Five Projects Using Social Media as an Early Warning System for the Opioid Epidemic

Kristy A. Carpenter[1], Delaney A. Smith[2], Anna T. Nguyen[3], Shashanka Subrahmanya[4], Aadesh Salecha[4], Salvatore Giorgi[5,6], Issah A. Samori[7]
Mathew V. Kiang[3], Keith Humphreys[8], Anna Lembke[8], Johannes C. Eichstaedt[9], Russ B. Altman[1,7,10,11]

Departments of [1]Biomedical Data Science, [2]Biochemistry, [3]Epidemiology and Population Health, [7]Bioengineering, [8]Psychiatry and Behavioral Sciences, [9]Psychology, [10]Genetics, [11]Medicine; [4]Institute for Human-Centered Artificial Intelligence, Stanford University, Stanford, CA 94305, USA
[5]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104, USA
[6]Technology & Translational Research Unit, National Institute on Drug Abuse, National Institutes of Health, Baltimore, MD 21224, USA

Why Track the Opioid Epidemic?

- The opioid epidemic persists in the United States, with over 70,000 deaths annually since 2020 [1-2]

- Rapidly changing dynamics of non-medical opioid use, and variation across geographies and demographics, require reliable and timely information to monitor and address the crisis

Why Social Media?

- Academic and government surveys are standard practice for tracking opioid usage patterns

- Survey reliability is limited by reliance on self-reporting of stigmatized behavior [3]

- Surveys and overdose death tallies can take months to years to complete, making real-time monitoring of quickly evolving trends difficult

- Social media platforms include billions of user-generated posts that offer a less-filtered window into the relationship between individuals and their use of drugs, accessible in near real-time

- Previous work has established social media as a viable source to track opioid use and opioid-related mortality [4-6]

Project 1: Opioid Discussion on Reddit Correlates with Overdose Death Rates
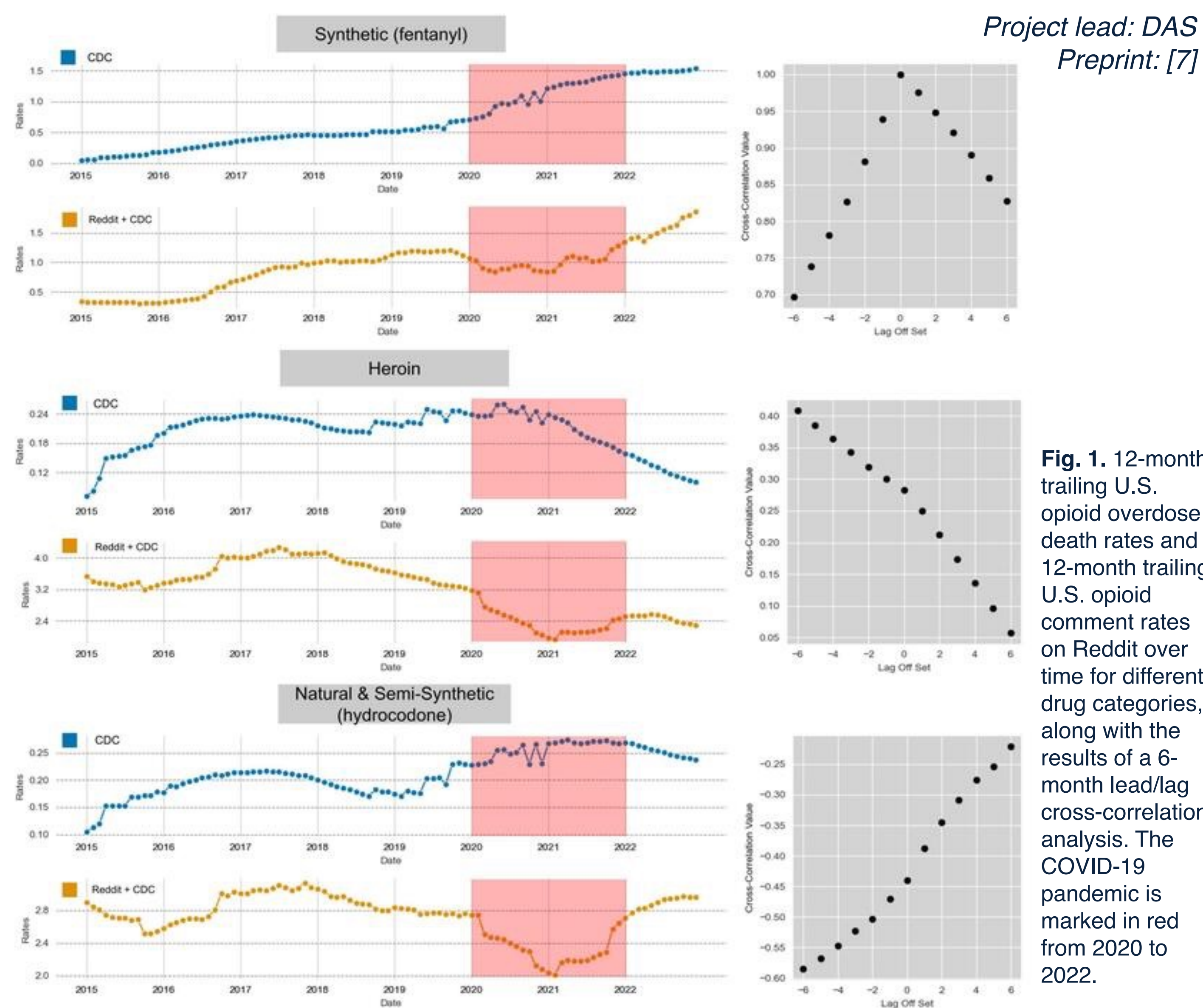
Project lead: DAS
Preprint: [7]

Fig. 1. 12-month trailing U.S. opioid overdose death rates and 12-month trailing U.S. opioid comment rates on Reddit over time for different drug categories, along with the results of a 6-month lead/lag cross-correlation analysis. The COVID-19 pandemic is marked in red from 2020 to 2022.

Project 2: Reddit Topic Modeling Informs Digital Ethnography

Project leads: DAS, ATN

Using/Addiction
Risky drugs, risky way, self-focused, present-focused, don't take responsibility, delayed discounting

Treatment/Early Recovery
Recent, negative affect, pain, withdrawal

Long-Term Recovery
Commitment, values, citizen, meaning, purpose, community, connection, support, future-looking, past-looking

Fig. 2. Topic modeling of Reddit posts, followed by annotation by opioid use experts, identified opioid discussion contexts falling under three themes (in green): active use/addiction, treatment/early recovery, and long-term recovery. The word clouds are constellations of semantically-related words (Latent Dirichlet Allocation topics) from Reddit text data.

Project 3: Topic Modeling Reveals Differential Trends in Public Discourse
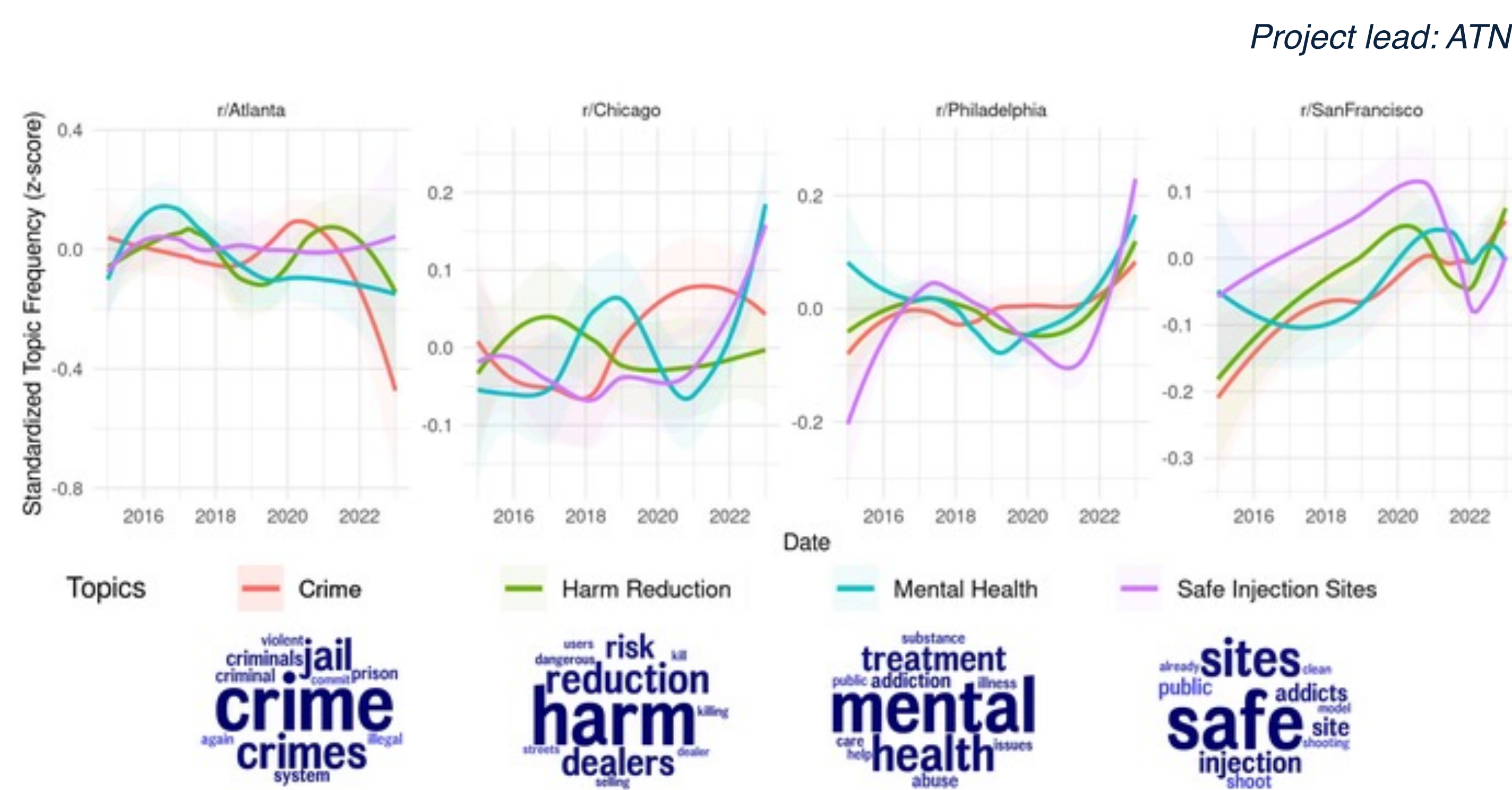
Project lead: ATN

Fig. 3. Frequency of topics over time detected in opioid-related discussions on regional subreddits.

Project 4: A Variety of Diverse Social Platforms Have Opioid-Related Content
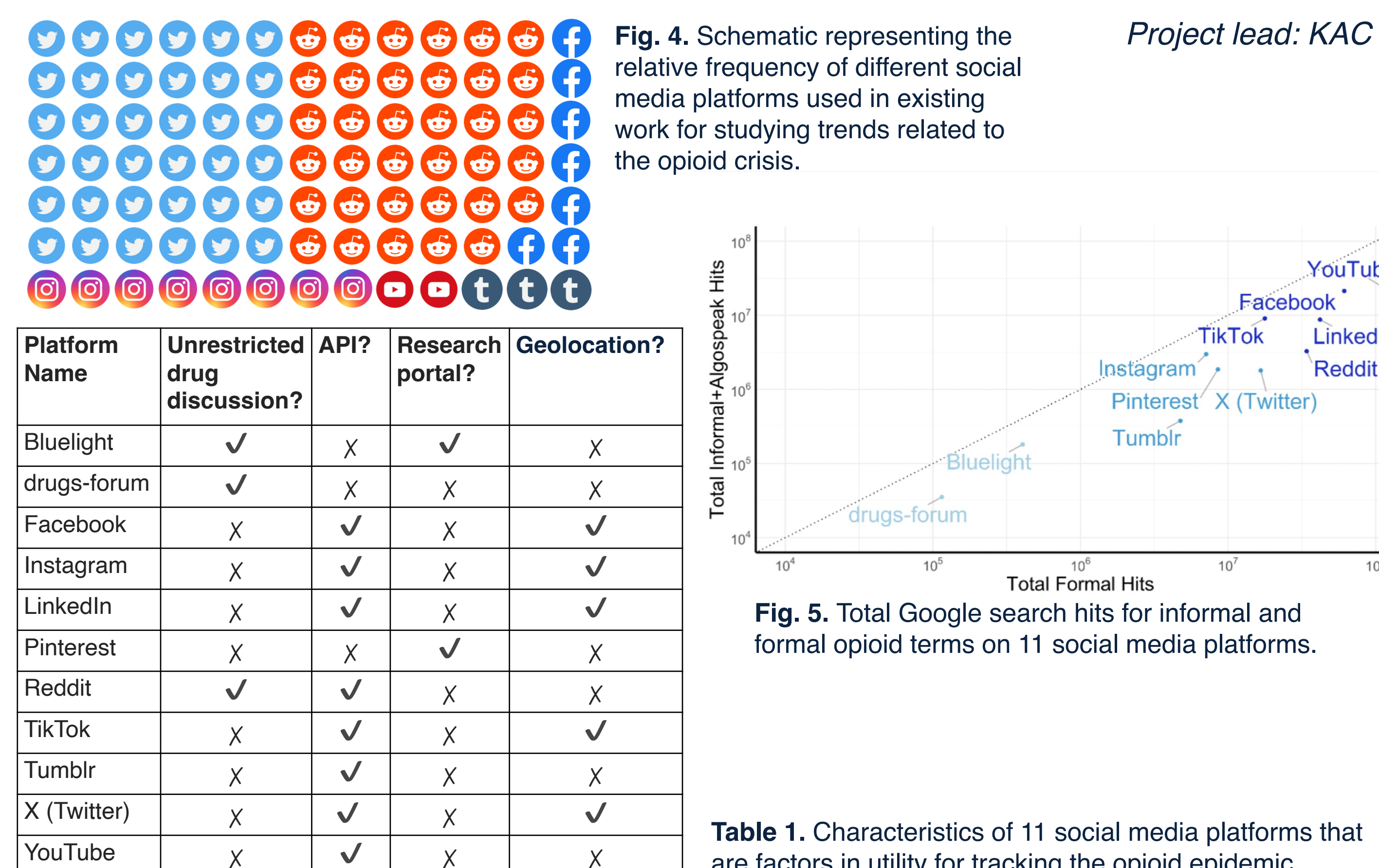
Project lead: KAC

Fig. 4. Schematic representing the relative frequency of different social media platforms used in existing work for studying trends related to the opioid crisis.

Fig. 5. Total Google search hits for informal and formal opioid terms on 11 social media platforms.

Platform Name	Unrestricted drug discussion?	API?	Research portal?	Geolocation?
Bluelight	✓	✗	✓	✗
drugs-forum	✓	✗	✗	✗
Facebook	✗	✓	✗	✓
Instagram	✗	✓	✗	✓
LinkedIn	✗	✓	✗	✗
Pinterest	✗	✗	✓	✗
Reddit	✓	✓	✗	✗
TikTok	✗	✗	✗	✓
Tumblr	✓	✓	✗	✗
X (Twitter)	✗	✓	✗	✓
YouTube	✗	✓	✗	✗

Table 1. Characteristics of 11 social media platforms that are factors in utility for tracking the opioid epidemic.

Project 5: Large Language Models Aid Identification of Opioid-Related Content

Project lead: KAC

How can you tell if **fenty** is laced with zines → **Fentanyl**

My **fenty** one is expired too -- ugh but canadian drug stores be wilding with their prices → **Not Fentanyl**

Can't believe MJF came back representing so many eras of Triple **H** at once → **Not Heroin**

Remember when **H** was H and you could walk into a docs office and walk out with real oxy? → **Heroin**

honestly you'd think you'd put an old **blues** record on for credibility points → **Not Oxycodone**

I've dabbled in the **blues** for the past year or so, I kinda want to quit now, how should I go about this? → **Oxycodone**
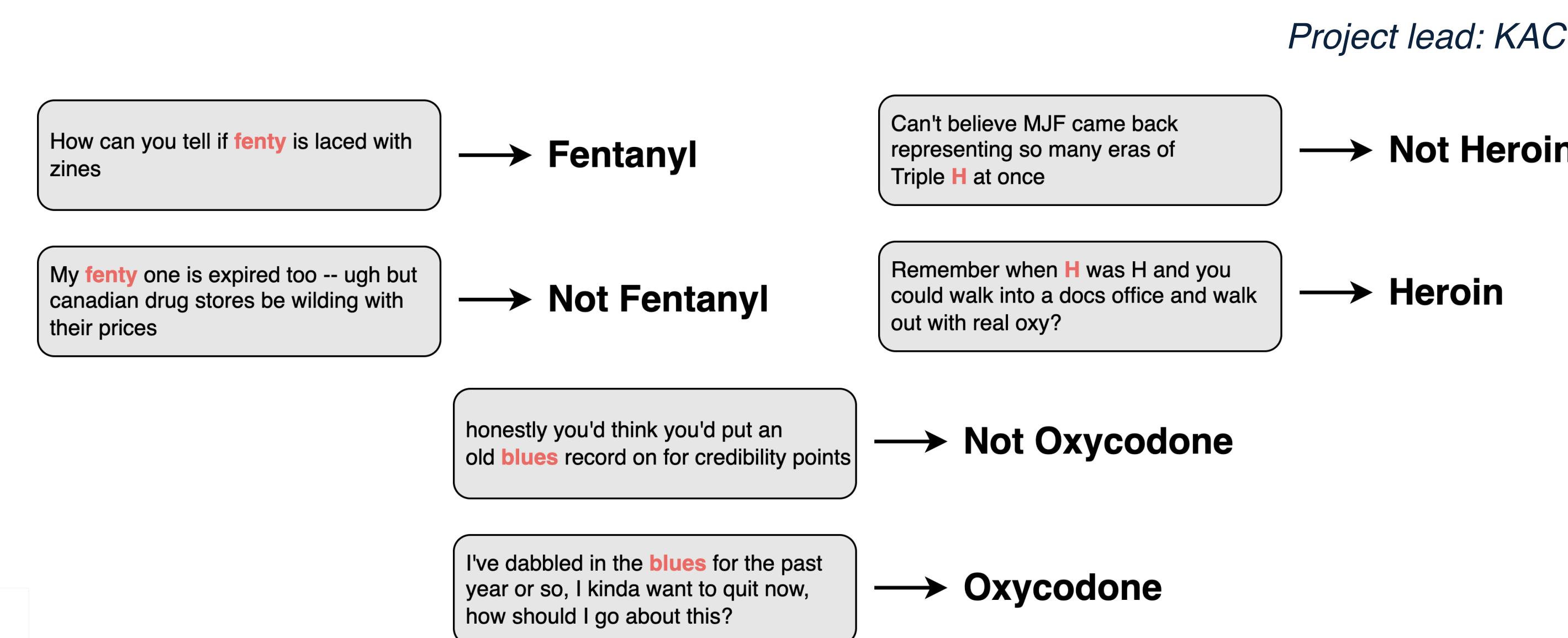
Fig. 6. Three examples of social media posts using ambiguous terms potentially related to opioids. The examples evaluate the use of slang terms "fenty," "H," and "blues" as slang terms for fentanyl, heroin, and oxycodone, respectively.

Manual label

GPT4 label	Yes (is fentanyl)	No (is not fentanyl)	Unclear	TOTAL
Yes (is fentanyl)	30	0	4	34
No (is not fentanyl)	3	414	40	457
ContentRestrictionError	0	1	0	1
TOTAL	33	415	44	492

Table 2. Confusion matrix of GPT4 classification versus human annotation of "fenty" as a slang term for fentanyl in a set of 492 tweets from September 2022. Accuracy = 99.3%, sensitivity = 90.9%, specificity = 100.0%, PPV = 100.0%.

REFERENCES
[1] www.cdc.gov/drugoverdose/data/statedeaths.html (2021).
[2] Humphreys, et al. The Lancet Commissions 399(10324):555-604 (2022).
[3] Palamar. Am J Drug Alcohol Abuse 45(2):117-123 (2019).
[4] Hanson, et al. J Med Internet Res 15(9):e189 (2013).
[5] Chary, et al. J Med Toxicol 13(4):278-286 (2017).
[6] Giorgi, et al. Sci Rep 13(1):9027 (2023).
[7] Lavertu, Hamamsy, & Altman. medRxiv 2021.04.0121254815 (2021).

HAI FIVE
Celebrating 5 Years of Impact

Stanford University

Stanford MEDICINE