CIVIS

Testing ad content with survey experiments

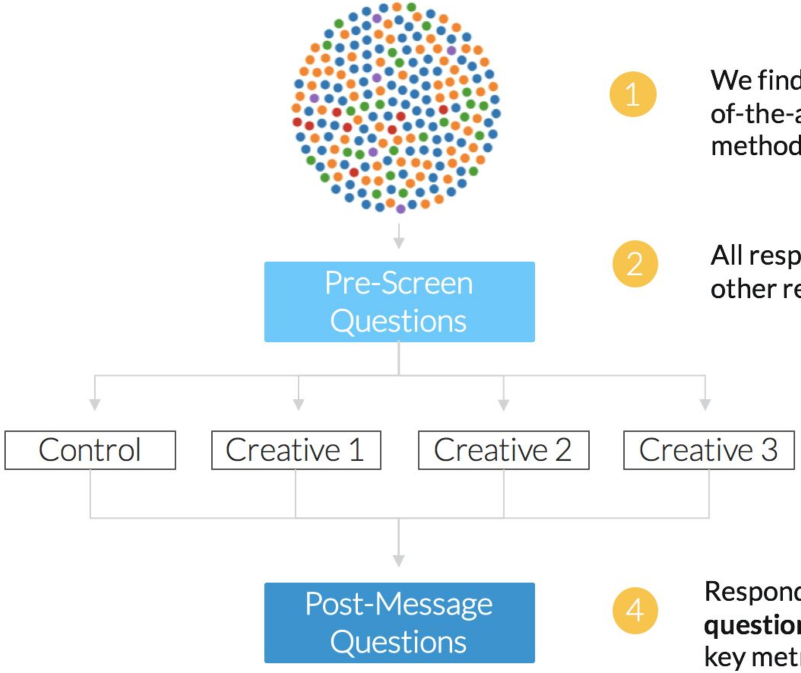
Patrick Miller

pmiller@civisanalytics.com

@patr1ck_mil

Charmin
enjoy the go
SNL

Testing ad content with a Survey Experiment



Pre-Screen Questions

Control

Creative 1

Creative 2

Creative 3

Post-Message Questions

1 We find a **representative sample** using our state-of-the-art sampling and weighting correction methods.

2 All respondents were asked demographic and other relevant **Pre-Screen questions**.

3 Respondents were then **randomly split** into treatment and control groups. Those in a treatment group were shown a single creative.

4 Respondents were then asked **Post-Message questions** to measure message effectiveness on key metrics.

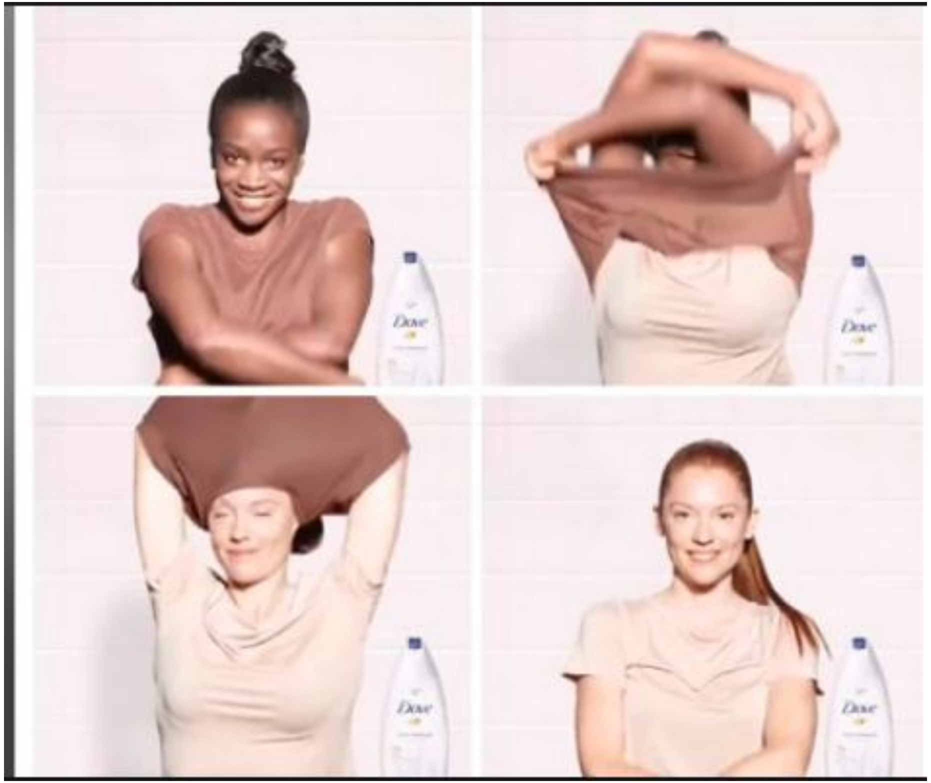
QUESTIONNAIRE
Very often
Often
Sometimes
Rarely

U.S. MAIL

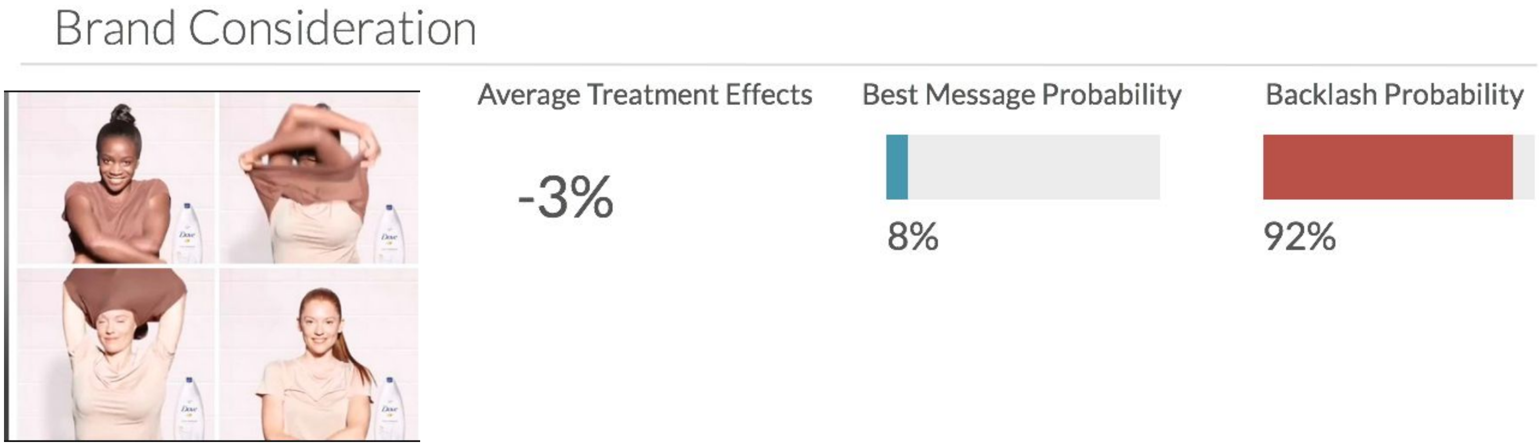
Examples

Tests we learned concrete things from

Dove



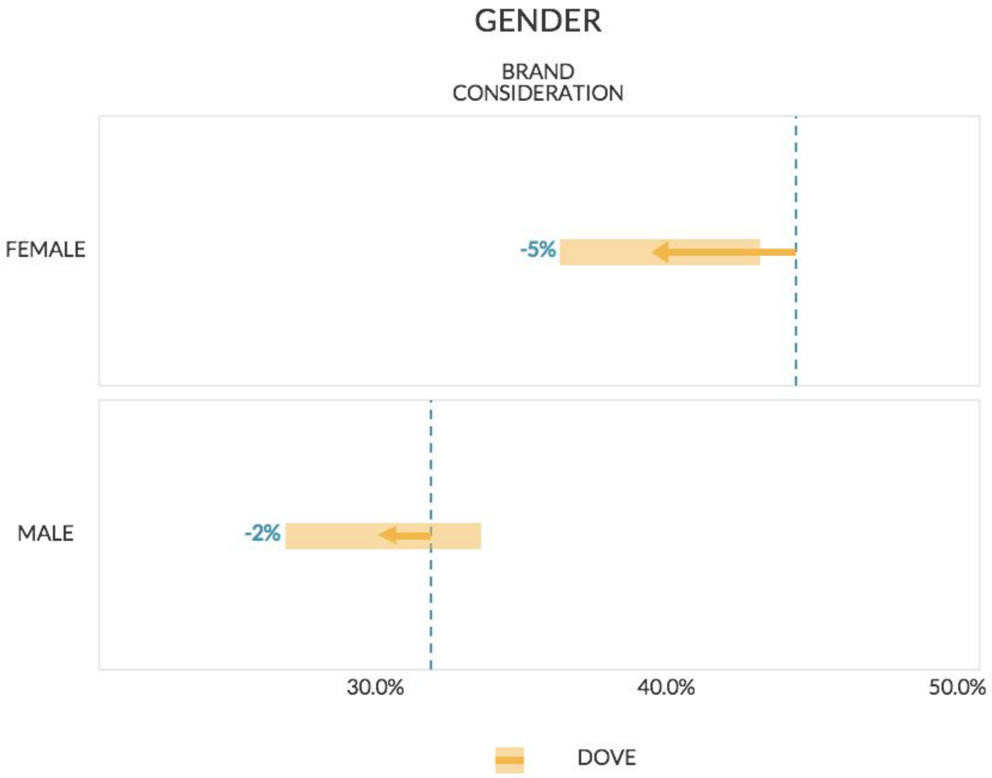
Overall Treatment Effects



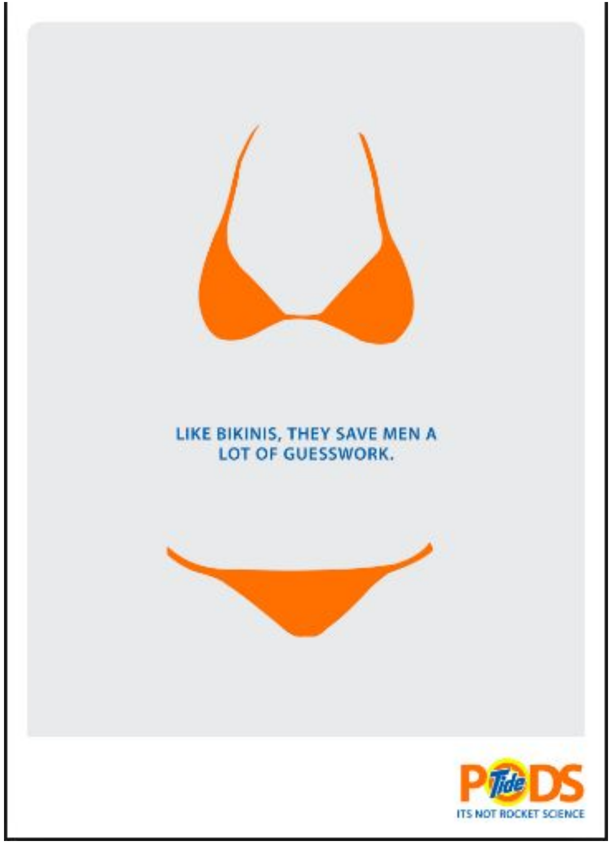
Brand Consideration

Average Treatment Effects	Best Message Probability	Backlash Probability
-3%	8%	92%

Treatment Effects by Gender



Tide

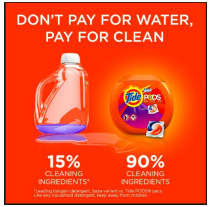

LIKE BIKINIS, THEY SAVE MEN A LOT OF GUESSWORK.
PODS
ITS NOT ROCKET SCIENCE

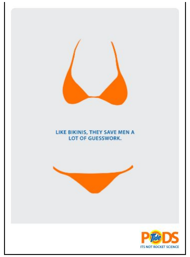
Tide
PODS
Tide
PROUD PARTNER

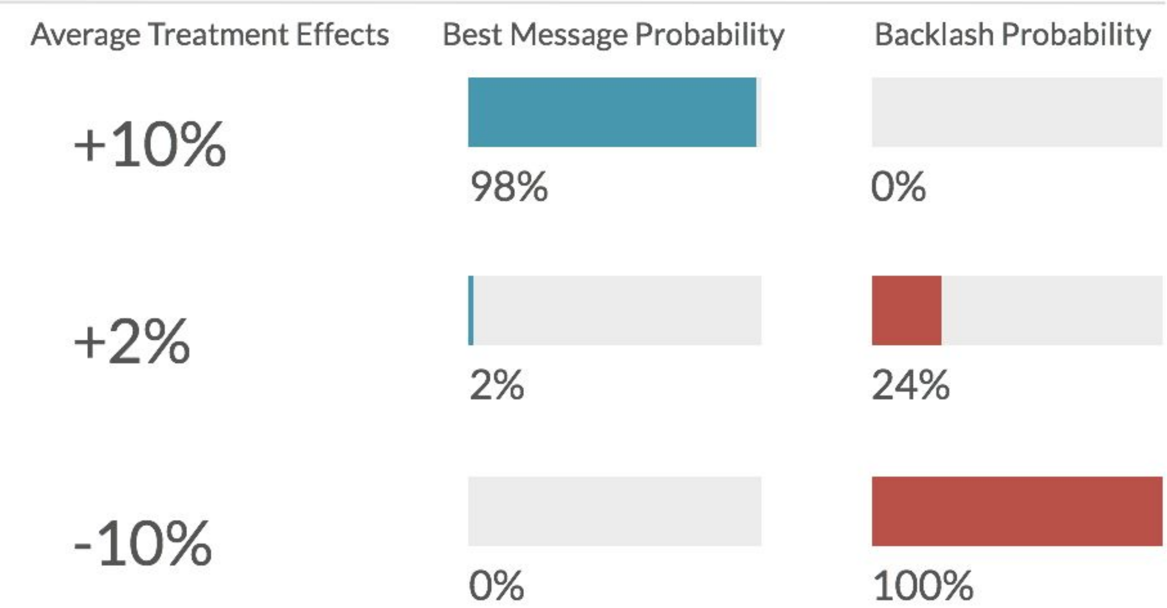
DON'T PAY FOR WATER, PAY FOR CLEAN
Tide PODS
15% CLEANING INGREDIENTS*
90% CLEANING INGREDIENTS
*Leading bargain detergent, base variant vs. Tide PODS® pacs. Like any household detergent, keep away from children.

Overall Treatment Effects

Brand Favorability

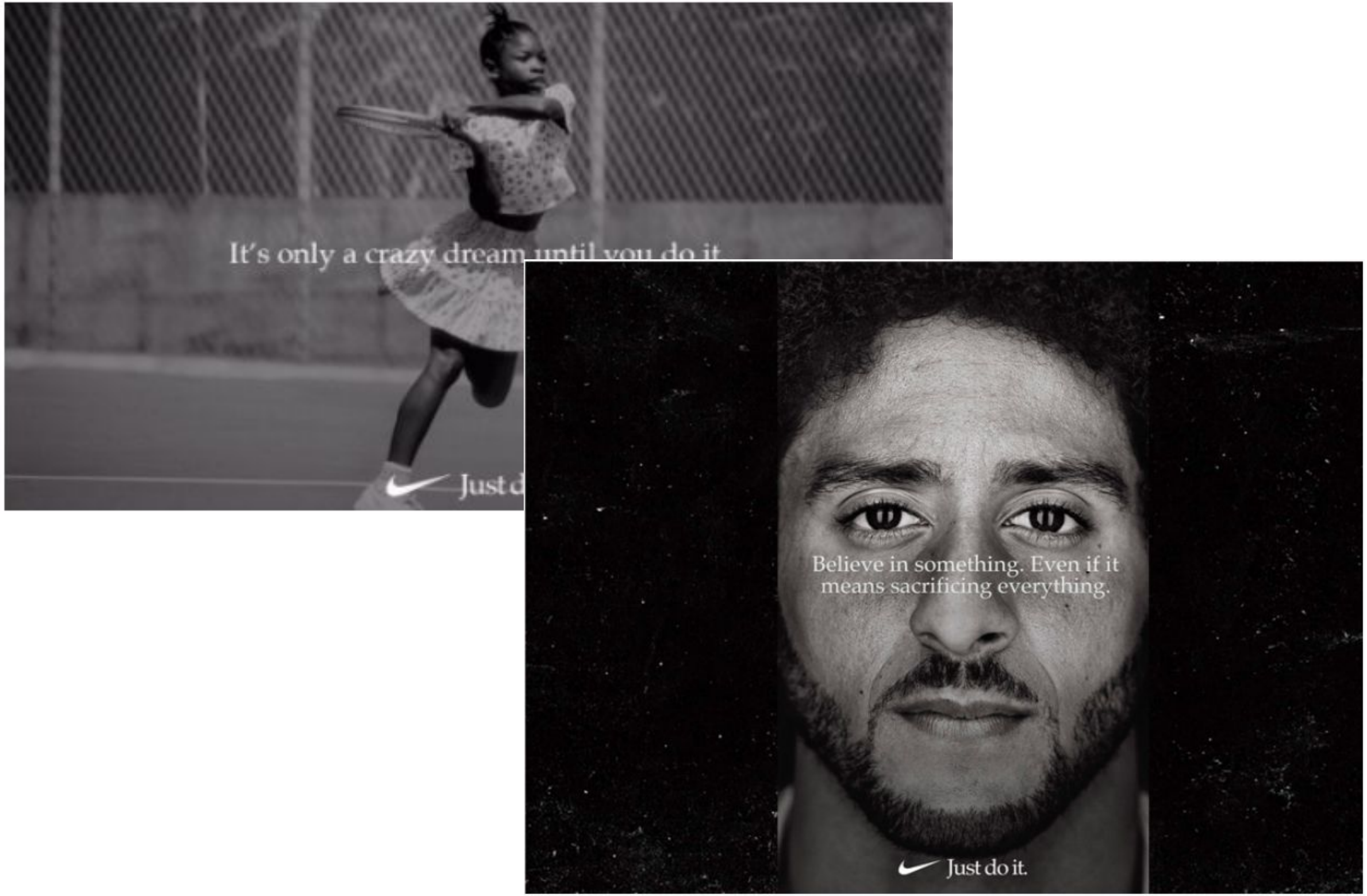

DON'T PAY FOR WATER, PAY FOR CLEAN
15% CLEANING INGREDIENTS*
90% CLEANING INGREDIENTS
*Leading bargain detergent, base variant vs. Tide PODS® pacs. Like any household detergent, keep away from children.





Average Treatment Effects	Best Message Probability	Backlash Probability
+10%	98%	0%
+2%	2%	24%
-10%	0%	100%

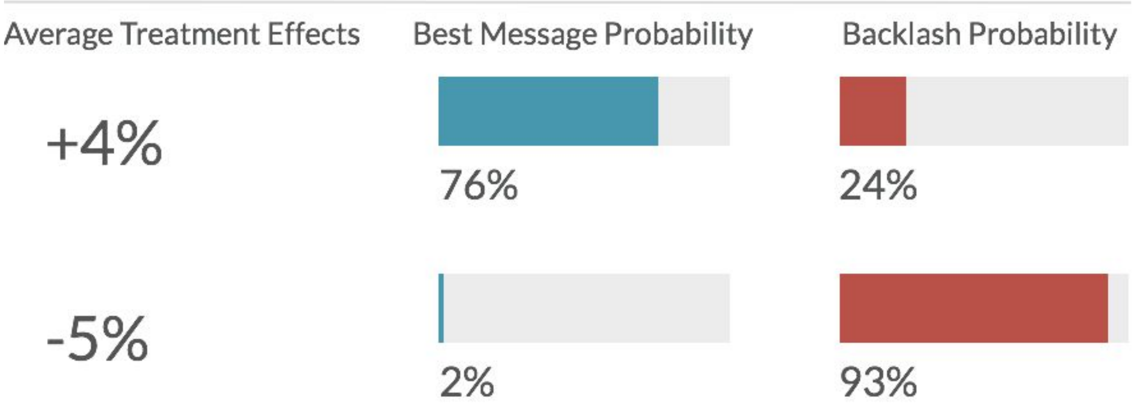
Nike


It's only a crazy dream until you do it.
Just d
Believe in something. Even if it means sacrificing everything.
Just do it.

Overall Treatment Effects

Brand Consideration

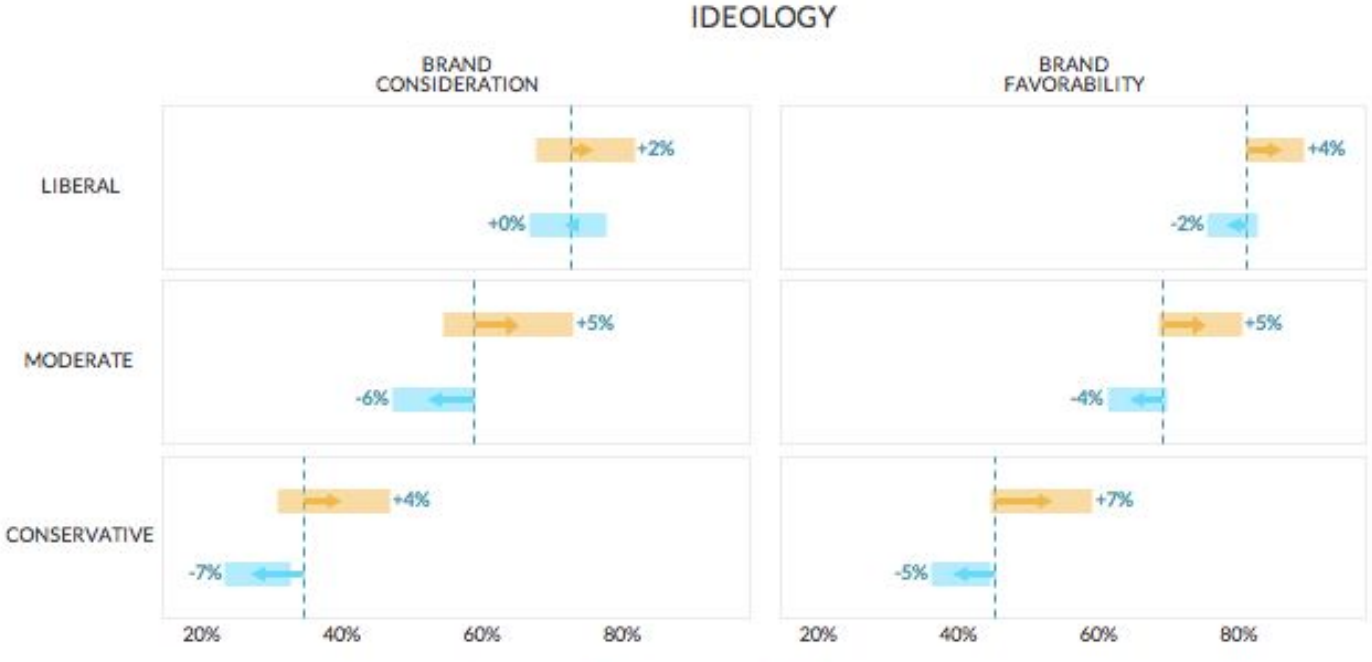
Believe in something. Even if it means sacrificing everything.
Just do it.



Average Treatment Effects	Best Message Probability	Backlash Probability
+4%	76%	24%
-5%	2%	93%

Treatment effects by Ideology

Conservatives had the most backlash **(-7%, -5%**) and had the lowest consideration (38%) while **Liberals** showed no backlash at all and had the **highest** consideration (76%)



▬ SERENA ▬ KAEPERNICK

Shaded areas represent 80% credible interval.

Meta analysis

Most ads are ineffective, but testing improves efficiency

Some ads are definitely ineffective

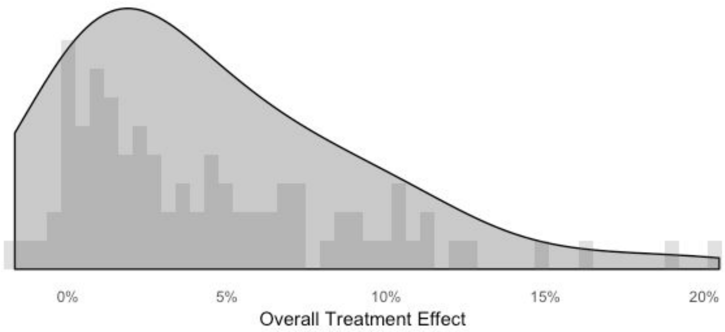
- **11%** of ads have backlash

A lot of ads are probably ineffective

- **26%** of ads have a treatment effect < 1pp
- **43%** of ads have treatment effects not conclusively different from 0pp

Testing multiple ads improves efficiency overall

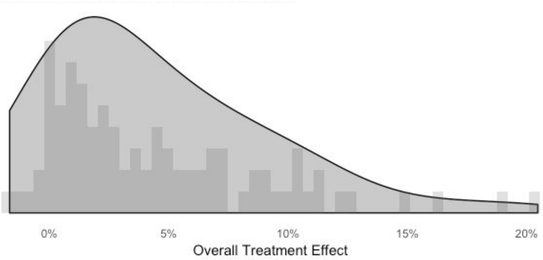
- The best ad was **13%** better on average than the worst ad in the experiment



Summary

EMPIRICAL DATA
empirical
data
©Study.com

Believe in something. Even if it means sacrificing everything.
Just do it.


0%
5%
10%
15%
20%
Overall Treatment Effect

Implementation

What we learned the hard way

Overview

Steps

1. Data collection
2. Survey weighting
3. Modeling
4. Reporting

Goals

1. Accurate
2. Interpretable
3. Trustworthy
4. Reusable

1. Data Collection

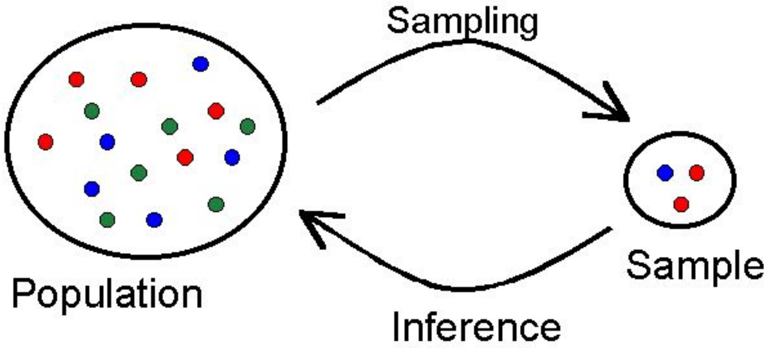
It’s small data, and measurement matters

Goals

1. **Accurate**
2. Interpretable
3. Trustworthy
4. Reusable

2. Survey Weighting

Your sample is biased, correct it with weighting



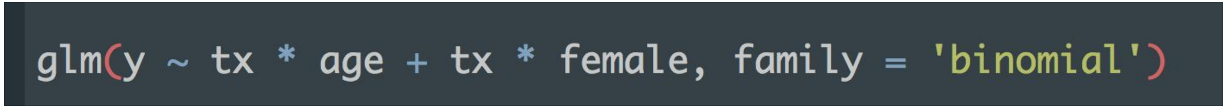
Goals

1. **Accurate**
2. Interpretable
3. Trustworthy
4. Reusable

3. Modeling

Keep it simple with a parametric model

"It's just logistic regression"



```
glm(y ~ tx * age + tx * female, family = 'binomial')
```

Goals

1. Accurate
2. **Interpretable**
3. **Trustworthy**
4. Reusable

freshspectrum

… and make it a service

https://devrant.com/rants/1854993/package-tsunami

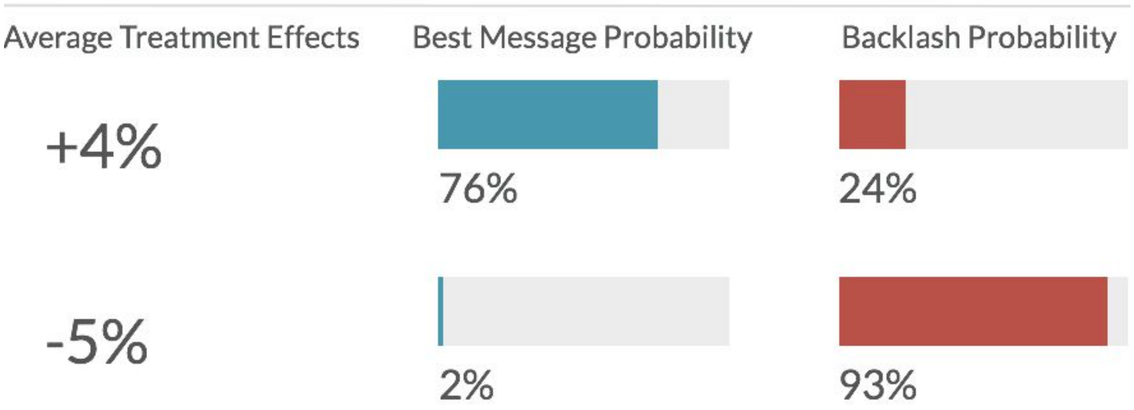
Goals

1. Accurate
2. Interpretable
3. Trustworthy
4. **Reusable**

4. Reporting

Overall Treatment Effects

Brand Consideration

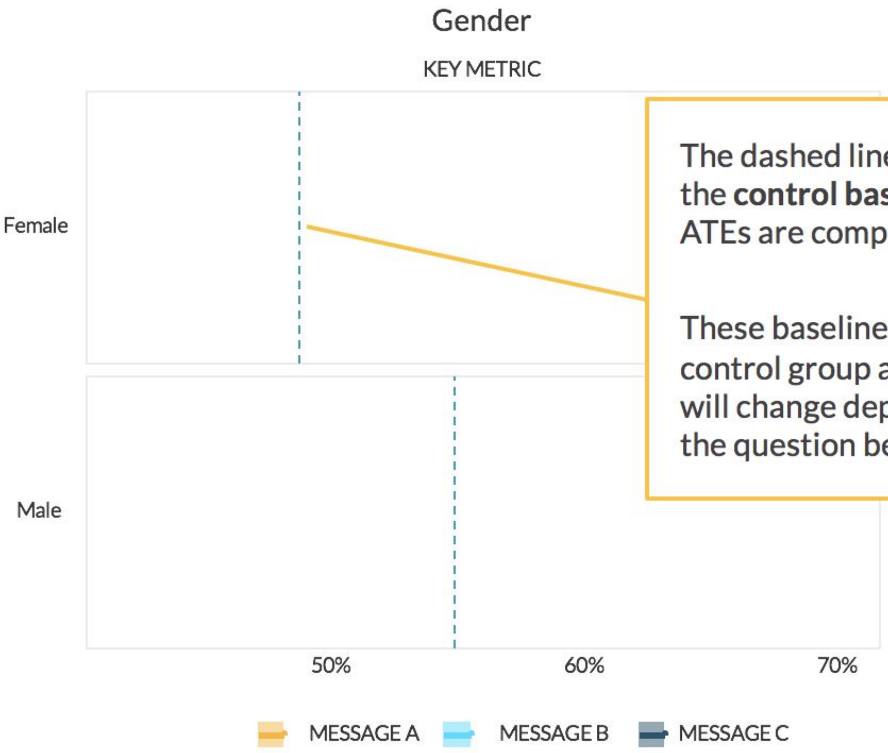


Average Treatment Effects	Best Message Probability	Backlash Probability
+4%	76%	24%
-5%	2%	93%

Goals

1. Accurate
2. **Interpretable**
3. Trustworthy
4. Reusable

Baselines



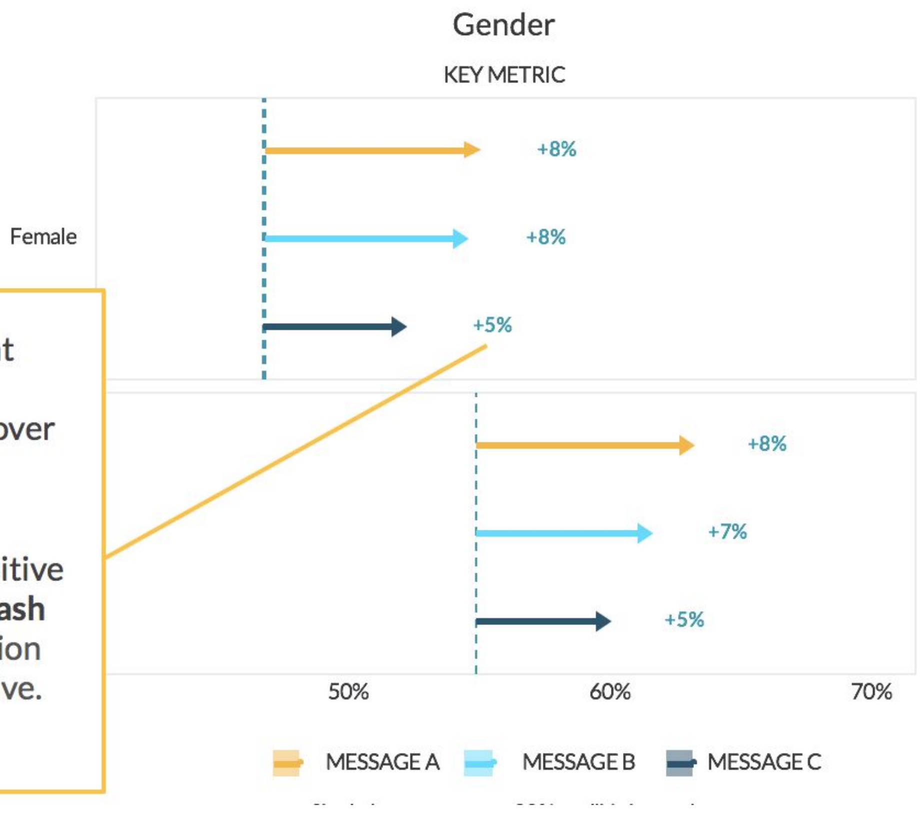
The dashed line represents the **control baseline** that ATEs are compared against.

These baselines are how the control group answered, so will change depending on the question being asked.

Goals

1. Accurate
2. Interpretable
3. **Trustworthy**
4. Reusable

Weighted Marginal Treatment Effects



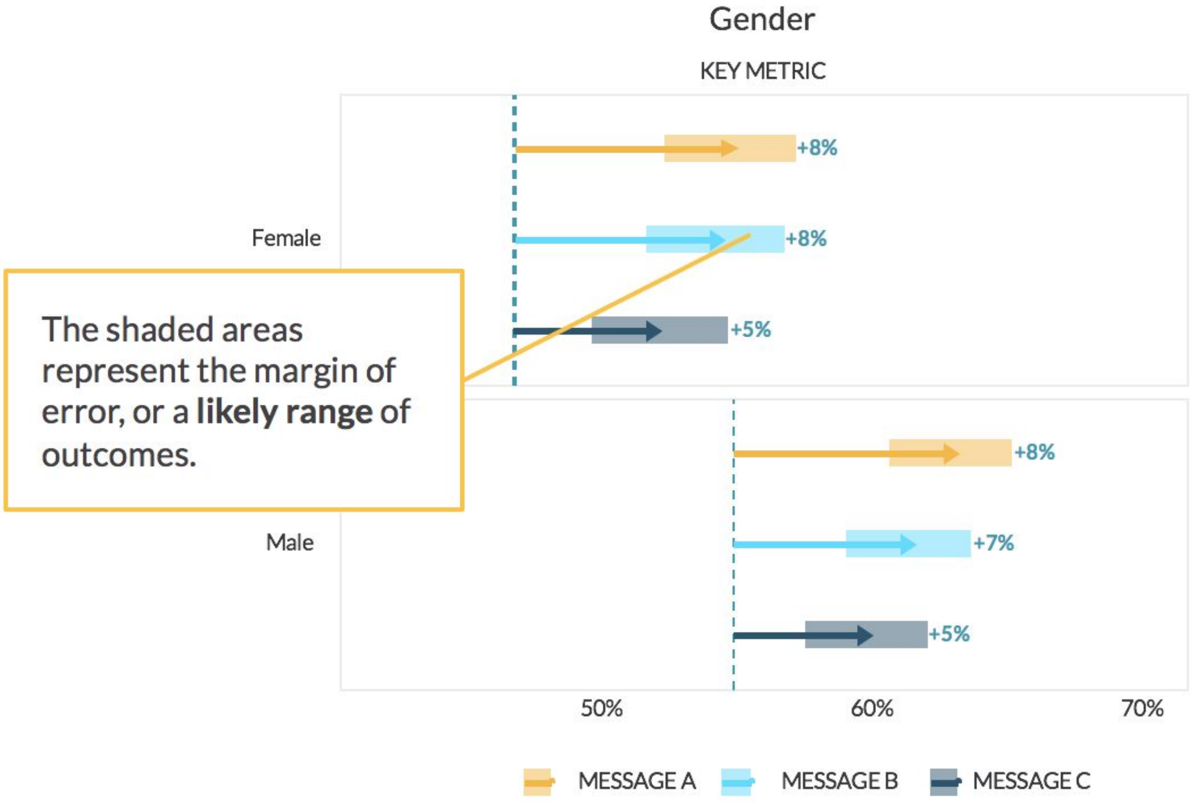
Average Treatment Effect (ATE) is the **incremental gain** over the control group.

An ATE can be positive or negative. **Backlash** is a negative reaction to a piece of creative.

Goals

1. **Accurate**
2. **Interpretable**
3. Trustworthy
4. Reusable

Uncertainty



The shaded areas represent the margin of error, or a **likely range** of outcomes.

Goals

1. Accurate
2. Interpretable
3. **Trustworthy**
4. Reusable

Testing ad content with survey experiments

Answer questions about ad effectiveness unambiguously, but testing allows your company to learn which ones are effective

Avoid bad ads that cause twitter/internet firestorms

For implementation prioritize trustworthiness and interpretability; make the model reusable by deploying as a service

Testing ad content with survey experiments

Patrick Miller

pmiller@civisanalytics.com

@patr1ck_mil