When Data Lies: Scientific Fraud and the Power of Measurement

Gov 51: Data Analysis and Politics

Scott Cunningham

Harvard University

Week 4, Thursday
February 19, 2026

A Study That Changed the Conversation

Can a single conversation change someone's mind about gay marriage?

LaCour and Green Claimed a 20-Minute Conversation Could Shift Attitudes

LaCour and Green (2014), published in *Science*:

- Door-to-door canvassing experiment in Los Angeles
- **Treatment**: 20-minute conversations about same-sex marriage
- **Control**: conversations about recycling
- Canvassers were either gay or straight (randomized)
- Outcome: feeling thermometer (0–100) toward gay people
- Measured at baseline and multiple follow-up waves

Michael LaCour was a graduate student at UCLA.
Donald Green was a senior professor at Columbia.

The Results Were Striking — and Persistent

What LaCour reported:

- Large shifts in attitudes after a single conversation
- Effects *persisted* across follow-up waves
- Gay canvassers produced larger effects than straight canvassers
- Effects even spread to other household members

This was unusual: most persuasion effects fade within days.
LaCour's effects seemed to last for months.

The Study Received Enormous Attention

- Published in *Science* — one of the most prestigious journals in the world
- Covered by the *New York Times*, *Washington Post*, NPR, and more
- Cited by political campaigns and advocacy organizations
- LaCour became a rising star — received a job offer from Princeton

But some researchers noticed something odd...

Something Doesn’t Add Up

Broockman and Kalla Tried to Build on LaCour's Work

David Broockman (Stanford) and **Joshua Kalla** (Berkeley) wanted to run a similar experiment.

- They contacted LaCour for methodological details
- As they dug into the data, they found irregularities
- They recruited **Peter Aronow** (Yale, statistician) to help investigate

What followed was one of the most important examples of *scientific self-correction* in recent history.

Your Measurement Tools Can Detect Fabricated Data

The tools you learned this week and last week are enough:

- **Histograms** reveal the shape of a distribution
- **Summary statistics** (mean, SD) describe its center and spread
- **Comparing distributions** across datasets flags anomalies

If you know what real data looks like, fake data stands out.

Fabricated data tends to look "too clean" or to match the wrong source.

The Baseline Data Matched an Existing Survey Exactly

Irregularity #1: The key finding that broke the case.

- The **CCAP** (Cooperative Campaign Analysis Project) is a large national survey from 2012
- LaCour's baseline feeling thermometer data was *statistically indistinguishable* from the CCAP data
- But LaCour claimed to have surveyed a separate sample in Los Angeles

Two independent samples from different populations should **not** produce nearly identical distributions.

Let's See This for Ourselves

```r
library(tidyverse)

## Load the data
gay <- read_csv("gayreshaped.csv")
ccap <- read_csv("ccap2012.csv")

## LaCour's Study 1 baseline thermometer
lacour_baseline <- gay %>%
  filter(study == 1) %>%
  pull(therm1)

## CCAP gay feeling thermometer
ccap_therm <- ccap$gaytherm
```

These are the same datasets from QSS Exercise 3.9.1.

The Summary Statistics Are Nearly Identical

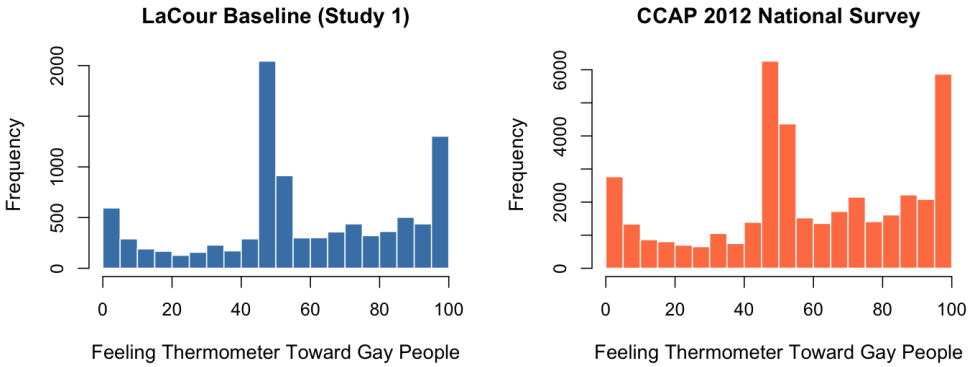
```
## Compare summary statistics
summary(lacour_baseline)
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     0.00   43.00   60.00   58.38   79.00  100.00    2441

summary(ccap_therm)
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.00   40.00   60.00   58.71   83.00  100.00

## Standard deviations
sd(lacour_baseline, na.rm = TRUE)  ## 28.50
sd(ccap_therm)                     ## 29.37
```

Same mean, same SD, same quartiles. From a "different sample"?

Histograms Make the Match Visually Obvious



Two independent samples should not produce the same histogram.

Two independent samples should not produce the same distribution.
If they do, something is wrong.

The Treatment Effects Were Too Uniform

Irregularity #2: The changes over time were suspiciously smooth.

- In real experiments, treatment effects are noisy — some people respond a lot, some not at all
- LaCour's treatment effects showed unusually uniform changes
- The variance of the within-person changes was *too small*
- Real human responses are messier than what LaCour reported

Real data is messy. If your data isn't messy, ask why.

The Reckoning

Broockman, Kalla, and Aronow Published Their Findings in May 2015

Their working paper, "Irregularities in LaCour (2014)," documented **eight** separate statistical anomalies:

1. Baseline data matched the CCAP survey
2. Treatment effects were too uniform
3. Response rates were implausibly high
4. The survey firm LaCour named denied involvement
5. LaCour could not produce the raw data
6. . . . and three more

Each irregularity was damning alone.
Together, they were conclusive.

Green Requested Retraction; *Science* Retracted in June 2015

- Donald Green (the senior author) immediately requested retraction
- Green had not collected the data himself — he trusted LaCour
- *Science* retracted the paper without LaCour's agreement
- LaCour's Princeton job offer was rescinded
- The survey firm LaCour claimed to have used denied any involvement

LaCour never admitted to fabrication, but could not produce any evidence the study was real.

Should Green have caught the fraud earlier?
What does this tell us about trust and verification in science?

The Real Experiment

Broockman and Kalla Ran the Study That LaCour Faked

After exposing the fraud, Broockman and Kalla did something remarkable:

- They actually ran a real canvassing experiment — **Broockman and Kalla (2016)**, published in *Science*
- Door-to-door conversations about **reducing prejudice toward transgender people**
- Used a verified survey firm with transparent data collection
- **Pre-registered** the study design before collecting data

They didn't just catch the fraud.
They did the science right.

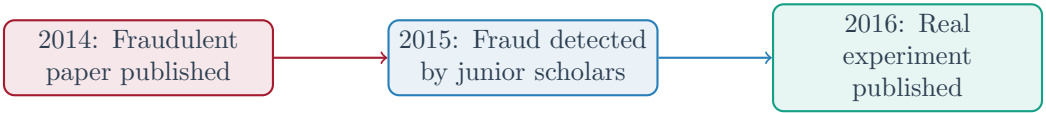
Genuine Conversations Do Change Minds — but Differently Than LaCour Claimed

The real experiment found:

- Durable attitude change toward transgender people
- Effects lasted at least 3 months (verified with real follow-up surveys)
- The mechanism was **analogic perspective-taking** — canvassers sharing personal stories prompted voters to reflect on their own experiences
- Effects were more modest than LaCour's fabricated results

The real science is more nuanced —
and more interesting — than the fraud.

This Is What Self-Correcting Science Looks Like



- The fraud was caught using **basic measurement tools**
- The retraction happened quickly once evidence was presented
- The real experiment used **pre-registration** and **open data**
- Two graduate students held a senior professor accountable

Skepticism + replication = science that works.

What This Means for You

The Tools You Already Have Are Powerful Enough to Catch Fraud

You don't need fancy statistics. You need **careful measurement**:

Histograms | **Summary Stats** | **Comparisons**

- Histograms reveal distributional anomalies
- Summary statistics (mean, SD) flag implausible patterns
- Comparing distributions across datasets is a basic but powerful check

These are the same tools you've been learning all semester.

Three Rules for Data Integrity

1. **Always look at your data before analyzing it**

   Plot histograms, run summary statistics, check for patterns. Never jump straight to regression.

2. **Compare your data to known benchmarks**

   If your survey data looks identical to an existing survey, ask why. Independent samples should differ.

3. **Be suspicious of results that are "too clean"**

   Real data is messy. Perfect results often mean something went wrong.

Replication is not an insult to the original researcher.
It is how science works.

Looking Ahead

Next week: Prediction and Linear Regression (QSS Chapter 4)

- How do we predict outcomes from data?
- The linear regression model
- Least squares estimation

Problem Set 2 is due Thursday, March 5.

PS2 uses `polls08.csv` for correlation and CI exercises. Start early — it has two parts.

Good measurement doesn't just describe data. It protects the integrity of science itself.

Questions?