

Generalizing from sample to population

Andrew Gelman

Department of Statistics and Department of Political Science
Columbia University, New York

Institute for Survey Research, University of Michigan
17 Nov 2014

me social tv video games music apps

A side-by-side portrait of Joe Biden on the left and Mitt Romney on the right, both in suits and ties, against a blue background.

Join the VP Debate
We're polling LIVE, 9 PM ET

UFC 153

Press Conf

A stylized, glowing image of a person's head and shoulders, likely representing a Kinect game or application.

Kinect Cent

A close-up image of a motorcycle handlebar and front wheel, likely from the game New Trials.

New Trials

A close-up portrait of a man with a beard and mustache, looking directly at the camera, likely an advertisement for a game or app.

ADVERTISEMENT



2/1

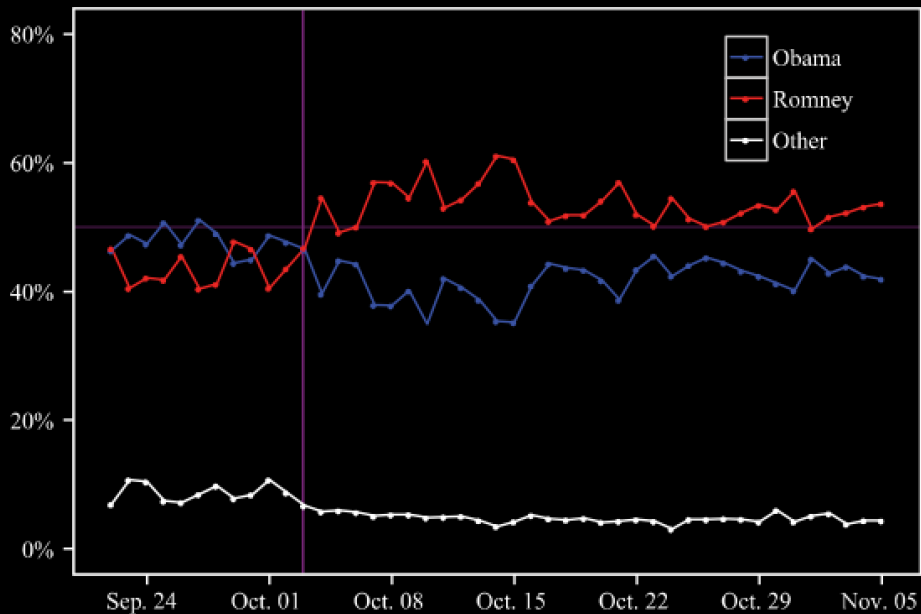
If the election were held today, who would you vote for?

Barack Obama

Mitt Romney

Other

Not sure



“This week, the New York Times and CBS News published a story using, in part, information from a non-probability, opt-in survey sparking concern among many in the polling community. In general, these methods have little grounding in theory and the results can vary widely based on the particular method used.”

— Michael Link,
President, American Association for Public Opinion Research

Michael W. Link is Chief Methodologist for Research Methods at The Nielsen Company, with a long base of experience in survey research, having worked in academia (University of South Carolina, 1999), not-for-profit research (RTI International, 1999-2004), government (Centers for Disease Control and Prevention, 2004-2007), and the private sector (Nielsen, 2007-present). He received his PhD in Political Science from the University of South Carolina. Michael's research centers around developing new survey methodologies for confronting some of the most pressing issues facing survey research, including new techniques for improving survey participation and data quality (use of address-based sampling, telephone call screening technologies), methodological issues involving use of multiple modes in data collection (mail, CATI, field, mobile, meters), and obtaining participation from hard-to-survey populations (e.g., isolated, racial and ethnic groups). His numerous research articles have appeared in *Public Opinion Quarterly* and other leading scientific journals.

An AAPOR member since 1993, Michael served as AAPOR Conference Chair in back-to-back years (2008 & 2010), a member of both the Cell Phone and Online task forces, an instructor for an AAPOR short-course, numerous short-courses, a reviewer for the student paper competition on several occasions, and a regular reviewer for *Public Opinion Quarterly*. He is a member of SAPOR, serving from 2008 to 2011 as President, Conference Chair, and Student Paper Competition Organizer and also a member of the AAPOR Council.

In 2011 he, along with several research colleagues, received AAPOR's Warren J. Mitofsky Award for their work on address based sampling designs. His current research focuses on new survey technologies, such as mobile and social platforms, as vehicles for measuring and understanding attitudes and behaviors. He will be teaching a short course on "The Role of New Technologies in Augmenting, or Replacing Traditional Surveys" at the 2012 AAPOR conference.

Nielsen feels the heat of competition as it flubs its ratings of news broadcasts, putting ABC ahead of NBC



BY DON KAPLAN

In spite of the goof, its global president took time to slam rival Rentrak, which collects different kind of data from viewers

NEW YORK DAILY NEWS / Sunday, October 19, 2014, 2:00 AM

AAA

MEDIA

TV Ratings by Nielsen Had Errors for Months

By BILL CARTER and EMILY STEEL OCT. 10, 2014



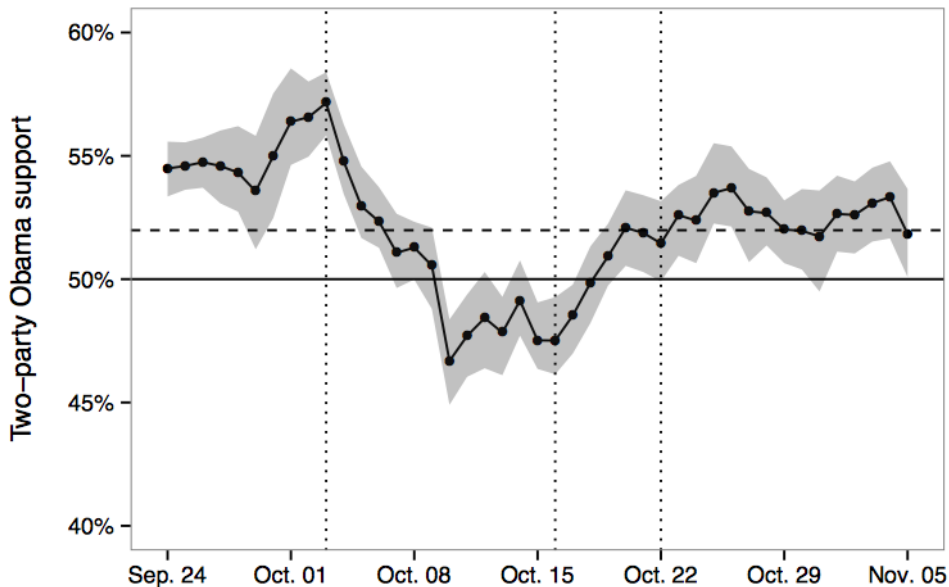
Email



Share

Nielsen, the television research firm, acknowledged on Friday that it had been reporting inaccurate ratings for the broadcast networks for the last seven months, a mistake that raises questions about the company's increasingly criticized system for measuring TV audiences.

Xbox estimates, adjusting for demographics

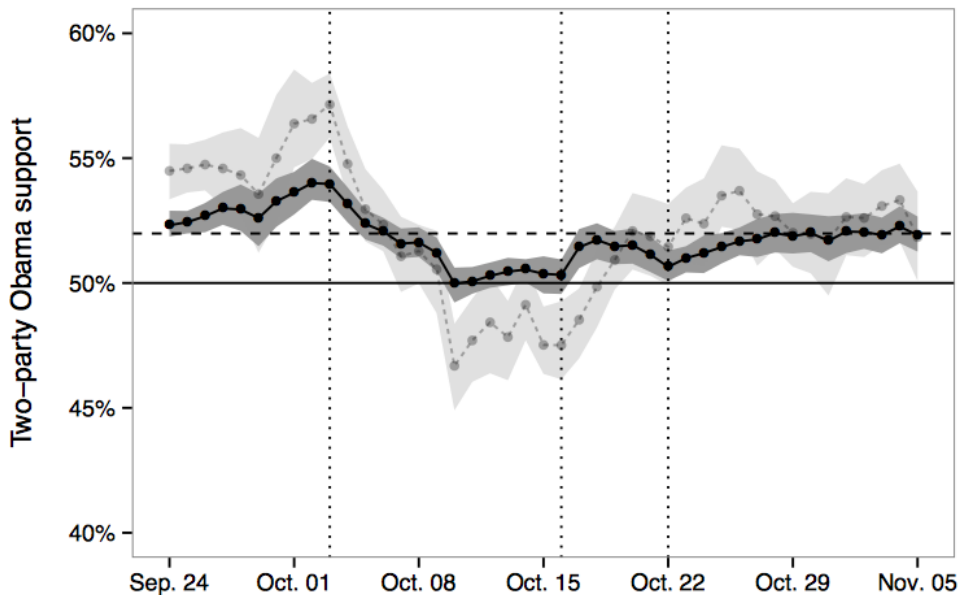






- ▶ Karl Rove, *Wall Street Journal*, 7 Oct: “Mr. Romney’s bounce is significant.”
- ▶ Nate Silver, *New York Times*, 6 Oct: “Mr. Romney has not only improved his own standing but also taken voters away from Mr. Obama’s column.”

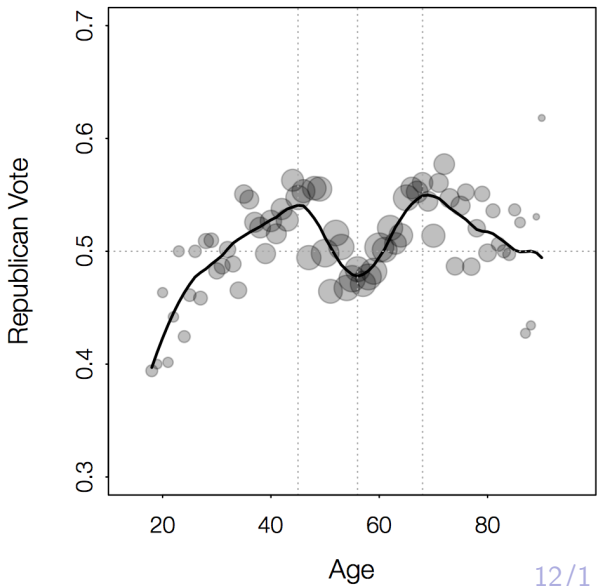
Xbox estimates, adjusting for demographics and partisanship



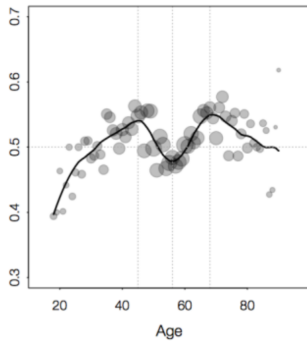
Jimmy Carter Republicans and George W. Bush Democrats



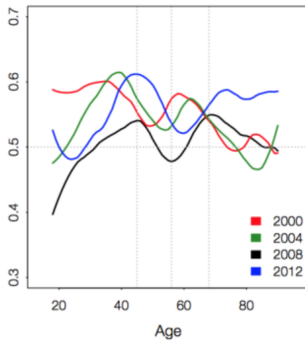
Non-Monotonic Age Curve in 2008



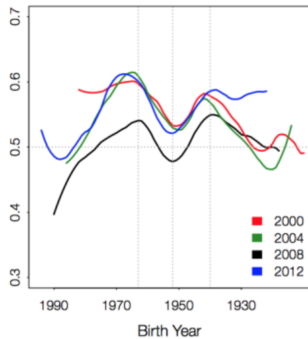
Non-Monotonic Age Curve in 2008



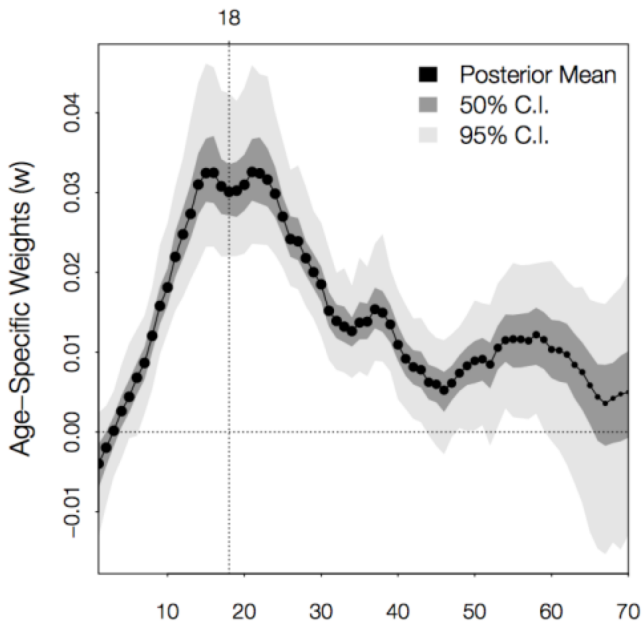
Non-Monotonicity in Other Elections



Lining up by Birth Year



The Formative Years



Stan is a probabilistic programming language implementing full Bayesian statistical inference with

- MCMC sampling (NUTS, HMC)

and penalized maximum likelihood estimation with

- Optimization (BFGS)

Stan is coded in C++ and runs on all major platforms (Linux, Mac, Windows).

Stan is freedom-respecting, open-source software (new BSD core, GPLv3 interfaces).

Interfaces

Download and getting started instructions, organized by interface:

- [RStan v2.5.0](#) (R)
- [PyStan v2.5.0](#) (Python)
- [CmdStan v2.5.0](#) (shell, command-line terminal)
- [MatlabStan](#) (MATLAB)
- [Stan.jl](#) (Julia)

[Home](#)[RStan](#)[PyStan](#)[CmdStan](#)[MatlabStan](#)[Stan.jl](#)[Manual](#)[Examples](#)[Groups](#)[Issues](#)[Contribute](#)[Source](#)

The Fluctuating Female Vote: Politics, Religion, and the Ovulatory Cycle

Psychological Science
XX(X) 1–10
© The Author(s) 2013
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797612466416
pss.sagepub.com



**Kristina M. Durante¹, Ashley Rae¹, and
Vladas Griskevicius²**

¹College of Business, University of Texas, San Antonio, and ²Carlson School of
Management, University of Minnesota

Abstract

Each month, many women experience an ovulatory cycle that regulates fertility. Although research has found that this cycle influences women's mating preferences, we proposed that it might also change women's political and religious views. Building on theory suggesting that political and religious orientation are linked to reproductive goals, we tested how fertility influenced women's politics, religiosity, and voting in the 2012 U.S. presidential election. In two studies with large and diverse samples, ovulation had drastically different effects on single women and women in committed relationships. Ovulation led single women to become more liberal, less religious, and more likely to vote for Barack Obama. In contrast, ovulation led women in committed relationships to become more conservative, more religious, and more likely to vote for Mitt Romney. In addition, ovulation-induced changes in political orientation mediated women's voting behavior. Overall, the ovulatory cycle not only influences women's politics but also appears to do so differently for single women than for women in relationships.

CHOOSE YOUR OWN ADVENTURE™ · 1

YOU'RE THE STAR OF THE STORY!
CHOOSE FROM 40 POSSIBLE ENDINGS

THE CAVE OF TIME

BY EDWARD

WARNING ! ! ! !

Do not read this book straight through from beginning to end! These pages contain many different adventures you can go on in the Cave of Time. From time to time as you read along, you will be asked to make a choice. Your choice may lead to success or disaster!

The adventures you take are a result of your choice. *You* are responsible because *you* choose! After you make your choice, follow the instructions to see what happens to you next.

Remember—you cannot go back! Think carefully before you make a move! One mistake can be your last . . . or it *may* lead you to fame and fortune!



Women are more likely to wear red or pink at peak fertility.

Beall AT, Tracy JL.

University of British Columbia.

Abstract

Although females of many species closely related to humans signal their fertile v
manner, often involving red or pink coloration, no such display has been found fo
evidence that men are sexually attracted to women wearing or surrounded by red
women show a behavioral tendency toward wearing reddish clothing when at pea
samples (N = 124), women at high conception risk were more than 3 times more
pink shirt than were women at low conception risk, and 77% of women who wore
to be at high, rather than low, risk. Conception risk had no effect on the prevalen
color. Our results thus suggest that red and pink adornment in women is reliably
and that female ovulation, long assumed to be hidden, is associated with a salie

[Home](#)[OnlineFirst](#)[All Issues](#)[Subscribe](#)[RSS](#) [Email Alerts](#)

The Ancestral Logic of Politics



Upper-Body Strength Regulates Men's Assertion of Self-Interest Over Economic Redistribution


Over human evolutionary history, upper-body strength has been a major component of fighting ability. Evolutionary models of animal conflict predict that actors with greater fighting ability will more actively attempt to acquire or defend resources than less formidable contestants will. Here, we applied these models to political decision making about redistribution of income and wealth among modern humans. In studies conducted in Argentina, Denmark, and the United States, men with greater upper-body strength more strongly endorsed the self-beneficial policies. A meta-analysis of these studies found that men with greater upper-body strength endorsed self-beneficial policies more strongly than men with lower upper-body strength. This effect was observed in all three studies and was not moderated by country or study design. These findings suggest that upper-body strength may be a key factor in the evolution of human political behavior.

Journal's Paper on ESP Expected to Prompt Outrage

By [BENEDICT CAREY](#)

Published: January 5, 2011

One of psychology's most respected journals has agreed to publish a paper presenting what its author describes as strong evidence for extrasensory perception, the ability to sense future events.

 [Enlarge This Image](#)



Heather Ainsworth for The New York Times

Work by Daryl J. Bem on extrasensory perception is scheduled to be published this year.

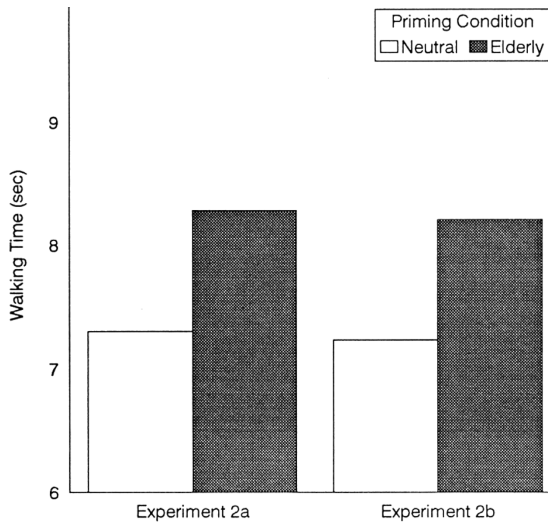
The decision may delight believers in so-called paranormal events, but it is already mortifying scientists. Advance copies of the [paper](#), to be published this year in The Journal of Personality and Social Psychology, have circulated widely among psychological researchers in recent weeks and have generated a mixture of amusement and scorn.

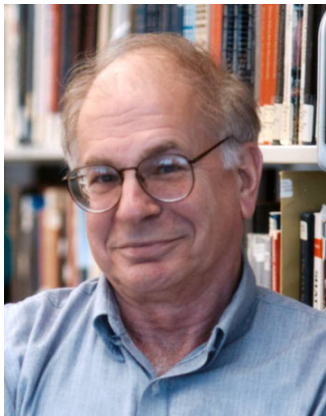
The paper describes nine unusual lab experiments performed over the past decade by its author, [Daryl J. Bem](#), an emeritus professor at Cornell, testing the ability of college students to accurately sense random events,

Results

Experiment 2a. A t test was computed to ascertain the effect of the priming manipulation on walking speed. Participants in the elderly priming condition ($M = 8.28$ s) had a slower walking speed compared to participants in the neutral priming condition ($M = 7.30$ s), $t(28) = 2.86$, $p < .01$, as predicted.

Experiment 2b. In the replication, analyses revealed that participants in the elderly priming condition ($M = 8.20$ s) again had a slower walking speed compared to participants in the neutral priming condition ($M = 7.23$ s), $t(28) = 2.16$, $p < .05$.





Daniel Kahneman (2011):

"When I describe priming studies to audiences, the reaction is often disbelief ... The idea you should focus on, however, is that disbelief is not an option. The results are not made up, nor are they statistical flukes. You have no choice but to accept that the major conclusions of these studies are true."

Elderly-Related Words Prime Slow Walking (#15)

[Return to View Chart](#)

[How to Cite this Report](#)

Reference to Original Report of Finding

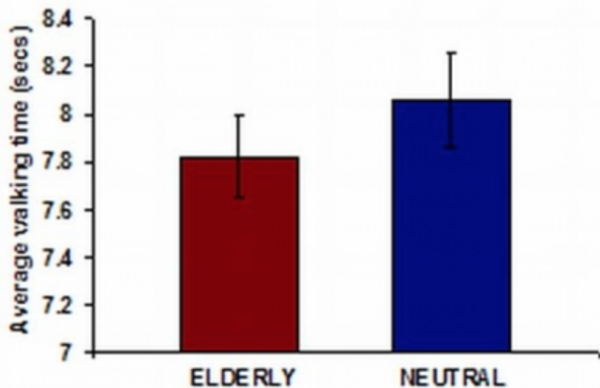
Bargh, J. A., Chen, M., & Burrows, L. (1996). Automatic action. *Journal of Personality and Social Psychology*, 71

Title

Elderly-Related Words Prime Slow Walking

The attempted replication

Average time (secs) to walk 32ft, as a function of priming stereotype condition



Daniel Kahneman (2011):

“When I describe priming studies to audiences, the reaction is often disbelief . . . The idea you should focus on, however, is that disbelief is not an option. The results are not made up, nor are they statistical flukes. You have no choice but to accept that the major conclusions of these studies are true.”

Wagenmakers et al. (2014):

“[After] a long series of failed replications . . . disbelief does in fact remain an option.”

Alan Turing (1950):



"I assume that the reader is familiar with the idea of extra-sensory perception, and the meaning of the four items of it, viz. telepathy, clairvoyance, precognition and psycho-kinesis. These disturbing phenomena seem to deny all our usual scientific ideas. How we should like to discredit them! Unfortunately the statistical evidence, at least for telepathy, is overwhelming."



Psychological SCIENCE

Research, Theory, & Application in
Psychology and Related Sciences

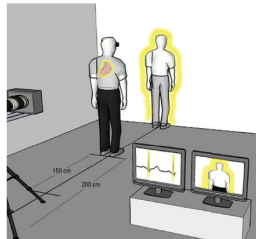
A Journal of the Association for Psychological Science

This Week in *Psychological Science* (TWiPS)

The links below take you to the journal via the APS website. If not already logged in, you will be redirected to log-in using your last name (Gelman) and Member ID (8167).

[Turning Body and Self Inside Out: Visualized Heartbeats Alter Bodily Self-Consciousness and Tactile Perception](#)

Jane Elizabeth Aspell, Lukas Heydrich, Guillaume Marillier, Tom Lavanchy, Bruno Herbelin, and Olaf Blanke



Studies of body perception have mostly focused on manipulations of exteroceptive cues (e.g., vision and touch); however, interoceptive cues (i.e., representations of internal bodily states) may be just as important for self-perception. Participants viewed a virtual body or a rectangle, each of which had a flashing outline that was synchronous or asynchronous with the participant's own heartbeat. Self-identification was stronger for people viewing the virtual body with the synchronous flashing outline than for those viewing the body with the asynchronous flashing outline or for those viewing the rectangles. This suggests that both interoceptive and exteroceptive cues play

important roles in bodily self-perception.

[Aging 5 Years in 5 Minutes: The Effect of Taking a Memory Test on Older Adults' Subjective Age](#)

Matthew L. Hughes, Lisa Geraci, and Ross L. De Forrest

Subjective age – how old people feel – is related to psychological and physical well-being. In this study, the researchers examined whether common memory-testing procedures influence adults' subjective age. Older and younger adults rated their subjective age before and after taking a memory test. Older adults reported feeling older after taking the memory test, but younger adults did not. A follow-up study found that simply anticipating taking a memory test increased older adults' subjective age. These findings bring to light changes in self-perception that can occur when older adults undergo routine memory testing.



Call for Submissions

This week in Psychological Science

- ▶ “Turning Body and Self Inside Out: Visualized Heartbeats Alter Bodily Self-Consciousness and Tactile Perception”
- ▶ “Aging 5 Years in 5 Minutes: The Effect of Taking a Memory Test on Older Adults’ Subjective Age”
- ▶ “The Double-Edged Sword of Grandiose Narcissism: Implications for Successful and Unsuccessful Leadership Among U.S. Presidents”
- ▶ “On the Nature and Nurture of Intelligence and Specific Cognitive Abilities: The More Heritable, the More Culture Dependent”
- ▶ “Beauty at the Ballot Box: Disease Threats Predict Preferences for Physically Attractive Leaders”
- ▶ “Shaping Attention With Reward: Effects of Reward on Space- and Object-Based Selection”
- ▶ “It Pays to Be Herr Kaiser: Germans With Noble-Sounding Surnames More Often Work as Managers Than as Employees”

This week in Psychological Science

- ▶ $N = 17$
- ▶ $N = 57$
- ▶ $N = 42$
- ▶ $N = 7,582$
- ▶ $N = 123 + 156 + 66$
- ▶ $N = 47$
- ▶ $N = 222,924$

Labor Market Returns to Early Childhood Stimulation: a 20-year Followup to an Experimental Intervention in Jamaica

Paul Gertler, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, Sally Grantham-McGregor

We find large effects on the earnings of participants from a randomized intervention that gave psychosocial stimulation to stunted Jamaican toddlers living in poverty. The intervention consisted of one-hour weekly visits from community Jamaican health workers over a 2-year period that taught parenting skills and encouraged mothers to interact and play with their children in ways that would develop their children's cognitive and personality skills. We re-interviewed the study participants 20 years after the intervention. Stimulation increased the average earnings of participants by 42 percent. Treatment group earnings caught up to the earnings of a matched non-stunted comparison group. These findings show that psychosocial stimulation early in childhood in disadvantaged settings can have substantial effects on labor market outcomes and reduce later life inequality.

The “That which does not destroy my statistical significance makes it stronger” fallacy

Charles Murray: “To me, the experience of early childhood intervention programs follows the familiar, discouraging pattern . . . small-scale experimental efforts [$N = 123$ and $N = 111$] staffed by highly motivated people show effects. When they are subject to well-designed large-scale replications, those promising signs attenuate and often evaporate altogether.”

James Heckman: “The effects reported for the programs I discuss survive batteries of rigorous testing procedures. They are conducted by independent analysts who did not perform or design the original experiments. The fact that samples are small works *against* finding any effects for the programs, much less the statistically significant and substantial effects that have been found.”



- ▶ Use MRP to adjust for known differences between sample and population
- ▶ “Crisis of replication” as a sampling problem, between and within studies

Application to a recent project in sociology

- ▶ Start with model $y = X\beta + \text{error}$
- ▶ Need to adjust for variables Z that are correlated with nonresponse
- ▶ New model, $y = X\beta + Z\gamma + \text{error}$
- ▶ Interaction model:
$$y = X\beta + Z\gamma + \delta(X\beta)(Z\gamma) + \text{error}$$
- ▶ Weighting adjustments as special case

Adjusting for known differences between sample and population

- ▶ Include more predictors
- ▶ Multilevel regression
- ▶ Poststratification

Big Data need Big Model

