

Fighting Fire with Fire: Risks & Benefits of AI to Data Quality

July 17, 2024

IRONWOOD
INSIGHTS
—group, LLC



Brian Cash

VP, Research Services Ironwood Insights



Brad Larson

Founder & CEO, Ironwood Insights

IRONWOOD INSIGHTS

group, LLC

**Integrated Full-Service
Solutions/AI**

**Data Collection
(Multi-Modal)**

Ironwood Online

Product Development Customer Experience
Qual OL Message Boards A-I Mail Surveys
Dashboards IDI's Quant Online
Branding / Coms Segmentation Multivariate Surveys
Communities Discrete Choice OL Focus Groups Strategy
Focus Groups Telephone Surveys



PersonaPanels



Some Perspective

A Legacy Problem: Fraud has posed a significant threat to online survey research quality for a decade or more.

Now an Existential Threat to the Industry: The proliferation of technology-led fraud—Bots, AI/Chat GPT—has undeniably and dramatically accelerated the downturn of...

- **Online sample quality**
- **Data integrity**
- **Reliability of insights and business decisions**



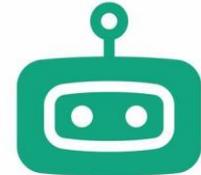
Primary Sources of Survey Fraud

However, AI is still just a piece of the fraud puzzle

1 AI-Led Fraud: Bots, bot farms



2 Human Fraud: “Traditional” bad respondents, survey farms



CHAT GPT

3 Blended Fraud: i.e., Fraudsters using Chat GPT



What is the Impact of Survey Fraud?



Timing/Study Delays...

Replacing bad surveys extends field time, project duration



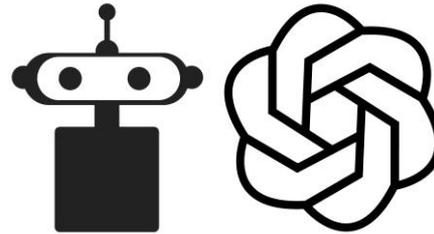
Cost Uncertainty/Overruns...

Replacing bad surveys adds unforeseen costs to the project



Data/Insights Quality...

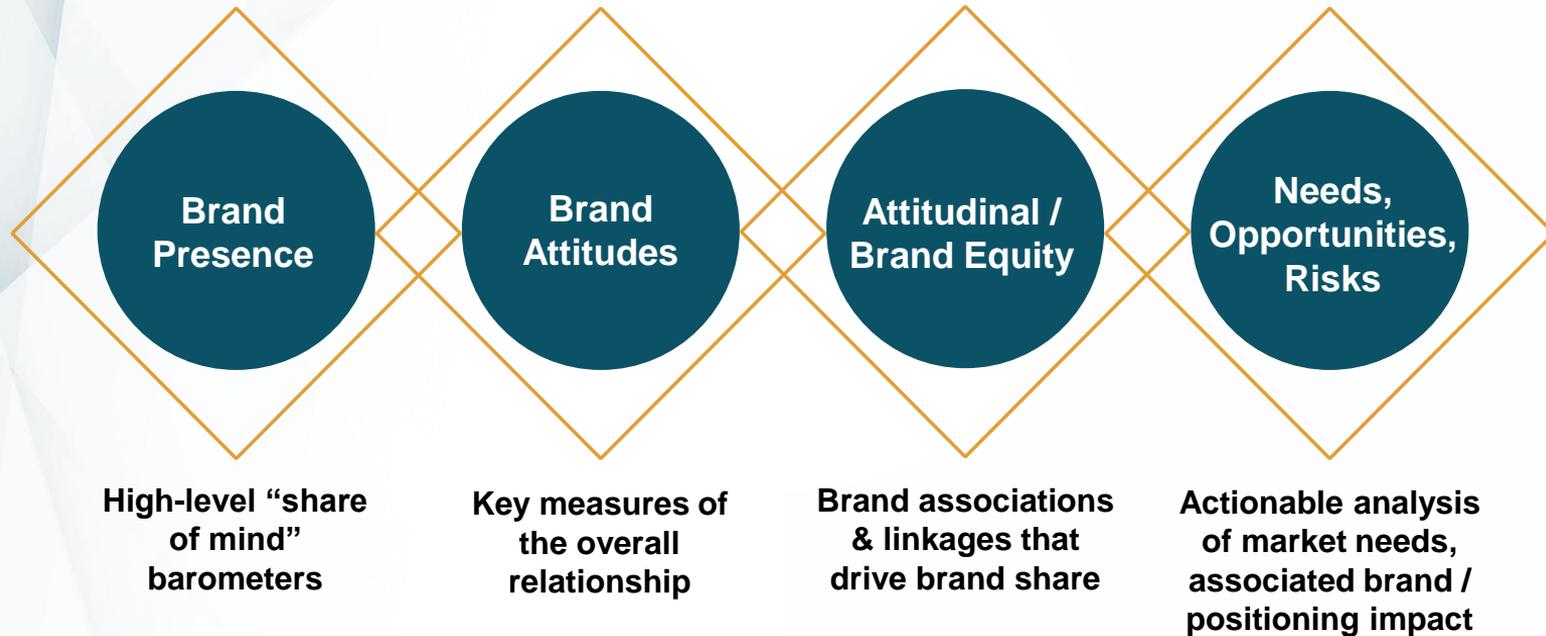
Unchecked fraud erodes sample quality, undermines data integrity



Illustrating the Impact of Fraud on Data Quality

The process...

- A recent brand health market study of Chicago hospitals and health systems.





Illustrating the Impact of Fraud on Data Quality

The process...

- Evaluated sample and data quality of those passing or failing 7 fraud countermeasures; “Good Guys” vs. “Fraudsters.”

Highlights of the Test

	Good Guys (Passed)	Fraudsters (Failed)
Fraud Countermeasures	Avg. <1 fail	Avg. 3.6 fails
Length of Interview	16 minutes	11.5 minutes
Sample Size	n=196	n=255
DQ Rate	130%	NA

Fraud Countermeasures Used

Most fraudsters are serial offenders!

- **IP Duplication Block** (Screening)
- **Visual Trap** (Screening)
- **Red Herring** (Closed-end Choice Lists)
- **Multiple Attention Testing/Speed Traps/Timers** (Attributes)
- **Bot Sentry Trap & Other Open-End Review**
- **Back-end Data Review** (Integrated Analysis)

Visual Trap



Attention Testing

[RANDOMIZED ATTRIBUTE LISTS]

- a. The quality of overall patient experience
- b. Transparent communications between clinical staff & patients – open, truthful, honest
- c. **For quality purposes, please enter a four**
- d. Has state-of-the-art technology/medical devices
- e. Innovative in their treatments/solutions

Red Herring

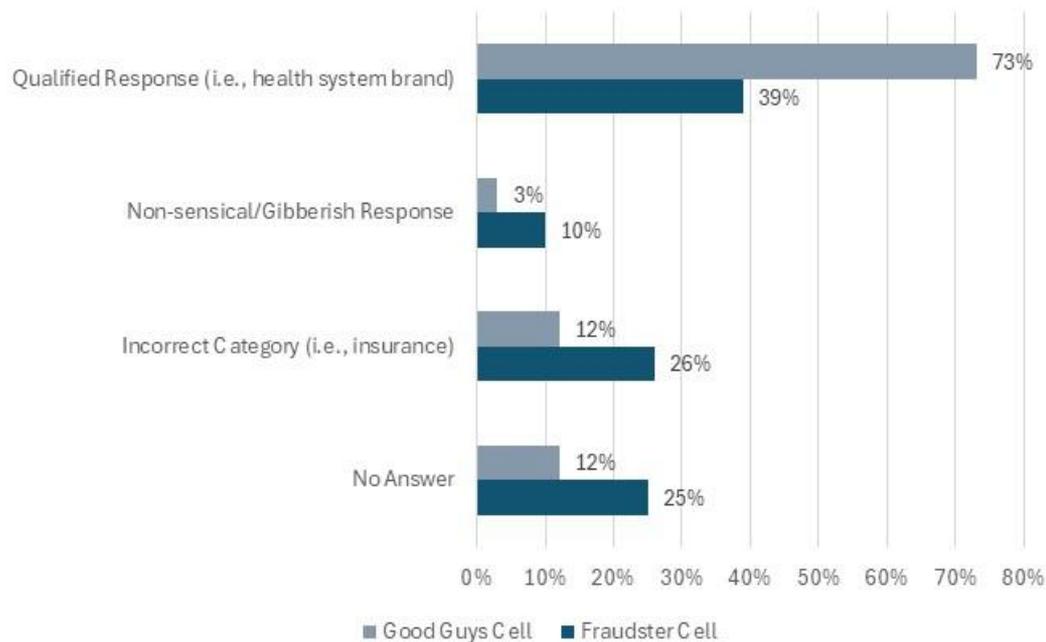
[RANDOMIZED PICK LISTS]

- a. Orthopedic care
- b. Cardiovascular care
- c. Oncology/radiology
- d. Gastroenterology
- e. Women's services
- f. **Dinner and a movie**
- g. Neurology
- h. General surgery
- i. Labs, diagnostic imaging
- j. Other (Specify: _____)
- k. All of the above [ANCHOR]

The Impact of Fraud on Data Quality: Some Examples

Write-ins: Make the respondent work a little harder.

Unaided Brand Awareness



Risks: Spotlights lack of attention/seriousness fraudsters exhibit throughout the survey

The Impact of Fraud on Data Quality: Some Examples

Ratings: Fraudsters tend to “rate high.”

Brand Familiarity

	Good Guys	Fraudsters
Scale:		
Know Nothing 1	22%	15%
2	13%	8%
B2B	35%	23%
3	12%	7%
4	16%	17%
5	13%	14%
T2B	24%	39%
6	8%	15%
Very Familiar 7	16%	24%

Risks: Distorts brand presence, contaminates subsequent brand image, brand equity outcomes.

The Impact of Fraud on Data Quality: Some Examples

Brand Attributes

[RANDOMIZE]

- a. The quality of overall patient experience
- b. Ease of accessing healthcare services when needed
- c. Ethical in what they do
- d. Organization's reputation in the community
- e. Trustworthy in what they do
- f. The ease of doing business (billing, follow-up, insurance, treatment, etc.)
- g. Transparent communications between clinical staff and patients – open, truthful, honest
- h. Clinical staff act as a team to address my healthcare needs
- i. Provides personalized care –caters to my needs throughout my healthcare journey.
- j. The quality/medical expertise of Doctors
- k. The Doctors, Nurses and staff treat patients and family with respect
- l. Consistent/reliable/dependable care
- m. For quality purposes, please enter four**
- n. Has state-of-the-art technology/medical devices
- o. Innovative in their treatments/solutions
- p. Uses alternative medicine and treatment options when called for
- q. Treats a wide variety of health conditions.
- r. Doctors, Nurses are caring and compassionate
- s. Ease of visiting with Doctors/Nurses online or with a mobile device
- t. Availability of virtual diagnostic and treatment options (online, mobile)
- u. Provides good value for the cost of care
- v. Has the widest network of doctors to choose from
- w. Offers a wide range of wellness and preventive care programs
- x. Facilities are conveniently located around the area
- y. Has any type of specialists physician I'd need to address my health



- **Bots Always Know:** rarely use the *don't know* response.
- **Bots Cover their Tracks:** don't universally straight-line; interrupt pattern/vary 1-2 ratings.

- **Red Herring Fails EXCEEDINGLY Positive:** top box use 2-3x “good guys”/other cheaters.
- **Fast & Not Furious:** Inattentives (3+ attention fails) also rate consistently high.

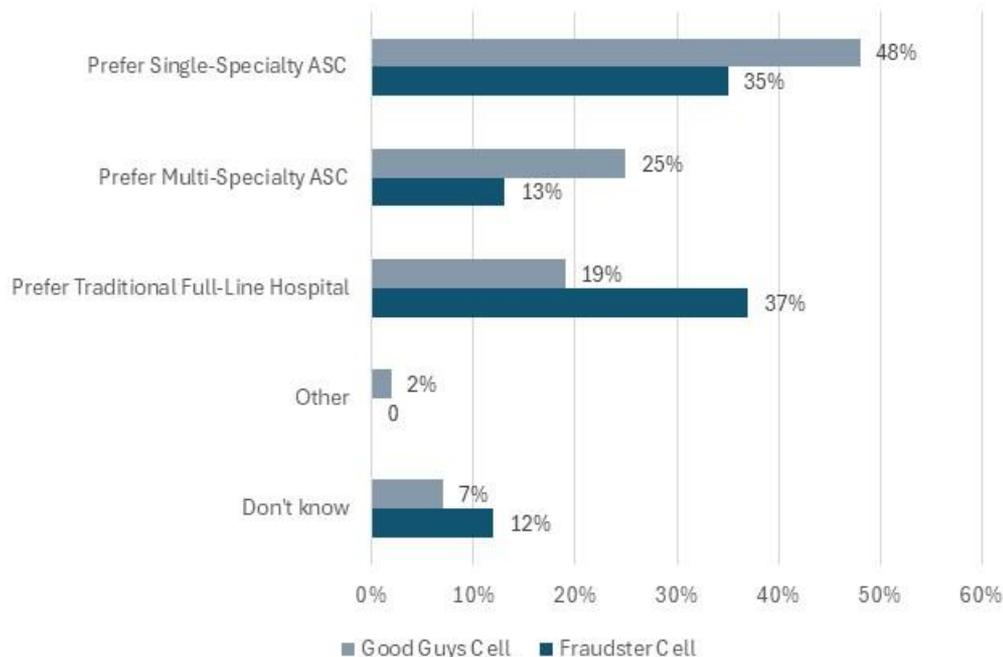
Risks: Distorts brand image, contaminates subsequent brand equity outcomes.

Some fraudsters answered 2, 24 attribute batteries in less than 50 seconds!; a rational respondent takes closer to 4 minutes.

The Impact of Fraud on Data Quality: Some Examples

Concept Review: Fraudsters may not comprehend stimuli

Healthcare Facility Preference

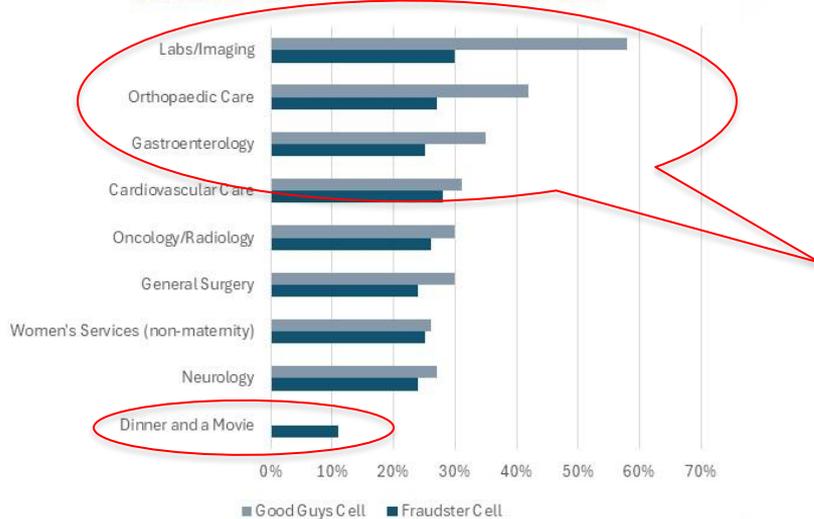


Risks: Points to the wrong pathway for facility/network decisions, capital investment.

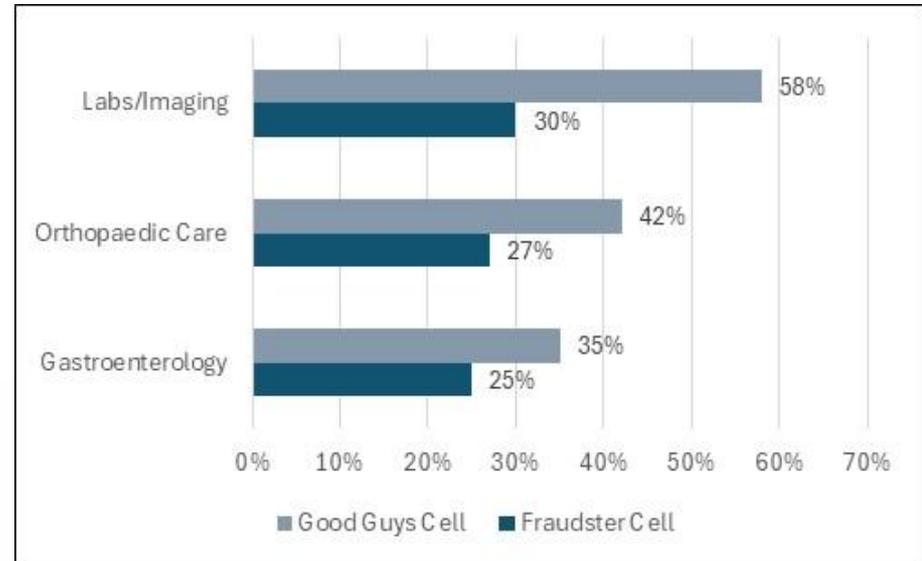
The Impact of Fraud on Data Quality: Some Examples

Choice/Pick Lists: Fraudsters overuse pre-lists, pick less thoughtfully.

Services Would Consider at an ASC



Risks: Points to the wrong pathway for facility/network decisions, capital investment.



The Role of Open Ends

Triangulate, analyze in conjunction with other countermeasures

“Regular” & “specialized” open ends—a thoughtful answer, a feeling/emotion OR real-time, timed responses (i.e., Eden AI)—can trip up bots/AI and flag lazy/unserious/inattentive human respondents.



Bot Sentry (OE Trap)

Think about your most recent interaction with a healthcare provider.

- What aspects of the interaction made you happy & why?
- What aspects of the interaction frustrated you & why?

Please be as specific as possible with your answer



Example Answers

- YES, I am awoken of the day I have.
- What factors are the most important to you when selecting a health system? (prior OE question)
- This is very good and very nice
- 0ž0• 0ž0«0š0¥0¥ 0ž0† 0š0¥0•
- Thank you for the heads up on the 6th grade teacher and the and the and the and the and the

Data Quality: Open-End Fraud Archetypes

Beavis & Butthead (Human)

Non-sensical: Kid, Salami, Banana, Clementine

Potty Mouth: i.e., poop, pee, a smorgasbord of swear words, male/female anatomy

The Cheerful Bot

Good to hear that you're having fun with the new year in Colorado Springs this year and the world.

Good morning and thank you

Ok I will be there in about a hourly mood and I will be there around the same time

Take Me to your Leader (Bot)

0200 020200020200...
4 0020200?

Babblers (Bot & Human)

Ditis que es ubks dules mju buenos h
qur se mr antija ta tldos por xok9leto
me gysra

Fgyuyu

The Pontificator (ChatGPT + Human)

Question

What do you like about CONCEPT NAME?

Answer (~1/2 the response shown)

CONCEPT NAME wearable is appealing to users interested in monitoring and improving their health. It provides real-time data on vital signs like heart rate, blood pressure, and oxygen levels. Helps in early detection of potential health issues. Tracks physical activities such as steps, distance, and calories burned. Encourages regular exercise and helps users meet fitness goals. Monitors sleep patterns, including duration and quality. Offers insights to improve sleep hygiene and overall restfulness. Continuous improvements in technology lead to more accurate and diverse health monitoring capabilities. Integration with AI for predictive analytics and personalized health insights.

ETC.

Summary: the Pathway to Quality Business Decisions

Poor sample & data quality can send you down the wrong decision pathway

Business Decision Risks

Survey Fraud: Unchecked

- Inaccurate view of market landscape, relevant competitors
- Distorted view of market presence
- Misappropriated ad spend
- Inaccurate view of brand positioning, brand equity/health
- Focus on the WRONG fixes to brand experience
- Misappropriated CapEx investment



Survey Fraud: Mitigated



What can we do about Online Survey Fraud?

Ensuring Sample Quality & Data Integrity is a Process!

 | **Prevention**

 | **Near-Time Mitigation**

 | **Real-Time Mitigation**

 | **Back-end Analysis**

Summary...The Pathway to Better Data



Prevention

- **Vet Online Sample Partners (multiple dimensions)**
 - Panel makeup & composition
 - Recruitment & hygiene practices
 - Panel management/fraud tech
 - Incentive policies & practices
- **Active Partner Monitoring & Dialogue**
 - Monitor study-to-study fraud metrics *and share back*
 - Quarantine or remove chronically sub-standard players



Summary...The Pathway to Better Data



Real-Time Mitigation

- **Screening Detection & Elimination**
 - CAPTCHA, meaningful visual traps to catch bots, lazy human respondents
 - IP/duplication check technology to head off repeaters/cheaters before they enter your survey
- **Screener Design**
 - Early screen outs
 - Mask screenouts to mitigate unqualified respondents from sneaking into survey

Please describe this picture. Be as descriptive as possible.

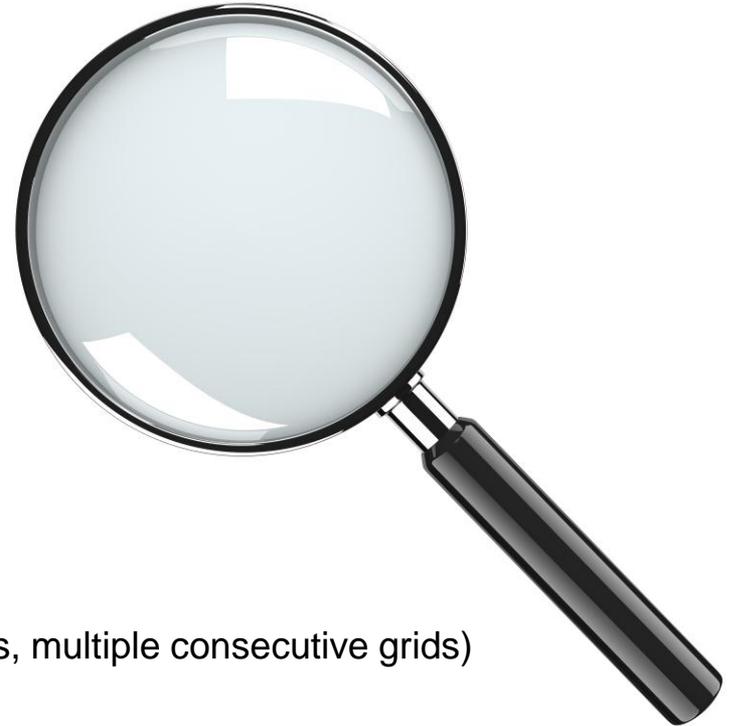


Summary...The Pathway to Better Data



Near-Time Mitigation

- **In-survey detection & elimination**
 - Speed traps & bumps, i.e.: attention testing
 - Red herring questions
 - Layered or emotion-based OE's
 - Real-time open-end validation (i.e., Eden AI, Chat GPT)
- **Survey Design**
 - **“Respondent Bill of Rights”**: consider respondent experience in survey design, compensate them fairly
 - Better question design
 - Question layout (limit length of attribute lists, multiple consecutive grids)
 - Design with mobile in mind



Summary...The Pathway to Better Data



Back-end Analysis

- **Integrated review of data file**
 - Duplicate I.P. addresses
 - Demographic comparisons: i.e., detecting identical ratings from those with identical demographics
 - Automated: i.e., speeding thresholds
 - Straight-lining/patterning of responses
 - Use of scales
 - Review of open-ends

Alternative Sampling & Data Collection Methods

Online research can be overused by researchers and marketers compulsively chasing fast turn, lowest-cost solution.

However, we sometimes see the limits of online research stretched and the integrity of project samples compromised...

- **Narrow geographies/small markets**
- **Hyper-targeted respondent profiles**
- **Some segments of the general pop, i.e., older, lower socioeconomic consumers**
- **Niche product categories**
- **Need to authenticate respondent, i.e.: role (B2B, Healthcare), geography**



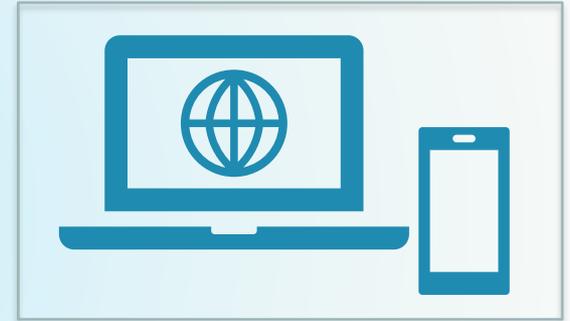
Consider Alternative Sampling & Data Collection Methods

Pursuing online “at all costs” can work against sample quality and ultimately data integrity.

Alternative or mixed methodologies—such as telephone, text-to-web (TTW)—can be used in place of or in conjunction with online to capture true probability samples and more representative samples limited constituents, markets, geographies.

Telephone is still the gold standard in a number of disciplines...

- **Government/public policy (mail, phone, TTW)**
- **Healthcare/Public health (phone, OL, mail) focused on specific cohorts**
- **Political polling (phone)**



Thank You!



briancash@ironwoodinsights.com



bradlarson@ironwoodinsights.com

**For more information contact Brad or Brian
OR stop by booth #415**