

Concrete eaters, bleach drinkers and flat earthers:

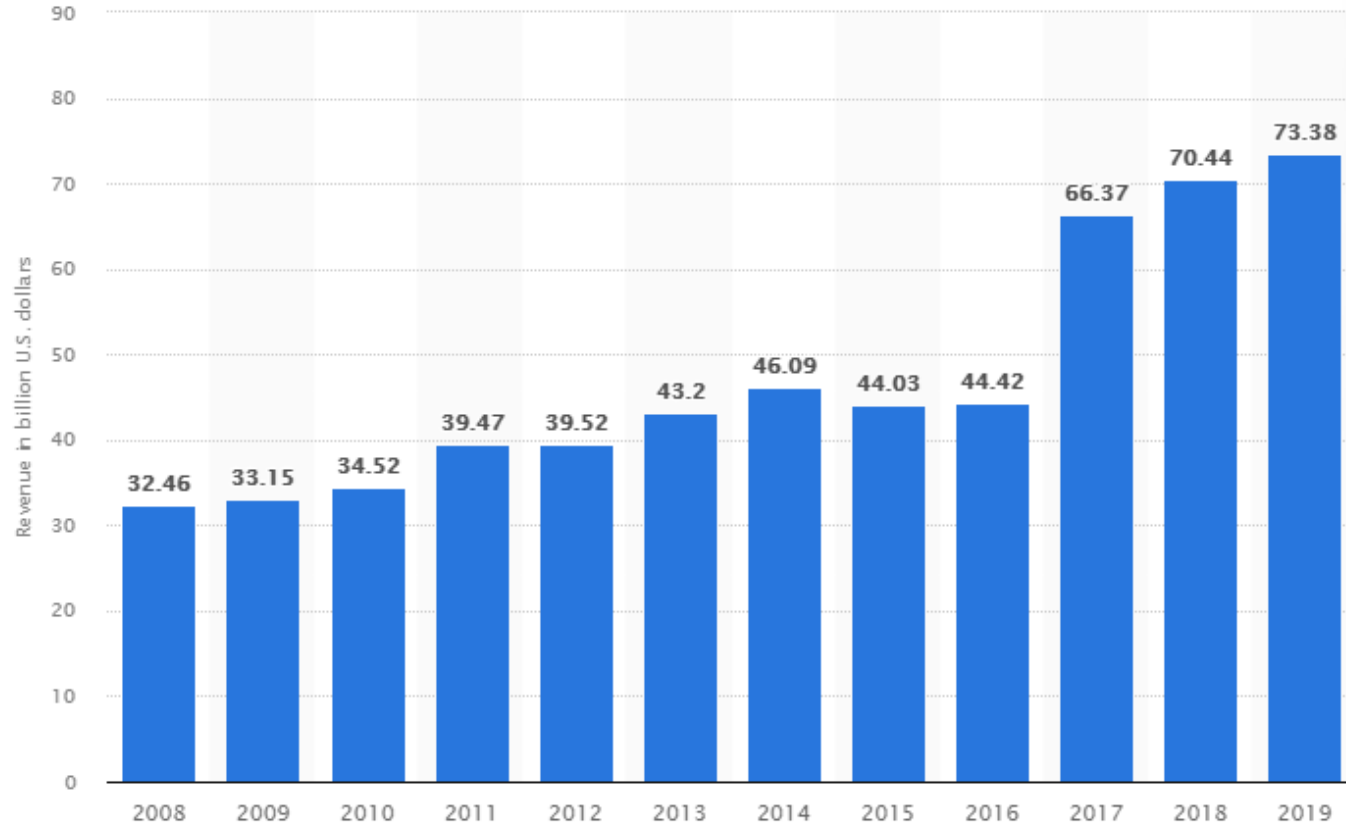
How to defend online research from bogus respondents

Leib Litman, PhD

Co-CEO CloudResearch



Rise of Online Research



Online Data Quality

Online Panels in Social Science Research: Expanding sampling methods beyond Mechanical Turk

**Behavior Research Methods,
September 11, 2017**

Jesse Chandler, Cheskie Rosenzweig,
Aaron J. Moss, Jonathan Robinson, and
Leib Litman

Assessing the Risks to Online Polls From Bogus Respondents:

**Pew Research Center
February 18, 2020**

Courtney Kennedy, Nick Hatley, Arnold Lau,
Andrew Mercer, Scott Keeter, Joshua Ferno, and
Dorene Asare-Marfo

Dirty Data: The Effects of Screening Respondents Who Provide Low-Quality Data in Survey Research

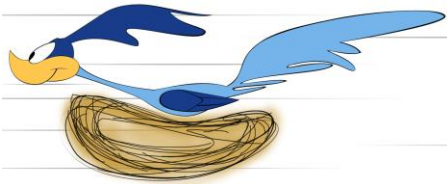
**Journal of Business and
Psychology, September 2, 2017**

Justin A. DeSimone & P. D. Harms

THREATS TO DATA QUALITY

Speedster

- Real Human
- Inconsistent
- Unengaged
- Inattentive
- Speeding
- Reward-Focused
- Honest When Possible



Fraudster

- Dishonest
- Reward-Focused
- Eligible 'For All Studies'
- Often a Speedster



Bot

- Human or Machine
- Uses Scripts or Code
- Provides Bad Data
- On Continuum of Sophistication

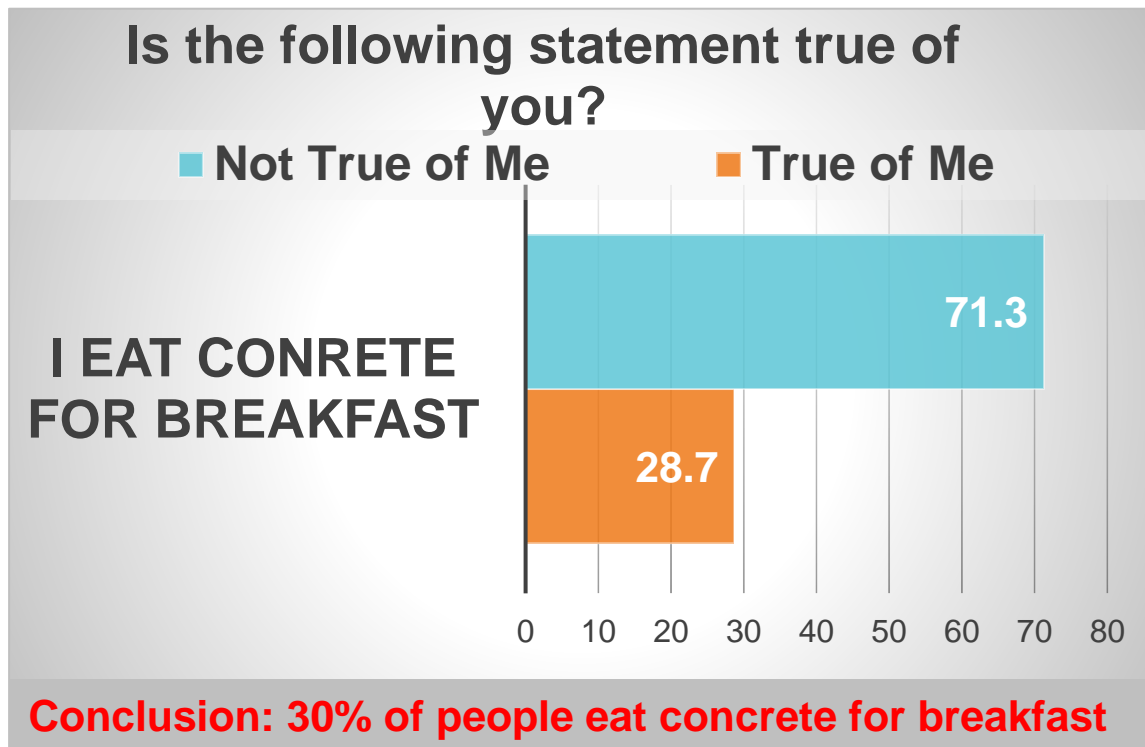


Golden Ticket

- Real Human
- Attentive
- Engaged
- Effortful
- Honest
- Internally Consistent

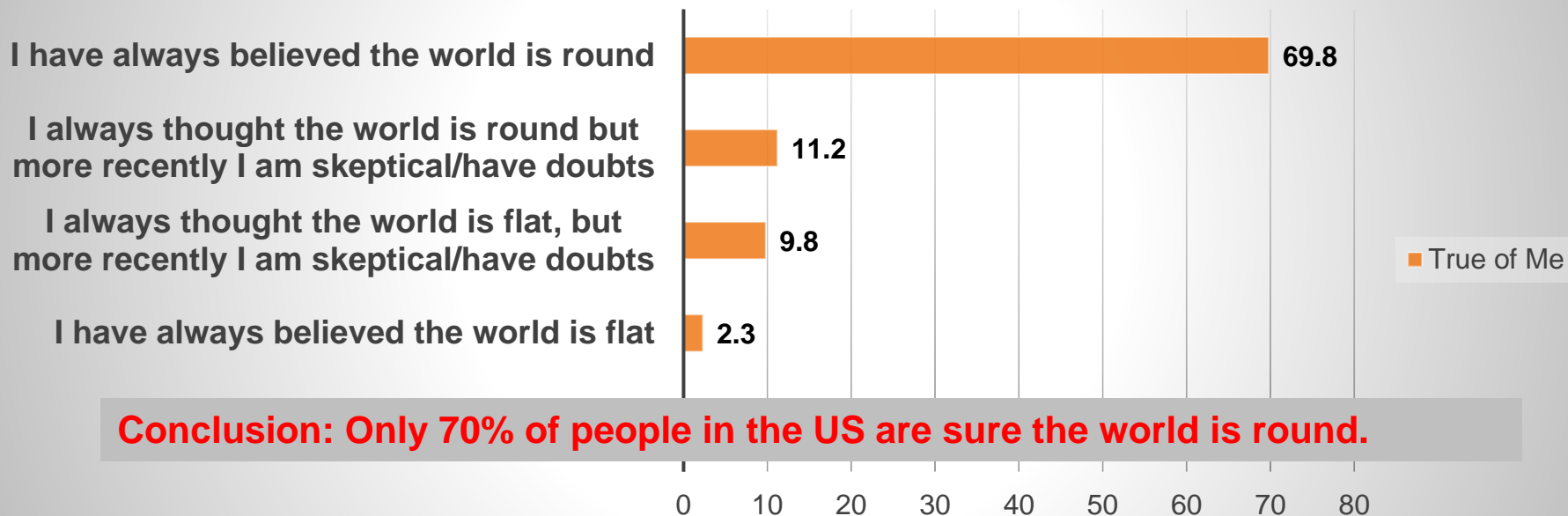


Bogus Respondents, Bogus Conclusions?



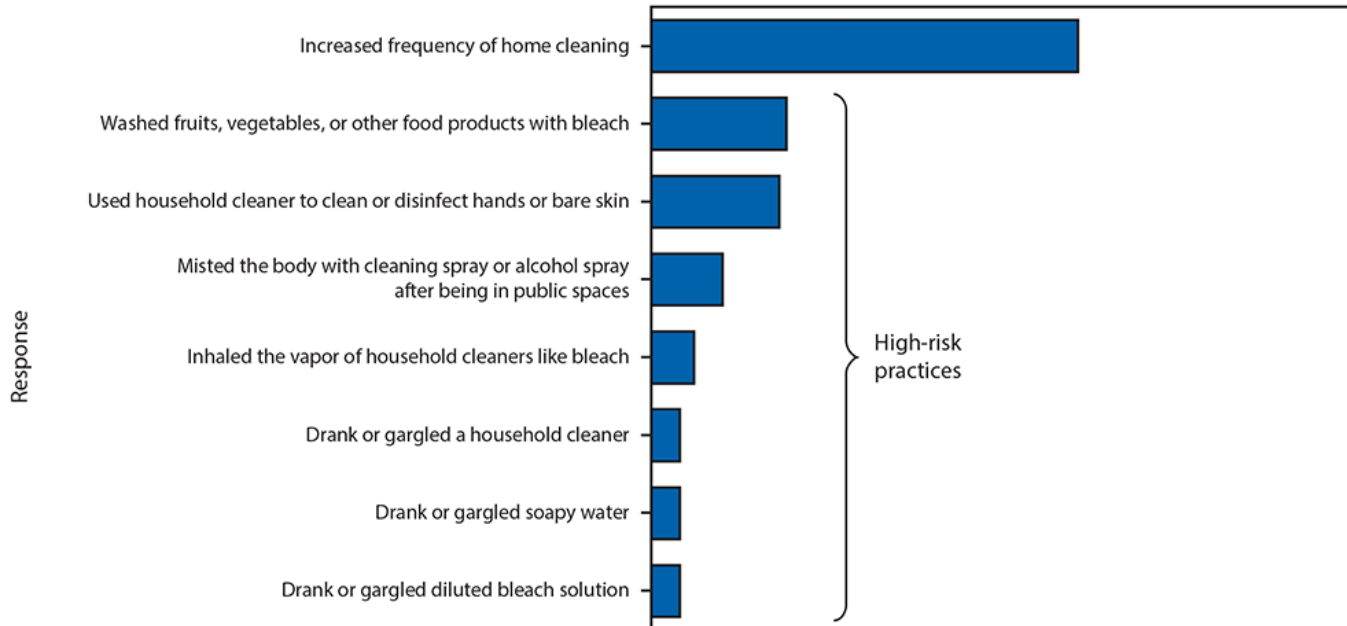
Bogus Respondents, Bogus Conclusions?

Do you believe that the world is round or flat?



Bogus Respondents, Bogus Conclusions?

FIGURE 2. Cleaning and disinfection practices in the previous month with the intent of preventing SARS-CoV-2 infection,^{*,†} based on responses to an opt-in Internet panel survey^S (N = 502 respondents) — United States, May 2020



Conclusion: “Thirty-nine percent reported intentionally engaging in at least one high-risk practice not recommended by CDC for prevention of SARS-CoV-2 transmission”.

Weighted % selecting response

* In response to the question “In the past month, which of the following cleaning behaviors have you or a household member engaged in to prevent coronavirus?”

Traditional Post-Survey Review:

Post Survey Review

In-Survey Attention Checks

Open End Review

Speed of Completion

Unrealistic Profiles

Total Flags

Cons

- Inconsistent across firms and surveys
- Unscientific: How many flags? What is a good attention check
- Very resource/time intensive
- Return to field multiple times
- Complex reconciliation process
- Checks to catch all kinds of bad respondents
- Slower to complete data collection

What about Pre-Survey Vetting?

Open Access | Published: 11 September 2019

Online panels in social science research: Expanding sampling methods beyond Mechanical Turk

[Jesse Chandler](#), [Cheskie Rosenzweig](#), [Aaron J. Moss](#), [Jonathan Robinson](#) & [Leib Litman](#) 




Behavior Research Methods **51**, 2022–2038 (2019) | [Cite this article](#)

14k Accesses | **48** Citations | **26** Altmetric | [Metrics](#)

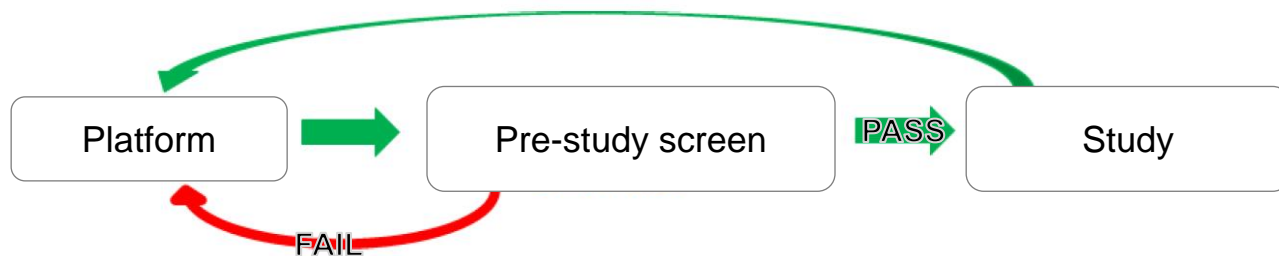
Evaluating CloudResearch's Approved Group as a Solution for Problematic Data Quality on MTurk

[D Hauser](#), [AJ Moss](#), [C Rosenzweig](#), [SN Jaffe](#)... - 2021 - [psyarxiv.com](#)

Maintaining data quality on Amazon Mechanical Turk (MTurk) has always been a concern for researchers. CloudResearch, a third-party website that interfaces with MTurk, assessed ~100,000 MTurkers and categorized them into those that provide high (~ 65,000, Approved) and low (~ 35,000, Blocked) quality data. Here, we examined the predictive validity of CloudResearch's vetting. Participants (N= 900) from the Approved and Blocked groups, along with a Standard MTurk sample, completed an array of data quality measures ...

☆   All 3 versions 

Showing the best result for this search. [See all results](#)



Sentry: Behavioral + Technological DQ System

General Participant Insights

9,276

Respondents

8,187

Passed

88.26%

172

Dropped

1.85%

917

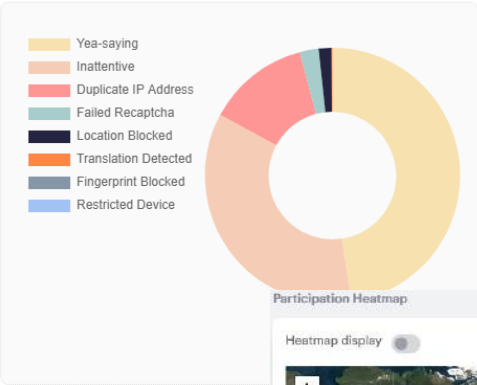
157 Blocked

1.69%

 | 760 Failed

8.19%

Failure Breakdown Analysis



Failure Reason	Type	Participants	Percentage
Yea-saying	Failed	437	47.66%
Inattentive	Failed	323	35.22%
Duplicate IP Address	Security Blocked	119	12.98%
Tech Blocked - Recaptcha	Security Blocked	22	2.4%
Tech Blocked - Location Blocked	Security Blocked	15	1.64%
Tech Blocked - Translation Detected	Security Blocked	1	0.11%
		0	0%
		0	0%

Sentry Stimuli

- Question creation and scoring based on associative network models
- Extensive testing, stabilization of optimal difficulty level

 CloudResearch

Question 1 of 1

Collect is most associated with

A	headline
B	quarterback
C	rhythm
D	gather

Please select answer to continue

Next →

 CloudResearch

Question 1 of 1

From memory, can you recall the name of every child of every President in the history of the U.S.?


A	Yes
B	No
C	I remember them all
D	Completely

Please select answer to continue

Next →

Sentry tech: Event Streamer

Assignment: 27fe6602-3135-447e-8b35-01be45287844




Instructions




The next few questions ask about your knowledge of the world, and your understanding of words.

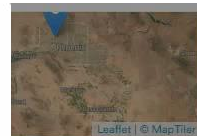
For each question, please pick the answer that seems most correct to you.

[Next →](#)

© 2021 Prime Research Solutions LLC

00:35  03:08

 **1x** 2x 4x 8x  skip inactive 



289 Longitude: -112.2443

address:

print:

ore:

2-46b9-84d5-
4

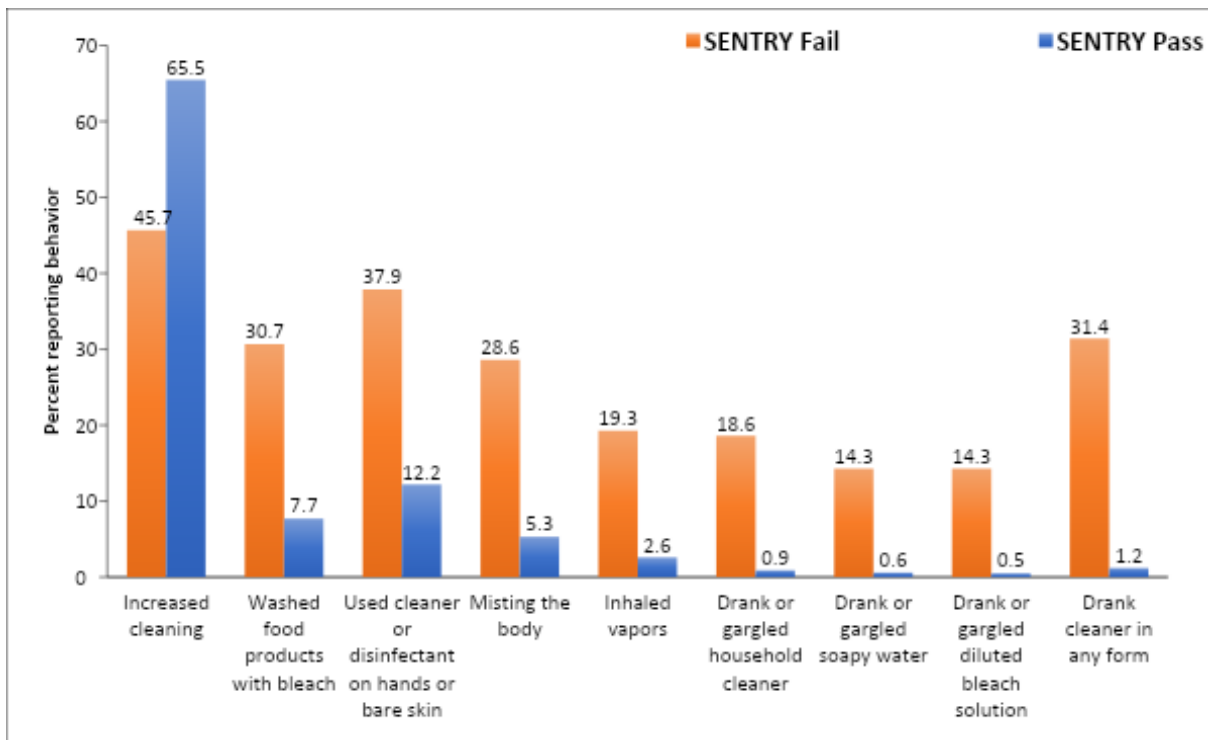
Session Recording

Download Events CSV

e For

Sentry Results: BLEACH DRINKING?

Conclusion: “Thirty-nine percent reported intentionally engaging in at least one high-risk practice not recommended by CDC for prevention of SARS-CoV-2 transmission”.



CLEAN DATA: Concrete Eaters

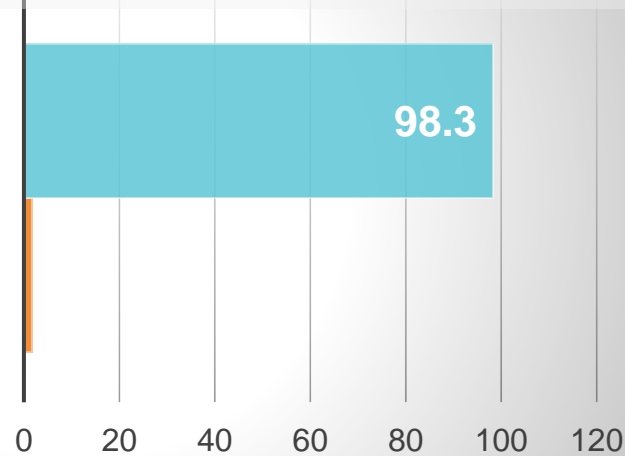
Conclusion: 30% of people eat concrete for breakfast

Is the following statement true of
you?

■ Not True of Me

■ True of Me

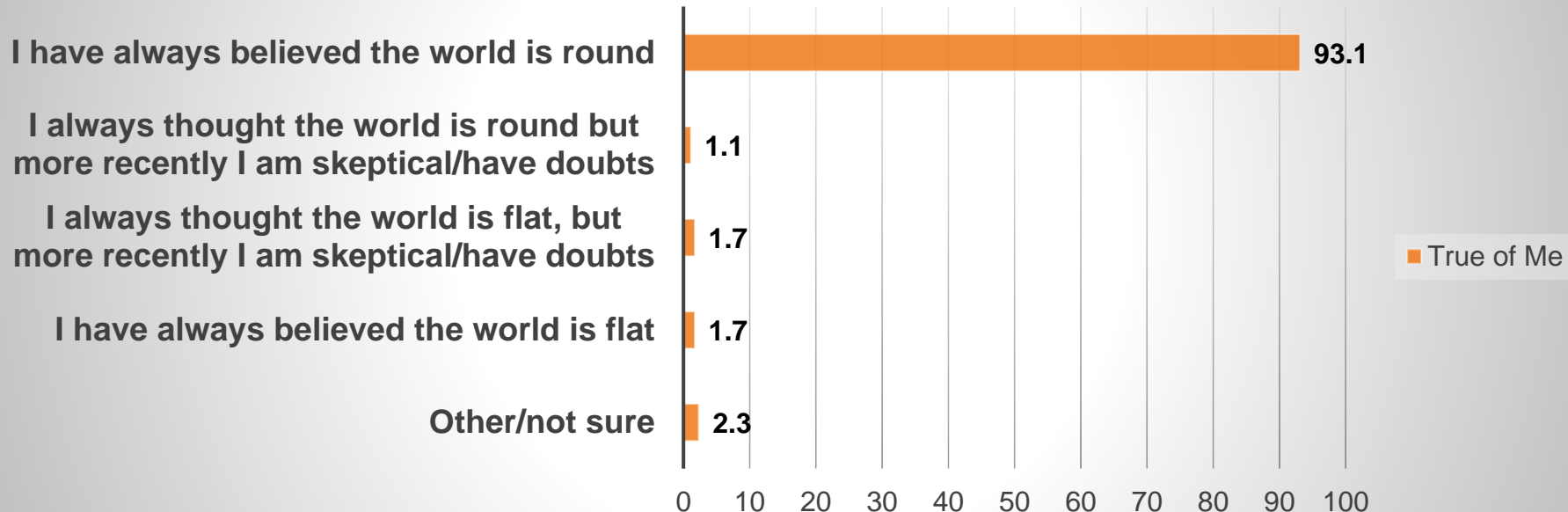
**I EAT CONCRETE
FOR BREAKFAST**



DATA CLEANED: Flat Earthers

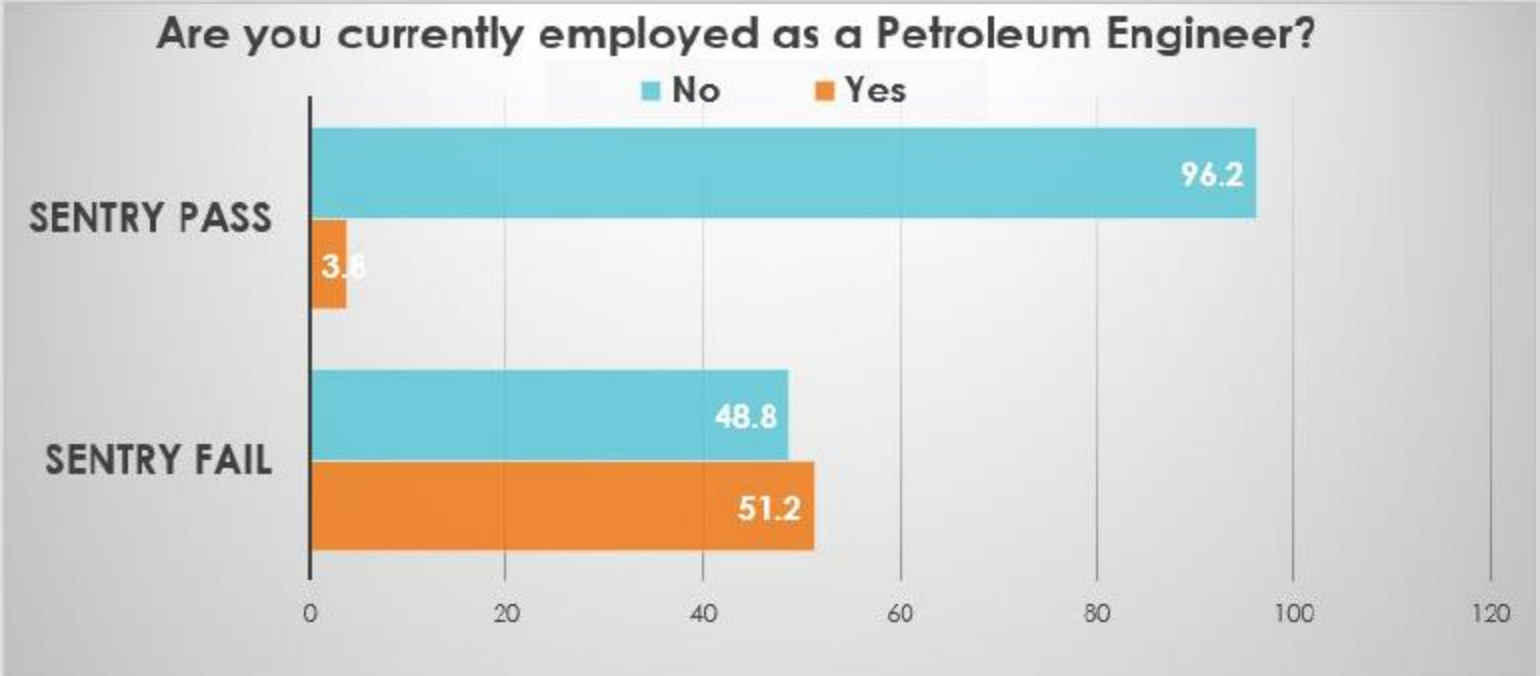
Conclusion: Only 70% of people in the US are sure the world is round.

Do you believe that the world is round or flat?



Sentry Results:

How do respondents who pass Sentry compare to those that don't?



Presurvey Screening Performance

Presurvey Screens	Estimated True Status			
		Flag	Pass	Total
	Flag	374	215	589
	Pass	73	1432	1505
	Total	447	1647	2094

- Presurvey screens quarantine 84% of problematic respondents
- Using presurvey screens alone, 95% of the sample is valid
- False positive rate is nearly equal to full postsurvey review

	Presurvey	Postsurvey
Accuracy	86%	91%
F1 Score	72%	81%
Cohen's Kappa	0.63	0.75
True Positive	84%	92%
True Negative	87%	90%
False Positive	13%	10%
False Negative	16%	8%

Summary

- All researchers must be aware of the different kinds of Bogus Participants, and use scientific, adaptable methods to keep their data quality high
- Robust Prescreen solutions can remove Low Quality Respondents and Provide Trustworthy Data, in line with leading Post-Survey Methods
- As threats continue to evolve, so must solutions