, etc.)

Don’t incentivize people to lie
Are there differences in data quality?

Unsupervised participants are **less likely** to pay attention to instructions

Oppenheimer, Meyvis, & Davidenko, 2009

Are M-Turk subjects worse than supervised students?

“Research in decision making shows that people, when making decisions and answering questions, prefer not to pay attention and minimize their effort as much as possible. Some studies show that over 50% of people don’t carefully read questions. If you are reading this question and have read all the other questions, **please select the box marked ‘other’ and type ‘Decision Making’ in the box below.** Do not select “predictions of your own behavior.” Thank you for participating and taking the time to read through the questions carefully!

What was this study about?
A Predictions of your own behavior
B Predictions of your friends’ behavior
C Political preferences
D Other _______________________

66.2% versus 88.5%, p < .001

Goodman, Cryder, Cheema, 2013
Are there differences in data quality?

YES. How big of a problem is this?

_It's no biggie_ (personal view): Increased sample size mitigates increased loss of statistical power. E.g. in Gillan et al., eLife 2016

_But to help things:_ establish study-specific _a priori_ exclusion criteria e.g. implausibly fast RTs, missing trials, ‘catch’ questions, comprehension test

Save yourself some money by requiring basic performance to play!
Data quality solutions

COMPREHENSION TEST

• Include a quiz on the basics of the task instructions
• If subjects do not pass, they must repeat instructions until they do
• This deals with bots
• This deal with people just hammering the keys randomly

• Crump et al., 2013 found this improved the issue, but didn’t resolve it fully.

• We noted similar issues in our own work, and found the quiz greatly improved the data quality. Again, doesn’t resolve it fully.
Internet based testing is not a panacea.

- Sometimes you need to have control over the testing environment

  Are addictive traits linked to problems with response inhibition?... Acute intoxication might confound the data

- Can people accurately self report on their mental status? (in some cases, e.g. schizophrenia, not always.)

- Is there an incentive to lie (e.g. monetary)?

- Are the findings relevant to ‘real patients’?
Are findings applicable to diagnosed patients?

OCD

All-Comers Internet-Based
Internet-Based Clinical Collaboration

OCD (N=110) - OCD+GAD (N=92) - GAD (N=83)

Gillan et al., under review
1. Can we reproduce the *dimensional* association between goal-directed planning and compulsivity in diagnosed patient sample?

2. What fits the cognitive data better… *Dimension or Diagnosis?*

*caveat: we use clinical controls not ‘healthy controls’*
Are findings applicable to diagnosed patients?

B

Compulsivity

Model-Based Planning

raw score

-2 -1 0 1 2 3

OCD (N=110)
OC&D+GAD (N=92)
GAD (N=83)

*Replicated after average of 413 days in N=110

Gillan et al., under review
How informative is the diagnosis itself?

**Kindest interpretation is p = .18**

Diagnosis

Goal-Directed Learning

- GAD
- OCD
- OCD + GAD

Strength of association

- Compulsions
- OCD

Gillan et al., under review
Major Depressive Disorder

The problem: DSM disorders are the “ground-truth” for research

Obsessive Compulsive Disorder

We need prediction and we need meaningful clinical outcomes

Social Anxiety Disorder

Major Depressive Disorder

Bipolar Disorder
Internet-Based Treatment Prediction

Can we use Internet-based methods to predict and understand treatment response?

Goal N=1000

Kevin Lynch
Internet-Based Treatment Prediction

Can we use Internet-based methods to predict and understand treatment response?

Kevin Lynch
Final take home message.

Internet-based research is awesome.

• large samples
• rare populations
• longitudinal, predictive research
• sometimes frictionless participation
• exploratory and confirmatory

QUESTIONS?
Thank You

Trinity College
Dublin
Tricia Seow
Andrew Pringle
Kevin Lynch
Eoghan Gallagher
Sean Kelley

Multi-Site Study
Eyal Kalanthroff
Michael Eans
Hilary Weingarden
Ryan Jacoby
Marina Gershkovich
Ivar Snorrason
Raphael Campeas
Cynthia Cervoni
Nick Crimarco
Yosef Sokol
Sarah Garnaat
Nicole McLaughlin
Liz Phelps
Anthony Pinto
Christina Boisseau
Sabine Wilhelm
Nathaniel Daw
Blair Simpson