



PREVENTING AND DETECTING FRAUD IN ONLINE SURVEYS

Elizabeth Wu, MPH
Director of Research Programs
UCLA Hub for Health Intervention, Policy & Practice

What is fraud in online research?

BOT

Short for “robot”

A programmed digital imposter that can be automated to infiltrate online research studies in high volume

Deceptive Respondent

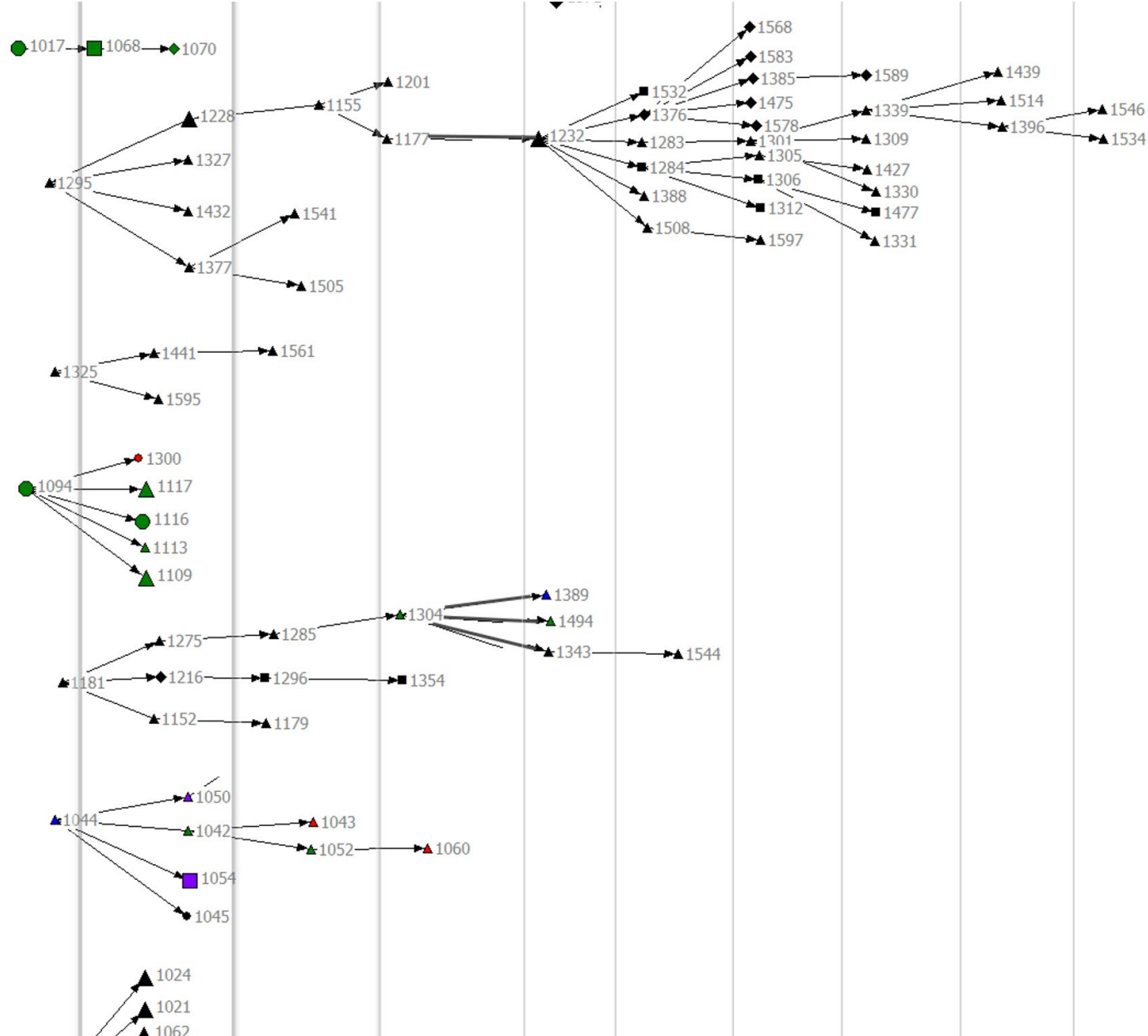
A real person who is responding to a survey with deceptive (untrue) responses

Case Study: Military Acceptance Project

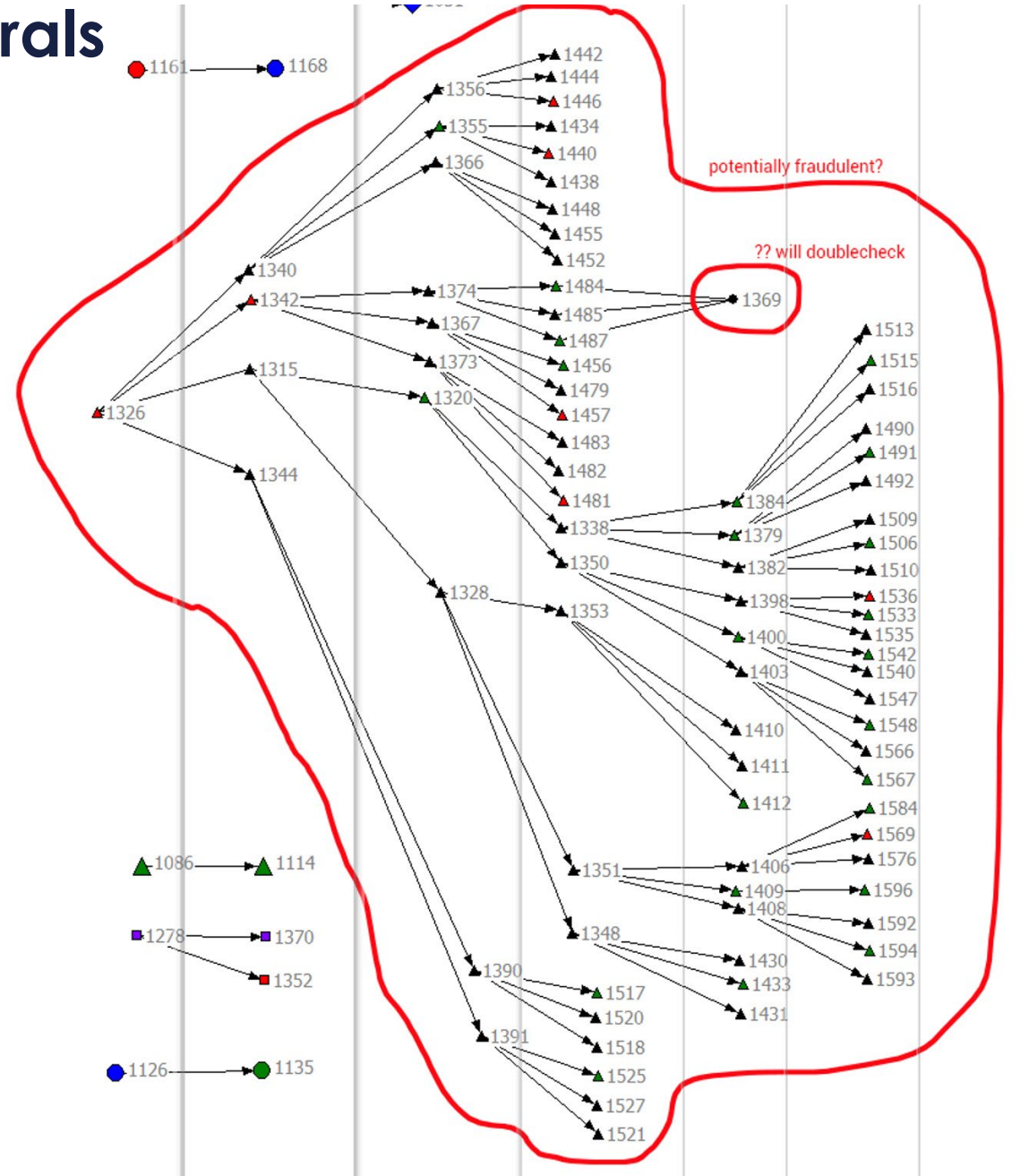
Defenses already in place:

- Screener included lots of “insider knowledge” questions that are difficult to research
- Needs referral code to access screener
- Quota on number of times a referral code can be used

Referral network analysis: Real referrals



Suspicious referrals



What we discovered when we analyzed these surveys

- GeoIP locations and IP addresses did not belong to a country with a U.S. military base (e.g., China)
- Google email addresses were all made up of the same username pattern of which there was a high volume

productivewigb@gmail.com

expansivethix@gmail.com

expirationuitg@gmail.com

contemplativeopxt@gmail.com

affirmativextyk@gmail.com

- Duplicate IP addresses used by high volume of survey takers in a short amount of time
- Duplicate email addresses as previous participants
- Rapid turnaround between referrals and survey completions

What we discovered when we analyzed these surveys

- Highly patterned and systematic responses (if all answers were 0 or patterns of 1-4, 4-1)
- Survey responses from same IP addresses had same answers

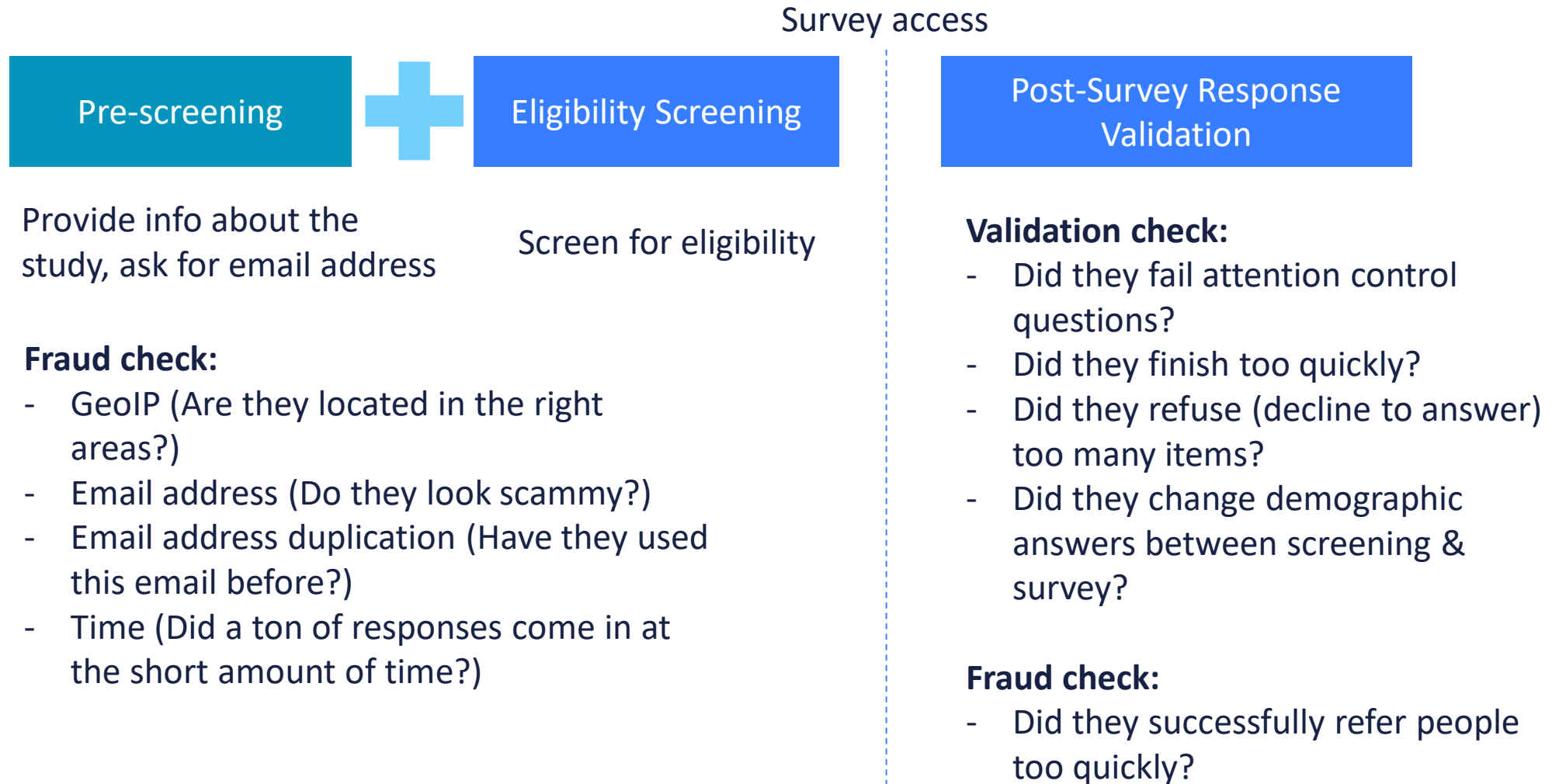
READ MORE

Recruiting Sexual and Gender Minority United States Military Personnel for Research Via Peer Networks: Addressing the Challenges

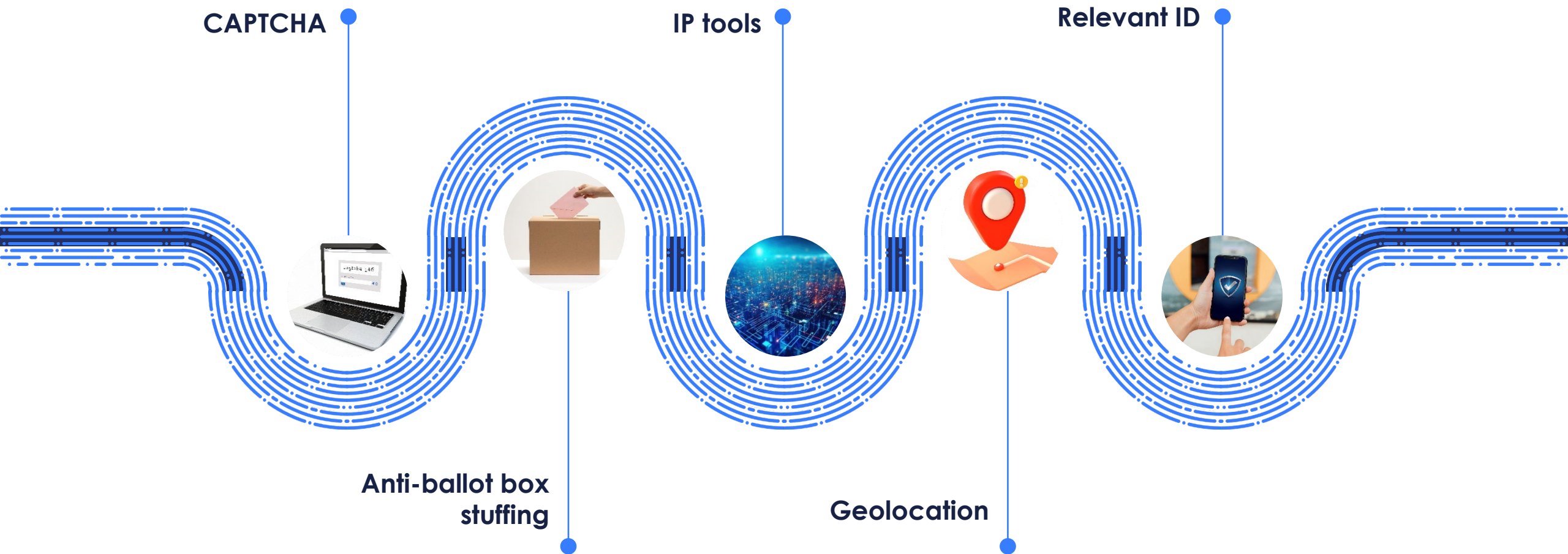
Raymond, Henry F. | Holloway, Ian W | Wu, Elizabeth | Tan, Diane | Mamey, Mary Rose | Goldbach, Jeremy T. | Castro, Carl A. | Schragger, Sheree M.

Annals of LGBTQ Public and Population Health Vol 3 Issue 3, Aug 2022, DOI: 10.1891/LGBTQ-2020-0002

What We did: Two-Step Screening & Validation Process



Before data collection: control access using automated fraud checks



How do bots & deceptive respondents get past these defenses?

Proxies & VPNs

A virtual private network that establishes a digital connect between their computer and a remote server (located somewhere else). A VPN does what a proxy does and hides the user's IP address.

VOIP

Voice Over Internet Protocol – allows people to make voice calls and send text messages using an internet connection

VOIP phone numbers can be chosen to make them look like a U.S.-based phone number.

Using third-party online tools to check IP addresses, use of VPNs, use of Proxies, etc.



404	IP: QuadraNet Enterprises LLC	2	3/25/22 6:26	3/25/22 6:29	155.94.216.13
405	IP: QuadraNet Enterprises LLC	2	3/25/22 6:31	3/25/22 6:35	155.94.216.13
406	IP: QuadraNet Enterprises LLC	2	3/25/22 6:36	3/25/22 6:39	155.94.216.13
407	IP: QuadraNet Enterprises LLC	2	3/25/22 6:41	3/25/22 6:43	155.94.250.93
408	IP: QuadraNet Enterprises LLC	2	3/25/22 7:45	3/25/22 7:53	155.94.250.93
409	IP: Highwinds Network Group, Inc.	2	3/25/22 9:17	3/25/22 9:19	64.145.67.38
410	IP: tzulo, inc.	2	3/25/22 9:37	3/25/22 9:43	68.235.33.100
411	IP: Cdn77 NYC	2	3/25/22 10:24	3/25/22 10:27	37.19.199.143
412	IP: Cdn77 NYC	2	3/25/22 10:27	3/25/22 10:31	37.19.199.143



Report Scammers Pricing Contact

Checking IP addresses

155.94.216.13 Fraud Risk

Very high Risk

← Lowest Risk Highest Risk →

Fraud Score: 100

We consider **web traffic** from IP address **155.94.216.13** to present a potentially **very high** fraud risk. This IP address is operated by **QuadraNet** whose web traffic we consider to present a potentially **medium** fraud risk, and is owned by **QuadraNet Enterprises LLC** whose web traffic we also consider to present a potentially **medium** fraud risk. In each of these cases, non-web traffic may present a different risk or no risk at all. Scamalytics see low levels of web traffic from this IP address across our global network, almost all of which we suspect to be potentially fraudulent. Considering only the web traffic where we have visibility, we apply a risk score of 100/100 to **155.94.216.13**, which reflects the proportion of this traffic which we suspect to be potentially fraudulent. If you see web traffic from this IP address there is potentially a very high risk that it is criminals engaged in fraudulent activity. Other types of traffic may present a different risk or no risk. The device on **155.94.216.13** is operating a public proxy, which could be proxying traffic from another geographical location. The geographical location of **155.94.216.13** is in **the United States**, however the geographical location of the user could be anywhere in the world.



Checking VPNs & Proxies

Location	
Country Name	United States
Country Code	US
State / Province	Georgia
District / County	Fulton
City	Atlanta

Proxies	
Anonymizing VPN	No
Tor Exit Node	No
Server	No
Public Proxy	Yes
Web Proxy	No
Search Engine Robot	No

Checking for VOIPs

RESULTS:

- **Phone Number:** 6576500894
- **Date of this Report:** April 10, 2024
- **Phone Line Type:** VOIP
- **Phone Company:** GOOGLE VOICE
- **Phone Location:** ORANGE, CA
- **Owner's Name & Address:** [Click Here](#)
Sponsored by PeopleFinders.com

(657) Area code = Anaheim, CA

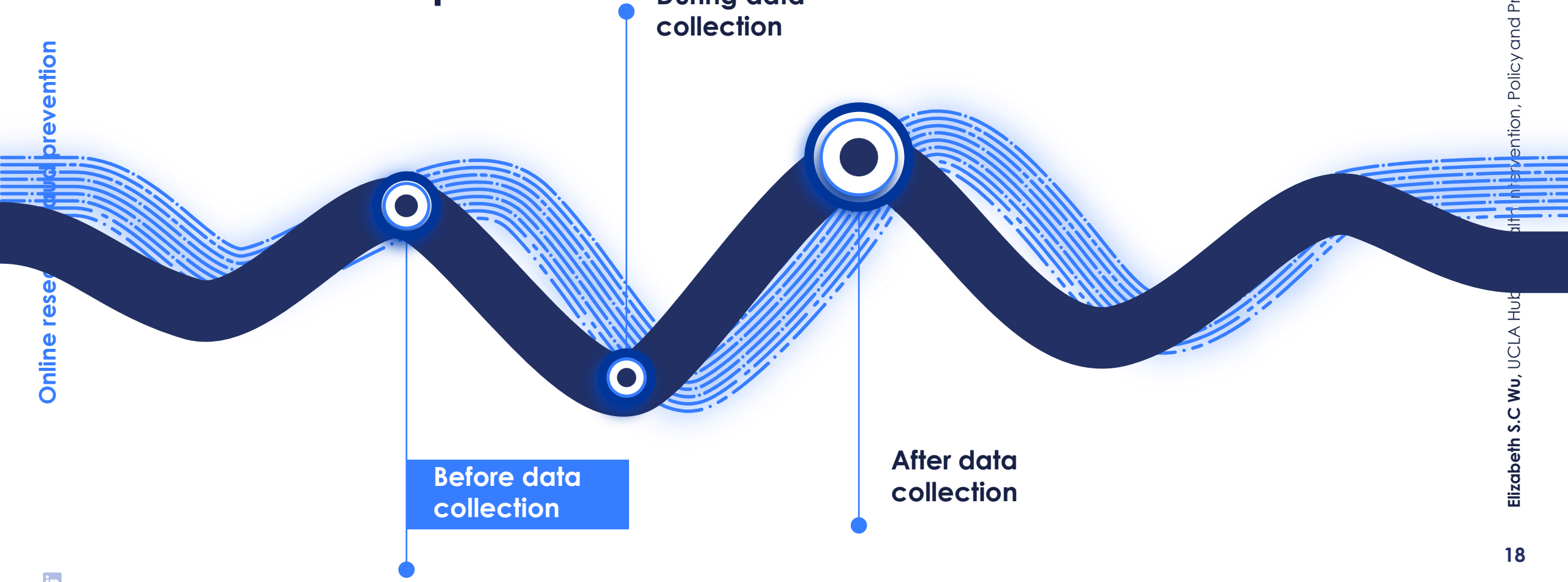
But this person can be *anywhere*

From PhoneValidator.com

...

Fraud can be stopped at these checkpoints

Online research and prevention



Control access to survey

Automated tools provided by more robust online survey platforms

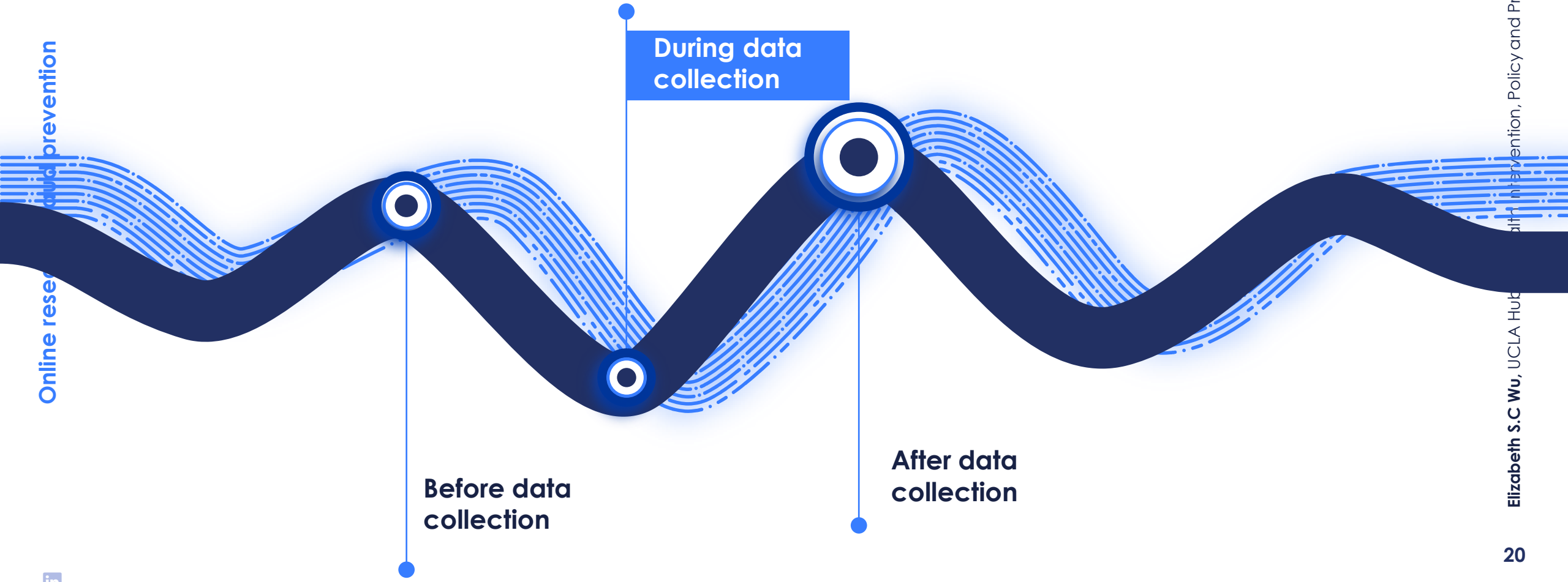
- CAPTCHA – detects non-human responses (i.e., bots)
- Anti-ballot box stuffing – prevents multiple submissions to a single survey
- Security scan monitoring – prevents email scanning software from starting survey sessions
- Relevant ID – assesses metadata to determine likelihood of duplicate respondents

Manual fraud checking:

- Check IP addresses to ensure respondents location is consistent with survey answers
- Check if phone number is a VOIP phone number
- Use personalized links
- Use insider knowledge questions in the screener
- Limit time of entry
- Watch for repeated patterns
- Limit advertising to in-person recruitment strategies and referral networks rather than posting online
- Keep a database of fraudulent/deceptive IP addresses and email addresses

...

When fraud can happen



Prevent bad data

Recheck:

- Eligibility criteria
- Geolocation

Utilize:

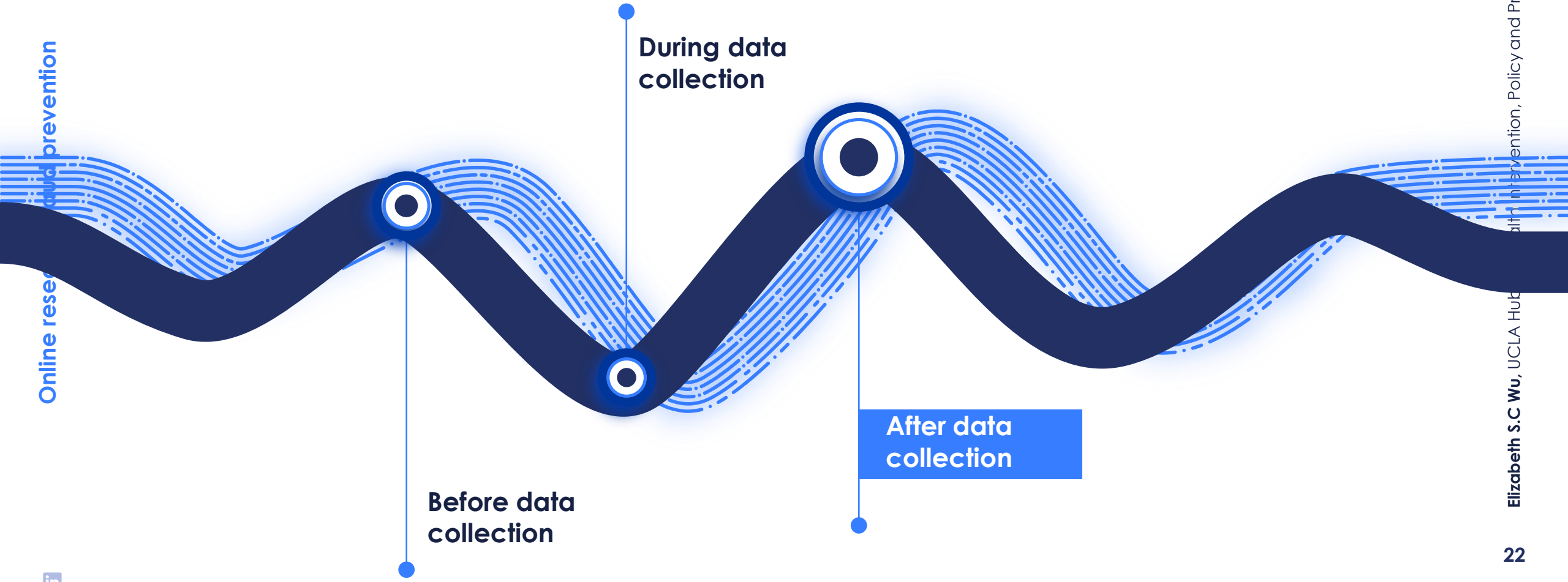
- Attention-control measures
- Duration controls

Analyze:

- Referrals for fraud patterns

...

When fraud can happen



Online research and prevention



Quality control

Check for fraud during payment process

- International deceptive respondents prefer Amazon gift cards and Paypal as opposed to Venmo, Cashapp, local grocery stores, local gas stations, etc.

Check for internal data inconsistencies

- Example: They said “yes” to using several drugs currently, and then answered elsewhere in the survey that they “never” used drugs.
- Example: In the demographic information, they selected “cisgender man” as their current gender identity, but then answered elsewhere in the survey that “they felt they were discriminated against for being a woman.”

Check for nonsensical answers in free-answer boxes



Questions?



Thank you!

Elizabeth S.C. Wu, MPH

[Email](#) | [LinkedIn](#)