Developing a lexicon for identification of experiences of bias, stigma, and discrimination in subreddit communities related to opioid use disorder

Drew Walker, MS, CHES,[1] Jennifer S. Love, MD,[2] Jeanmarie Perrone, MD,[3] Abeed Sarker, PhD[4]

1. Department of Behavioral, Social, and Health Education Sciences, Rollins School of Public Health, Emory University, Atlanta, GA. 2. Department of Emergency Medicine, Medical Toxicology, Mount Sinai, New York, NY. 3. Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA. 4. Department of Biomedical Informatics, Emory University School of Medicine

INTRODUCTION

- Social media data, particularly among MOUD-related communities, offer researchers novel opportunities to assess self-reported experiences of stigmatization and bias.

- This study involved the development of an expanded lexicon, guided by existing literature on stigmatizing language, clinical expertise and utilizing word embeddings models used for the identification and assessment of patient-reported experiences of stigma and bias in the context of online communities related to MOUD.

METHODS + RESULTS

Lexicon development occurred in a series of 7 steps. Figure 1 outlines this process graphically.

Step 1) Identification of initial stem word list from NIDA and research describing experiences of stigmatization, such as treatment coercion, imposing restrictions, desire for social distance, and decreased intention to help.
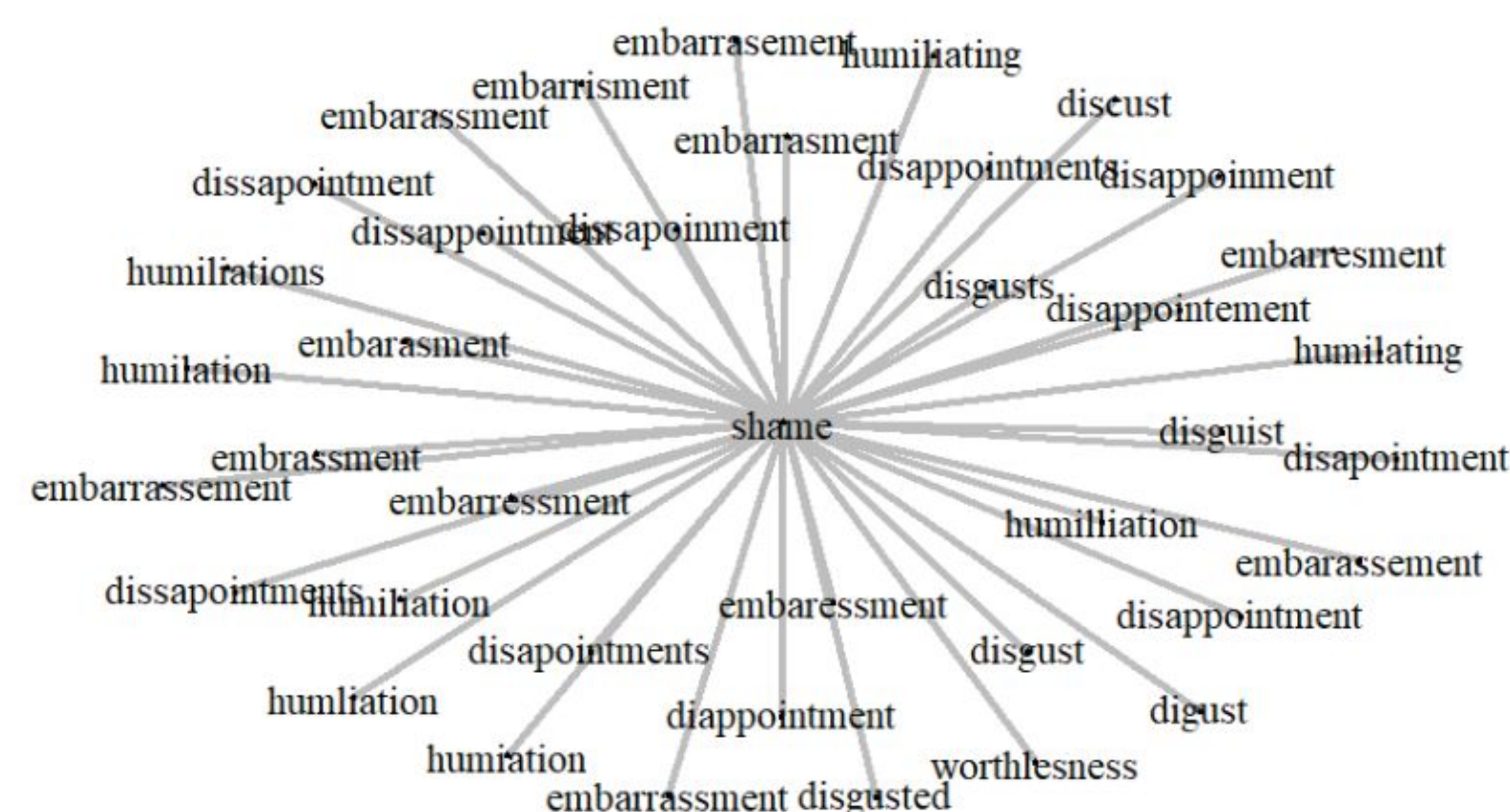
- **The final stem word list consisted of the following 13 words:** "addict", "user","abuser","junkie","alcoholic", "drunk", "habit", "dirty", "stigma","bias","stereotype","shame","blame".

Step 2) Identify top 10 most semantically-similar words for each word in the stem word list from an n-gram word embeddings model learned from drug chatter on social media data where at least one type of medication was mentioned. **(88 resulting words)**
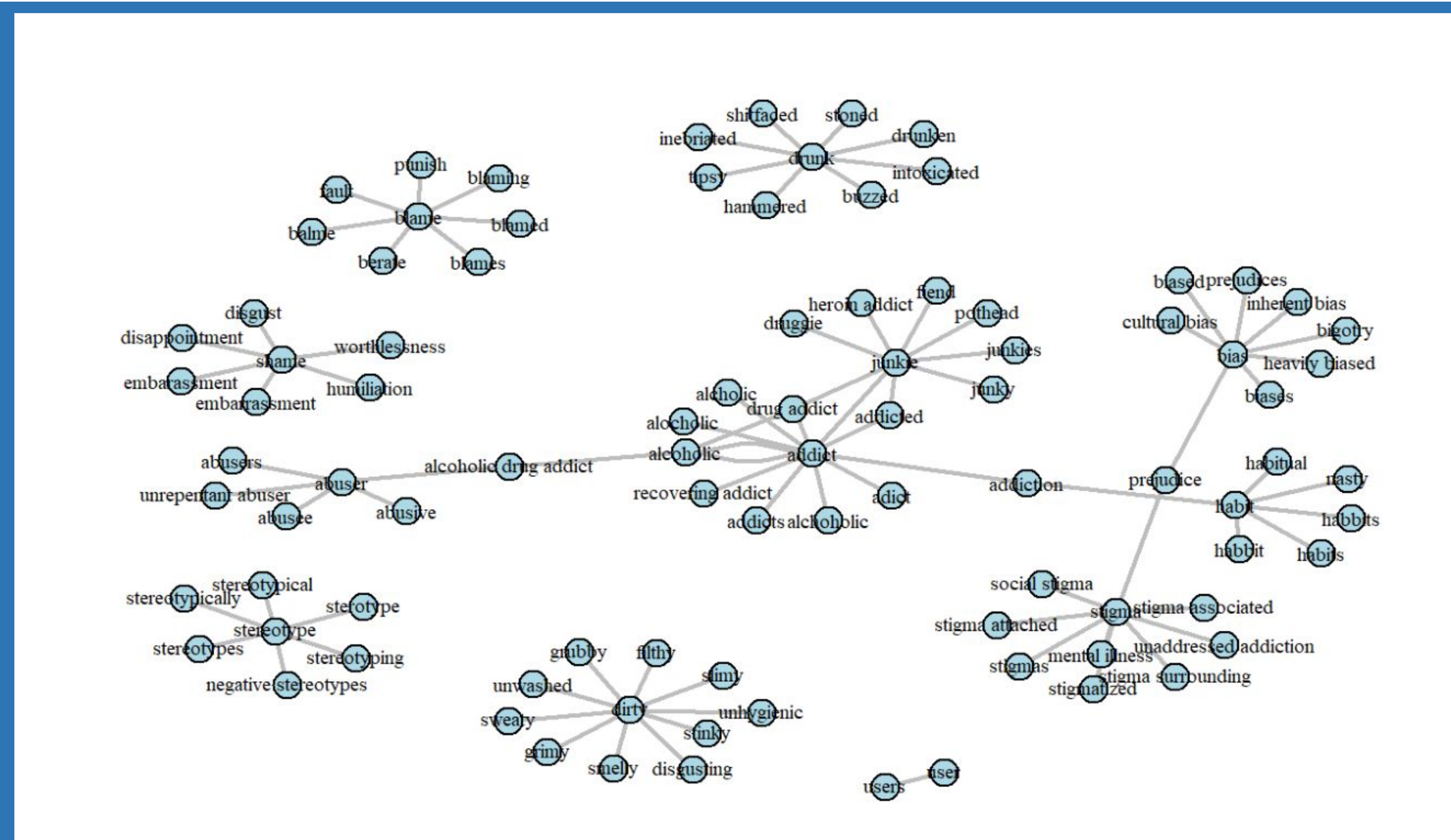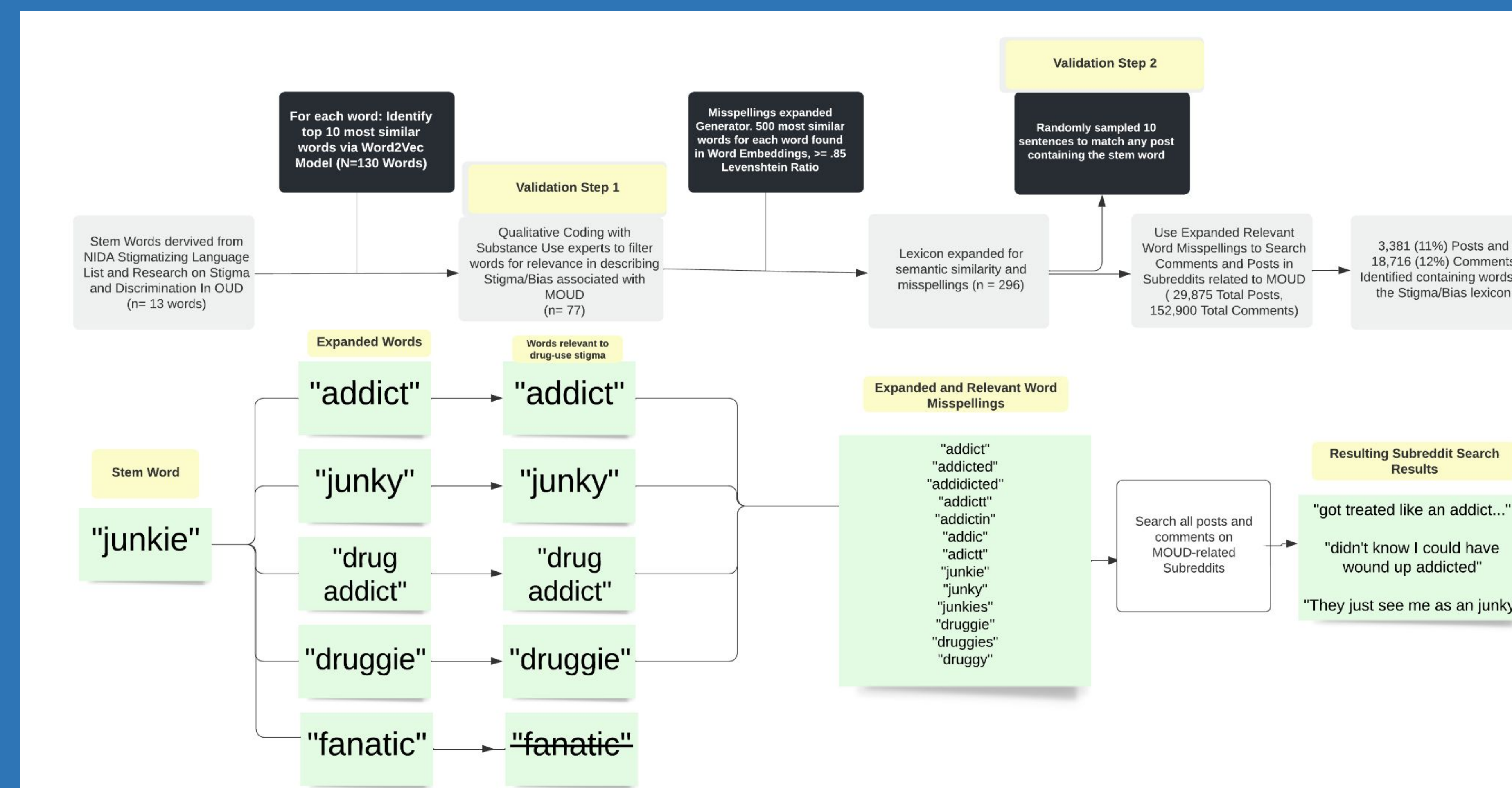
- We loaded a DSM-language-model-1B-LARGE generated from ~ 1B tweets from user timelines where at least 1 medication is mentioned.

Step 3) Terms to expand lexicon were manually assessed and validated by research team to incorporate only the most relevant terms into the next phase of lexicon expansion. Authors JL, DW, AS discussed and adjudicated on differences (Kappa = .87,) **(72 resulting words)**

Step 4) We then generated lists of close misspellings based off of the of manually pruned/validated similarity-expanded word list, by re-running embeddings model for top 500 lexical similarity, returning any words within >=.85 Levenshtein ratio of the expanded lexicon word. **(296 resulting words)**

Word embeddings natural language processing methods can be successfully applied to identify and evaluate patient-reported experiences of MOUD-related stigmatization and discrimination on social media.

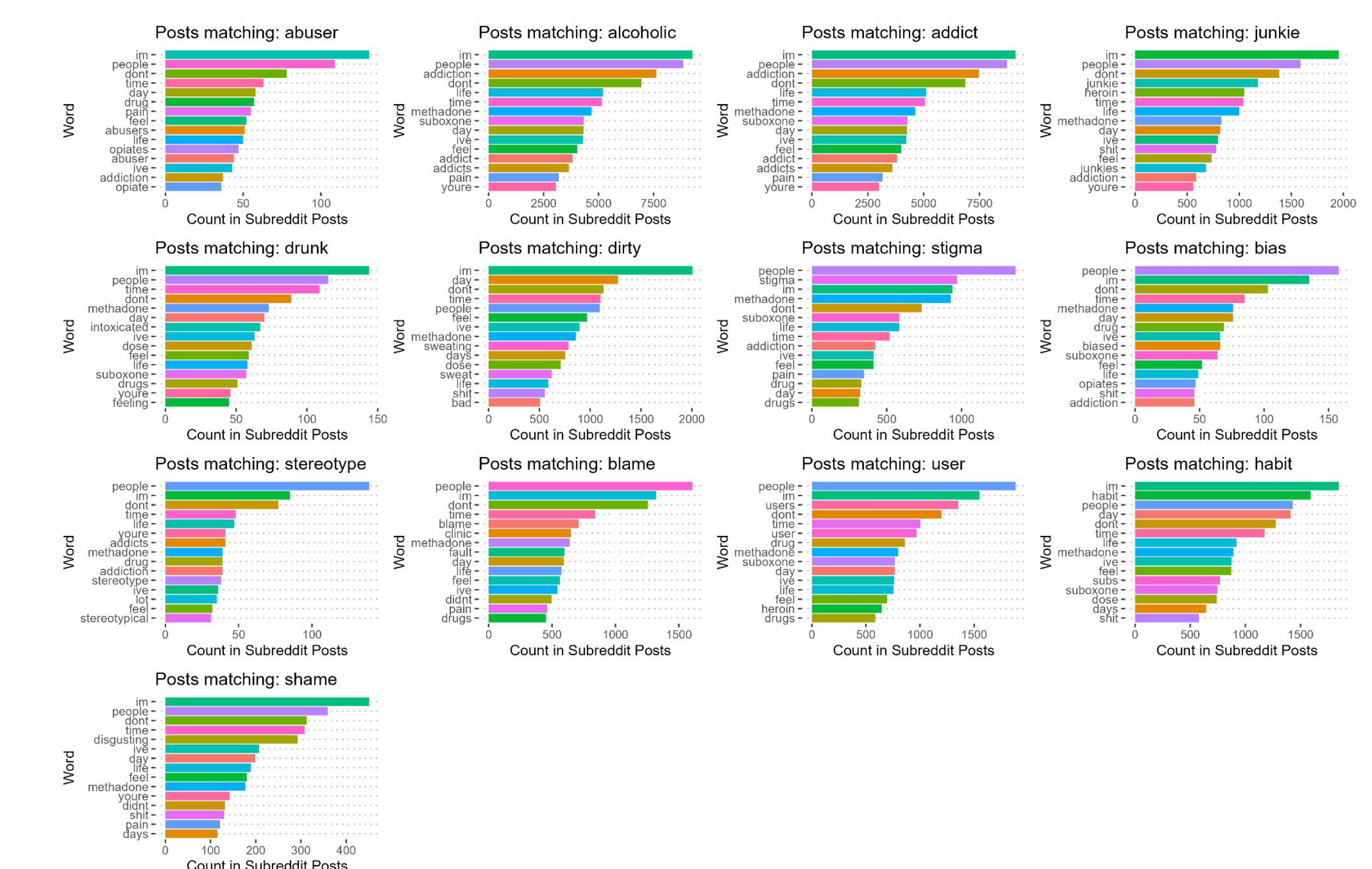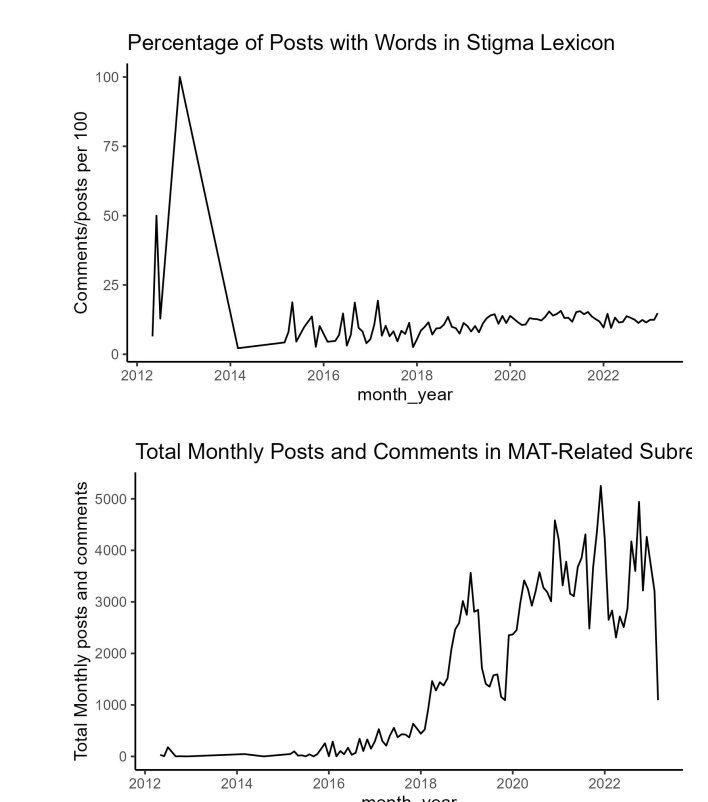**Step 5):** Collect corpus of **182, 761 posts and comments** from 4 MOUD-related Subreddits using PRAW

Step 6): We searched posts from MOUD subreddits containing all words expanded from each stem word in the dataset
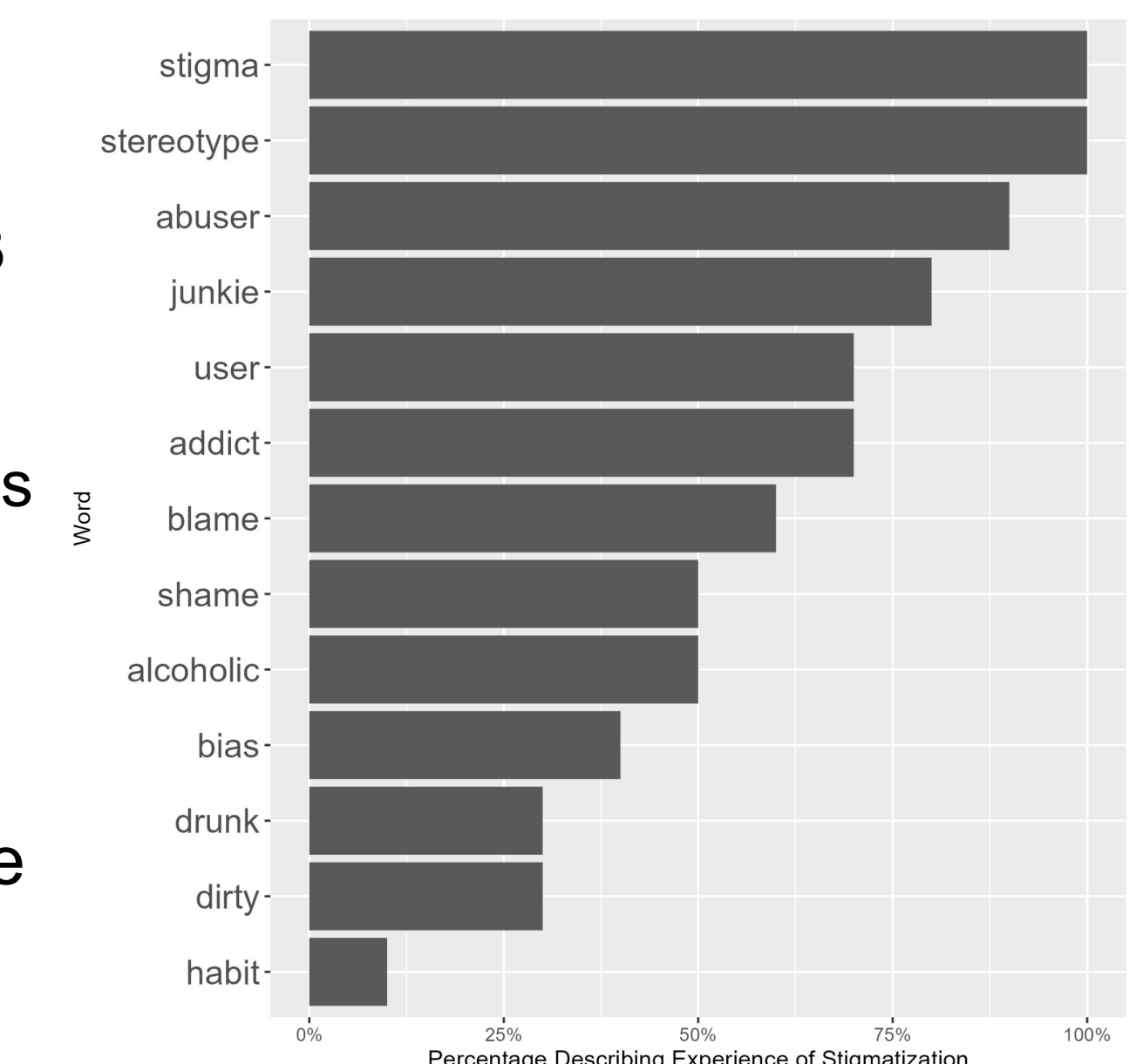
% of posts/comments containing lexicon words : 12.1% (22,097)

subreddit	Total Posts and Comments	Total Posts	Total Comments
opiates	66411	7157	59254
suboxone	60180	11830	48350
Methadone	54657	10324	44333
OpiateRecovery	1527	564	963

Step 7: We assessed a random sample of 10 sentences from posts and comments using each of the 13 stem words with expanded lexicon, and employed qualitative coding to assess the frequency at which terms were being used to describe experiences of bias and stigma.

Top words (>75%):
Stigma, stereotype, abuser, junkie
(<50%): Bias, drunk, dirty, habit

DISCUSSION

- We were able to successfully filter a large online dataset with the aim of efficiently identifying experiences of stigmatizaion among MOUD-related subreddit communities
- Stigmatization has been a prevalent topic of discussion in subreddit communities over years
- Lexicon can be used as a tool which can help identify experiences more rapidly, used in tandem with more in-depth qualitative methods for added nuance
- May be able to be used in other settings such as hospital patient reviews or EHR systems

Andrew.walker@emory.edu

EMORY | ROLLINS SCHOOL OF PUBLIC HEALTH