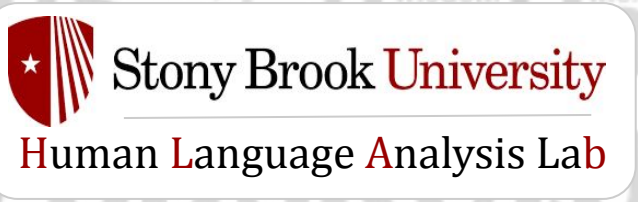
AI Beyond Standard Healthcare Data

H. Andrew Schwartz

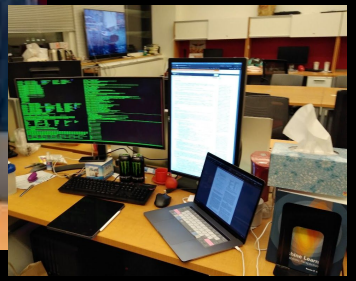
has@cs.stonybrook.edu



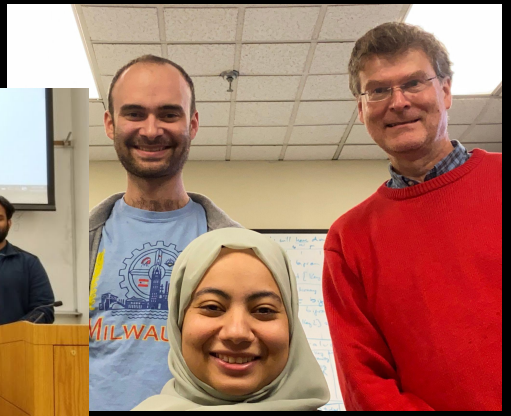
Natural Language Processing

Stony Brook University

nlp.cs.stonybrook.edu



nlp.cs.stonybrook.edu



NLP and Related Faculty

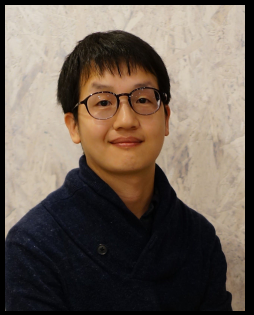
Niranjan Balasubramanian

Ritwik Banerjee

Owen Rambow

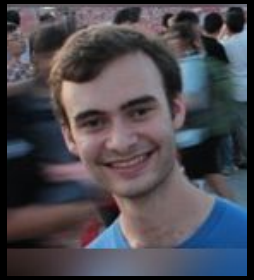
Andy Schwartz

Steven Skiena



Tengfei Ma

Paul Fodor



Jordan Kodner

Jeffrey Heinz

Jason Jones

Thomas Graf

Tuhin Chakrabarty

Yuan Gong

SBU NLP-related Research Groups

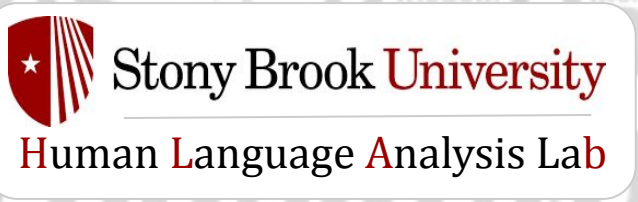
Stony Brook University

- Cognitive States Group - Owen Rambow
- Cognitive Science and Language Learning - Jordan Kodner
- DSL: Data Science Lab - Steven Skiena
- Grammatical Inference - Jeffrey Heinz
- HLAB: Human Language analysis Beings - H. Andrew Schwartz
- Ipseology and Identity Trends - Jason J. Jones
- KALM: Knowledge Authoring Logic Machine - Paul Fodor
- LAIR: Language and AI Research - Ritwik Banerjee
- LUNR: Language Understanding and Reasoning - Niranjan Balasubramanian

AI Beyond Standard Healthcare Data

H. Andrew Schwartz

has@cs.stonybrook.edu



Artificial Intelligence

Psychology + Health

Stony Brook University

Human Language Analysis Beings

Artificial Intelligence

Psychology + Health

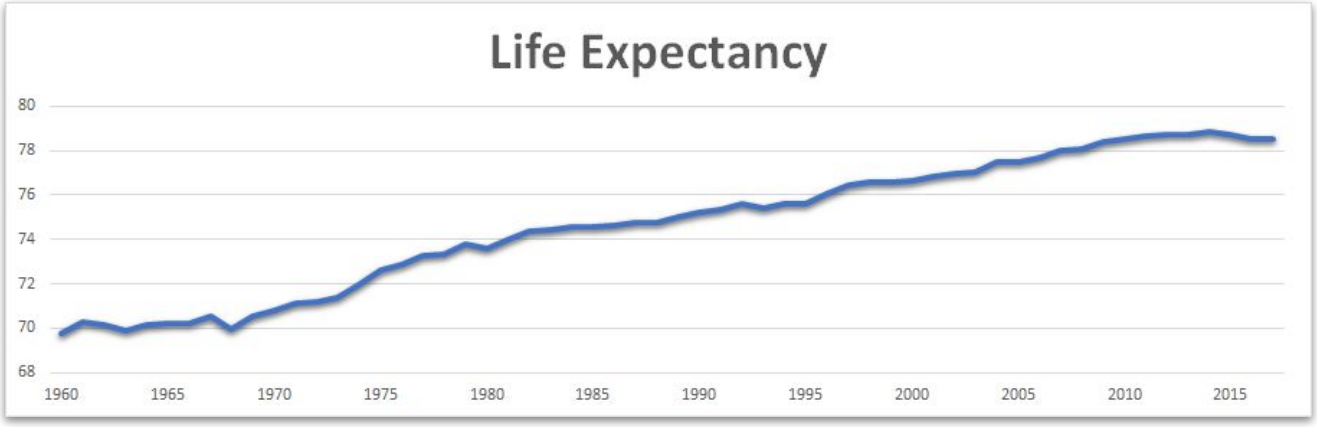
Stony Brook University

Human Language Analysis Beings

Artificial Intelligence

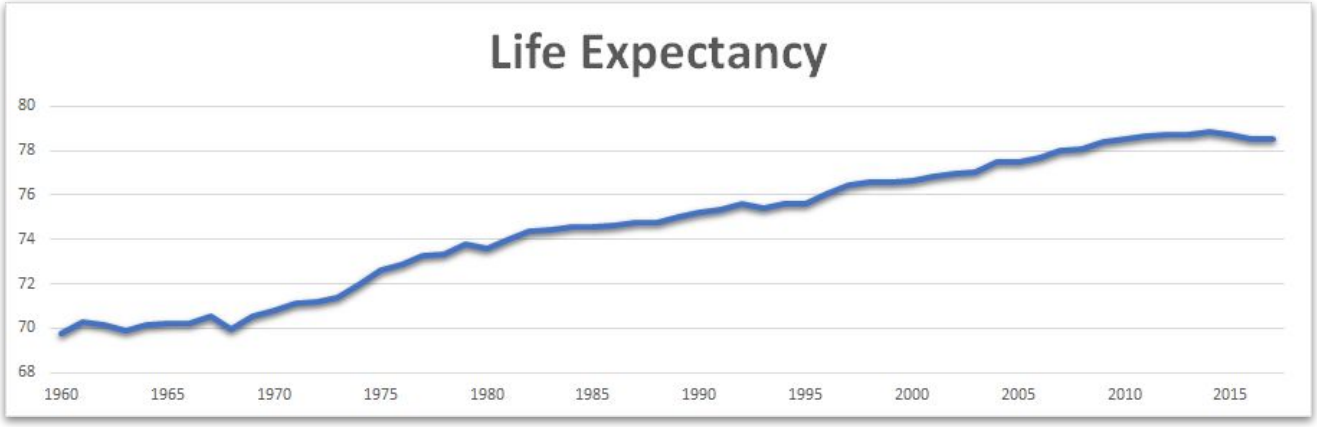
Psychology + Health

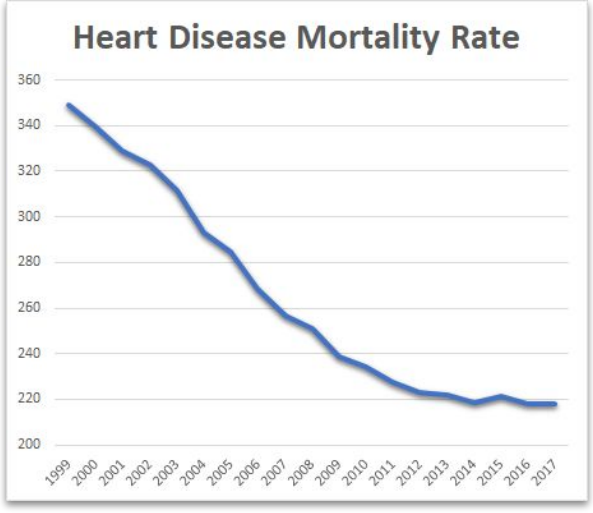
The Health of Individuals is Improving



(Source: US Centers for Disease Control; Age Adjusted Rates)

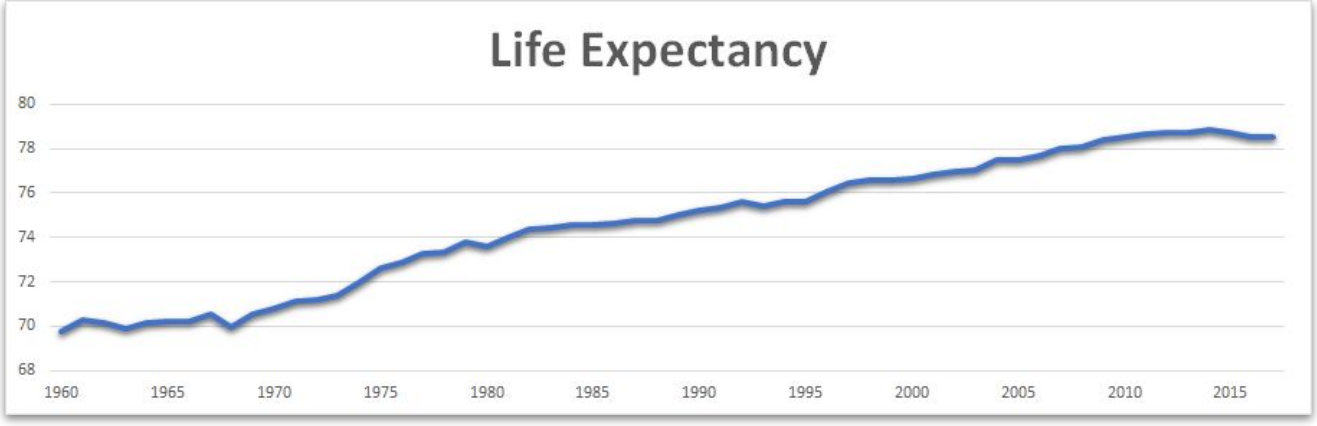
The Health of Individuals is Improving

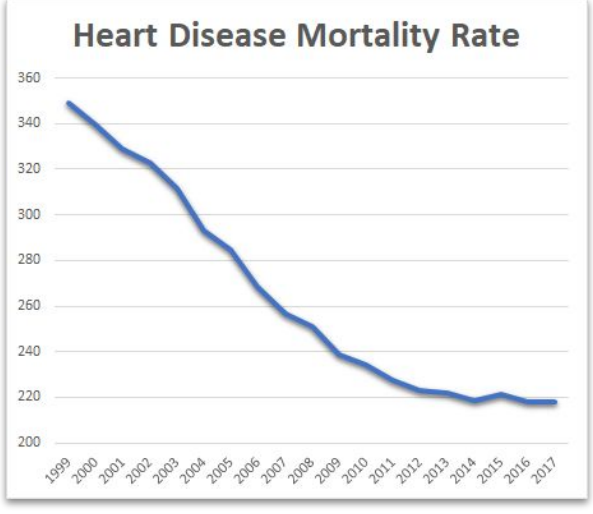


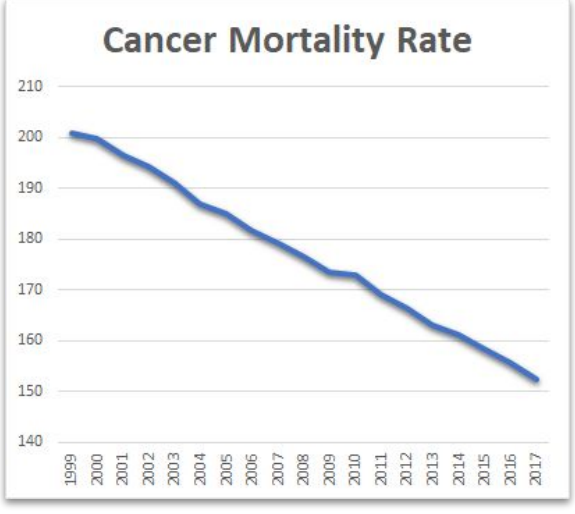


(Source: US Centers for Disease Control; Age Adjusted Rates)

The Health of Individuals is Improving

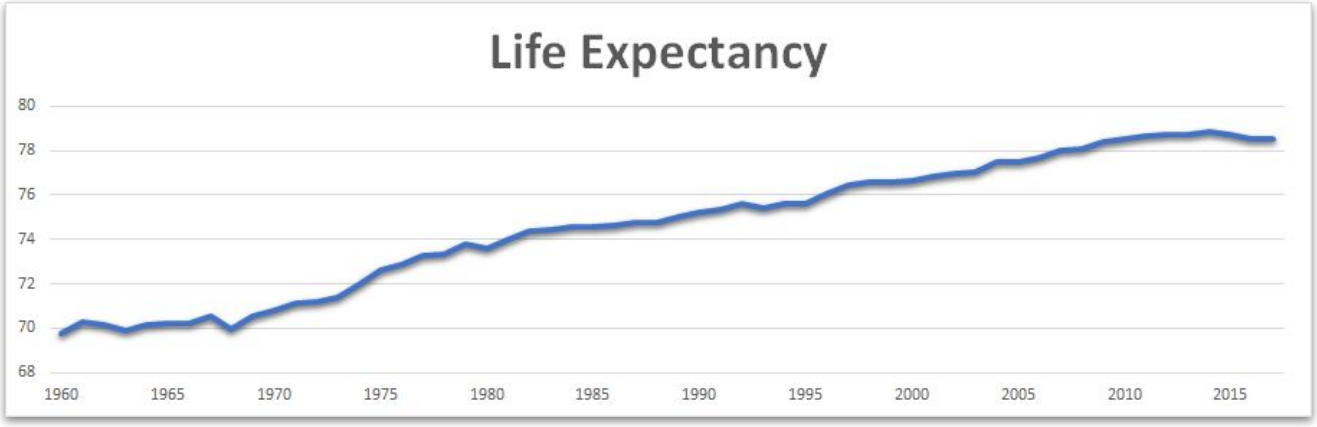


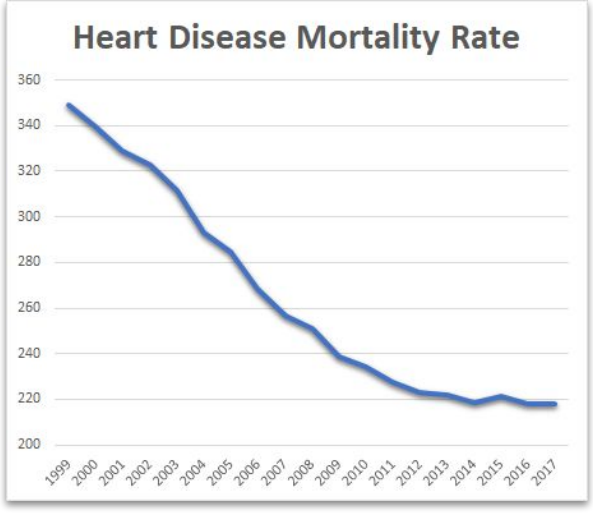


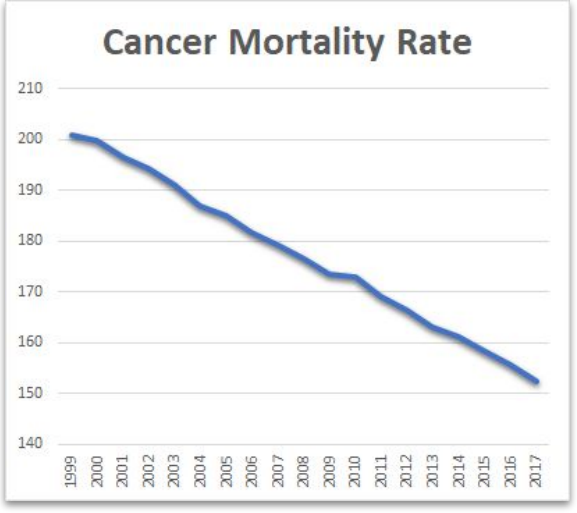


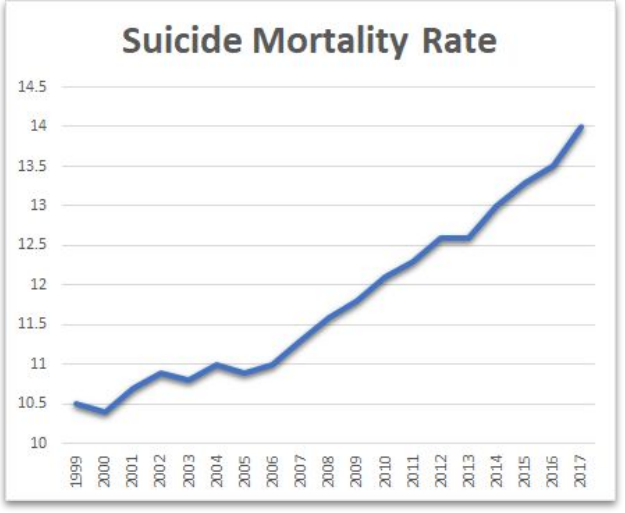
(Source: US Centers for Disease Control; Age Adjusted Rates)

The Health of Individuals is Improving



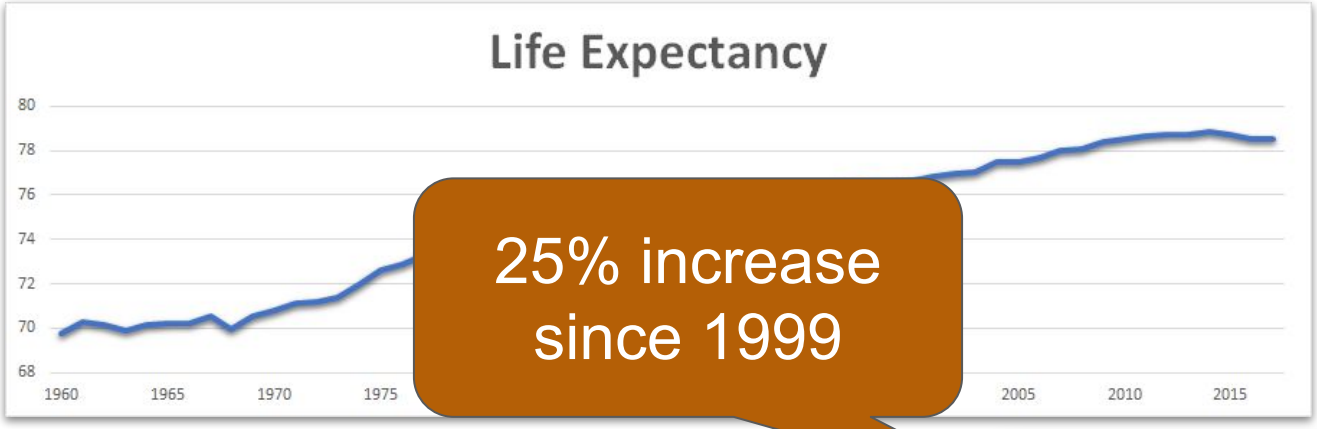


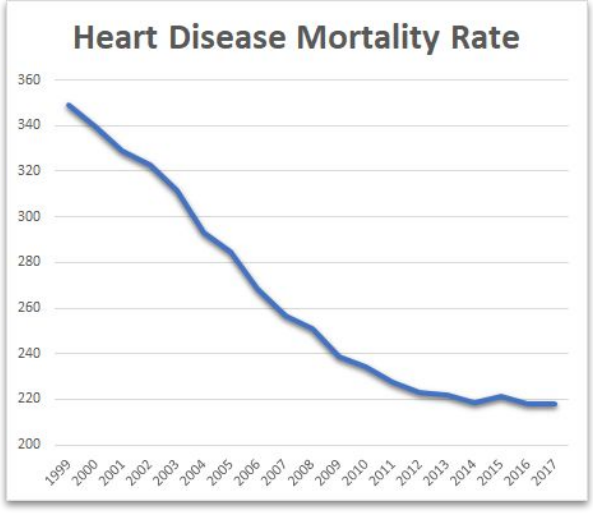


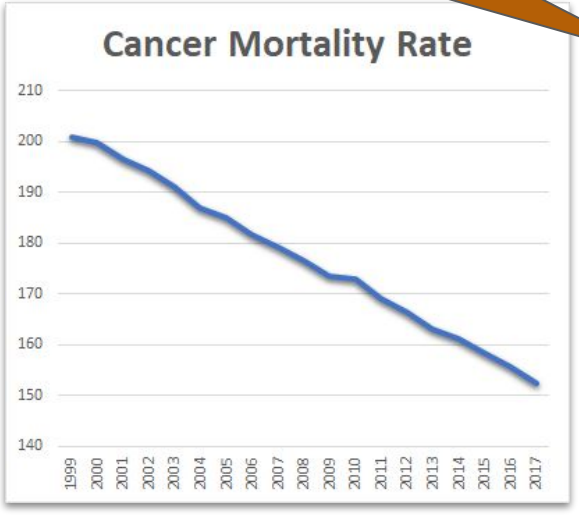


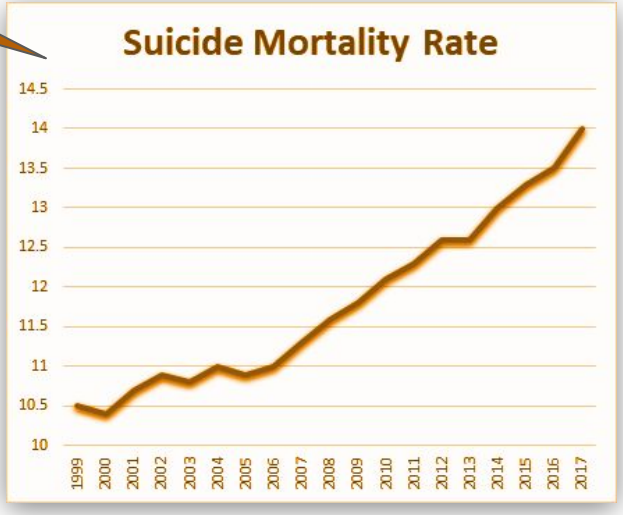
(Source: US Centers for Disease Control; Age Adjusted Rates)

The Health of Individuals is Improving







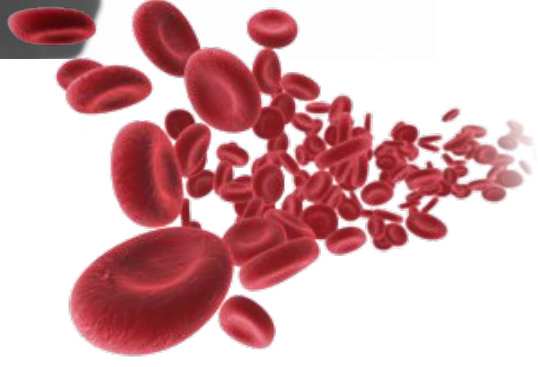


(Source: US Centers for Disease Control; Age Adjusted Rates)

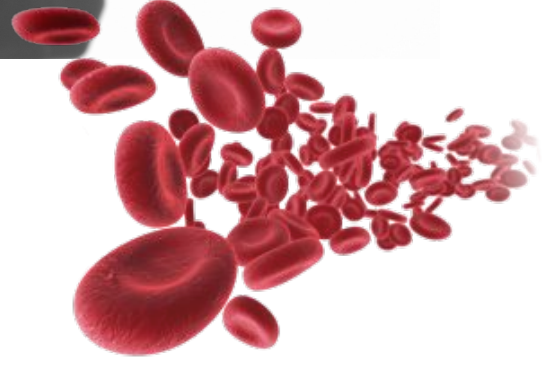
Data in Health Care

Data in Health Care

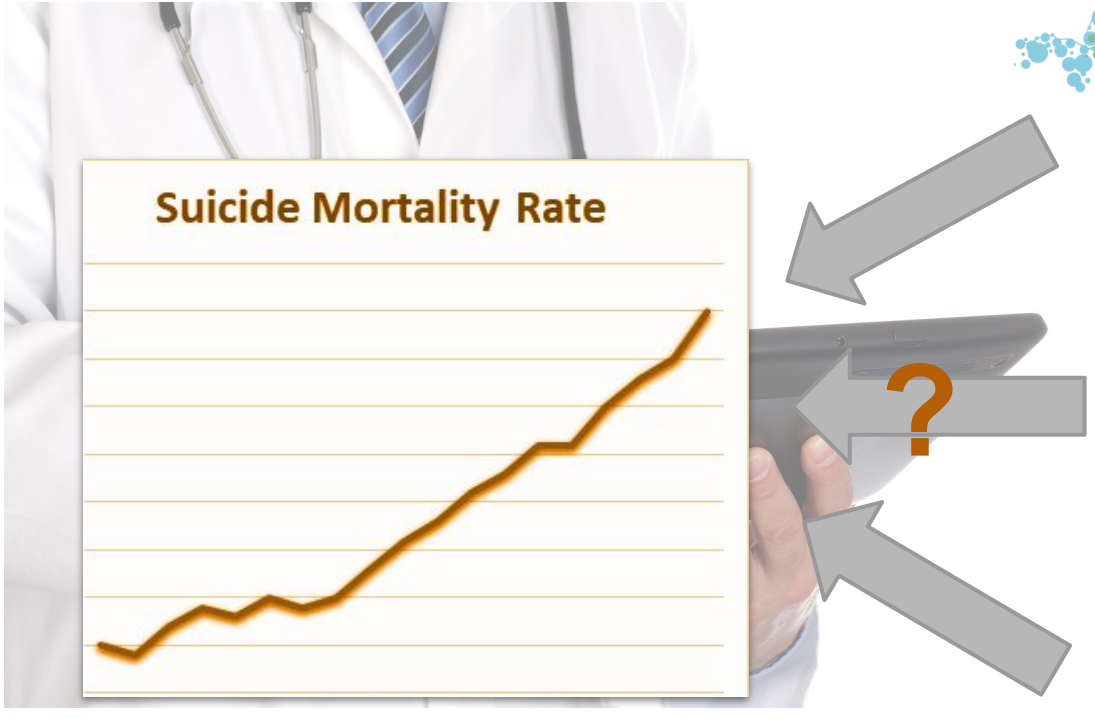
Data in Health Care

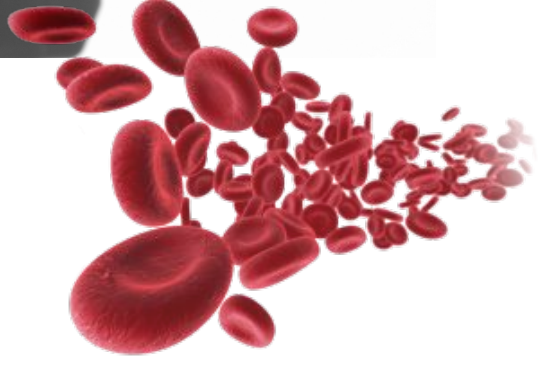


Data in Health Care

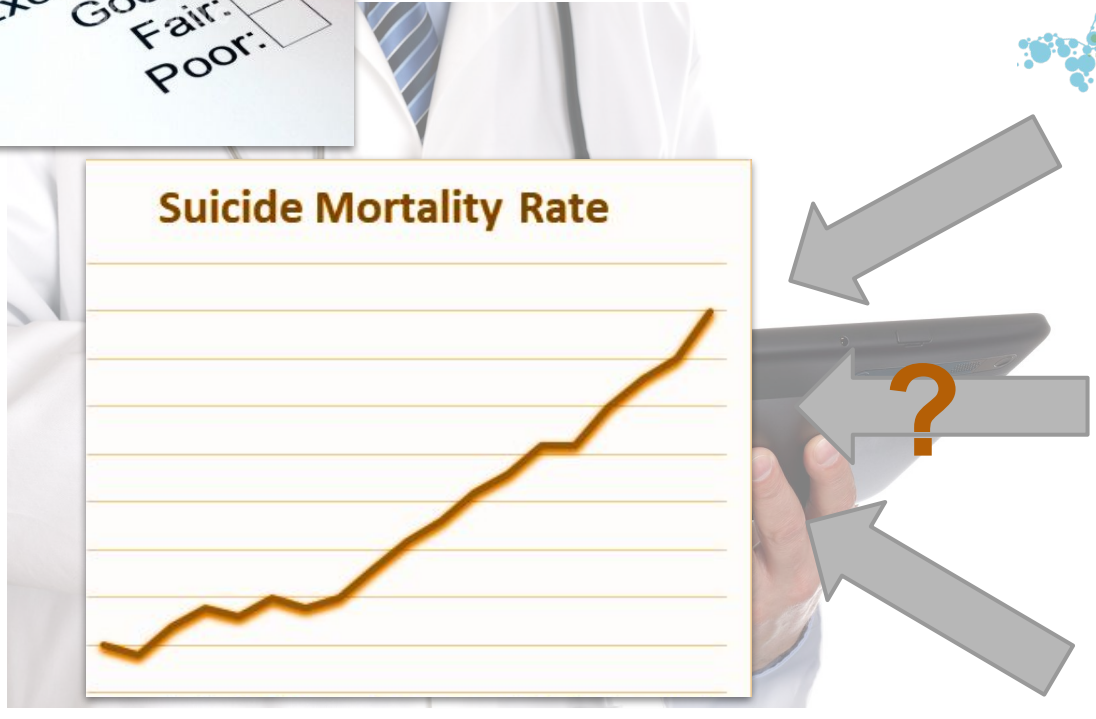


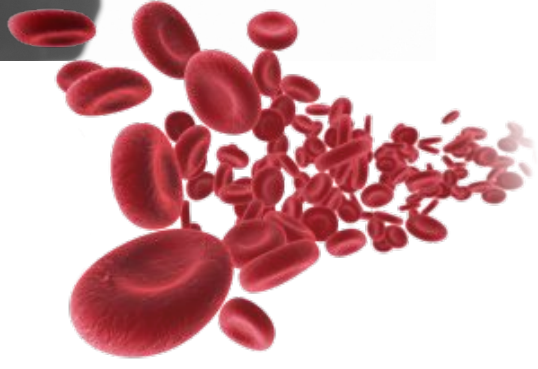
Data in Health Care





Data in Health Care





Data in Health Care

Limited across

- Time - *How frequent?*
- Spatial - *How many people?*
- Conceptual - *What aspects of daily life, who we are?*

Data in Health Care

Limited across

- Time - *How frequent?*
- Spatial - *How many people?*
- Conceptual - *What aspects of daily life, who we are?*

~1960s Technology

Data in Health Care

Limited across

- Time - *How frequent?*
- Spatial - *How many people?*
- Conceptual - *What aspects of daily life, who we are?*

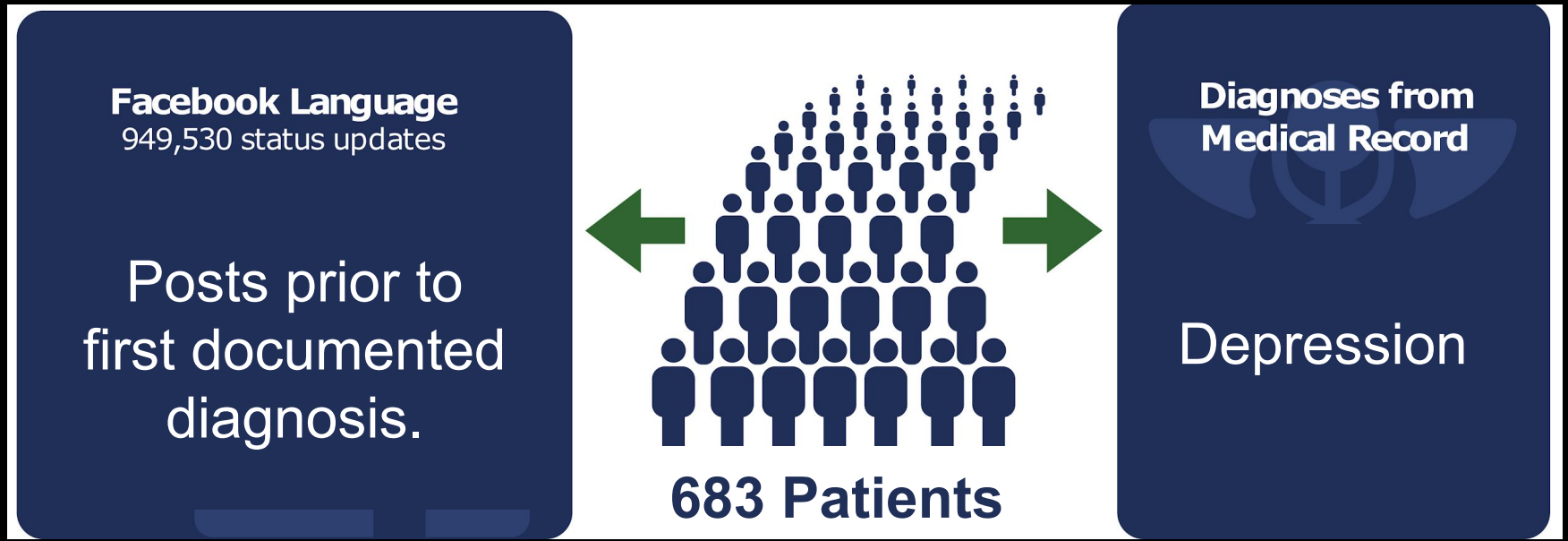
~1960s Technology

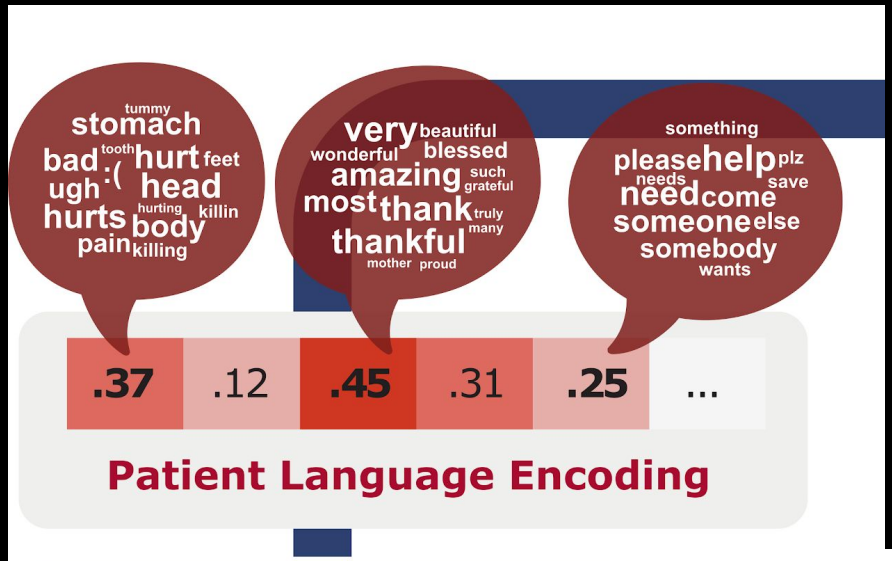
Data in Health Care

Data in Health Care

?

Johannes C. Eichstaedt, Robert J. Smith, Raina M. Merchant, Lyle H. Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A. Asch, **H. Andrew Schwartz**. 2018. Facebook language predicts depression in medical records. ***Proceedings of the National Academy of Sciences.*** DOI:10.1073/pnas.1802331115





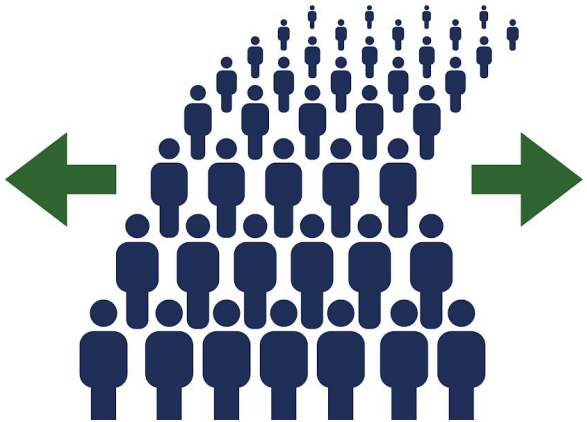
Facebook Language
949,530 status updates

*"Ugh **stomach hurts**, but still goin to the store later. **:(**"*

*"Sh**, **someone help** me!"*

*"I am **blessed** to spend all day with my daughter"*

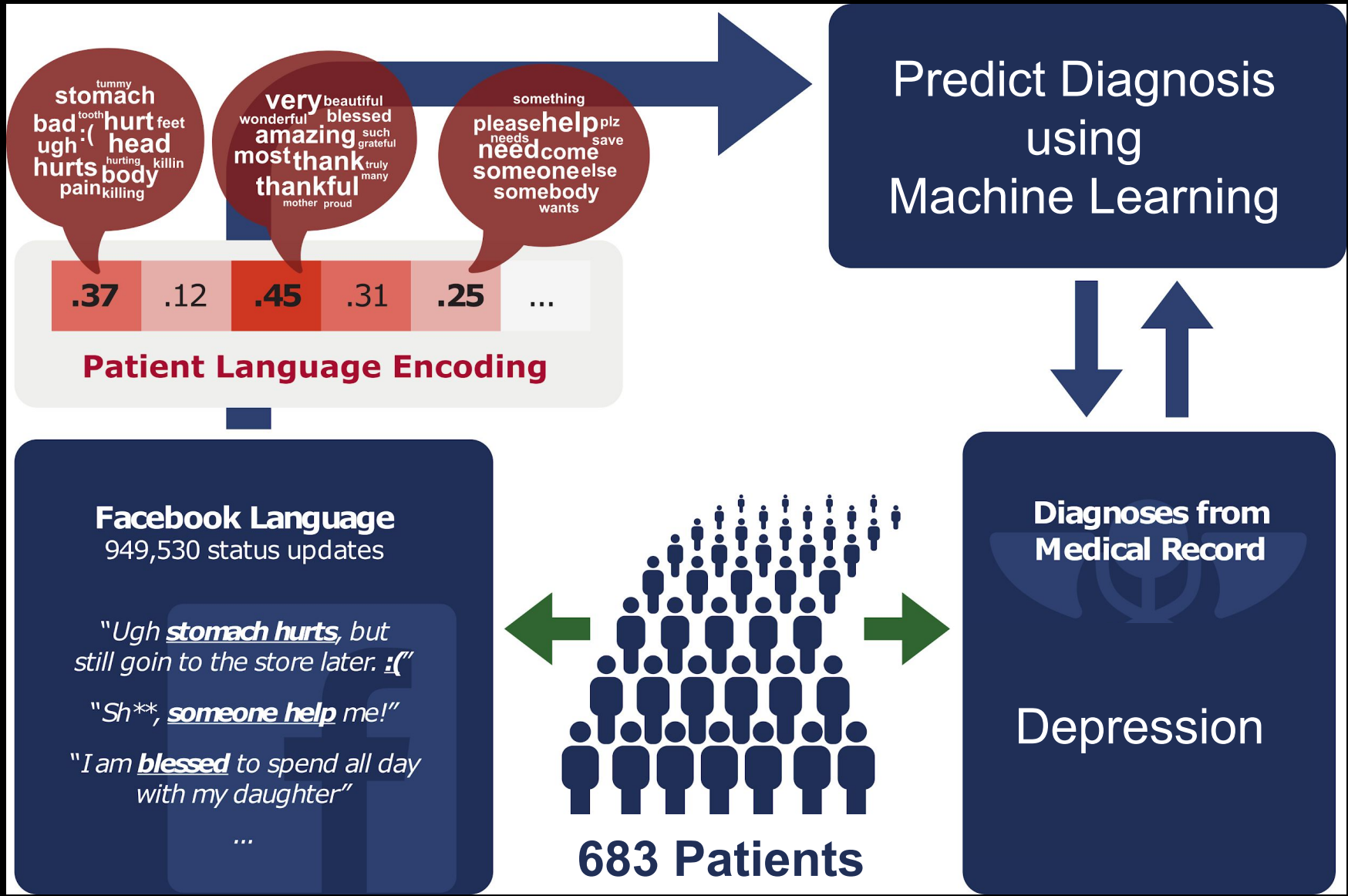
...

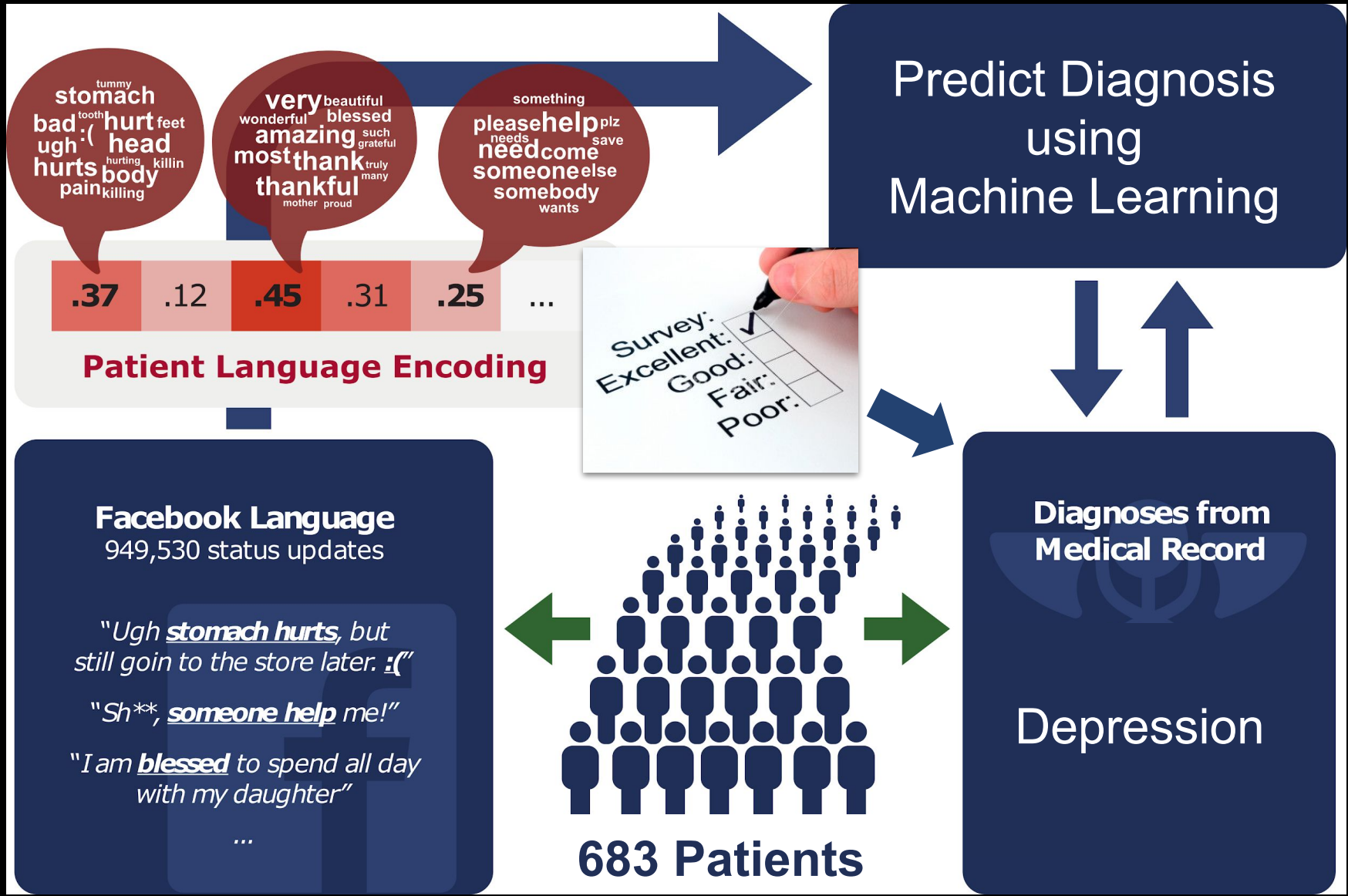


683 Patients

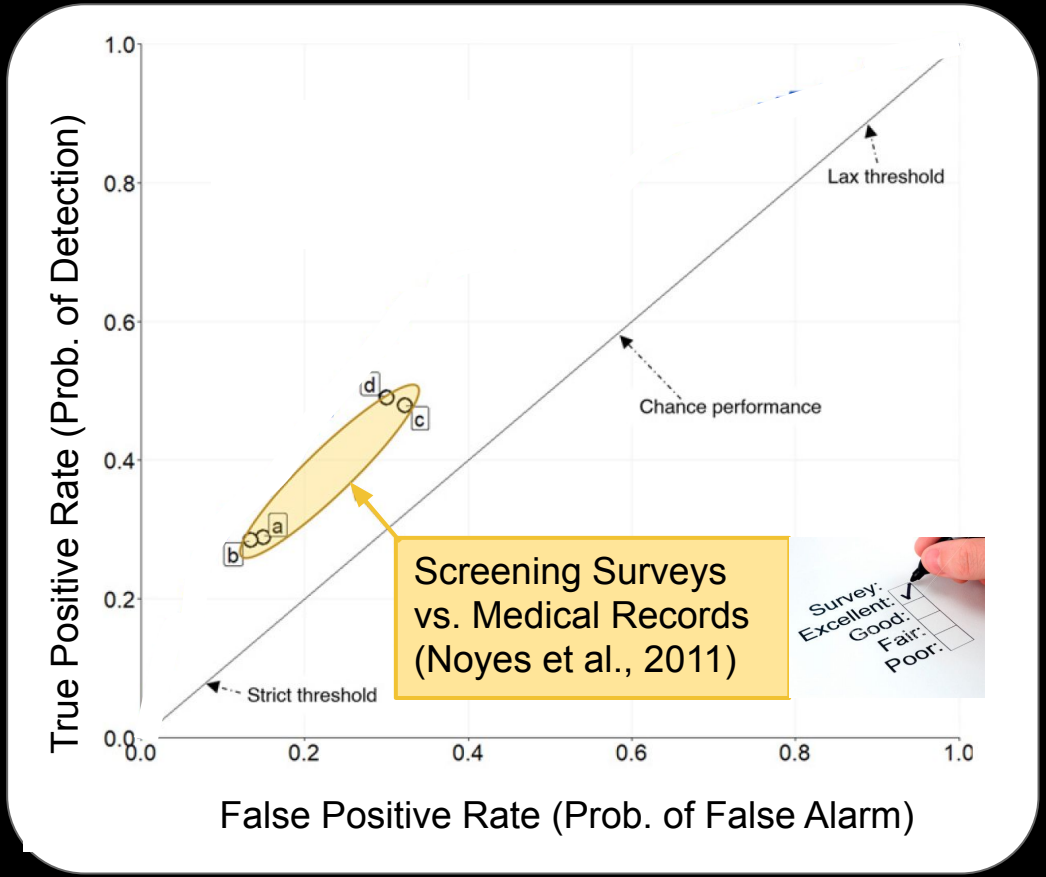
Diagnoses from Medical Record

Depression


tummy
stomach
bad
tooth
hurt
feet
ugh
:(
head
hurts
hurting
killin
body
pain
killing
very
beautiful
wonderful
blessed
amazing
such
grateful
most
thank
truly
many
thankful
mother
proud
something
please
help
plz
needs
save
need
come
someone
else
somebody
wants
.37
.12
.45
.31
.25
...
Patient Language Encoding
Predict Diagnosis using Machine Learning
Facebook Language
949,530 status updates
"Ugh stomach hurts, but still goin to the store later. :("
"Sh**, someone help me!"
"I am blessed to spend all day with my daughter"
...
683 Patients
Diagnoses from Medical Record
Depression

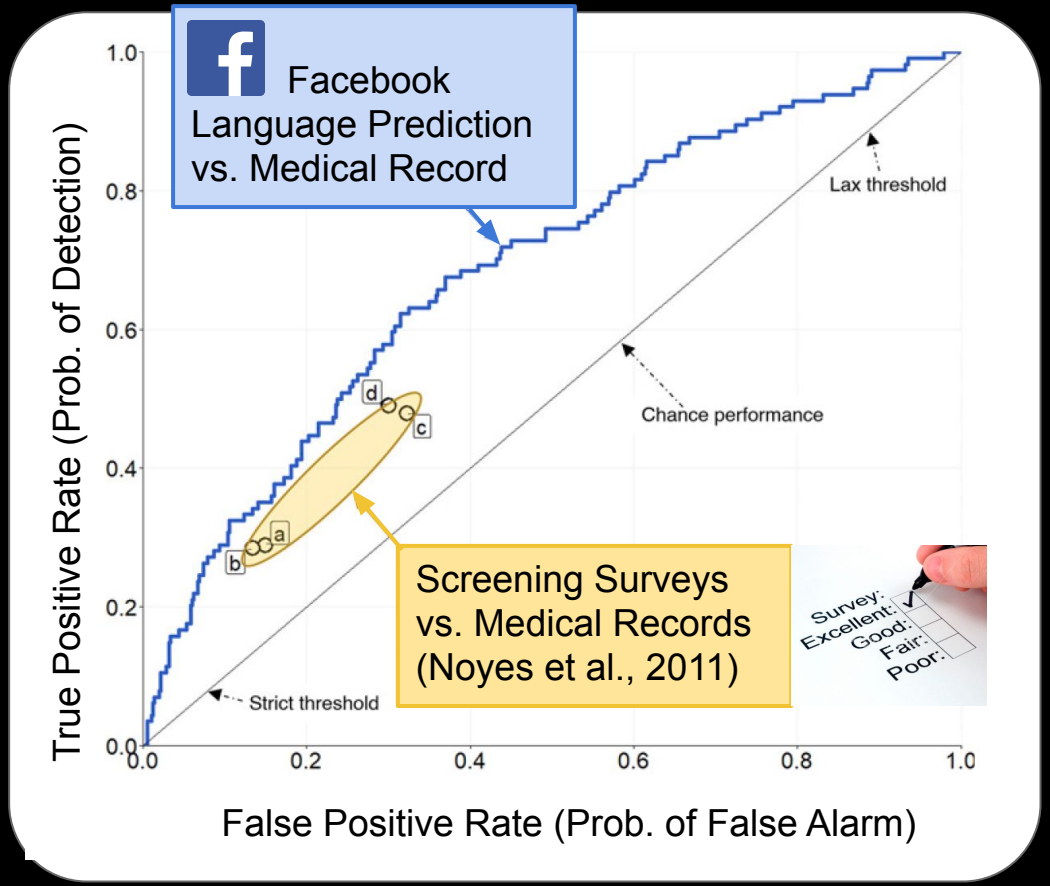

tummy
stomach
bad
tooth
:(
hurt
feet
ugh
head
hurts
hurting
killin
body
pain
killing
very
beautiful
wonderful
blessed
amazing
such
grateful
most
thank
truly
many
thankful
mother
proud
something
please
help
plz
needs
save
need
come
someone
else
somebody
wants
.37
.12
.45
.31
.25
...
Patient Language Encoding
Predict Diagnosis using Machine Learning
Survey:
Excellent:
Good:
Fair:
Poor:
Facebook Language
949,530 status updates
"Ugh stomach hurts, but still goin to the store later. :("
"Sh**, someone help me!"
"I am blessed to spend all day with my daughter"
...
683 Patients
Diagnoses from Medical Record
Depression

Prediction of Depression Diagnosis



Johannes C. Eichstaedt, Robert J. Smith, Raina M. Merchant, Lyle H. Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A. Asch, **H. Andrew Schwartz**. 2018. **Facebook language predicts depression in medical records.** ***Proceedings of the National Academy of Sciences.*** DOI:10.1073/pnas.1802331115

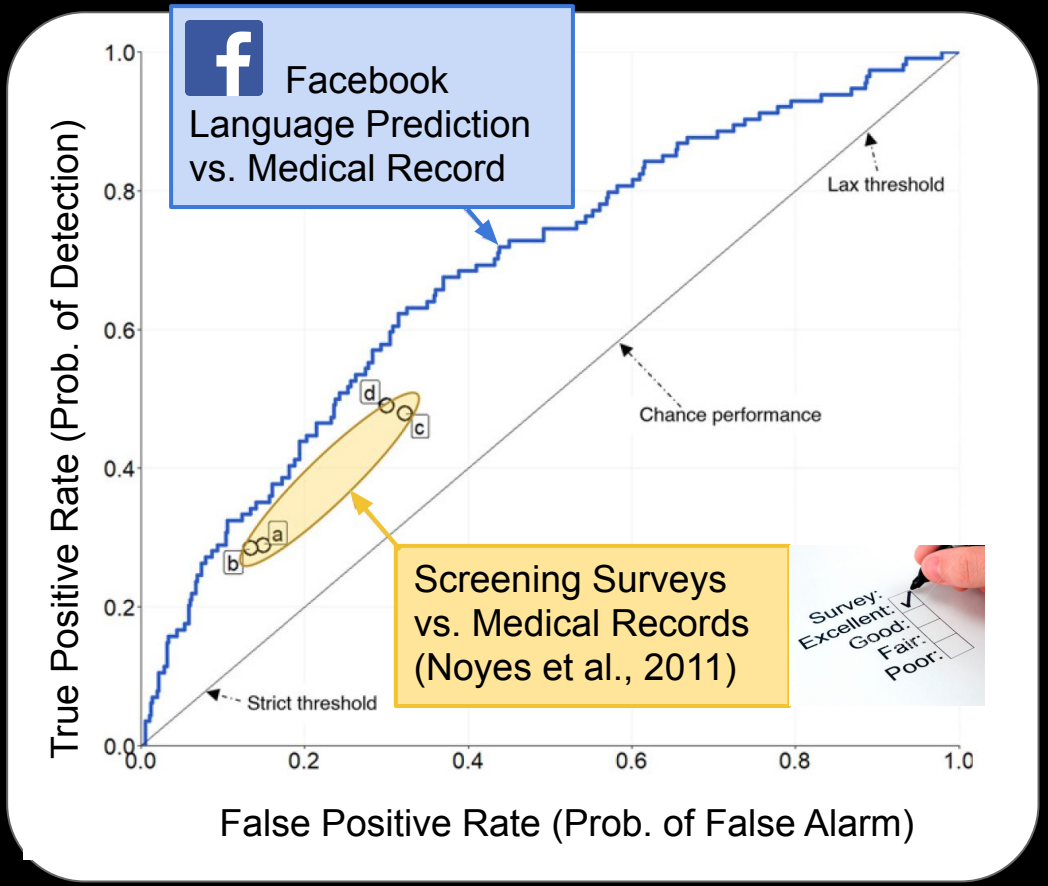
Prediction of Depression Diagnosis



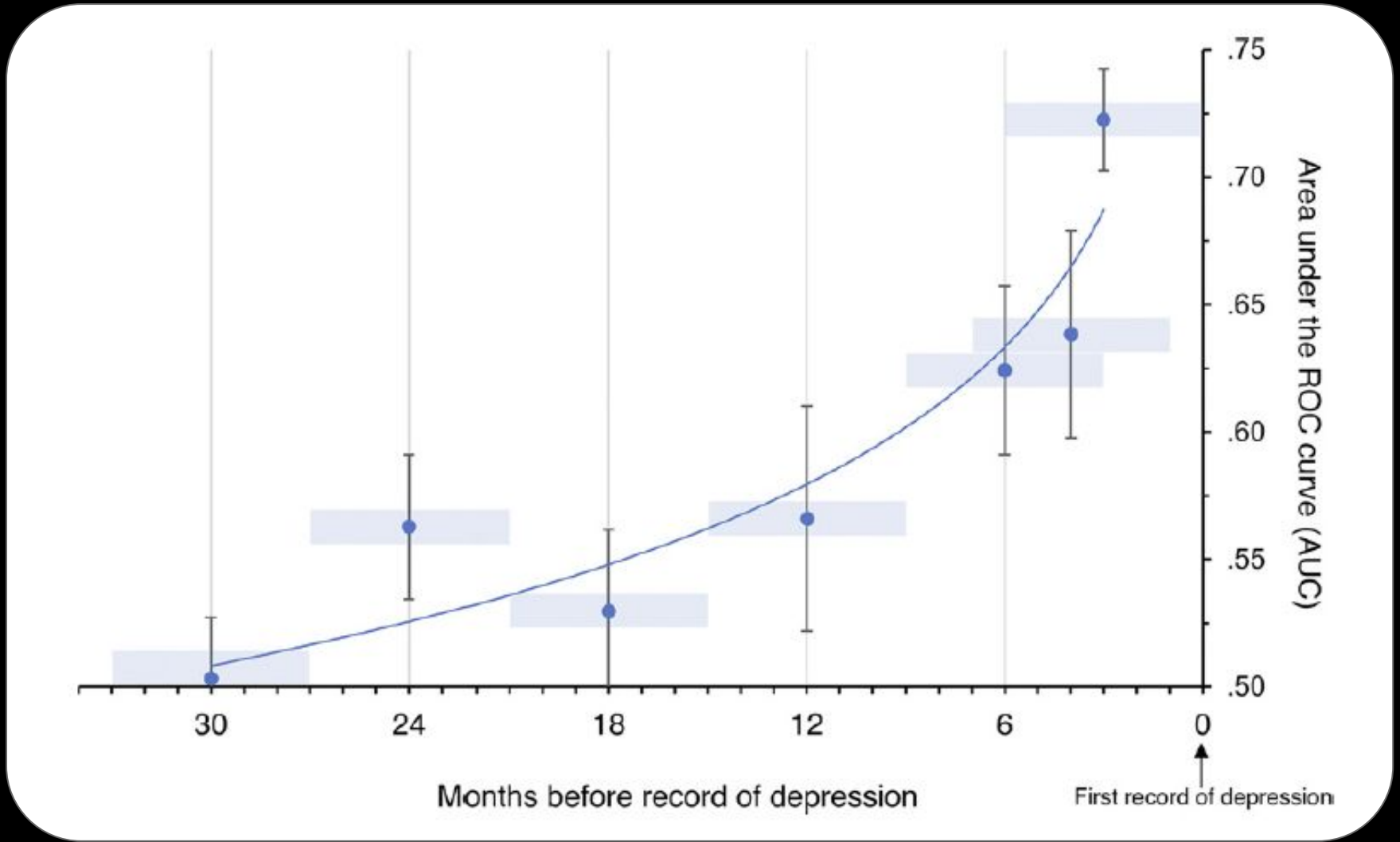
Johannes C. Eichstaedt, Robert J. Smith, Raina M. Merchant, Lyle H. Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A. Asch, **H. Andrew Schwartz**. 2018. **Facebook language predicts depression in medical records.** ***Proceedings of the National Academy of Sciences.*** DOI:10.1073/pnas.1802331115

Prediction of Depression Diagnosis

Topics that predict depression

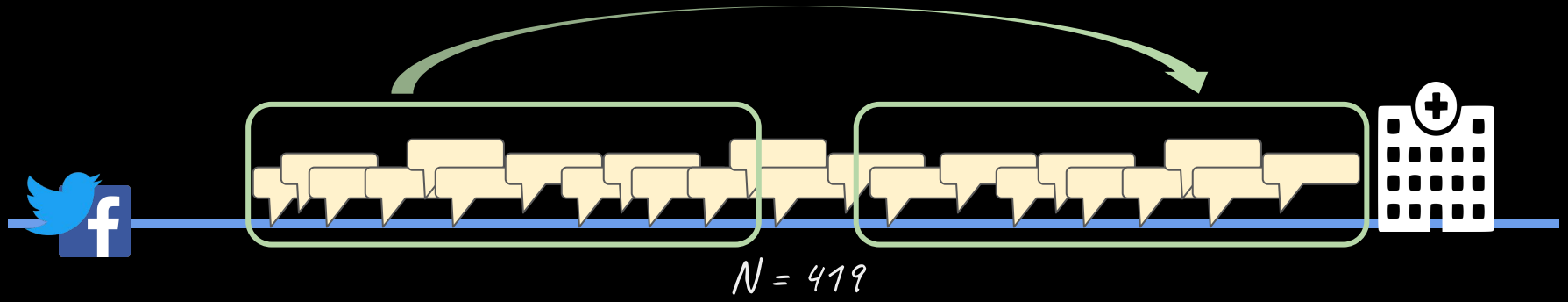


Johannes C. Eichstaedt, Robert J. Smith, Raina M. Merchant, Lyle H. Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A. Asch, **H. Andrew Schwartz**. 2018. **Facebook language predicts depression in medical records.** ***Proceedings of the National Academy of Sciences.*** DOI:10.1073/pnas.1802331115



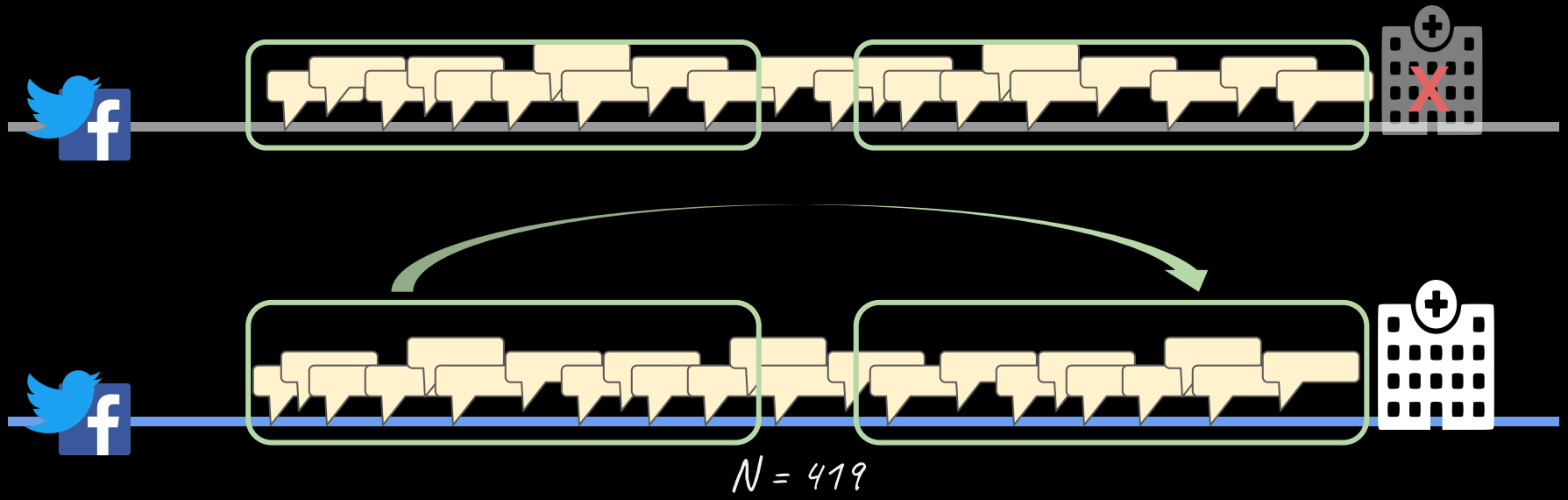
Eichstaedt, … Schwartz, 2018. Proceedings of the National Academy of Sciences.

Social media language change prior to hospital visit



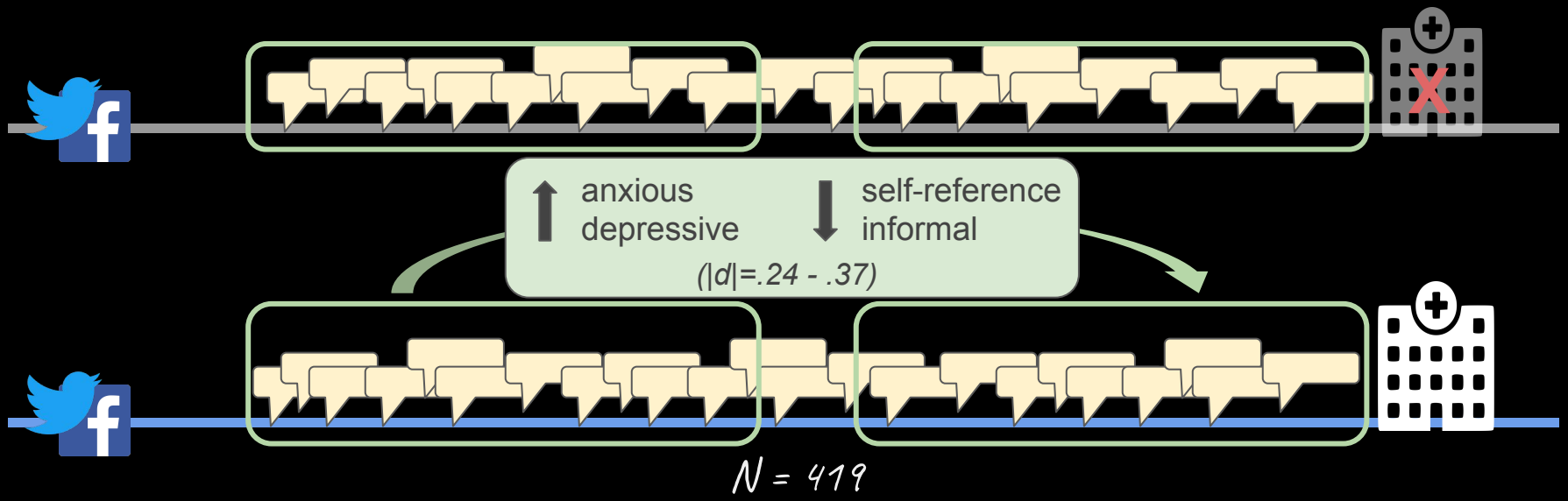
Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1),* 1-9.

Social media language change prior to hospital visit



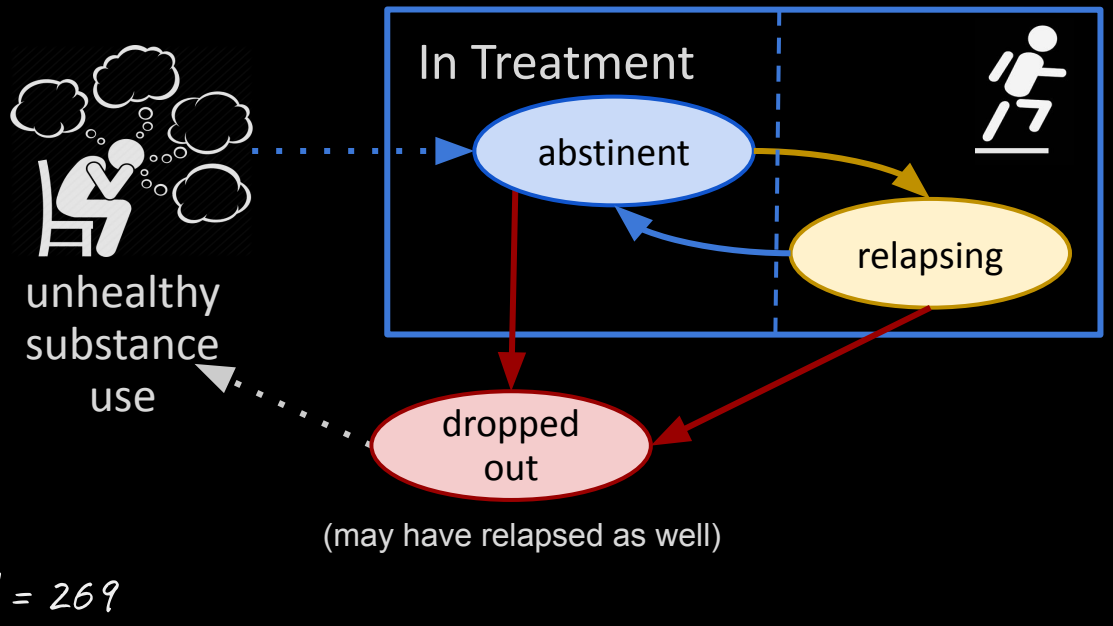
Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1),* 1-9.

Social media language change prior to hospital visit



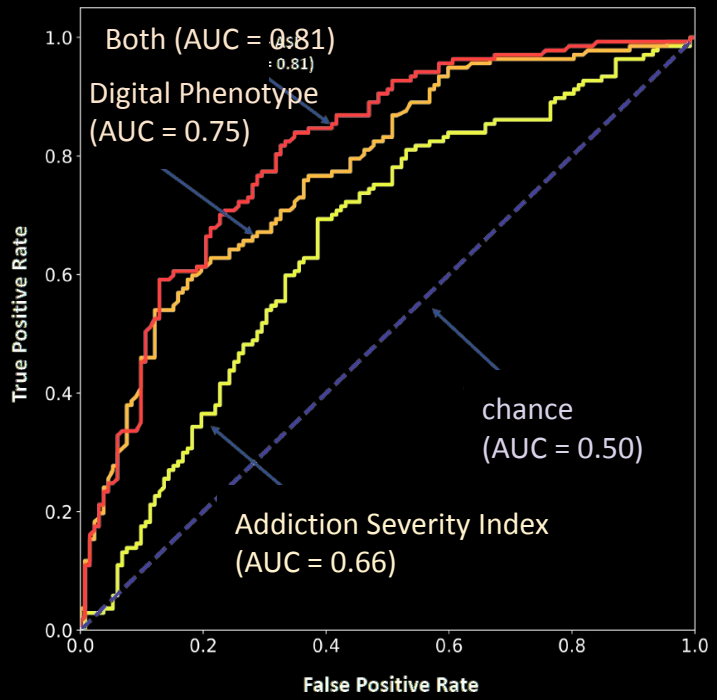
Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1),* 1-9.

Addiction Treatment Outcome Risk Assessment

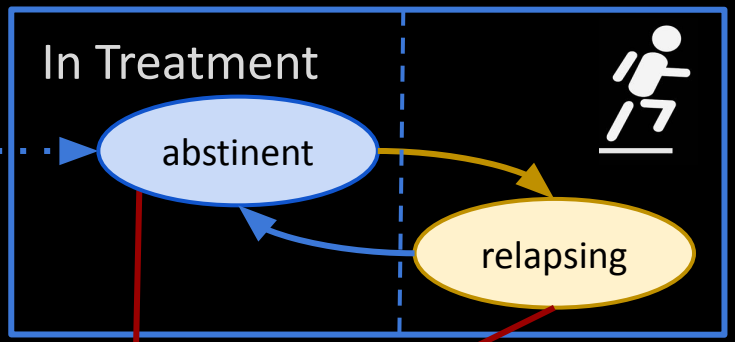


Curtis, B., Giorgi, S., Ungar, L., Vu, H., Yaden, D., Liu, T., ... & Schwartz, H. A. (2023). AI-based analysis of social media language predicts addiction treatment dropout at 90 days. *Neuropsychopharmacology*, *48*(11), 1579-1585.

Addiction Treatment Outcome Risk Assessment



unhealthy substance use

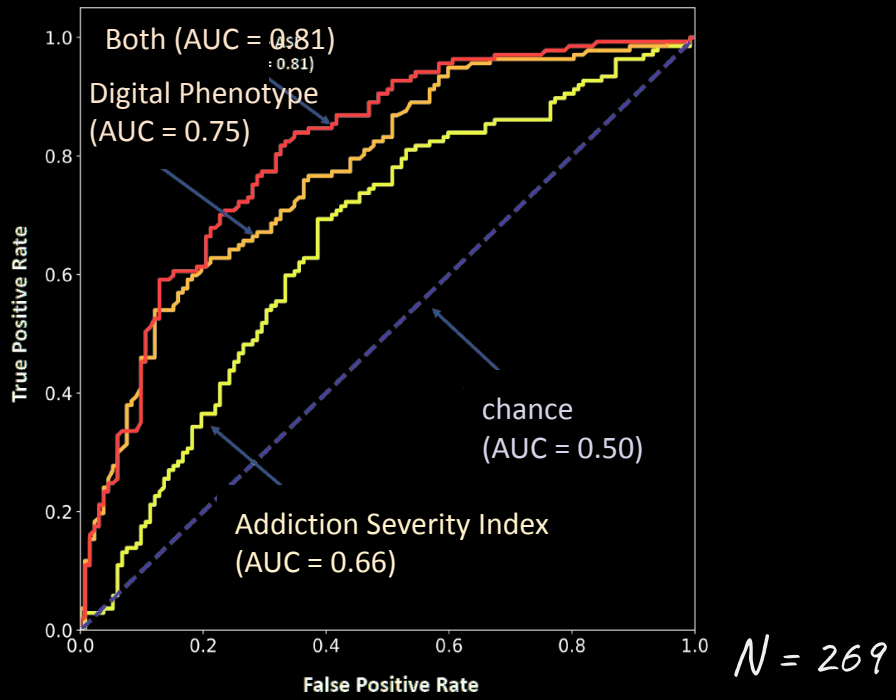


(may have relapsed as well)

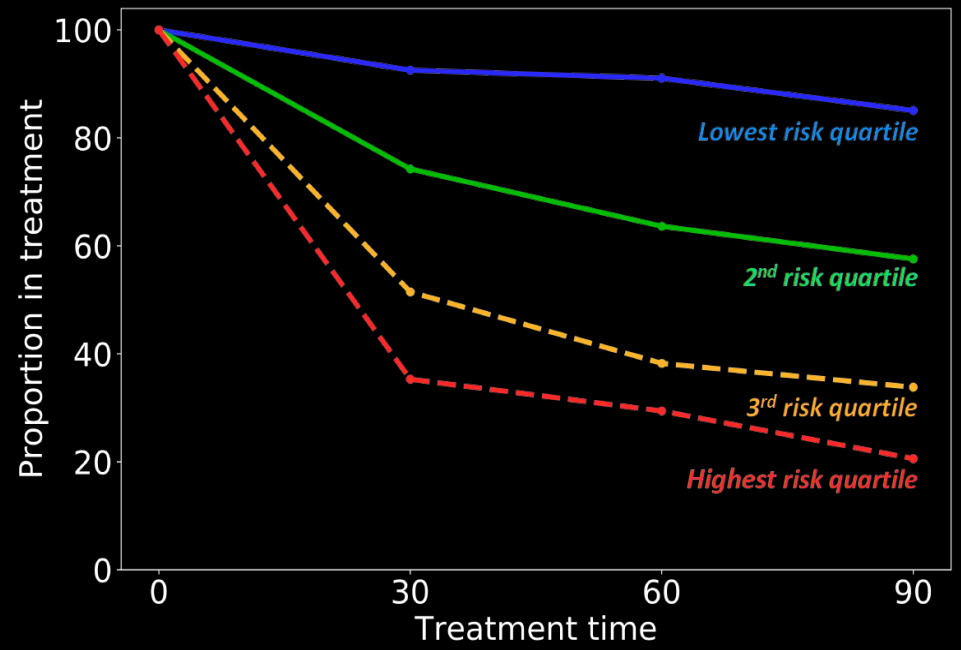
N = 269

Curtis, B., Giorgi, S., Ungar, L., Vu, H., Yaden, D., Liu, T., ... & Schwartz, H. A. (2023). AI-based analysis of social media language predicts addiction treatment dropout at 90 days. *Neuropsychopharmacology*, *48*(11), 1579-1585.

Addiction Treatment Outcome Risk Assessment



N = 269

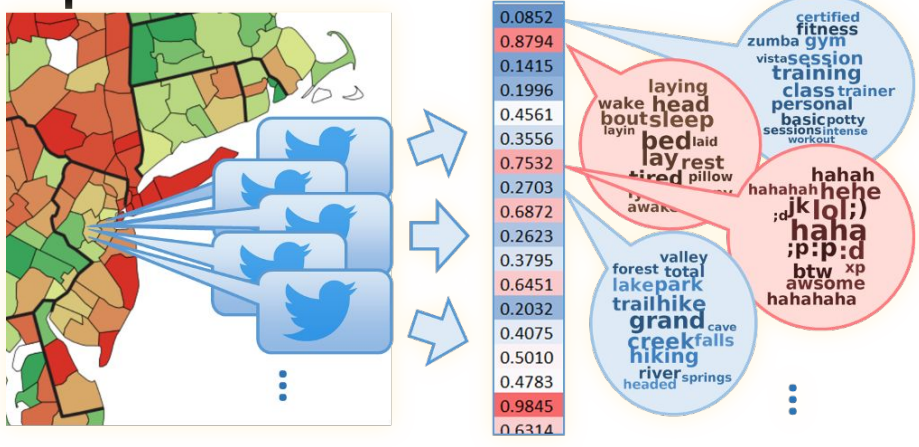


Curtis, B., Giorgi, S., Ungar, L., Vu, H., Yaden, D., Liu, T., ... & Schwartz, H. A. (2023). AI-based analysis of social media language predicts addiction treatment dropout at 90 days. *Neuropsychopharmacology*, *48*(11), 1579-1585.

1. Digital footprints

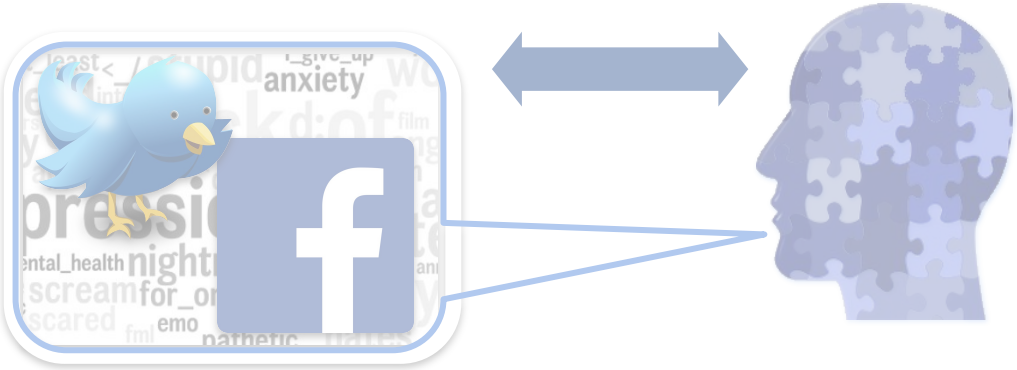
AI for Mental Health

2. Populations



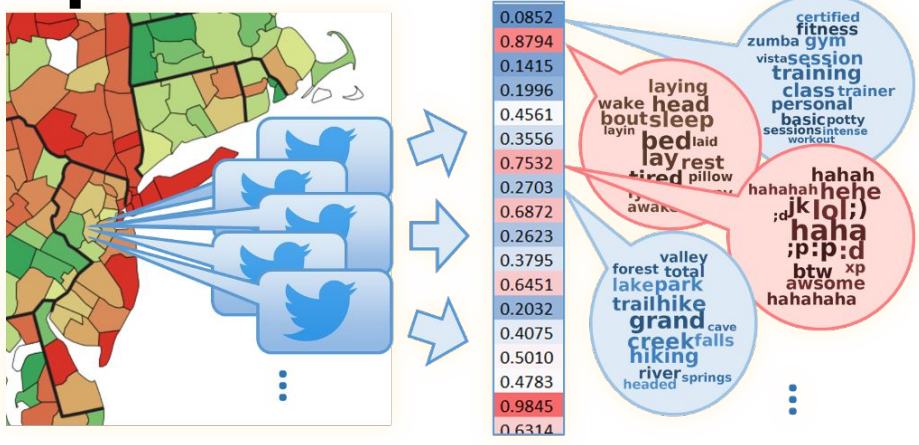
3. Automated Interviews

1. Digital footprints



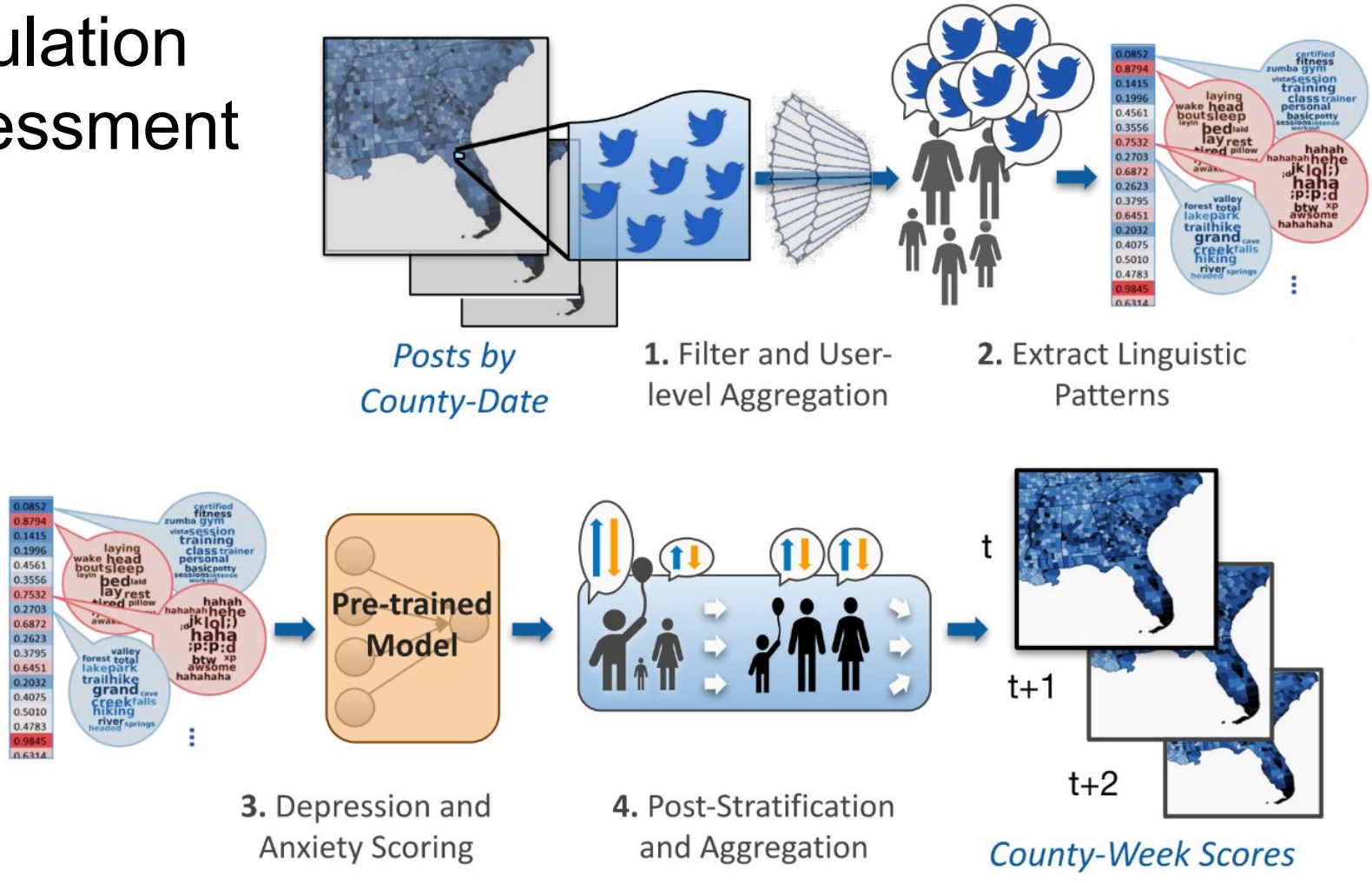
AI for Mental Health

2. Populations

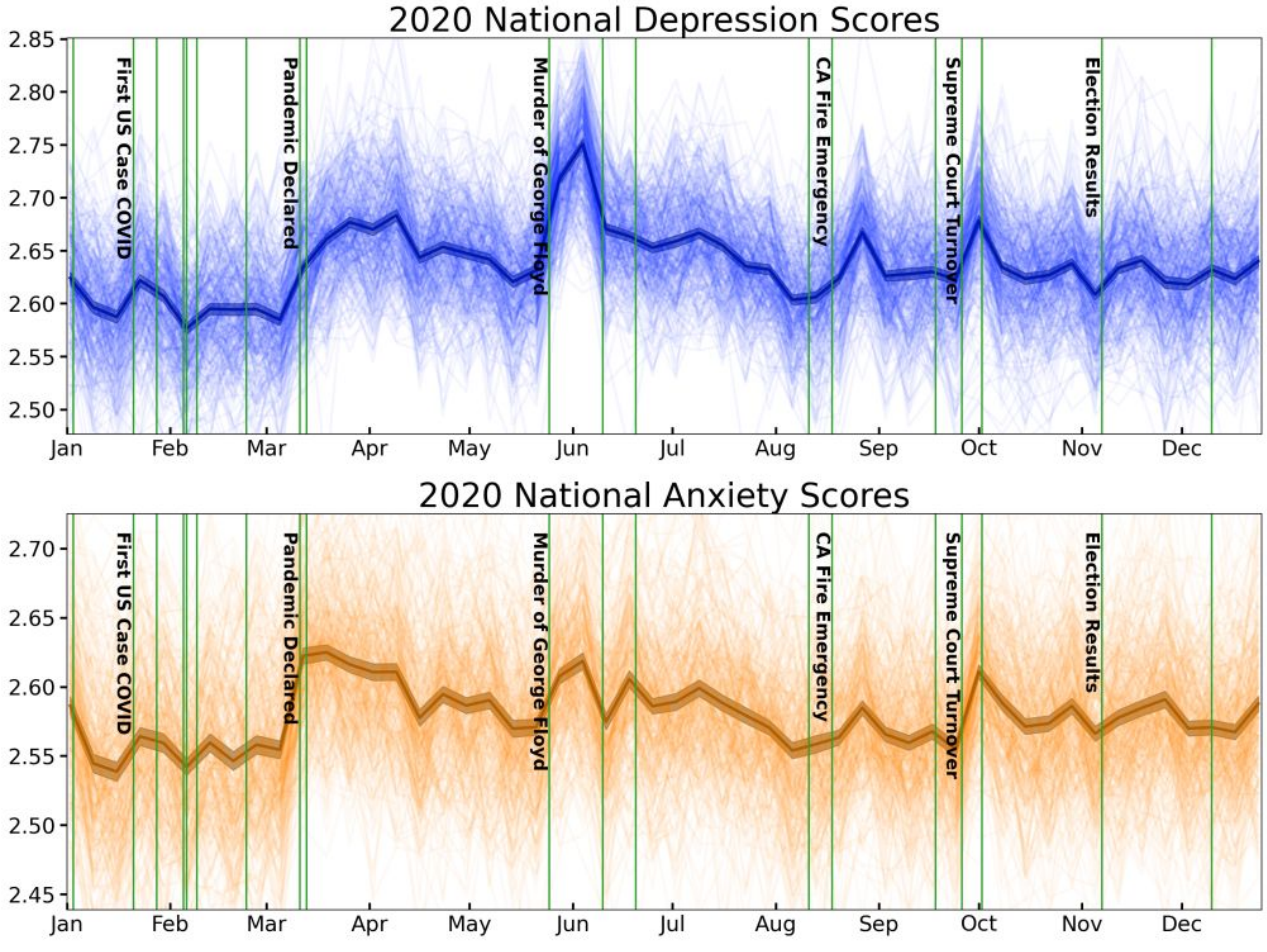


3. Automated Interviews

Population Assessment



Mangalik, S., Eichstaedt, J. C., Giorgi, S., Mun, J., Ahmed, F., Gill, G., ... & Schwartz, H. A. (2024). Robust language-based mental health assessments in time and space through social media. *npj Digital Medicine*, 7(1), 109.



(a)

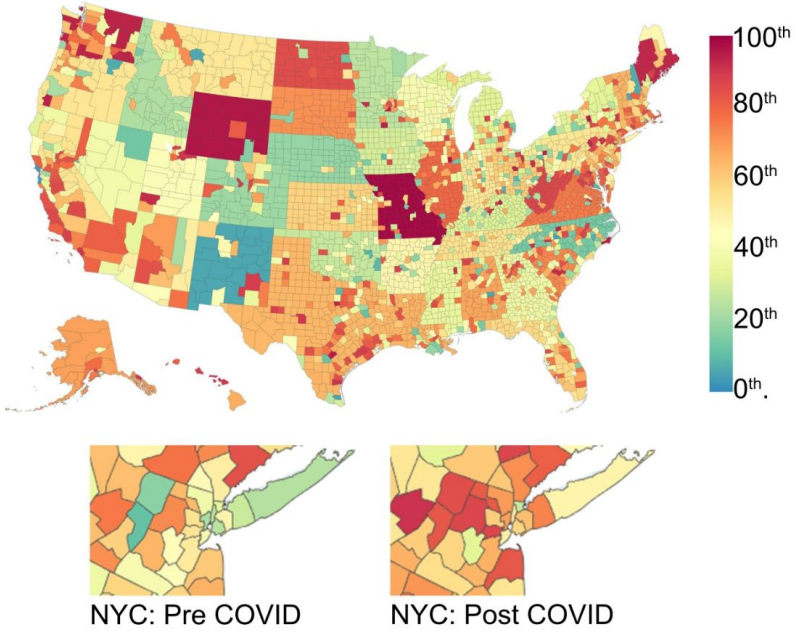
(b)

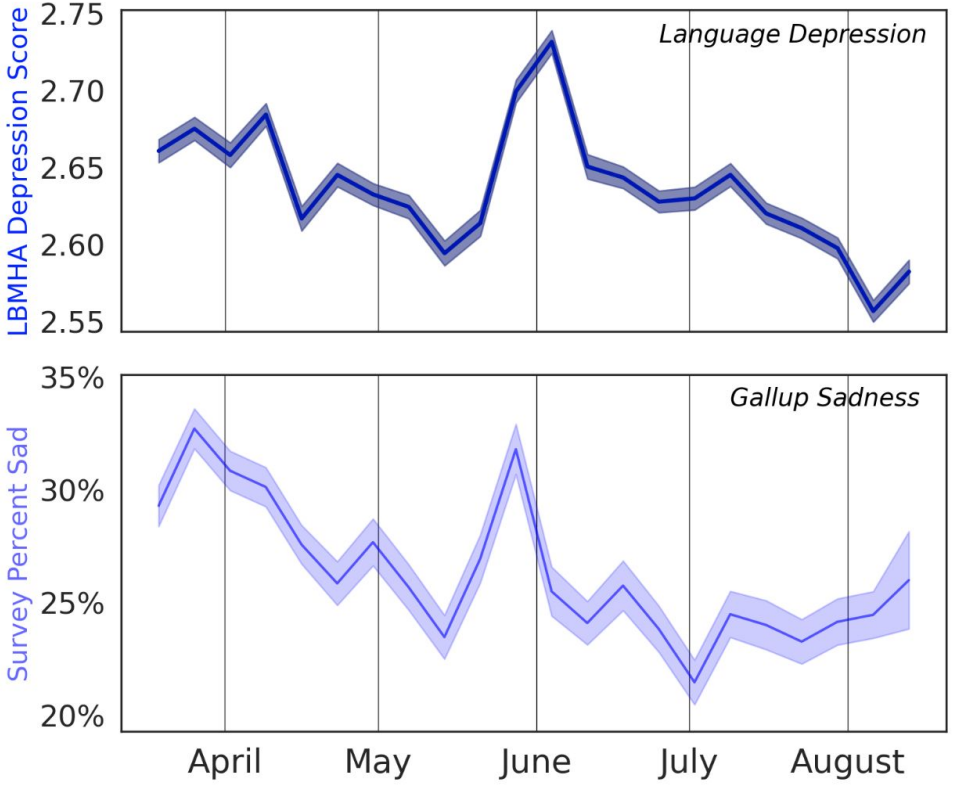
Mangalik, S., Eichstaedt, J. C., Giorgi, S., Mun, J., Ahmed, F., Gill, G., ... & Schwartz, H. A. (2024). Robust language-based mental health assessments in time and space through social media. *npj Digital Medicine*, 7(1), 109.

Population Assessment

Space (N)	Time (N)	Depression β	Anxiety β
National (1)	Weeks (22)	$0.763^{\dagger}$	$1.823^{\ddagger}$
Regions (4)	Weeks (22)	$0.759^{\ddagger}$	$1.817^{\ddagger}$
Counties (132)	Quarters (3)	$0.681^{\ddagger}$	$1.423^{\ddagger}$
Counties (132)	Weeks (22)	$0.410^{\ddagger}$	$0.343^{\ddagger}$

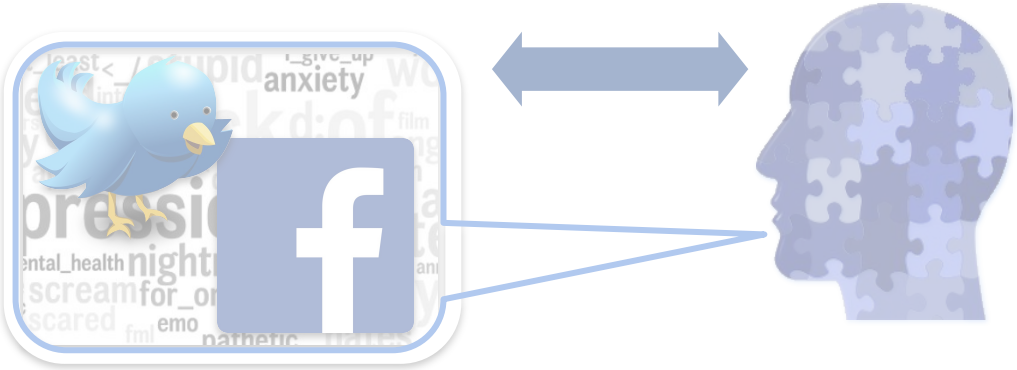
fixed effects betas





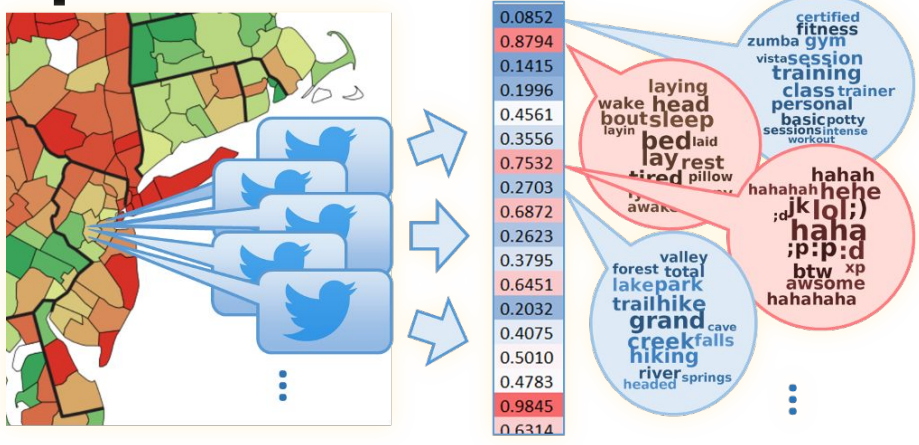
Mangalik, S., Eichstaedt, J. C., Giorgi, S., Mun, J., Ahmed, F., Gill, G., ... & Schwartz, H. A. (2024). Robust language-based mental health assessments in time and space through social media. *npj Digital Medicine*, 7(1), 109.

1. Digital footprints



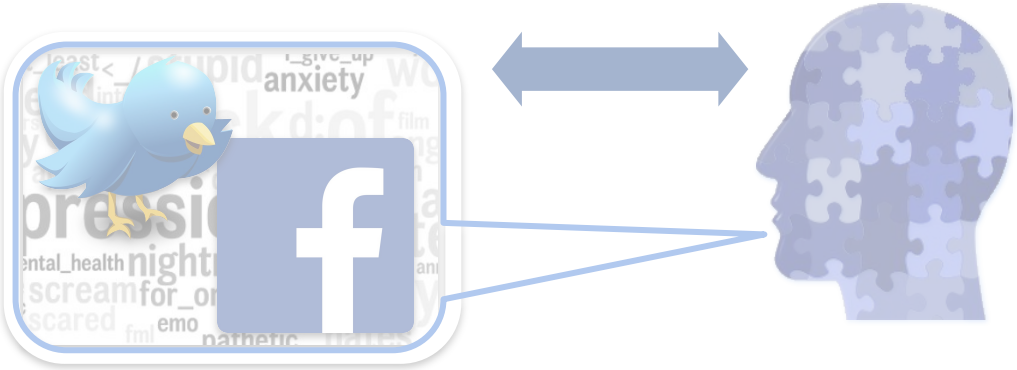
AI for Mental Health

2. **Population**



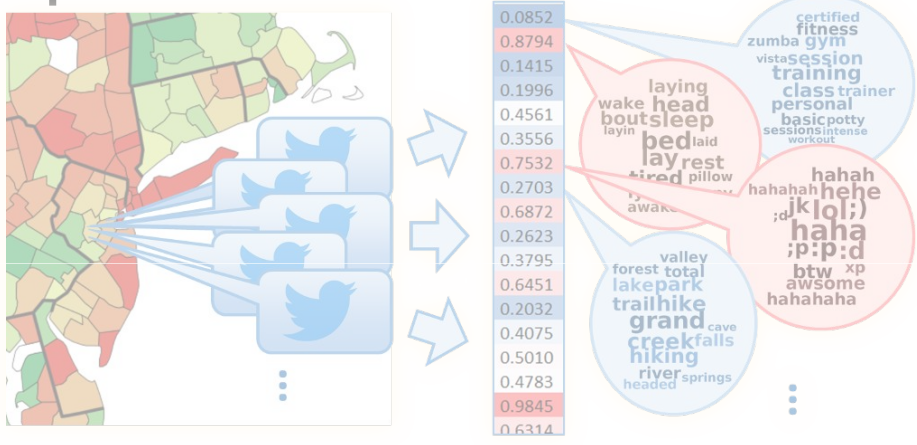
3. Automated Interviews

1. Digital footprints



AI for Mental Health

2. Population



3. Automated Interviews

r = 0.65 to 0.85

rating-scale assessments

"reliability"

Good old-fashion AI

r = 0.4 to 0.60

r = 0.65 to 0.85

rating-scale assessments

"reliability"

accepted psychological scores

Good old-fashion AI

r = 0.4 to 0.60

the theoretical upper-limit!

r = 0.65 to 0.85

rating-scale assessments

"reliability"

accepted psychological scores

Kjell, O. N., Sikström, S., Kjell, K., & Schwartz, H. A. (2022). **Natural language analyzed with AI-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy.** *Scientific reports*, *12*(1), 1-9.

Good old-fashion AI

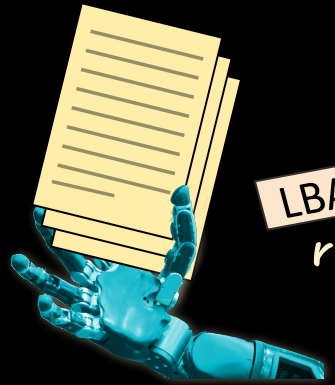
r = 0.4 to 0.60

r = 0.65 to 0.85

rating-scale assessments

"reliability"

accepted psychological scores



+ prompted language

LBAs with **Transformers**

r = ??

Lund University

Kjell, O. N., Sikström, S., Kjell, K., & Schwartz, H. A. (2022). **Natural language analyzed with AI-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy.** *Scientific reports*, *12*(1), 1-9.

Good old-fashion AI

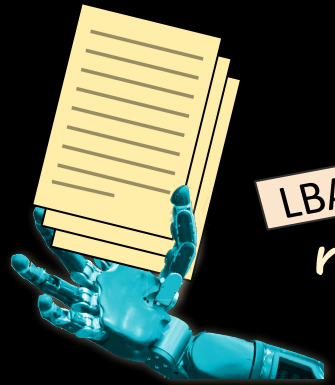
r = 0.4 to 0.60

r = 0.65 to 0.85

rating-scale assessments

"reliability"

accepted psychological scores



+ prompted language

LBAs with Transformers

r = 0.84

Lund University

Kjell, O. N., Sikström, S., Kjell, K., & Schwartz, H. A. (2022). **Natural language analyzed with AI-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy.** *Scientific reports*, *12*(1), 1-9.

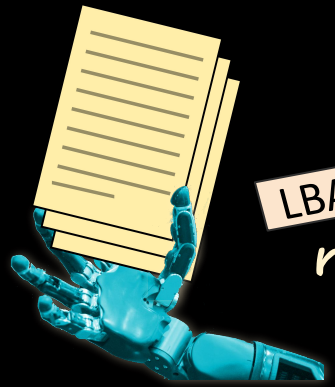
Good old-fashion AI

r = 0.4 to 0.60

r = 0.65 to 0.85

rating-scale assessments

"reliability"



+ prompted language

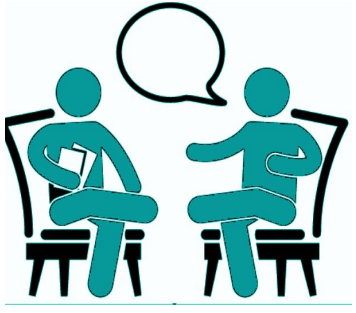
LBAs with **Transformers**

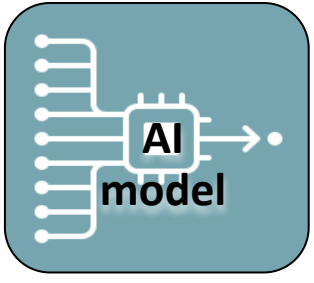
r = 0.84

matches the theoretical upper-limit!

Lund University

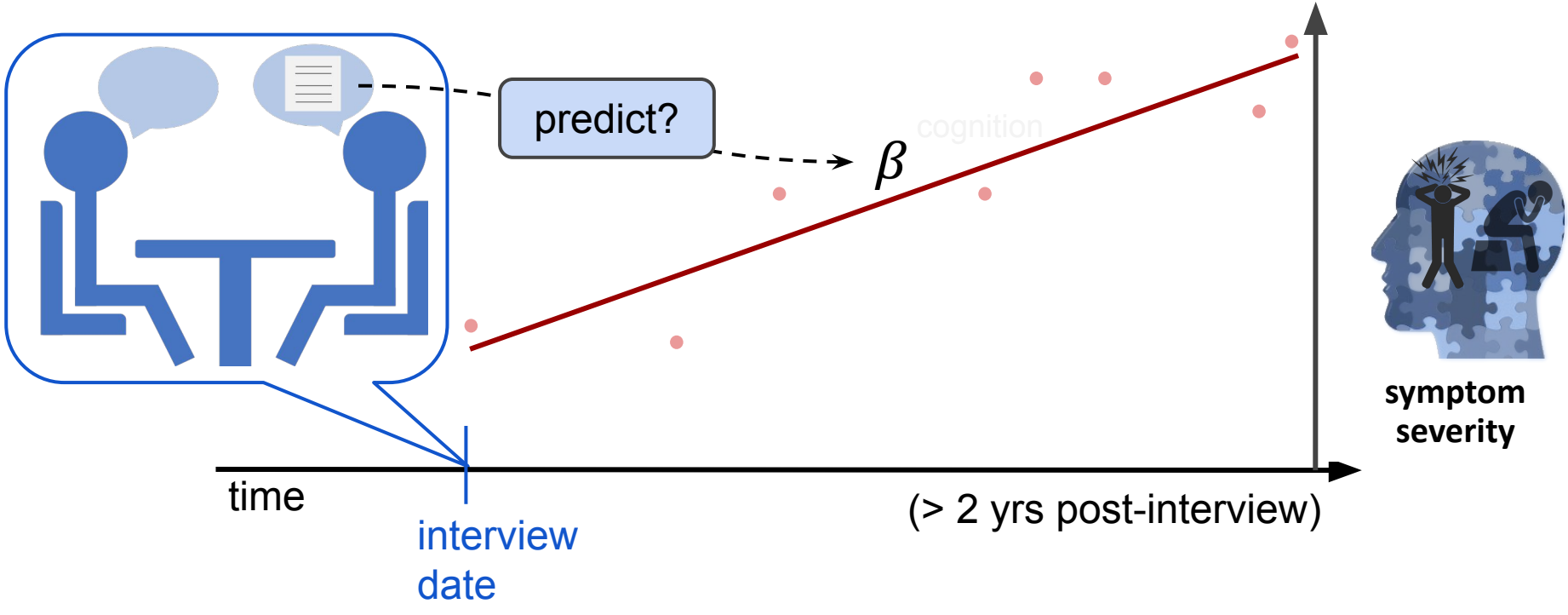
Clinic Application



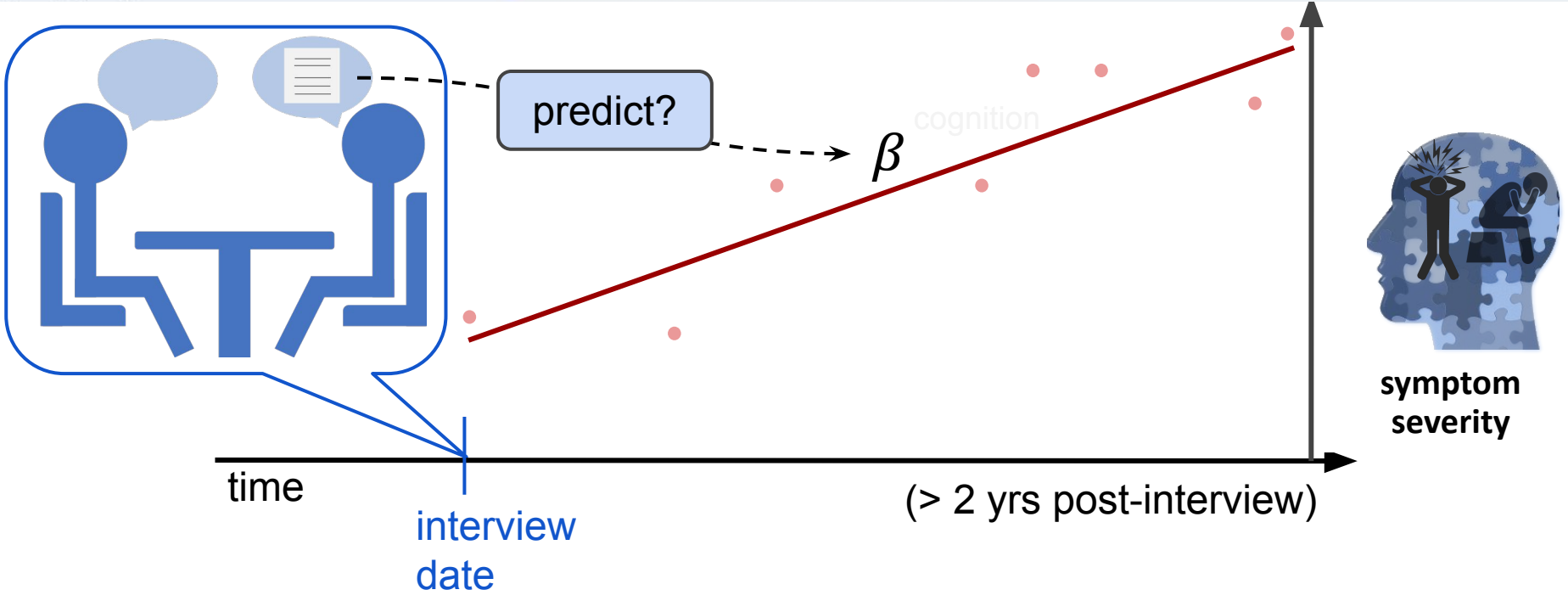


Current and Future PTSD Severity
(PCL Score)

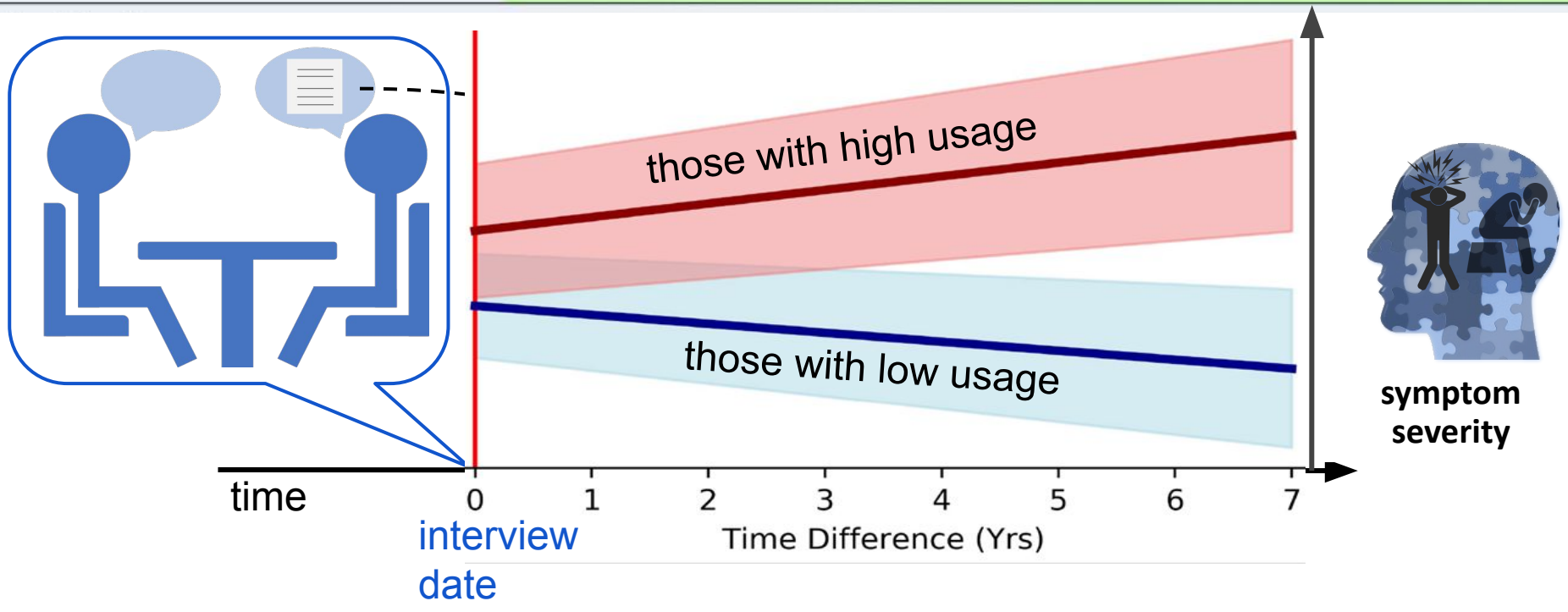
Son, Y., Clouston, S. A., Kotov, R., Eichstaedt, J. C., Bromet, E. J., Luft, B. J., & Schwartz, H. A. (2021). **World Trade Center responders in their own words: predicting PTSD symptom trajectories with AI-based language analyses of interviews.** *Psychological medicine, 1-9*.



Interview language features	PTSD symptoms future trajectories	
	r (direct correlation with symptom slope)	*β* (adjusted for age, gender, occupation, days since 9–11, Interview PCL)
Psychological traits		
Anxiety	0.16 (−0.07 to 0.37)	0.30* (0.08–0.49)
Depression	−0.00 (−0.23 to 0.22)	0.16 (−0.07 to 0.37)
Neuroticism	0.07 (0.29 to −0.16)	0.20 (−0.03 to 0.40)
Extraversion	0.17 (−0.06 to 0.38)	0.18 (−0.05 to 0.39)



Interview language features	PTSD symptoms future trajectories	
	r (direct correlation with symptom slope)	*β* (adjusted for age, gender, occupation, days since 9–11, Interview PCL)
Psychological traits		
Anxiety	0.16 (−0.07 to 0.37)	0.30* (0.08–0.49)
Depression	−0.00 (−0.23 to 0.22)	0.16 (−0.07 to 0.37)
Neuroticism	0.07 (0.29 to −0.16)	0.20 (−0.03 to 0.40)
Extraversion	0.17 (−0.06 to 0.38)	0.18 (−0.05 to 0.39)



PTSD-STOP:
Symptom Tracer and Outcome Prognosticator

communication

1. video

2. audio

3. language

new multimodal AI model (PTSD-STOP)

- hyper-arousal
- numbing
- avoidance
- re-experiencing
- *open-vocab markers*

Roman Kotov
(Psychiatry, SBU)

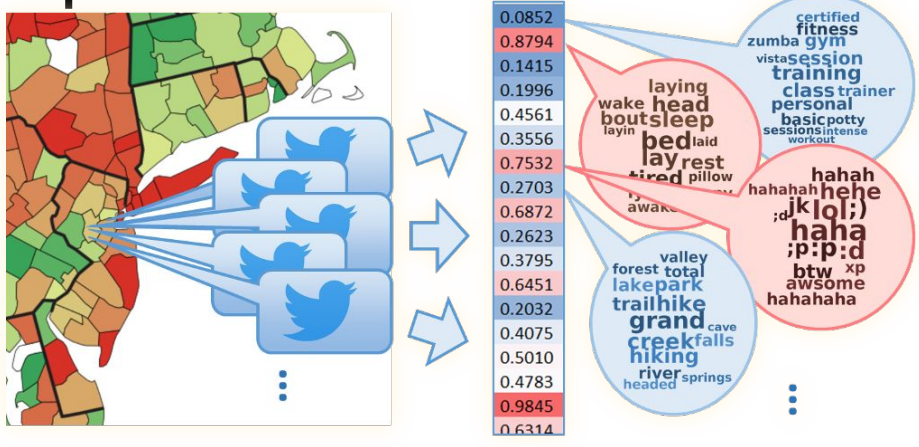
Dimitris Samaras
(CS, SBU - Vision)

Oscar Kjell
(Psychology, Lund Univ)

1. Digital footprints

AI for Mental Health

2. Populations



3. Automated Interviews

Thank You: Collaborators


Johannes Eichstaedt


Oscar Kjell


Ryan Boyd


Katarina Kjell


Niranjan Balasubramanian

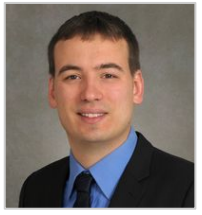

MZ Zamani


Veronica Lynn


Salvatore Giorgi


Matthew Matero


Rediet Abebe


Roman Kotov


Martin Seligman


Nikita Soni

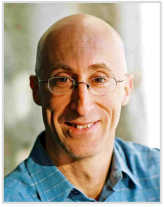

Youngseo Son


Vasudha Varadarajan


Huy Vu


Adithya Ganesan


Brenda Curtis

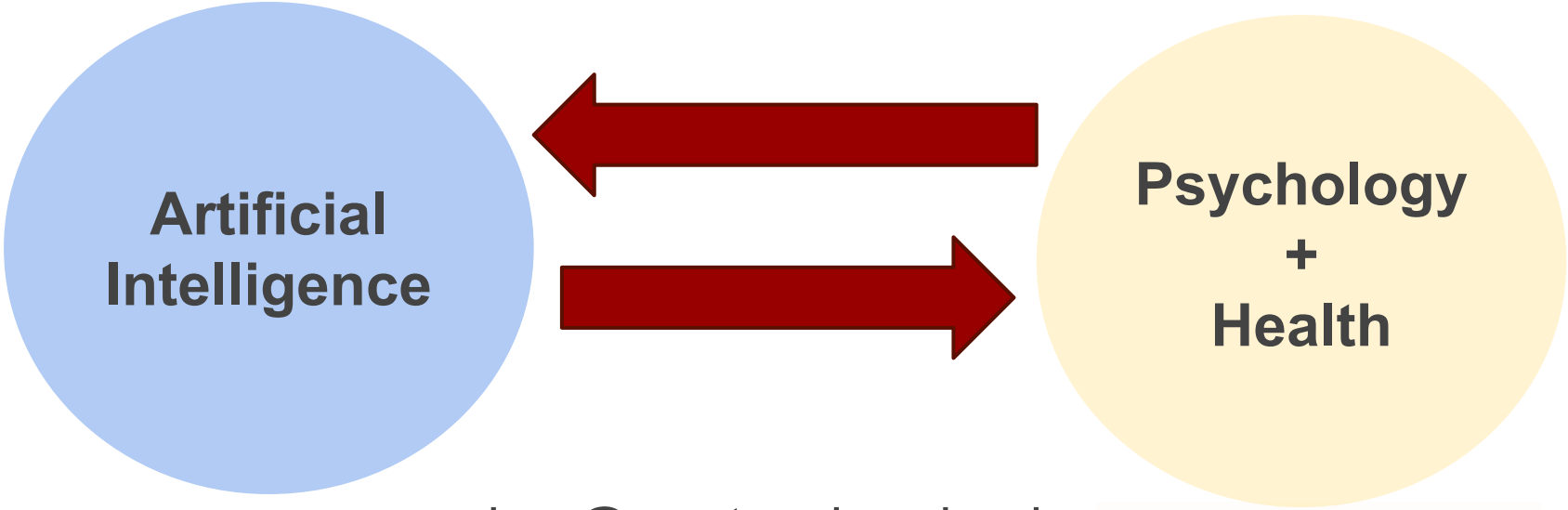

Lyle Ungar

World Well-Being Project

Stony Brook University

Human Language Analysis Beings

Thank You!



has@cs.stonybrook.edu

Stony Brook University

Human Language Analysis Beings