

# Pregnancy and Substance Use:

A computational discourse study using large language model-assisted annotation of Reddit narratives, 2020 to 2022

Eunhee Victoria Park, PhD, MPH  
Postdoctoral Scholar  
Primary Care Research Training (NRSA T32)  
School of Medicine and Health Sciences



---

THE GEORGE  
WASHINGTON  
UNIVERSITY

---

WASHINGTON, DC

# Postdoctoral Training Mentors



**Monica Ruiz, PhD, MPH**

**Primary Mentor**

*Associate Professor*

Department of Prevention and Community Health  
GW Milken Institute School of Public Health



**David Broniatowski, PhD, MS**

**Co-Mentor**

*Professor*

Department of Engineering  
GW School of Engineering and Applied Science  
*Deputy Director*  
NIST-NSF Institute for Trustworthy AI in Law and Society



**Trudy Mallinson, PhD, MS**

**Supervising Mentor**

*Professor*

Department of Clinical Research & Leadership  
GW SMHS  
*Associate Dean, Health Sciences Research*

# Acknowledgement

This work was supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) under Award T32HP42014 (Ruth L. Kirschstein NRSA).

The content is solely the responsibility of the authors and does not necessarily represent the official views of HRSA or HHS.

- Examines literature and policy implications of substance use among pregnant and parenting women
- Describes research on health effects, policies ensuring access to care, and barriers to treatment
- Highlights the impact of laws and policies on pregnant women and families
- Policies assume all substance use during pregnancy is harmful, leading to harsh penalties

## PREGNANT WOMEN AND SUBSTANCE USE

Overview of Research & Policy in the United States

Darla Bishop, MPH  
Liz Borkowski, MPH  
Megan Couillard, BS  
Amy Allina, BA  
Susannah Baruch, JD  
Susan Wood, PhD

**Bridging the Divide:**

A Project of the Jacobs Institute of Women's Health

February 2017

# PREGNANCY AND SUBSTANCE USE

## A HARM REDUCTION TOOLKIT



**NATIONAL  
HARM REDUCTION  
COALITION**



### FEDERAL LEGISLATION THAT MAY AFFECT YOU:

## CHILD ABUSE PREVENTION AND TREATMENT ACT

The Child Abuse Prevention and Treatment Act (CAPTA) is a federal law directed only to states, not to hospitals or individual healthcare providers.

CAPTA requires that if states want to have the benefit of certain federal funds, they must provide an assurance that their state has a system for notifying the child protective services system of the occurrence of certain conditions in infants.

These conditions are:

- when infants are born “**affected by substance abuse**” (a term not defined in the statute).
- when infants have “**withdrawal symptoms resulting from prenatal drug exposure.**”
- when infants are diagnosed with “**a Fetal Alcohol Spectrum Disorder.**”

There is no federal law that requires all pregnant people be tested for drugs.

[www.perinatalharmreduction.org](http://www.perinatalharmreduction.org)

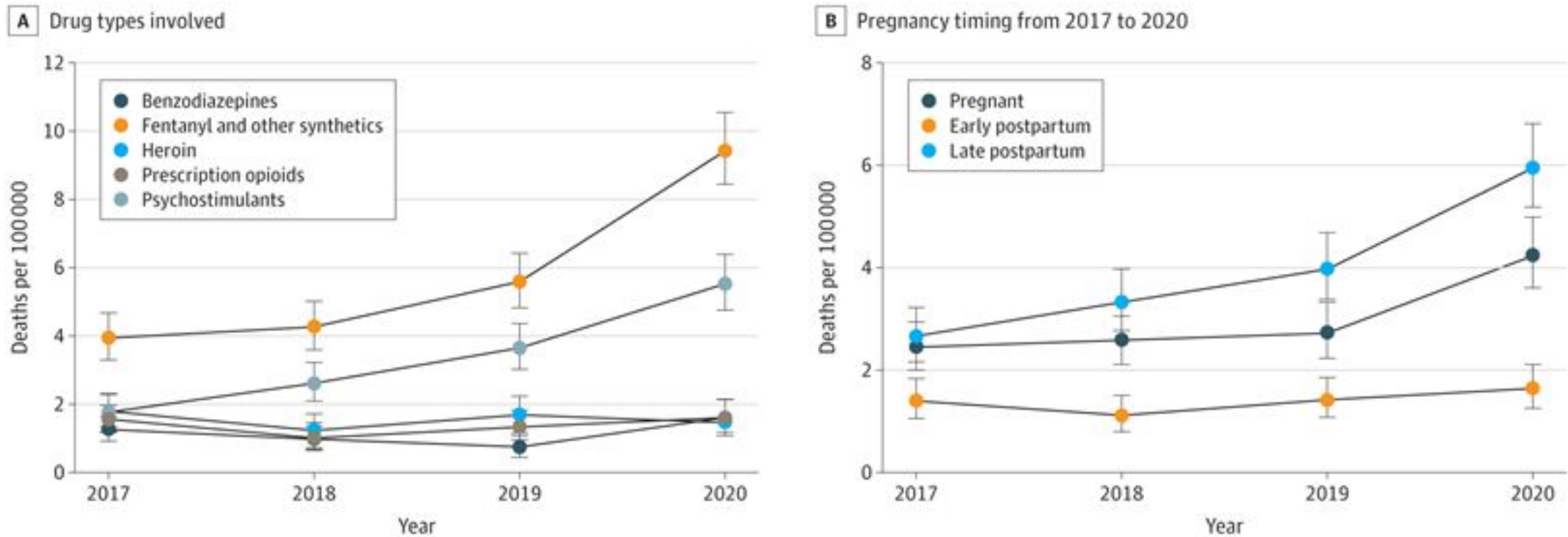
A 2016 federal amendment passed by Congress requires that a “**Plan of Safe Care**” (See [Appendix B](#)) **must be developed for all infants with these conditions.** These plans must include the needs of the affected family and/or caregiver.

**States can decide where reports are directed**, as well as who is responsible for developing and monitoring the Plan of Safe Care and may establish a reporting system that **does not make families vulnerable to allegations of or investigations** for child abuse or neglect.

For more information about how CAPTA relates to you, please see the fact sheet from National Advocates for Pregnant Women: [Understanding CAPTA and State Obligations.](#)

[www.harmreduction.org](http://www.harmreduction.org)

## Substance use in pregnancy is widespread and a growing driver of maternal mortality



- In 2019, 7% of US women used prescription opioids during pregnancy, with 20% misusing the drugs.
- Approximately 8.3% of pregnant women used illicit substances, and 8.4% used tobacco products in 2020.
- From 2017 to 2020, drug overdose deaths in the United States among pregnant and postpartum females increased by 81%.

# How do pregnant people talk about substance use and care?

Substance use during pregnancy is shaped by social learning, stigma, and policy environments that influence perceptions of risk and access to prenatal care.

Traditional public health surveillance and clinical data offer limited insight into how pregnant people interpret these issues in real time.

Social media platforms such as Reddit provide a large, naturally occurring discourse that captures peer-to-peer exchange and self-reported lived experience, complementing existing data sources.

## RESEARCH QUESTION

What barriers to care do pregnant people disclose in substance-use narratives, and how do these patterns vary across substance type?

*Are stigma-related mentions more common in opioid or stimulant narratives than in alcohol or cannabis narratives?*

# A multi-LLM consensus pipeline for narrative annotation

~2.1M

Reddit posts  
2020 to 2022

5

subreddits  
final corpus

3

locally deployed  
LLM annotators

100

human-coded  
validation samples

## PIPELINE

**Models compared:** Llama 3.3-70B, GPT-OSS-20B, Qwen3-30B. Qwen3 achieved the highest overall annotation accuracy (90.3%).

**Decision tree filter:** two subject matter experts developed a five-category classification (self-disclosure, substance use, pregnancy status, self-reported substance use during pregnancy, and treatment or care engagement).

**Three-model consensus:** yielded ~1,370 high-precision self-report labels related to the five categories.

# Annotation Schema

The system prompt defines seven non-mutually-exclusive barrier categories with explicit operational definitions, signal phrases, and inter-category boundary rules:

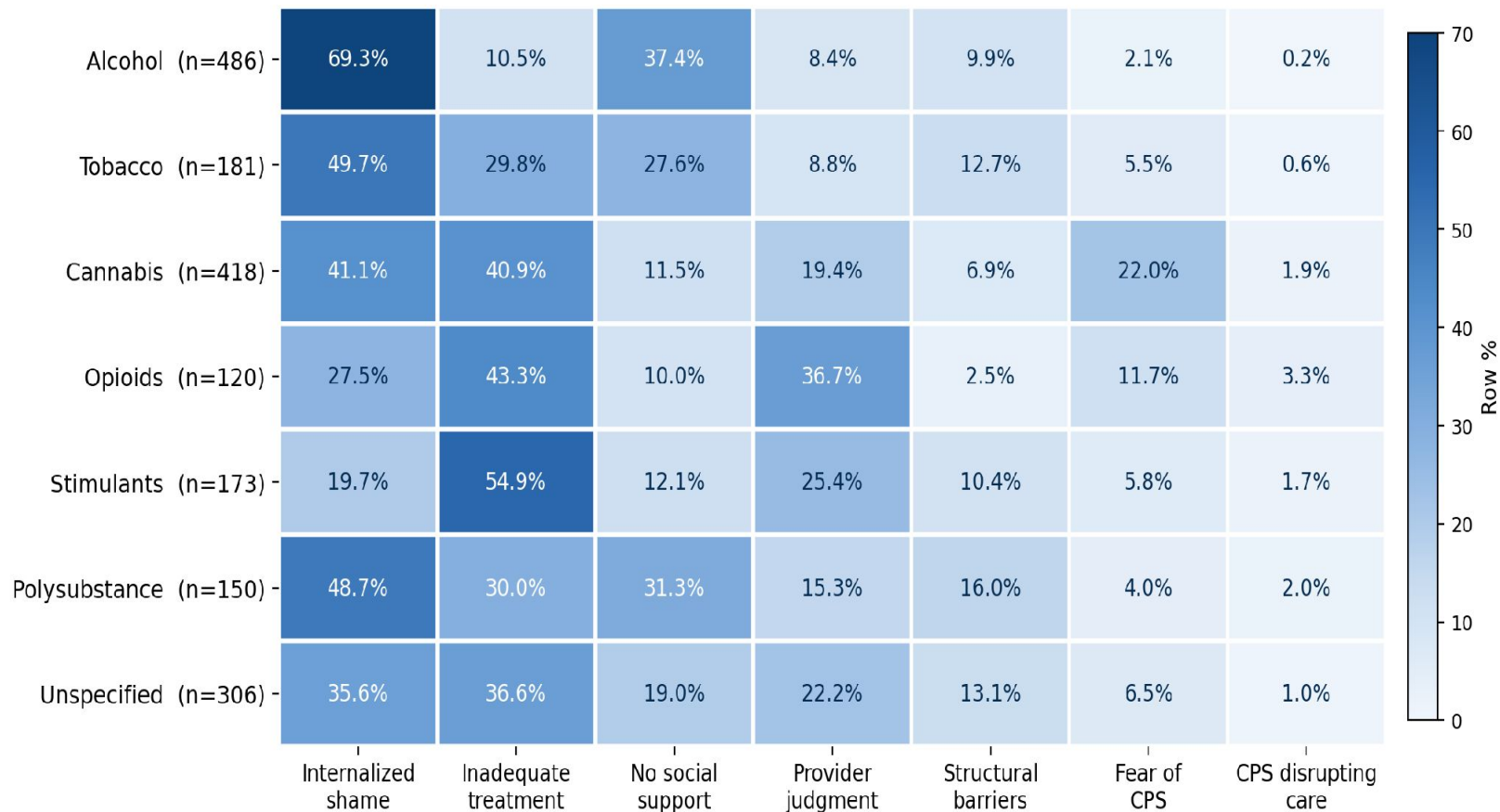
1. fear\_of\_child\_protective\_services\_or\_custody\_loss
2. child\_protective\_services\_disrupting\_treatment\_access
3. structural\_life\_barriers\_housing\_transport\_finances
4. inadequate\_or\_inaccessible\_treatment\_options
5. lack\_of\_trusted\_social\_or\_family\_support
6. internalized\_shame\_or\_anticipated\_community\_judgment
7. experienced\_or\_anticipated\_healthcare\_provider\_judgment

Each barrier definition includes a Boundary subsection explicitly distinguishing it from adjacent categories — the highest-payoff piece of the prompt, since adjacent-category confusion (cps\_fear vs cps\_disrupting\_treatment, self\_stigma vs provider\_stigma, etc.) is the dominant failure mode in this kind of annotation.

In addition to barriers, each comment receives:

1. substance (primary; one of tobacco / alcohol / cannabis / opioids / stimulants / polysubstance / unspecified / none)
2. substances\_mentioned (multi-label list)
3. legal\_context (licit / illicit / mixed / n/a)
4. care\_outcome (avoided / delayed / intermittent / engaged\_with\_disclosure / engaged\_without\_disclosure / unclear / n/a)
5. disclosure\_to\_provider (full / partial / none / n/a)
6. temporal\_context (current\_pregnancy / past\_pregnancy / anticipated\_pregnancy / unclear)
7. reasoning (3–5 sentence rationale citing specific phrases)
8. Inclusion criteria require all three of: first-person substance use disclosure, pregnancy context, and barrier framing (explicit or behavioral). Exclusion reasons are enumerated to enable downstream re-segmentation: no\_substance, not\_first\_person, no\_pregnancy\_context, no\_barrier\_framing, pure\_emotion\_no\_behavior.

# Barrier signatures differ sharply by substance type



### KEY OBSERVATIONS

#### Licit substances

Alcohol and tobacco posts cluster on internalized shame (alcohol 69%, tobacco 50%) and lack of social support.

#### Illicit substances

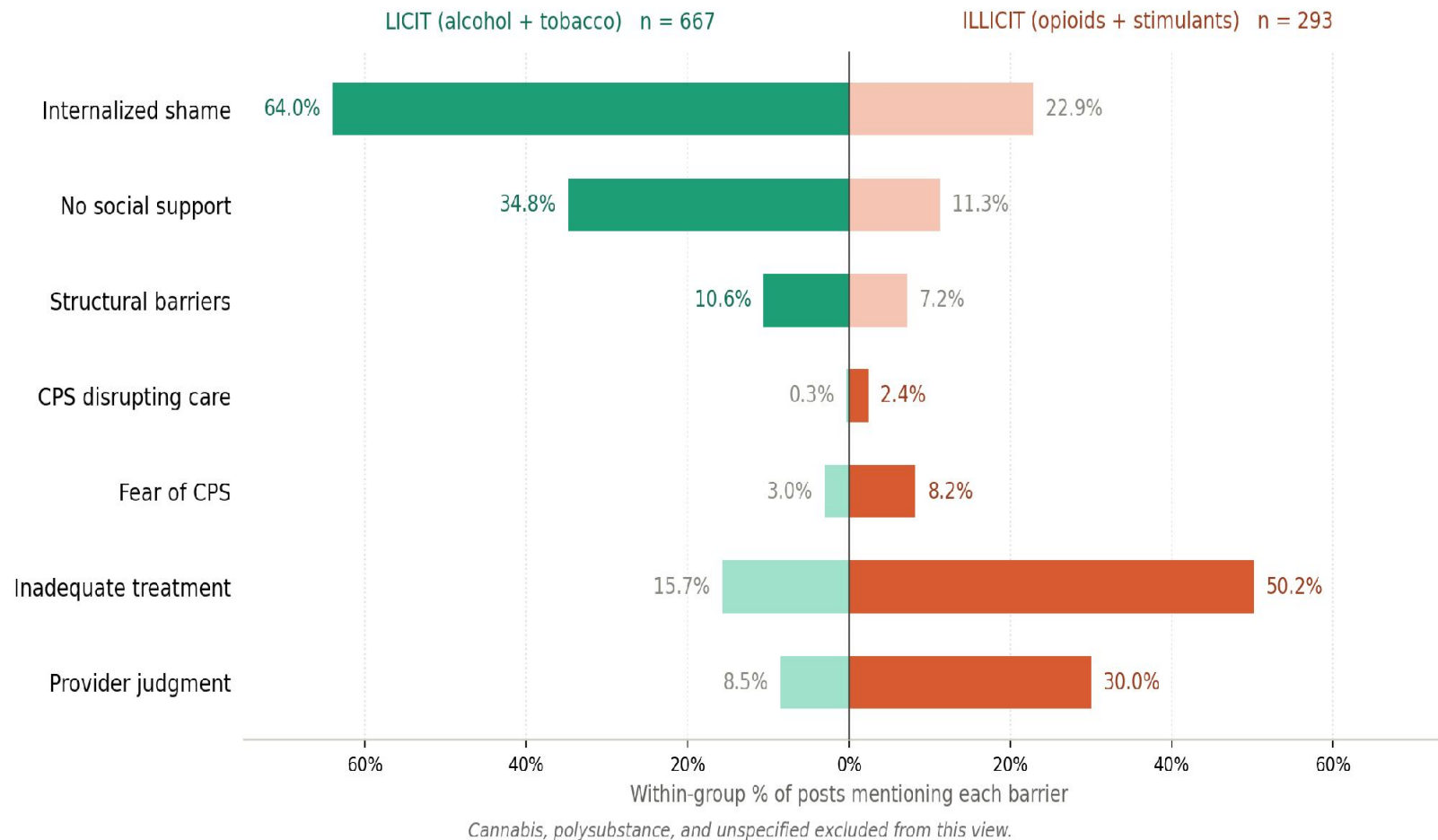
Opioids and stimulants cluster on inadequate treatment access (43% and 55%) and provider judgment.

#### Cannabis is distinctive

22% of cannabis posts mention fear of CPS or custody loss, far higher than any other substance.

Row %: share of posts within each substance mentioning each barrier. Posts can mention multiple barriers, so rows sum to >100%. n = 1,834.

## Two intervention targets: stigma versus system access



### WHAT THE CONTRAST SHOWS

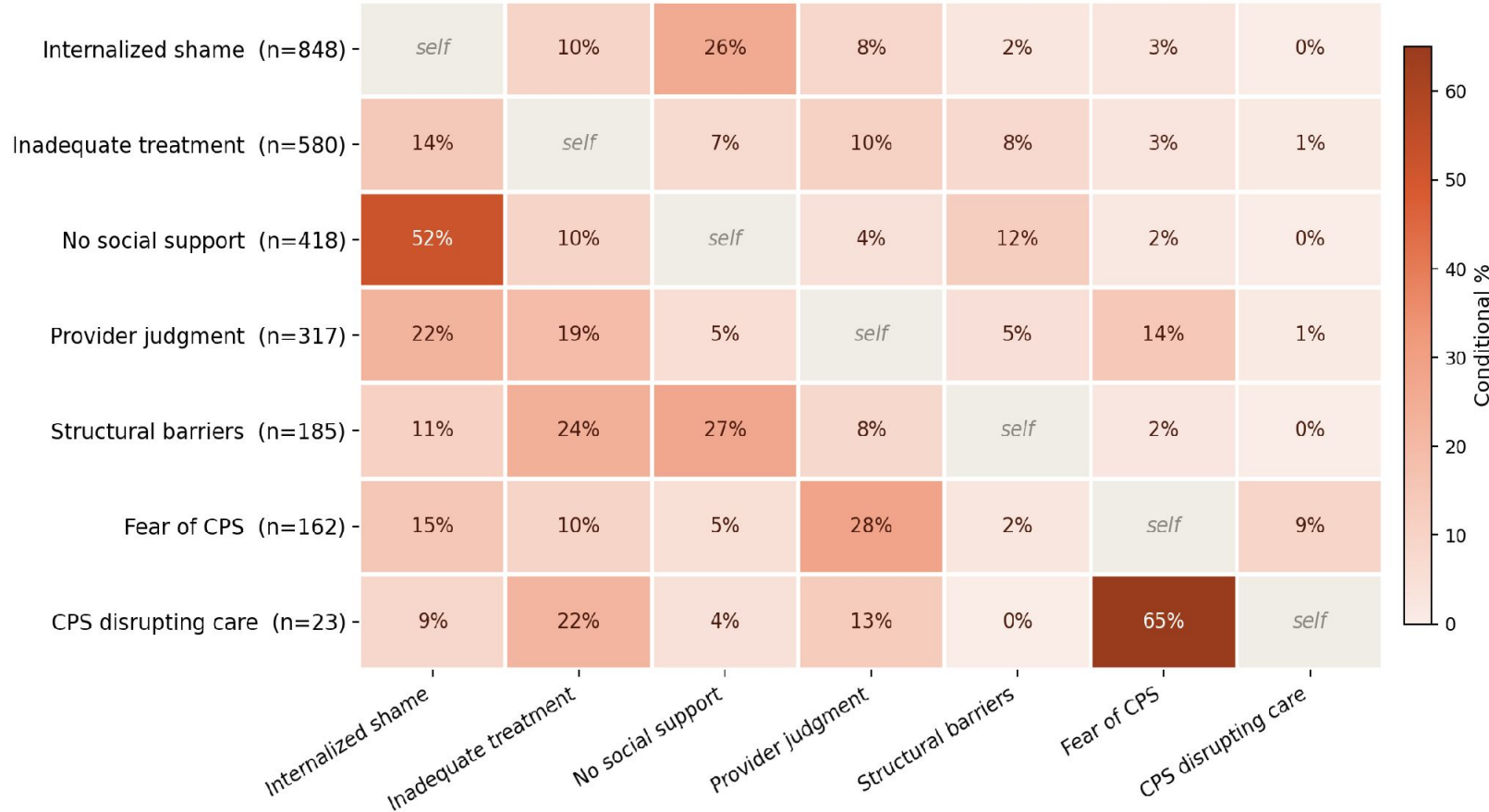
#### Licit barriers are stigma-driven.

Internalized shame (64%) and isolation from social support (35%) dominate alcohol and tobacco discourse.

#### Illicit barriers are system-driven.

Inadequate treatment (50%) and provider judgment (30%) dominate opioid and stimulant discourse.

# Barriers rarely travel alone



Read row-wise: given the row barrier appears, % of those posts also mentioning the column barrier. Asymmetric.

## WITHIN-POST CLUSTERS

### Shame is the connective tissue.

52% of posts citing lack of social support also disclose internalized shame.

### Provider judgment compounds CPS fear.

28% of fear-of-CPS posts also describe provider judgment.

### Methodological note

Single-label coding would have flattened a multi-dimensional experience. Multi-label annotation preserves clustered structure.

# LLM-assisted annotation extends maternal health surveillance into hard-to-reach populations

## Methods

Instruction-guided narrative interpretation by multi-LLM consensus reliably detects lived-experience discourse in unstructured online text, moving beyond conventional BERT-based classification.

## Findings

Barriers vary by substance: stigma and isolation dominate licit-substance discourse, while treatment access and provider judgment dominate illicit-substance discourse. Cannabis posts uniquely express CPS-related fear.

## Implications

Accurate self-report identification is the prerequisite for studying stigma, care avoidance, and policy-relevant concerns on Reddit. LLM-based annotation complements traditional surveillance and identify signals from populations underrepresented in clinical data.

# Thank you.

If you have questions or suggestions, please reach out to

Eunhee Park, PhD, MPH  
Web. [eunheepark.com](http://eunheepark.com)  
Email. [eunhee.park@gwu.edu](mailto:eunhee.park@gwu.edu)



---

THE GEORGE  
WASHINGTON  
UNIVERSITY

---

WASHINGTON, DC