

# Problematic Advertising and its Disparate Exposure on Facebook



**Muhammad Ali**

w. Angelica Goetzen, Alan Mislove, Elissa M. Redmiles, Piotr Sapiiezynski

USENIX Security '23

Anaheim, CA

August 11, 2023

# Ad delivery can have discriminatory outcomes

## Discrimination through optimization: How Facebook's ad delivery can lead to biased outcomes

MUHAMMAD ALI\*, Northeastern University

PIOTR SAPIEZYNSKI\*, Northeastern University

MIRANDA BOGEN, Upturn

ALEKSANDRA KOROLOVA, University of Southern California

ALAN MISLOVE, Northeastern University

AARON RIEKE, Upturn



Facebook's ad system seems to discriminate by race and gender

New research shows that Facebook's ad-distribution software is disturbingly biased

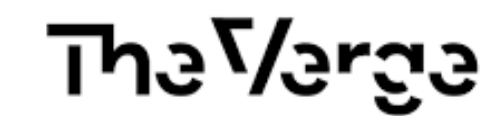


**Facebook's ad-serving algorithm discriminates by gender and race**

Even if an advertiser is well-intentioned, the algorithm still prefers certain groups of people over others.

By Karen Hao

April 5, 2019



TECH / REPORT / FACEBOOK

**Facebook's ad delivery could be inherently discriminatory, researchers say**



THE UNITED STATES  
DEPARTMENT of JUSTICE

FOR IMMEDIATE RELEASE

Tuesday, June 21, 2022

## Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising

Lawsuit is the Department's First Case Challenging Algorithmic Discrimination Under the Fair Housing Act; Meta Agrees to Change its Ad Delivery System

Meta also will develop a new system to address racial and other disparities caused by its use of personalization algorithms in its ad delivery system for housing ads.

# Expanding Our Work on Ads Fairness

June 21, 2022

By Roy L. Austin Jr, Vice President of Civil Rights and Deputy General Counsel



We are building into our ads system a method — referred to in the settlement as the “**variance reduction system**” — designed to make sure the audience that ends up seeing a housing ad more closely reflects the eligible targeted audience for that ad.



THE UNITED STATES  
DEPARTMENT *of* JUSTICE

FOR IMMEDIATE RELEASE

Tuesday, June 21, 2022

**Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising**

**Lawsuit is the Department's First Case Challenging Algorithmic Discrimination Under the Fair Housing Act; Meta Agrees to Change its Ad Delivery System**

## Expanding Our Work on Ads Fairness

June 21, 2022  
By Roy L. Austin Jr, Vice President of Civil Rights and Deputy General Counsel



## Are disparate outcomes of advertising solved?

- Maybe for housing ads alone...
- What about domains not protected by law? e.g. scams, clickbait, vulnerabilities?
- What about variances in individual experiences?



Kuala Lumpur Silent Trader

Sponsored · 🌐



Sly N Ashu-The Brand Maker

Sponsored · 🌐

Lea  
suc  
cha



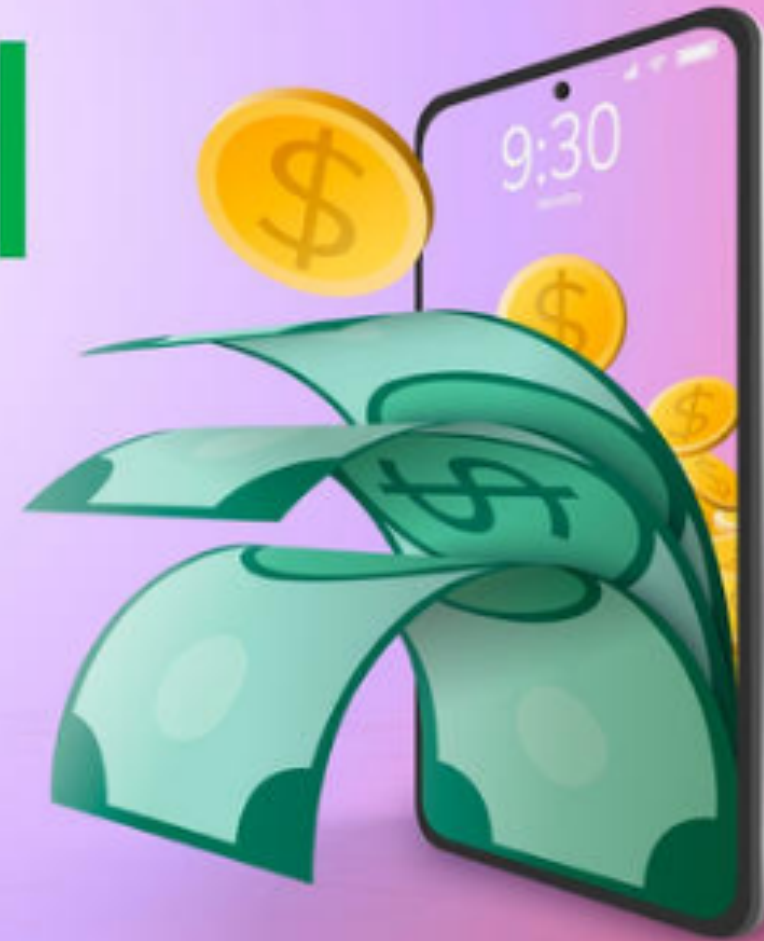
Cocktails dream

Sponsored · 🌐

Like

Earn up  
to 100\$ per month  
renting your account

Gmail



THI  
Lo



Digestinol



Tara Smith



HealthyWage

Sponsored · 🌐

Time to shed the winter coat! Join our FREE \$5,000 Wellness Challenge now and get paid to get fit!

Get Paid  
To Get  
Healthy!



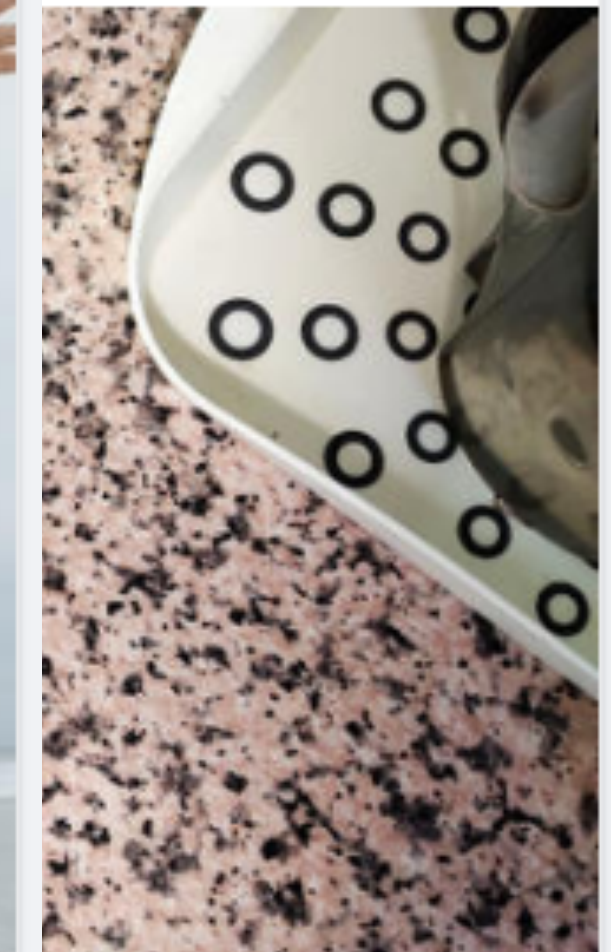
HEALTHYWAGE.COM

Limited Time: FREE \$5,000 Wellness Challenge

Limited Time: FREE \$5,000 Wellness Challenge

as

can finally enjoy  
en I retired was  
problem...I constantly  
ause I was 234+ lbs.  
and I could barely  
on't remember  
g my 3 kids it never



ars

# User-informed “bad” ads exist in the marketplace



[Liza Gak et al., CSCW '22]



[Eric Zeng et al., ConPro '20]

---

Clickbait

---

Deceptive, Untrustworthy

---

Don't Like the Product or Topic

---

Offensive, Uncomfortable, Distasteful

---

Politicized

---

Pushy, Manipulative

---

[Eric Zeng et al., CHI '21]

1. Gak et al.. "The Distressing Ads That Persist: Uncovering The Harms of Targeted Weight-Loss Ads Among Users with Histories of Disordered Eating." CSCW '22
2. Zeng et al."Bad News: Clickbait and Deceptive Ads on News and Misinformation Websites." ConPro '20
3. Zeng et al. 'What Makes a “Bad” Ad? User Perceptions of Problematic Online Advertising.' CHI '21

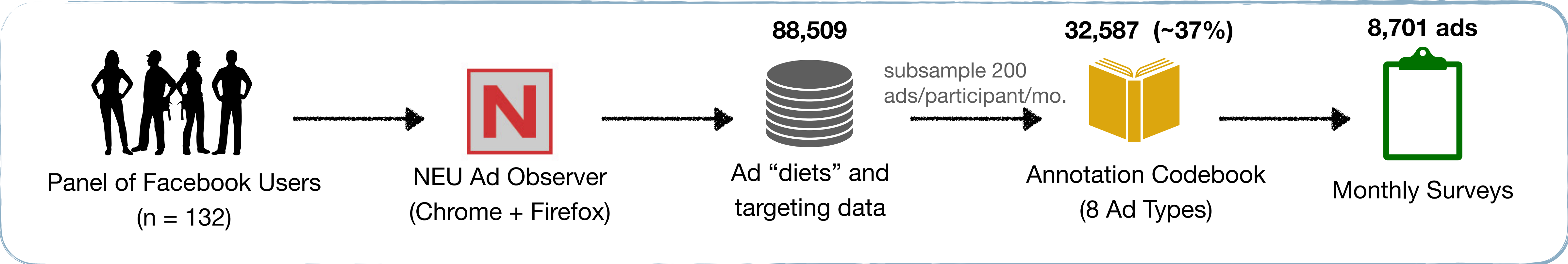
# In this talk: Problematic advertising and its disparate exposure on Facebook

## Research Questions

1. What types of ads do users consider problematic? **RQ1**
2. Are there skews in the distribution of such ads? **RQ2**
3. Who is responsible for skews?  
The advertisers or ad delivery/personalization? **RQ3**

# Methodology

Nov. 2021 — Sep. 2022 (11 months); rolling recruitment; each participant stays 3 months







# Categorizing ads with a codebook

Pilot Data Collection

Mixture of inductive qualitative coding from collected ads + deductive analysis of prior work and Facebook policies

## Prohibited or demoted

- Deceptive
- Potentially Prohibited
- Clickbait

[Zeng et al., CHI '21]

## Increased scrutiny

- Sensitive
  - Financial
  - Gambling
  - Alcohol
  - Weight loss
  - Online pharmacies
  - Prescription + over-the-counter drugs

[Gak et al., CSCW '22]


- Healthcare
- Opportunity
- Neutral



Deceptive

Kuala Lumpur Silent Trader  
Sponsored · 🌐

Tired of losing so much Money in Forex? Come Join my Channel <https://t.me/SilentTraderKL> <https://t.me/SilentTraderKL> <https://t.me/SilentTraderKL> See more




[HTTPS://T.ME/SILENTRADERKL](https://t.me/SilentTraderKL)  
<https://t.me/SilentTraderKL>  
<https://t.me/SilentTraderKL> See more

Clickbait

Grateful Neighbor  
Sponsored · 🌐

Unemployed Americans ( Aged 49 - 62 ) Without Disability Benefits Are Entitled To Monthly Assistance Thanks To This New Service. To Qualify You Must Meet 3 Requirements.- Must not be receiving Disability Benefits- Must be an US citizen - Must be between 49 - 62




DISABILITY-HI  
ALL 50 US !  
Free 30 Secon

Sensitive: Financial

Upstart  
Sponsored · 🌐

Pay off \$1,000-\$50,000 today. Checking your rate doesn't hurt your credit score!

"I was drowning in credit card debt"




Sensitive: Other

HealthyWage  
Sponsored · 🌐

Time to shed the winter coat! Join our FREE \$5,000 Wellness Challenge now and get paid to get fit!

Get Paid To Get Healthy!



HealthyWage

HEALTHYWAGE.COM  
Limited Time: FREE \$5,000 Wellness Challenge  
Limited Time: FREE \$5,000 Wellness Challenge

Code	Count	%
Neutral	20,596	68.52
Healthcare	3564	11.86
Opportunity	2267	7.54
Sensitive: Financial	1429	4.75
Sensitive: Other	631	2.10
Clickbait	1182	3.93
Deceptive	542	1.80
Potentially Prohibited	253	0.84

Neutral

kate spade new york  
Sponsored · 🌐

psst... we have some big news... everything's up to 75% off! it true! shop surprise sale now.



Pot. Prohibited



Become a Cybersecurity Professional Online in 24 Weeks

- Learn skills such as: Defensive and offensive cybersecurity, networking, systems, web technologies and databases
- 1:1 career services support
- Part-time schedule — keep your day job

SMU CONTINUING AND PROFESSIONAL EDUCATION

TECHBOOTCAMP.SMU.EDU  
Become a Cybersecurity Analyst in 24 Weeks (Download Our Curriculum Outline for Free!)  
Learn from a name employers trust

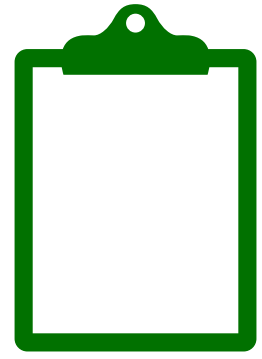
d all-natural Aloe based product which has  
by individuals for relief from digestive

diseases/disorders

Digestinol Research

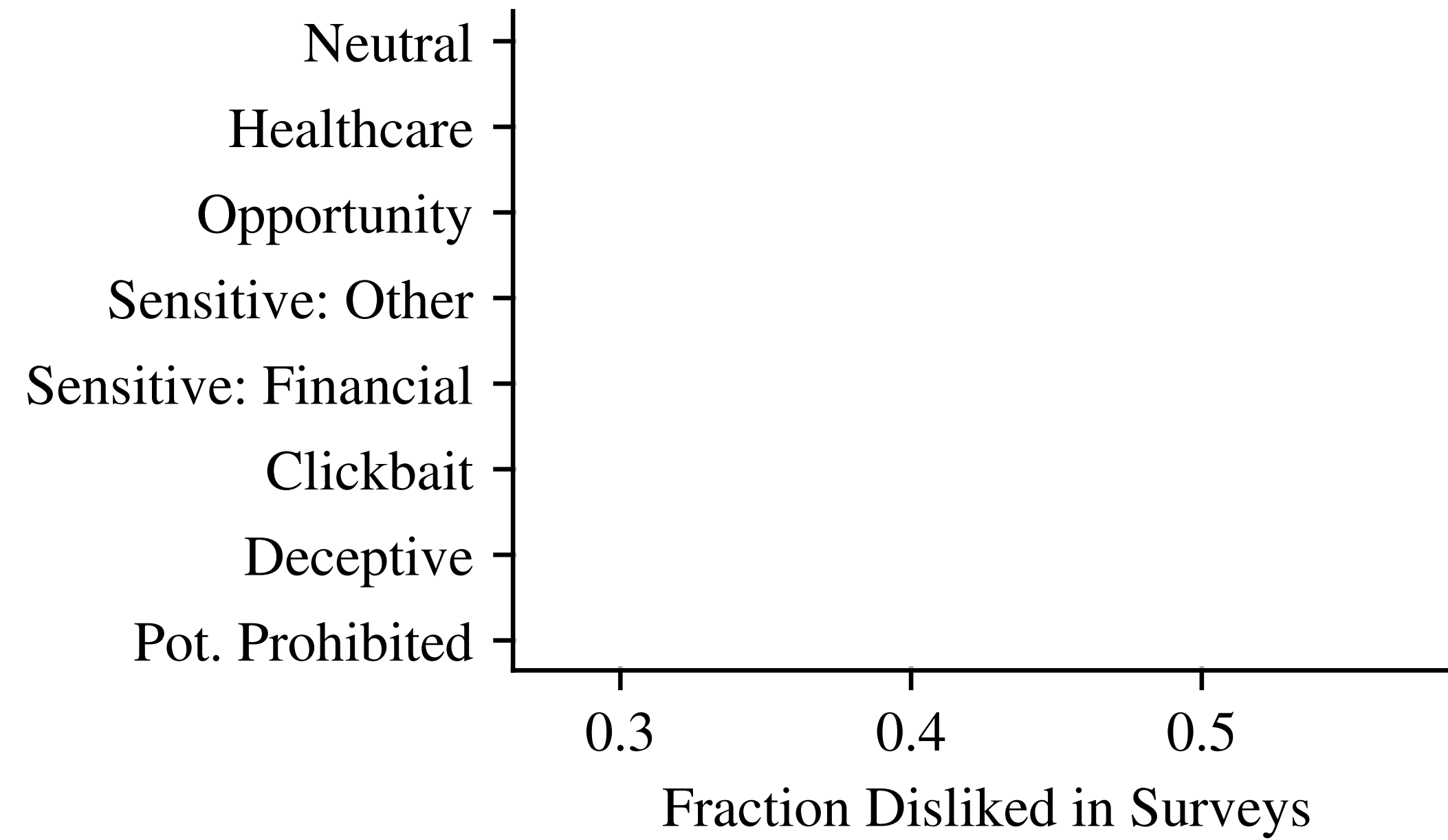


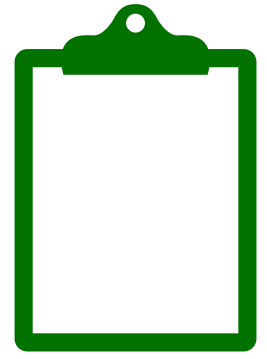
DIGESTINOL.COM  
Digestinol - Colon Health



# RQ1: Which categories of ads do participants perceive as problematic?

Monthly Surveys

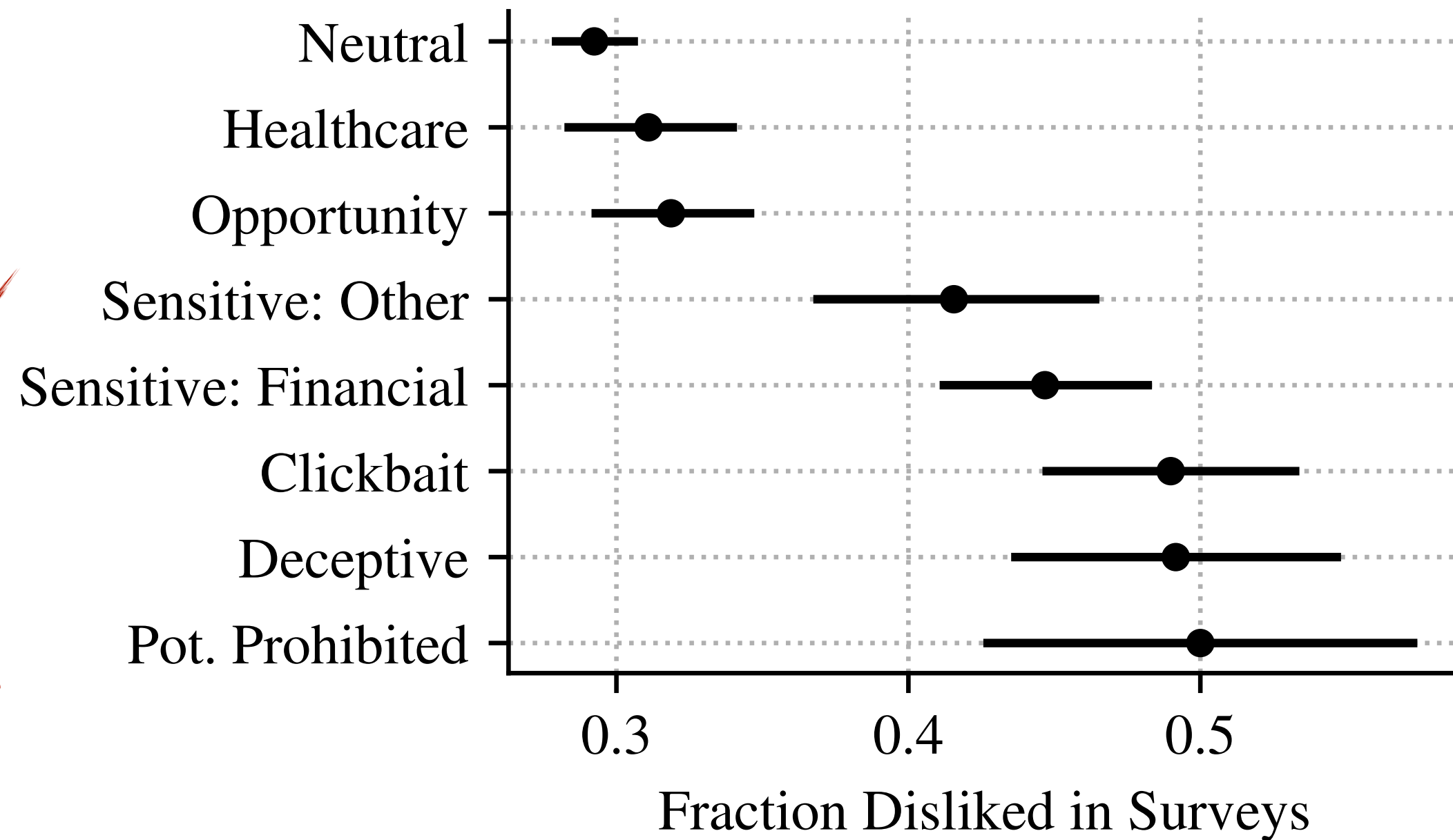




# RQ1: What categories of ads do participants perceive as problematic?

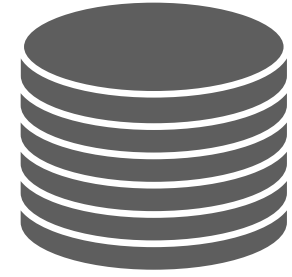
Monthly Surveys

Problematic



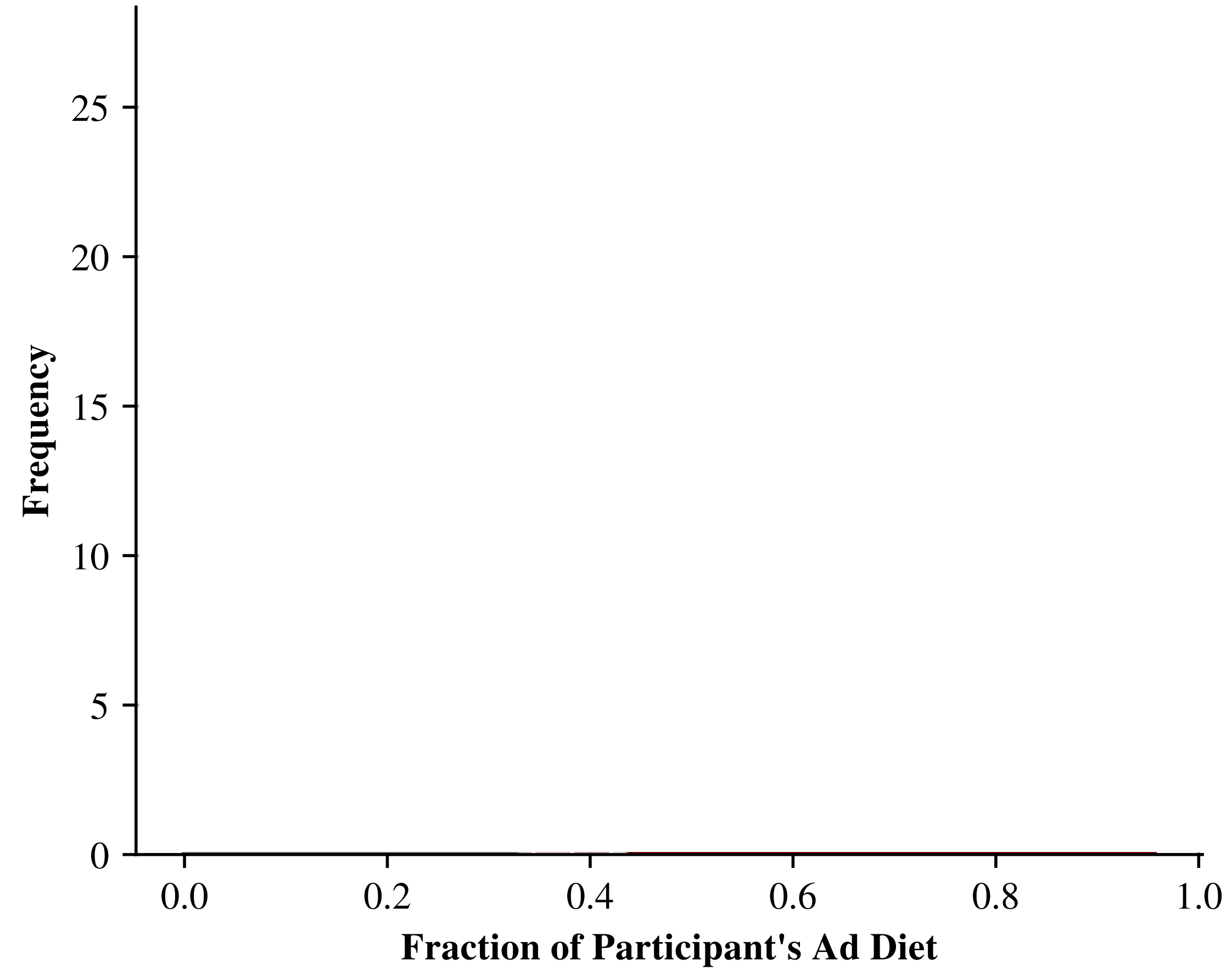
Reasons of dislike? (compared to Neutral)

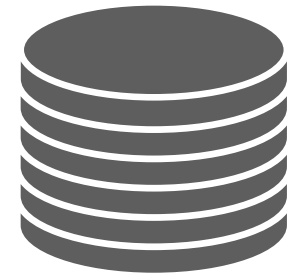
- They have higher odds of being considered **irrelevant, clickbait, scam.**
- Sensitive ads have higher odds of being disliked due to the **advertiser** of the **product.**



## RQ2: Are there skews in the distribution of problematic ads?

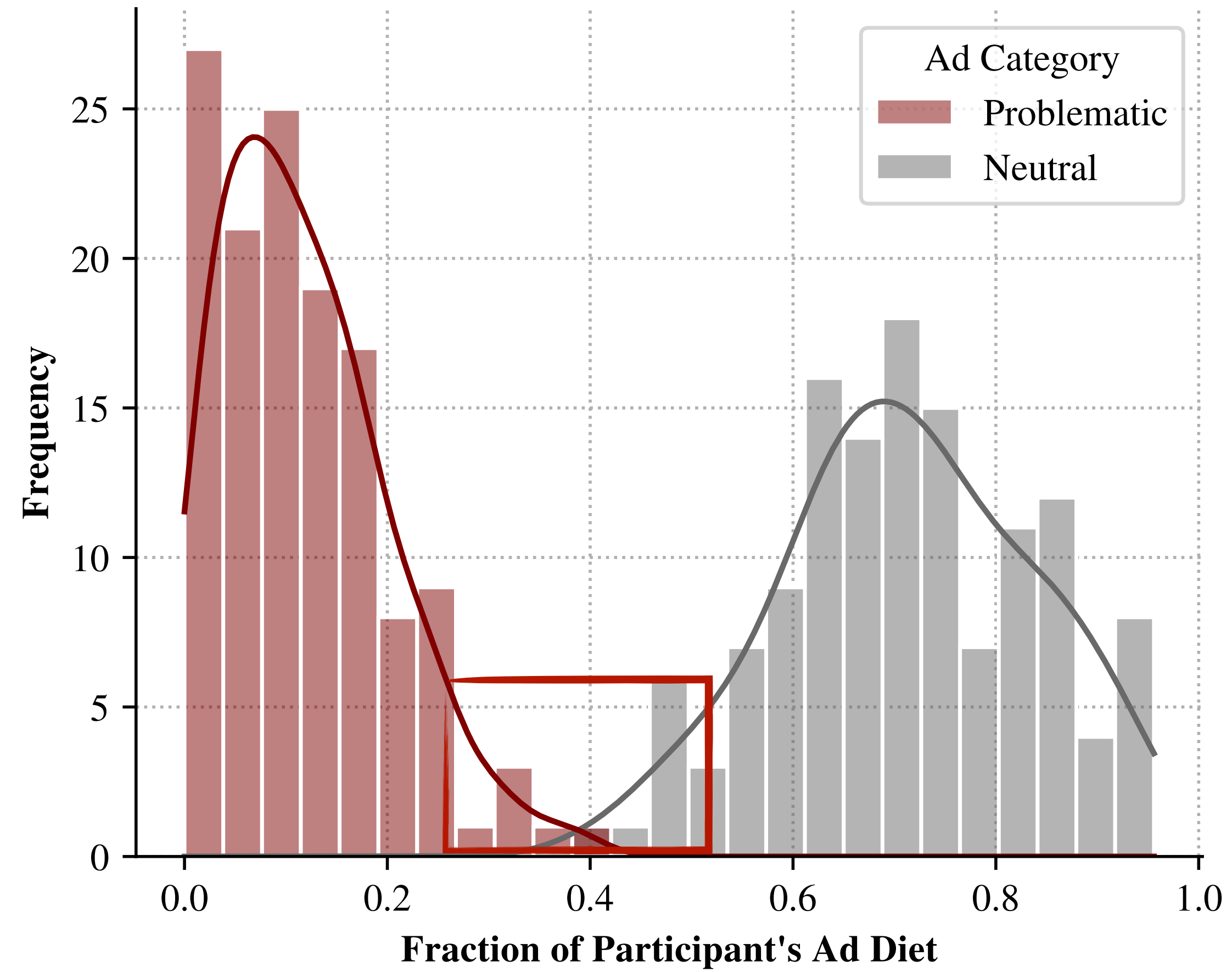
Participants' Ad "Diet"

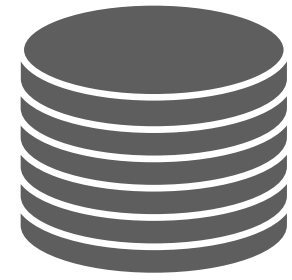




# RQ2: Are there skews in the distribution of problematic ads?

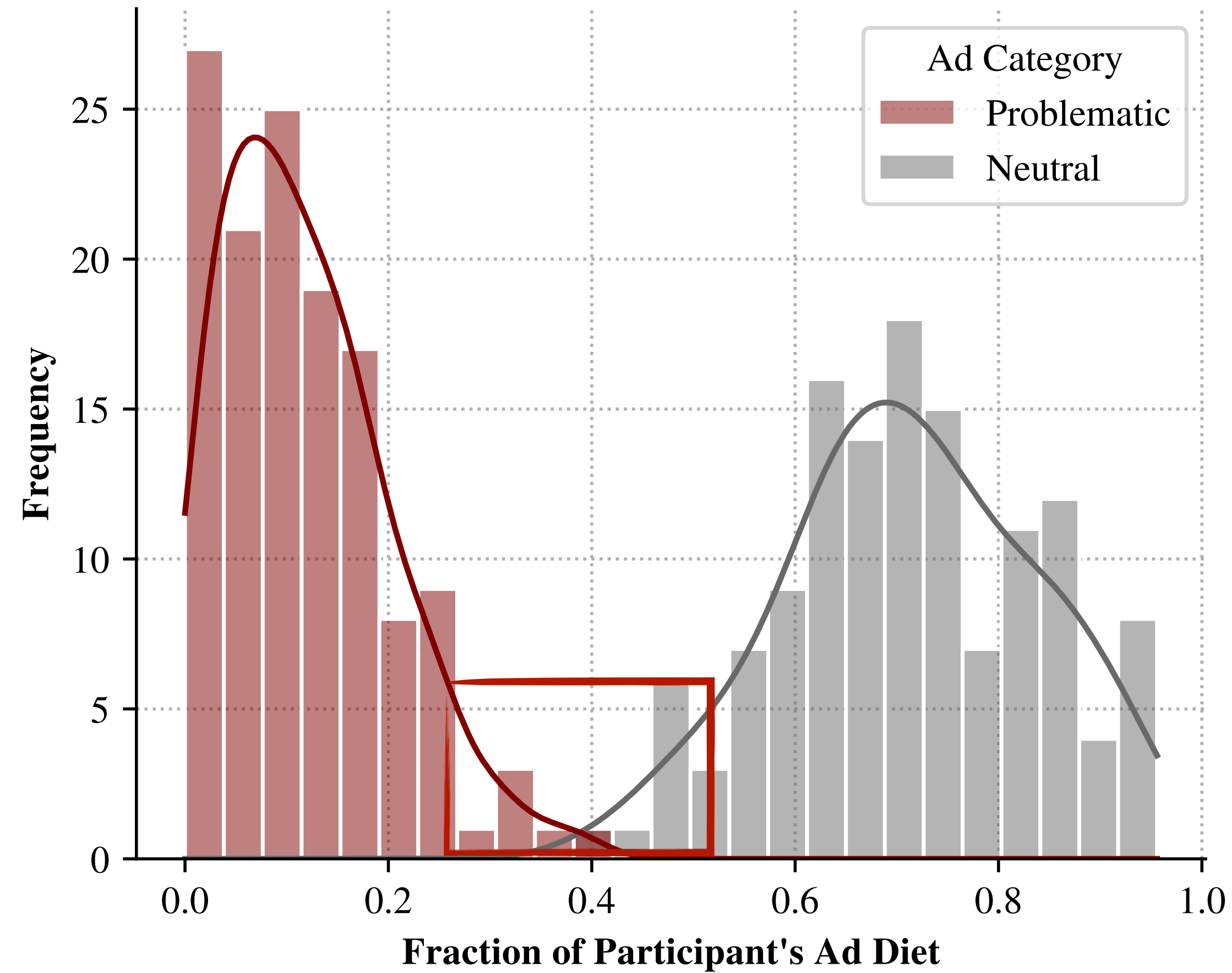
Participants' Ad "Diet"





# RQ2: Are there skews in the distribution of problematic ads?

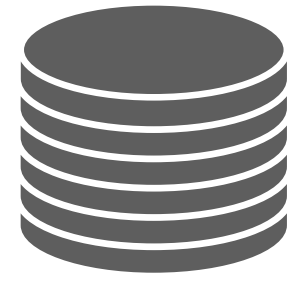
Yes. Do they relate to participant demographics?



Linear Regression

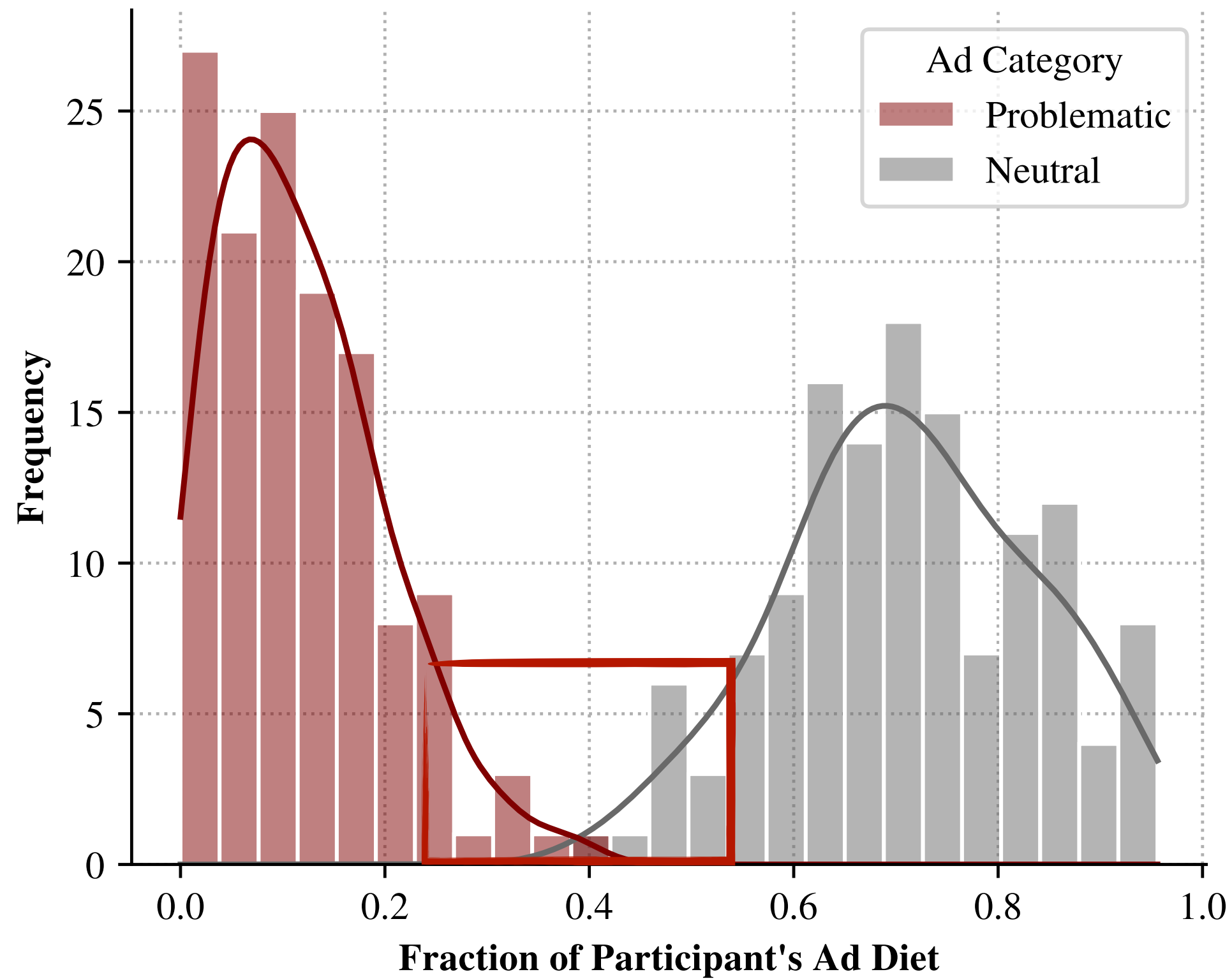
e.g. `fraction_clickbait ~ woman + hispanic + older + ...`



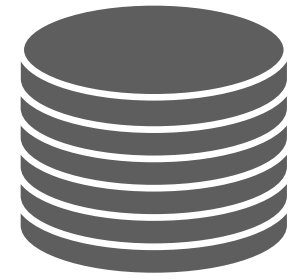


# How are problematic ad skews related to demographics?

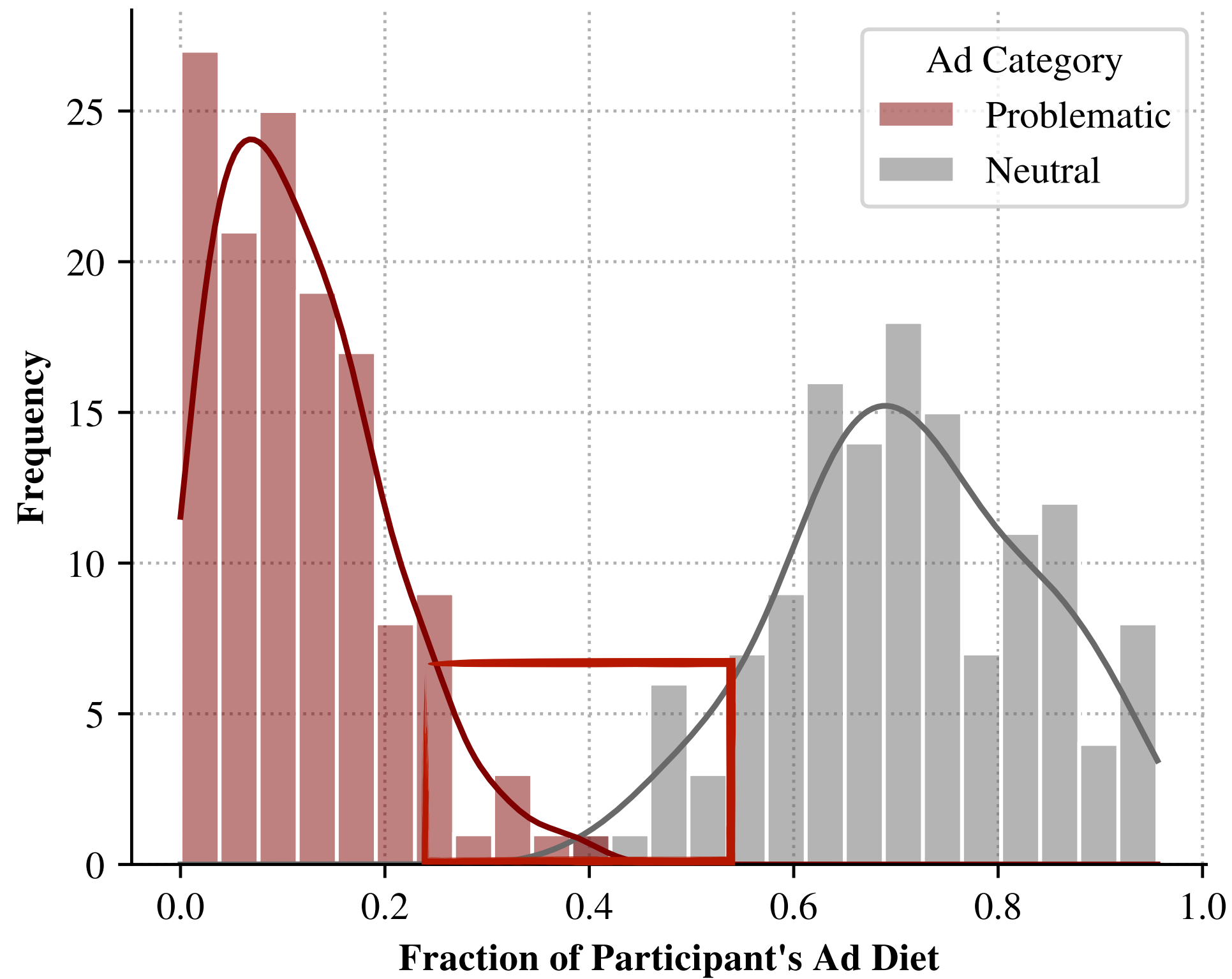
Ad Diets, Linear Regression



Variable	Estimate ( $\beta$ )					
	Problematic	Pot. Prohibited	Deceptive	Clickbait	Sensitive: Financial	Sensitive: Other
Intercept						
Gender: Woman						
Race: Black						
Race: Asian						
Ethnicity: Hispanic						
Education: college and above						
Age: Gen-X and older						

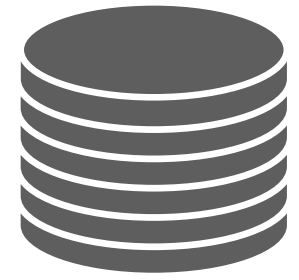


# How are problematic ad skews related to demographics?

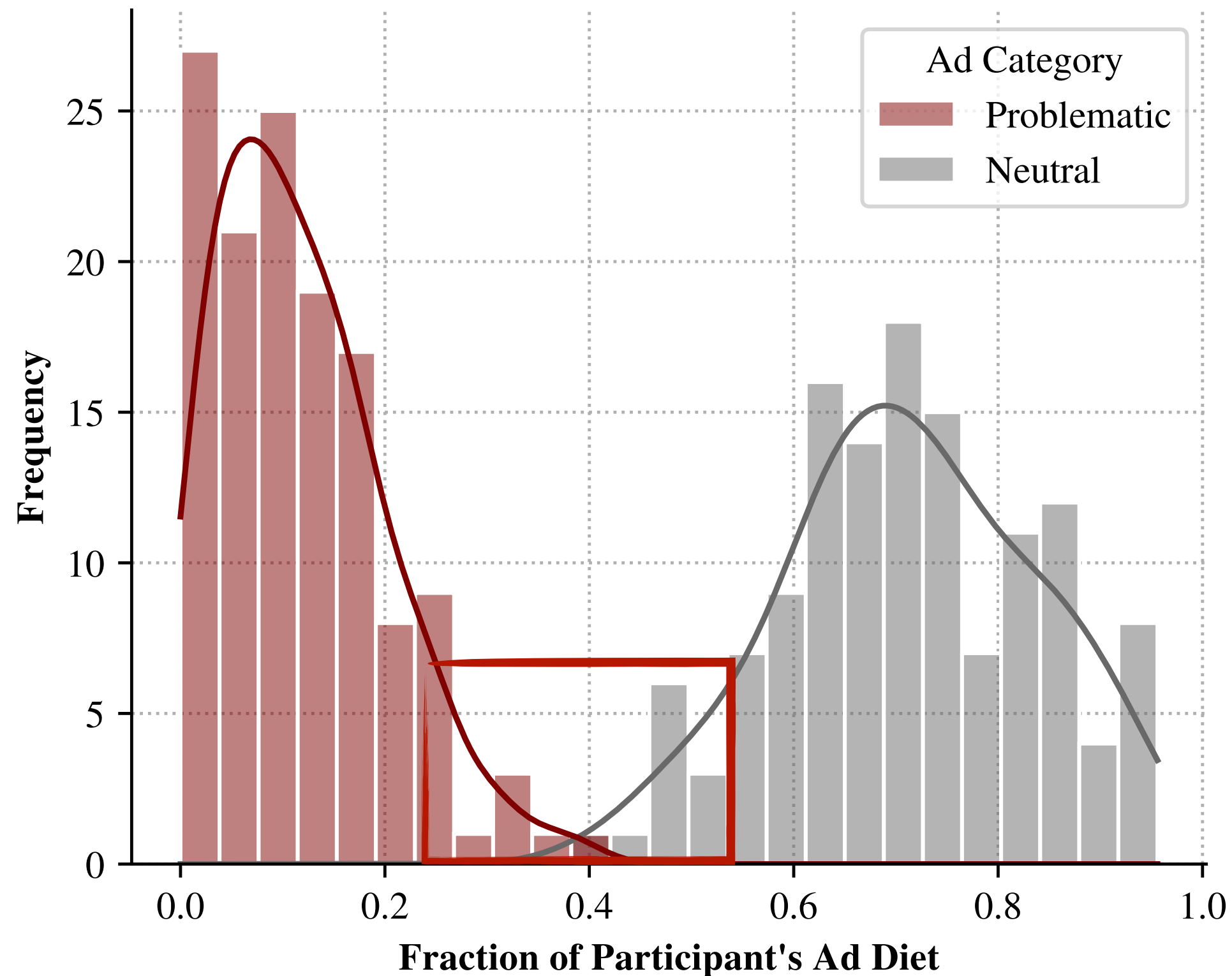


Variable	Estimate ( $\beta$ ) [95% CI]					
	Problematic	Pot. Prohibited	Deceptive	Clickbait	Sensitive: Financial	Sensitive: Other
Intercept	0.12*** [0.09, 0.15]					
<b>Gender: Woman</b>	<b>-0.064***</b> [-0.09, -0.04]					
<b>Race: Black</b>	0.025 [-0.01, 0.06]					
<b>Race: Asian</b>	-0.002 [-0.04, 0.04]					
<b>Ethnicity: Hispanic</b>	0.023 [-0.03, 0.08]					
<b>Education: college and above</b>	0.01 [-0.02, 0.04]					
<b>Age: Gen-X and older</b>	<b>0.051***</b> [0.02, 0.08]					

**Older participants** see **5.1 pp more** Problematic ads.  
 Participants identifying as **women** see **6.4 pp fewer** Problematic ads.



# How are problematic ad skews related to demographics?

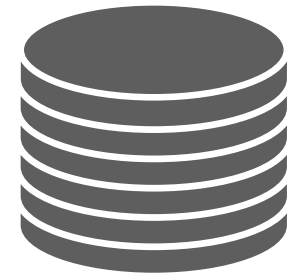


Variable	Estimate ( $\beta$ ) [95% CI]					
	Problematic	Pot. Prohibited	Deceptive	Clickbait	Sensitive: Financial	Sensitive: Other
Intercept	0.12*** [0.09, 0.15]	0.01*** [0.01, 0.01]	0.008 [0, 0.02]	0.012 [0, 0.02]		
Gender: Woman	<b>-0.064***</b> [-0.09, -0.04]	-0.002 [0, 0]	-0.005 [-0.01, 0]	-0.008 [-0.02, 0]		
Race: Black	0.025 [-0.01, 0.06]	-0.001 [0, 0]	0.006 [0, 0.02]	<b>0.013*</b> [0, 0.02]		
Race: Asian	-0.002 [-0.04, 0.04]	0.001 [0, 0.01]	-0.003 [-0.02, 0.01]	0.005 [-0.01, 0.02]		
Ethnicity: Hispanic	0.023 [-0.03, 0.08]	<b>-0.007*</b> [-0.01, 0]	0.005 [-0.01, 0.02]	-0.007 [-0.03, 0.01]		
Education: college and above	0.01 [-0.02, 0.04]	-0.002 [0, 0]	0.004 [0.01, 0.01]	0.01 [0, 0.02]		
Age: Gen-X and older	<b>0.051***</b> [0.02, 0.08]	<b>-0.003*</b> [-0.01, 0]	<b>0.011*</b> [0, 0.02]	<b>0.017**</b> [0.01, 0.03]		

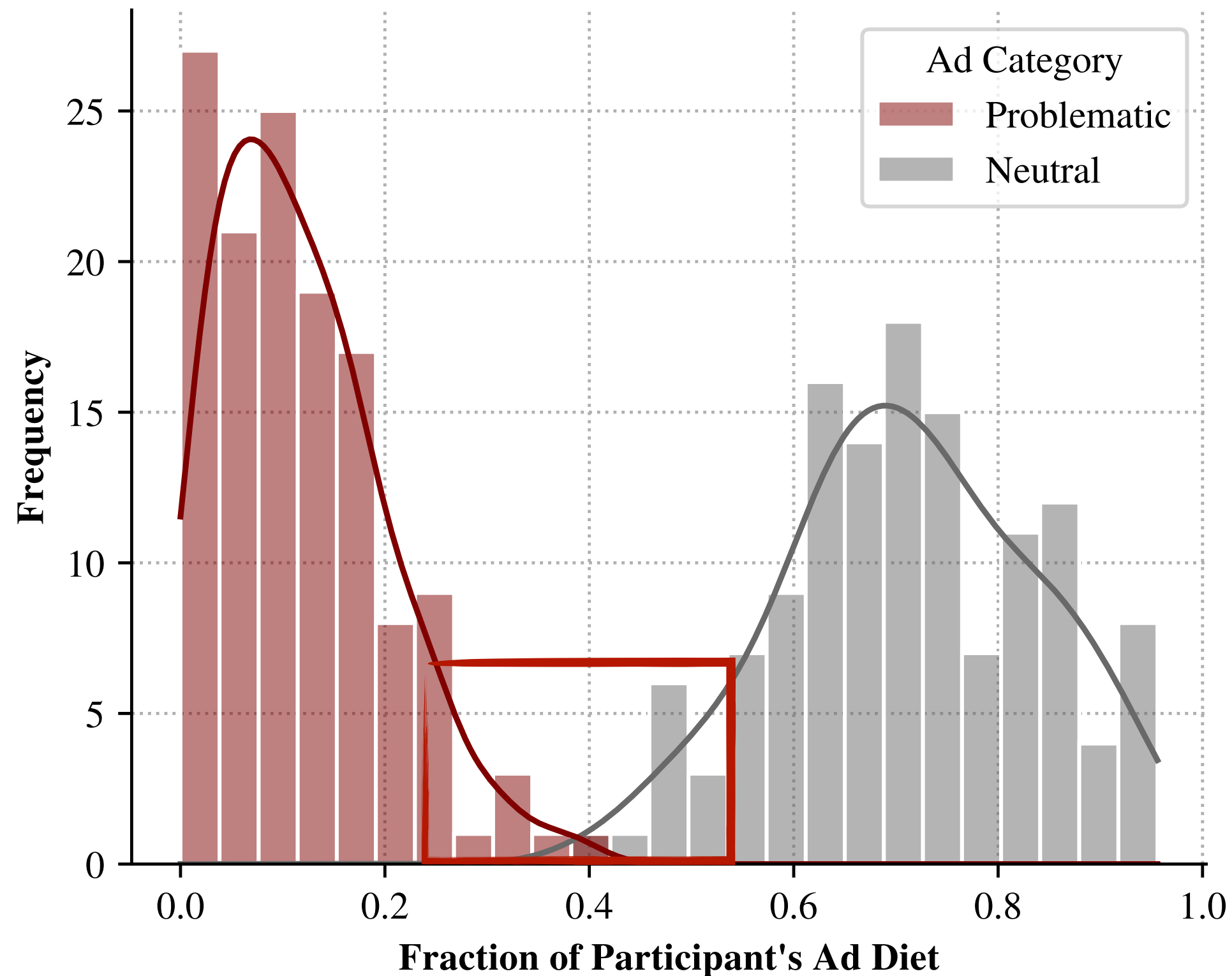
**Older participants** see **5.1 pp more** Problematic ads—including Deceptive and Clickbait content.

**Black participants** see **1.3 pp more** Clickbait than other races.

Participants identifying as **women** see **6.4 pp fewer** Problematic ads.

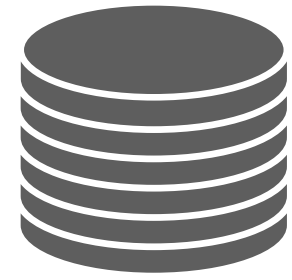


# How are problematic ad skews related to demographics?



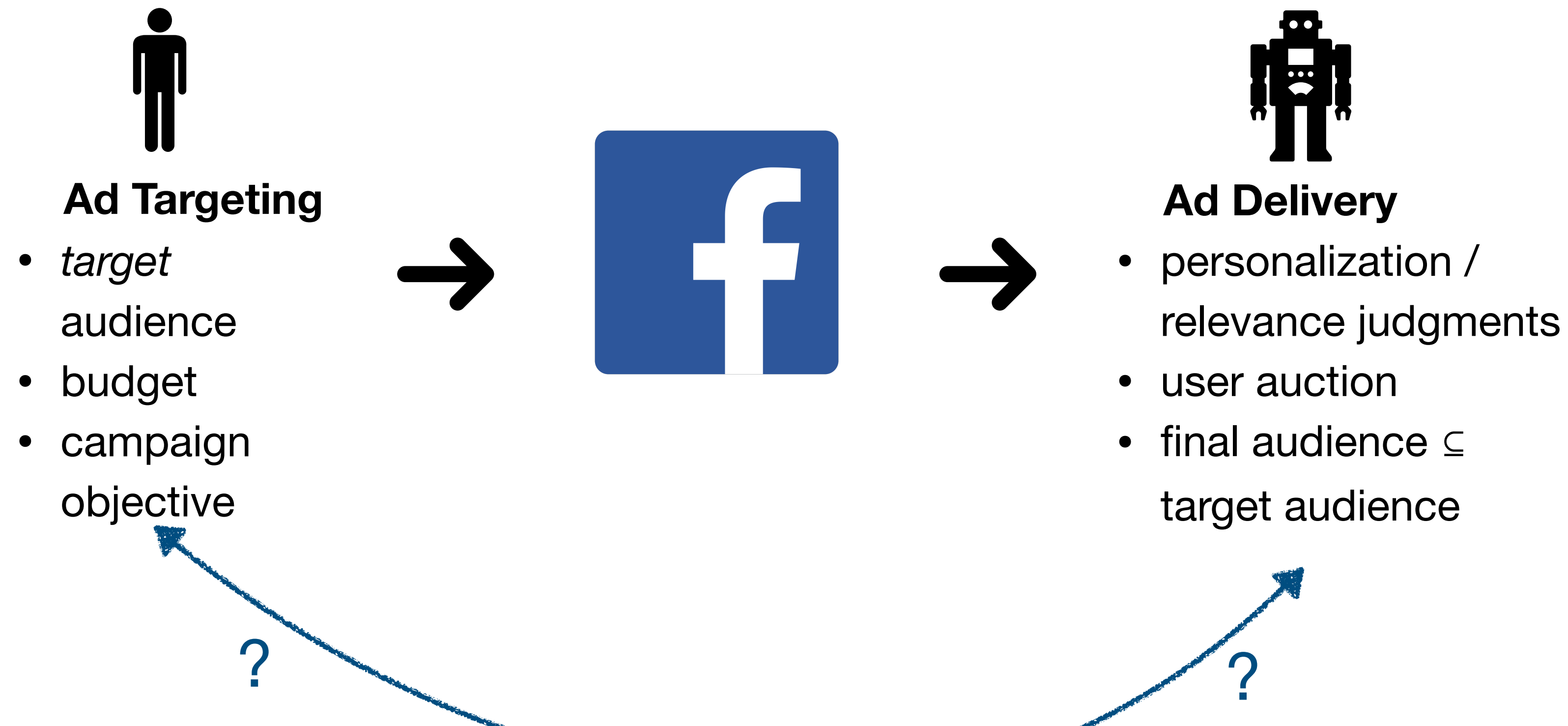
Variable	Estimate ( $\beta$ )					
	[95% CI]					
	Problematic	Pot. Prohibited	Deceptive	Clickbait	Sensitive: Financial	Sensitive: Other
Intercept	0.12*** [0.09, 0.15]	0.01*** [0.01, 0.01]	0.008 [0, 0.02]	0.012 [0, 0.02]	0.07*** [0.04, 0.1]	0.02** [0.01, 0.03]
Gender: Woman	<b>-0.064***</b> [-0.09, -0.04]	-0.002 [0, 0]	-0.005 [-0.01, 0]	-0.008 [-0.02, 0]	<b>-0.045***</b> [-0.07, -0.02]	-0.004 [-0.02, 0.01]
Race: Black	0.025 [-0.01, 0.06]	-0.001 [0, 0]	0.006 [0, 0.02]	<b>0.013*</b> [0, 0.02]	0.004 [-0.02, 0.03]	0.002 [-0.01, 0.02]
Race: Asian	-0.002 [-0.04, 0.04]	0.001 [0, 0.01]	-0.003 [-0.02, 0.01]	0.005 [-0.01, 0.02]	-0.007 [-0.04, 0.03]	0.002 [-0.02, 0.02]
Ethnicity: Hispanic	0.023 [-0.03, 0.08]	<b>-0.007*</b> [-0.01, 0]	0.005 [-0.01, 0.02]	-0.007 [-0.03, 0.01]	0.036 [-0.01, 0.08]	-0.003 [-0.02, 0.02]
Education: college and above	0.01 [-0.02, 0.04]	-0.002 [0, 0]	0.004 [-0.01, 0.01]	0.01 [0, 0.02]	-0.003 [-0.03, 0.02]	0 [-0.01, 0.01]
Age: Gen-X and older	<b>0.051***</b> [0.02, 0.08]	<b>-0.003*</b> [-0.01, 0]	<b>0.011*</b> [0, 0.02]	<b>0.017**</b> [0.01, 0.03]	0.017 [-0.01, 0.04]	0.009 [0, 0.02]

**Older participants** see **5.1 pp more** Problematic ads—including Deceptive and Clickbait content.  
**Black participants** see **1.3 pp more** Clickbait than other races.  
 Participants identifying as **women** see **6.4 pp fewer** Problematic ads—largely due to lower exposure to Financial ads.

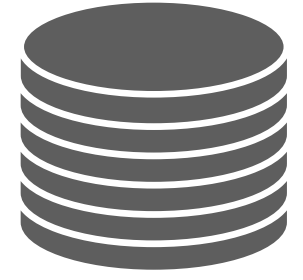


# RQ3: Who is responsible for skews? Advertisers or algorithms?

Ad targeting data, “Why am I seeing this?”



**Older participants** see **5.1 pp more** Problematic ads—including Deceptive and Clickbait content.

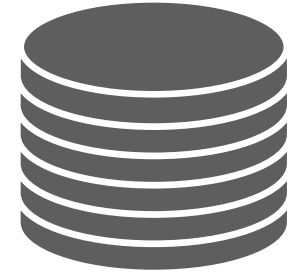


## RQ3: Who is responsible for skews? Advertisers or algorithms?

Deep-dive into age: are advertisers targeting older participants?



Age targeting has high usage in our data, 49.7% ads. Compared to only 12.1% ads using gender targeting.

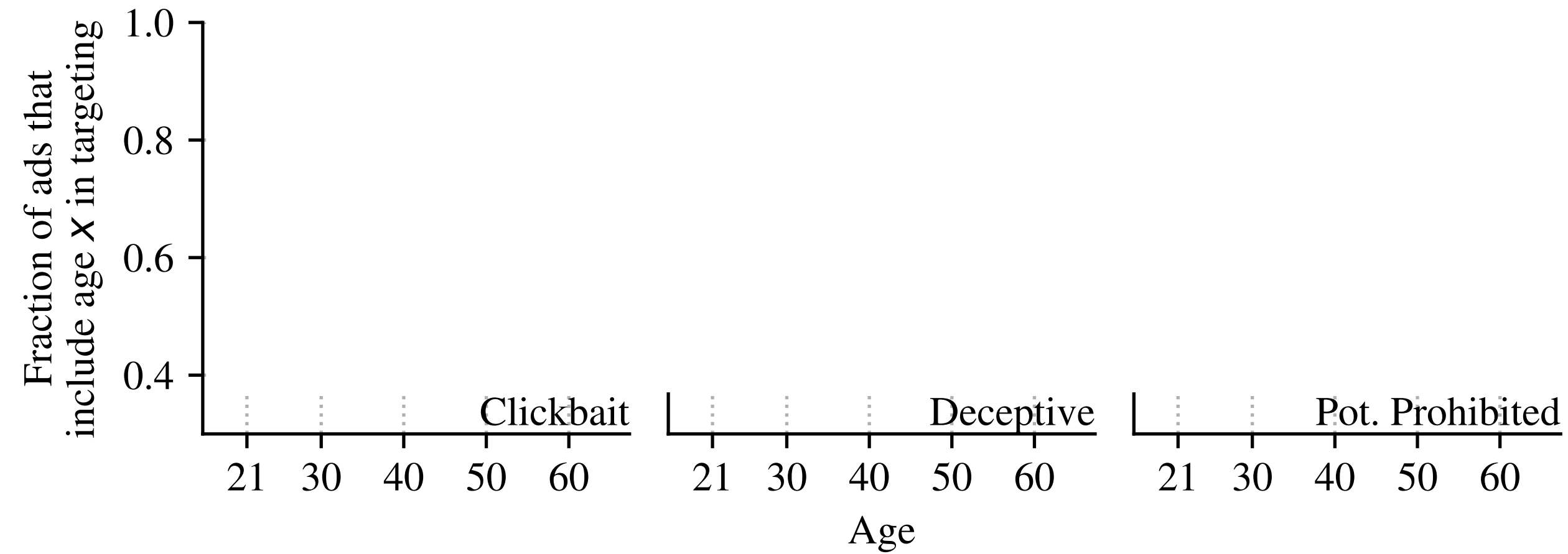


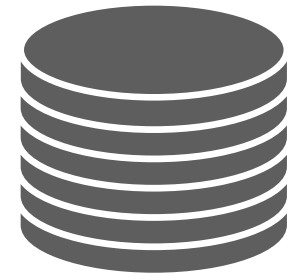
# RQ3: Who is responsible for skews? Advertisers or algorithms?

Deep-dive into age: are advertisers targeting older participants?



Age targeting has high usage in our data, 49.7% ads. Compared to only 12.1% ads using gender targeting.



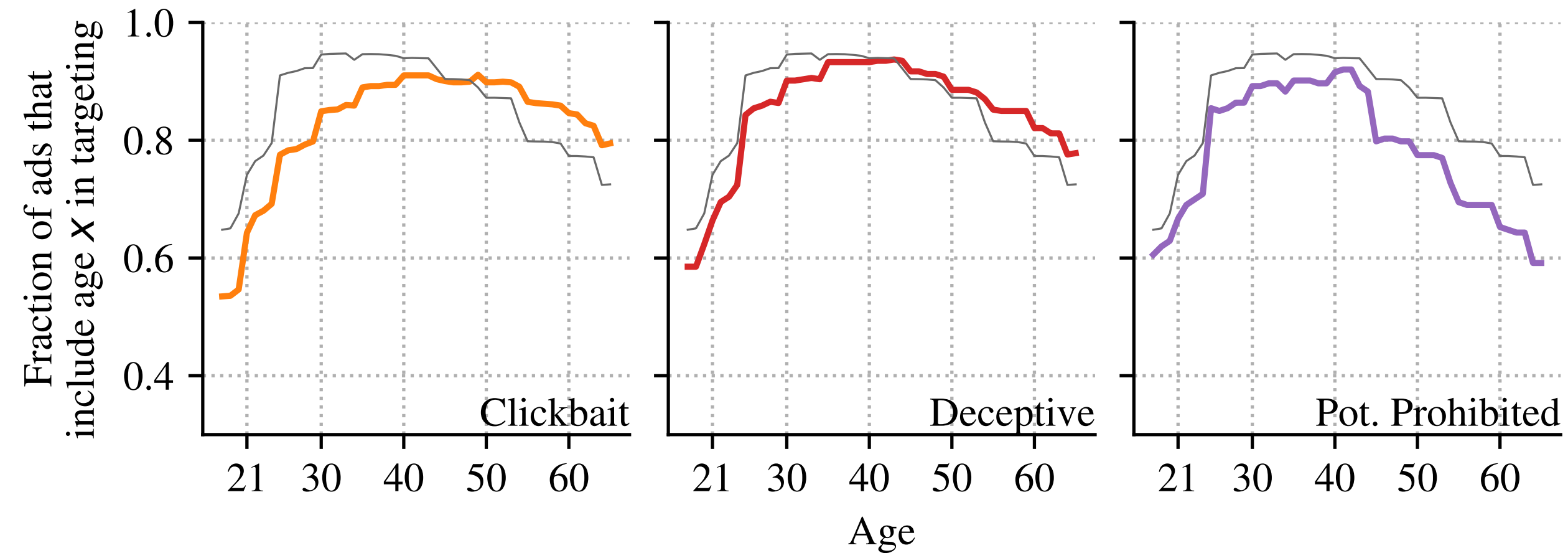


# RQ3: Who is responsible for skews? Advertisers or algorithms?

Deep-dive into age: are advertisers targeting older participants?



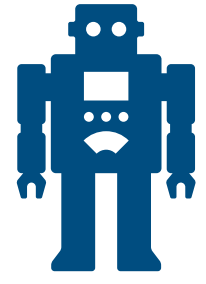
Age targeting has high usage in our data, 49.7% ads. Compared to only 12.1% ads using gender targeting.



Advertisers' targeting aligns with observed skews: **Clickbait** and **Deceptive** is actively **targeted to older users**. **Pot. Prohibited** is **targeted less** to older users.

So advertisers are clearly responsible, what about algorithms?





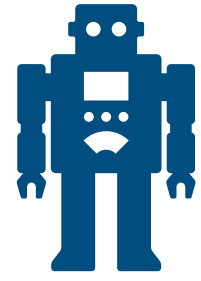
# Isolating algorithm's influence: ads with “default” targeting

Advertiser has no preference whatsoever

21.2% ads target to all adults in the US, i.e. 267 million users

Linear Regression (as before)  
on subset of default targeting ads

e.g. `fraction_clickbait` ~ woman + hispanic + older + ...



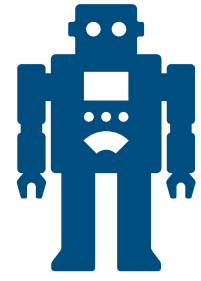
# Isolating algorithm's influence: ads with "default" targeting

Advertiser has no preference whatsoever

Linear Regression (as before)  
on subset of default targeting ads

e.g. `fraction_clickbait` ~ woman + hispanic + older + ...

Variable	Estimate ( $\beta$ )				
	Problematic	Pot. Prohibited	Deceptive	Clickbait	[95% CI]
Intercept					
<b>Gender: Woman</b>					
<b>Race: Black</b>					
race: Asian					
<b>Ethnicity: Hispanic</b>					
<b>Education: college and above</b>					
<b>Age: Gen-X and older</b>					



# Isolating algorithm's influence: ads with "default" targeting

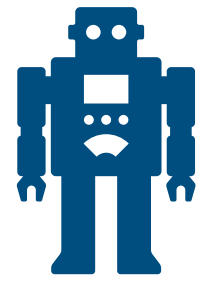
Advertiser has no preference whatsoever

Variable	Estimate ( $\beta$ )					
	[95% CI]					
	Problematic	Pot. Prohibited	Deceptive	Clickbait	Sensitive: Financial	Sensitive: Other
Intercept	0.191*** [0.13, 0.26]					
<b>Gender: Woman</b>	<b>-0.059*</b> [-0.11, -0.01]					
<b>Race: Black</b>	0.01 [-0.05, 0.07]					
<b>race: Asian</b>	-0.019 [-0.1, 0.06]					
<b>Ethnicity: Hispanic</b>	0.017 [-0.08, 0.12]					
<b>Education: college and above</b>	-0.033 [-0.09, 0.02]					
<b>Age: Gen-X and older</b>	<b>0.077**</b> [0.02, 0.13]					

Even within ads with the broadest possible targeting:

**Older participants** (still) see **7.7 pp more** Problematic ads.

Participants identifying as **women** (still) see **5.9 pp fewer** Problematic ads.



# Isolating algorithm's influence: ads with “default” targeting

Advertiser has no preference whatsoever

Variable	Estimate ( $\beta$ )					
	[95% CI]					
	Problematic	Pot. Prohibited	Deceptive	Clickbait	Sensitive: Financial	Sensitive: Other
Intercept	0.191*** [0.13, 0.26]	0.013*** [0.01, 0.02]	0.014* [0, 0.03]	0.023 [-0.01, 0.05]	0.133*** [0.08, 0.18]	0.009* [0, 0.02]
Gender: Woman	<b>-0.059*</b> [-0.11, -0.01]	<b>-0.006*</b> [-0.01, 0]	-0.007 [-0.02, 0]	-0.003 [-0.03, 0.02]	<b>-0.046*</b> [-0.09, 0]	0.004 [0, 0.01]
Race: Black	0.01 [-0.05, 0.07]	0.002 [0, 0.01]	0.007 [-0.01, 0.02]	0.011 [-0.02, 0.04]	-0.007 [-0.06, 0.04]	-0.003 [-0.01, 0]
race: Asian	-0.019 [-0.1, 0.06]	-0.005 [-0.01, 0]	-0.003 [-0.02, 0.01]	-0.007 [-0.04, 0.03]	-0.003 [-0.07, 0.06]	0 [-0.01, 0.01]
Ethnicity: Hispanic	0.017 [-0.08, 0.12]	-0.009 [-0.02, 0]	<b>0.028**</b> [0.01, 0.05]	-0.021 [-0.06, 0.02]	0.027 [-0.05, 0.11]	-0.008 [-0.02, 0]
Education: college and above	-0.033 [-0.09, 0.02]	-0.002 [-0.01, 0]	0 [-0.01, 0.01]	0.005 [-0.02, 0.03]	-0.036 [-0.08, 0.01]	-0.001 [-0.01, 0.01]
Age: Gen-X and older	<b>0.077**</b> [0.02, 0.13]	-0.003 [-0.01, 0]	0.011 [0, 0.02]	<b>0.041**</b> [0.02, 0.06]	0.034 [-0.01, 0.08]	-0.005 [-0.01, 0]

Even within ads with the broadest possible targeting:

**Older participants** (still) see **7.7 pp more** Problematic ads—**4.1 pp more Clickbait** content.  
 Participants identifying as **women** (still) see **5.9 pp fewer** Problematic ads—largely due to lower exposure to Financial ads.  
 New effect: **Hispanic participants** see **2.8 pp more** Deceptive than non-Hispanic participants.

# Summary + takeaways

- First study of real user experiences with problematic ads—provides an understanding of disparate exposure through lived experiences
- Malicious advertisers are aware of vulnerable populations, and do use tools at their disposal to run ads
- Even if advertisers are not aware, personalization will roll out the red carpet
- Personalization and malicious advertisers together can expose vulnerable users to harmful content
- In addition to moderation, platforms might need to limit optimization as well—*proposal*: stop personalization altogether for problematic content
- Transparency is valuable, despite platforms being resistant to studies

# Thank you, USENIX Security!



More results +  
discussion in full  
paper!

## Questions?

[ali.muh@northeastern.edu](mailto:ali.muh@northeastern.edu)  
[@lukshmichowk](https://twitter.com/lukshmichowk)



# Backup Slides

# Panel Demographics

Variable	Value	Recruited		Active		Census
		n	%	n	%	%
Gender	Female	96	52.17	71	53.79	50.5
	Male	86	46.74	59	44.70	49.5
	Non-binary	2	1.09	2	1.52	–
Age	Younger than Gen-X	134	72.83	88	66.67	33.6
	Gen-X and older	50	27.17	44	33.33	47.8
Race / Ethnicity	White	105	57.07	82	62.12	75.8
	Latino/Hispanic	21	11.41	16	12.12	18.9
	Black	53	28.80	32	24.24	13.6
	Asian	21	11.41	16	12.12	6.1
	Other	3	1.63	3	2.27	–
Education	Below Bachelor's	72	39.13	51	38.64	58.5
	Bachelor's or above	112	60.87	81	61.36	32.9
<b>Total</b>		<b>184</b>		<b>132</b>		

Table 1: Demographics of panel participants.



# Survey Instrument

**Q1.** How would you describe the advertised product/offer's relevance to you?

[Completely Irrelevant]

[Irrelevant]

[Neutral]

[Relevant]

[Completely Relevant]

**Q2.** Which of the following, if any, describe your reasons for disliking this ad?

- It is **irrelevant** to me, or does not contain interesting information.
- I do not like the **design** of the ad.
- It contains **clickbait**, sensationalized, or shocking content.
- I do not trust this ad, it seems like a **scam**.
- I dislike the **advertiser**.
- I dislike the type of **product** being advertised.
- I find the content **uncomfortable**, offensive, or repulsive.
- I dislike the **political** nature of the ad.
- I find the ad **pushy** or it causes me to feel anxious.
- I cannot tell what is being advertised. (**unclear**)
- I do not dislike this ad.

**Q3.** Which of the following, if any, describe your reasons for liking this ad?

- The content is engaging, clever or **amusing**
- It is well **designed** or eye-catching.
- I am **interested** in what is being advertised.
- It is **clear** what product the ad is selling.
- I **trust** the ad, it looks authentic or trustworthy.
- I trust the **advertiser**.
- It is **useful**, interesting, or informative.
- It clearly looks like an ad and can be **filtered** out.
- I do not like this ad

[Q2 and Q3 from Zeng et. al., CHI '21]