Taking Bayesian inference seriously

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

Harvard conference on Big Data, 22 Aug 2016

Birthdays!

Social Science & Medicine 73 (2011) 1246–1248

Contents lists available at SciVerse ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Short report

Influence of Valentine's Day and Halloween on Birth Timing

Becca R. Levy*, Pil H. Chung, Martin D. Slade

Yale University, School of Public Health, Division of Social & Behavioral Sciences, 60 College Street, New Haven, CT 06520-8034, United States

ARTICLE INFO

Article history:
Available online 28 July 2011

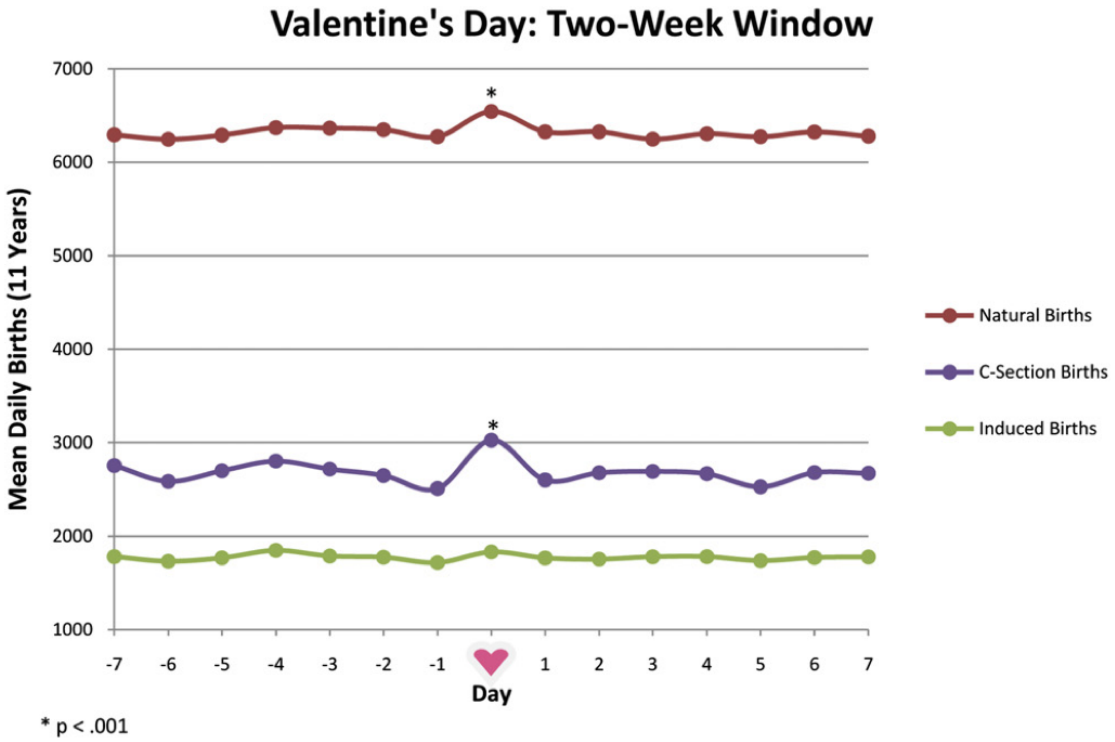
Keywords:
United States
Culture
Birth timing
Holidays
Pregnancy
Biocultural
Birth

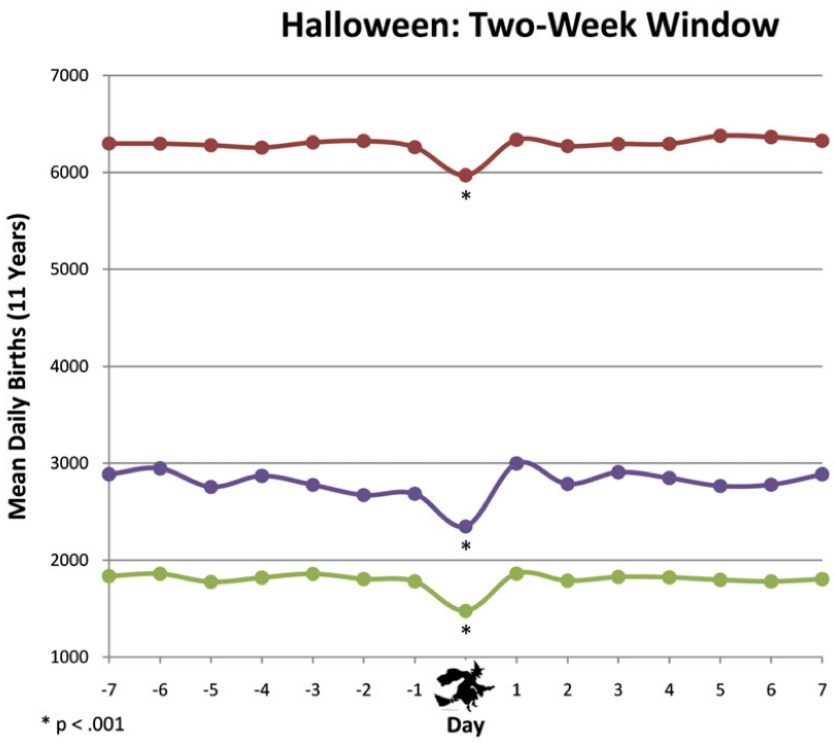
ABSTRACT

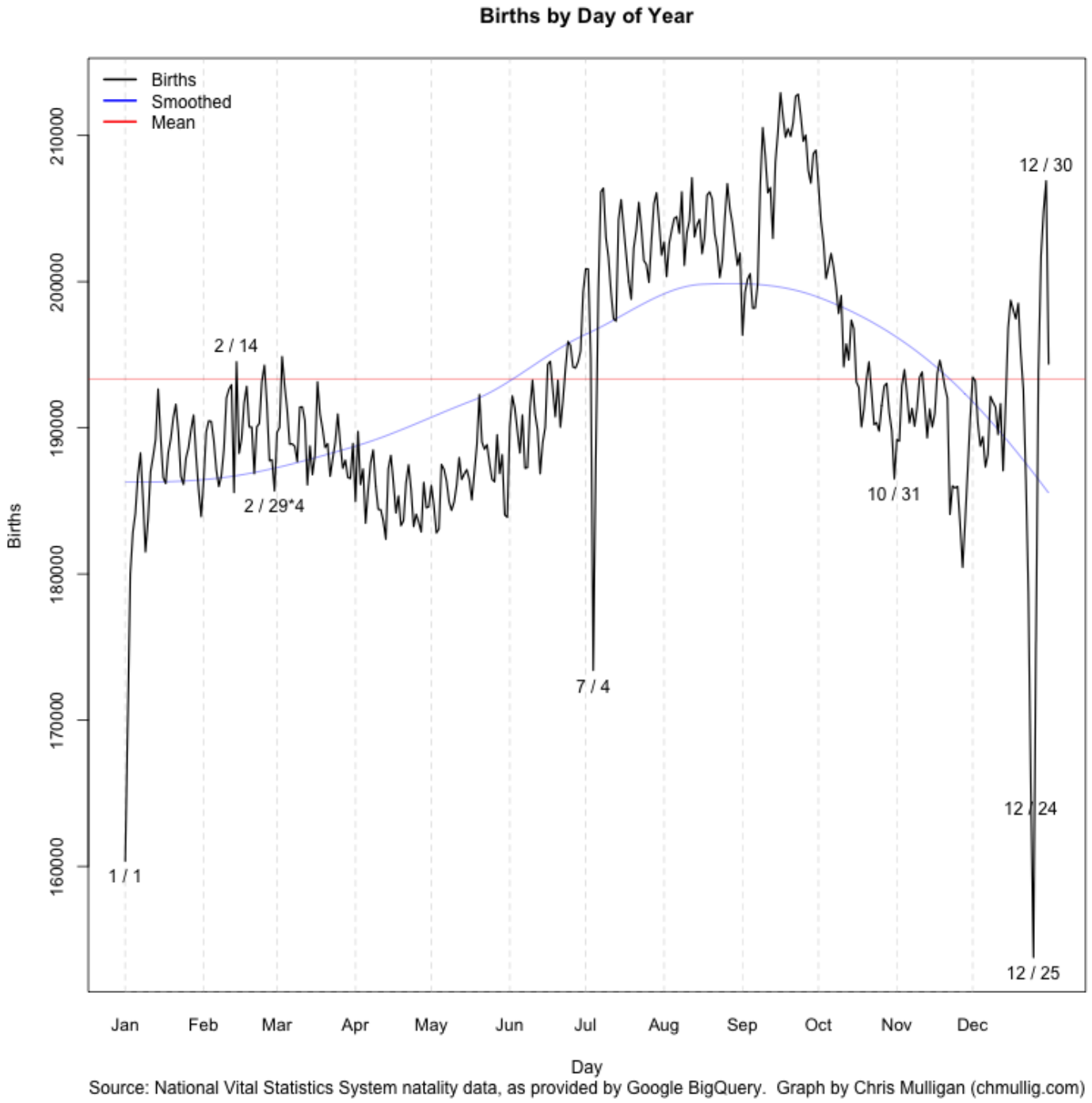
It is known that cultural representations, in the form of stereotypes, can influence functional health. We predicted that the influence of cultural representations, in the form of salient holidays, would extend to birth timing. On Valentine's Day, which conveys positive symbolism, there was a 3.6% increase in spontaneous births and a 12.1% increase in cesarean births. Whereas, on Halloween, which conveys negative symbolism, there was a 5.3% decrease in spontaneous births and a 16.9% decrease in cesarean births. These effects reached significance at $p < .0001$, after adjusting for year and day of the week. The sample was based on birth-certificate information for all births in the United States within one week on either side of each holiday across 11 years. The Valentine's-Day window included 1,676,217 births and the Halloween window included 1,809,304 births. Our findings raise the possibility that pregnant women may be able to control the timing of spontaneous births, in contrast to the traditional assumption, and that scheduled births are also influenced by the cultural representations of the two holidays.

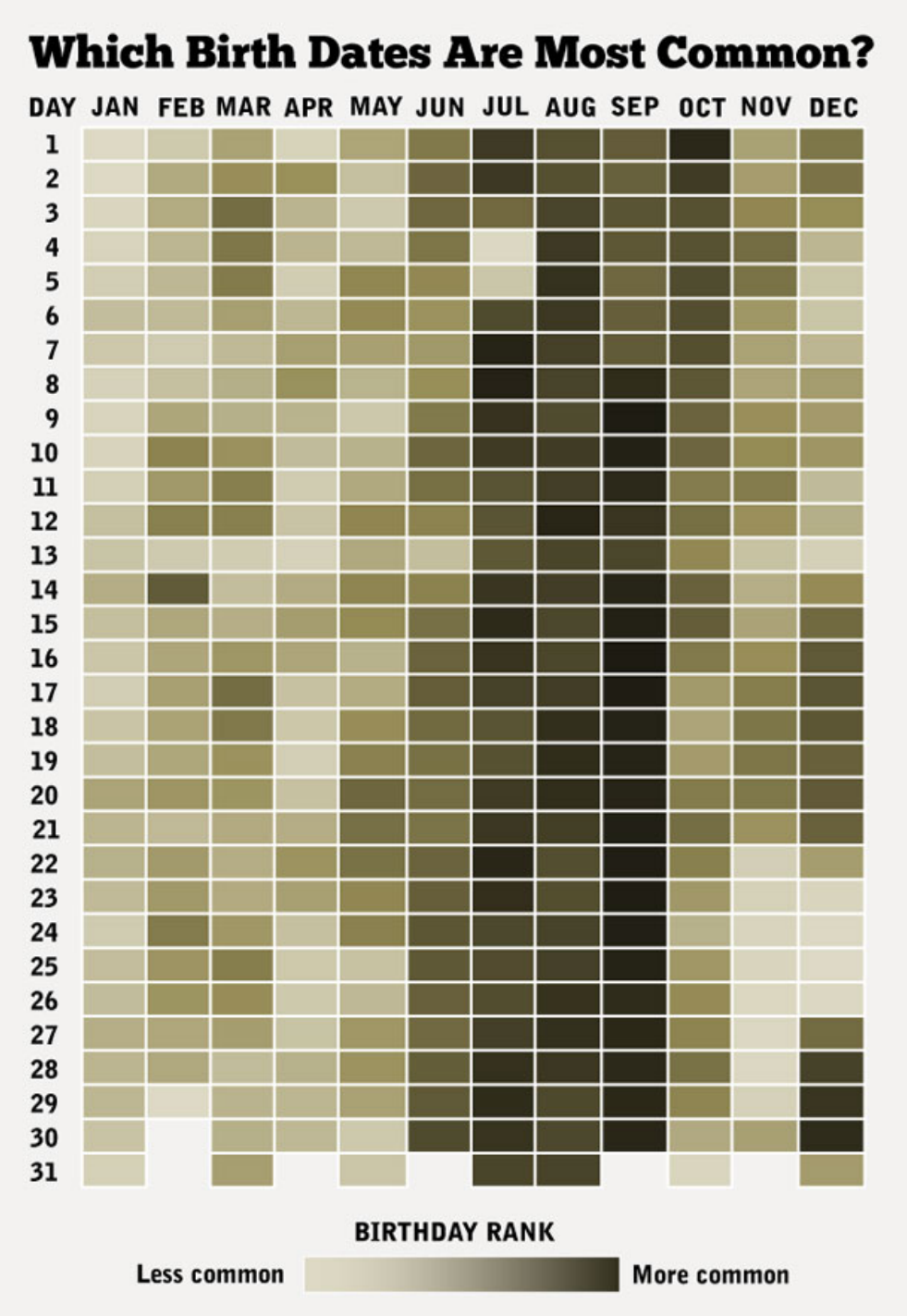
© 2011 Elsevier Ltd. All rights reserved.

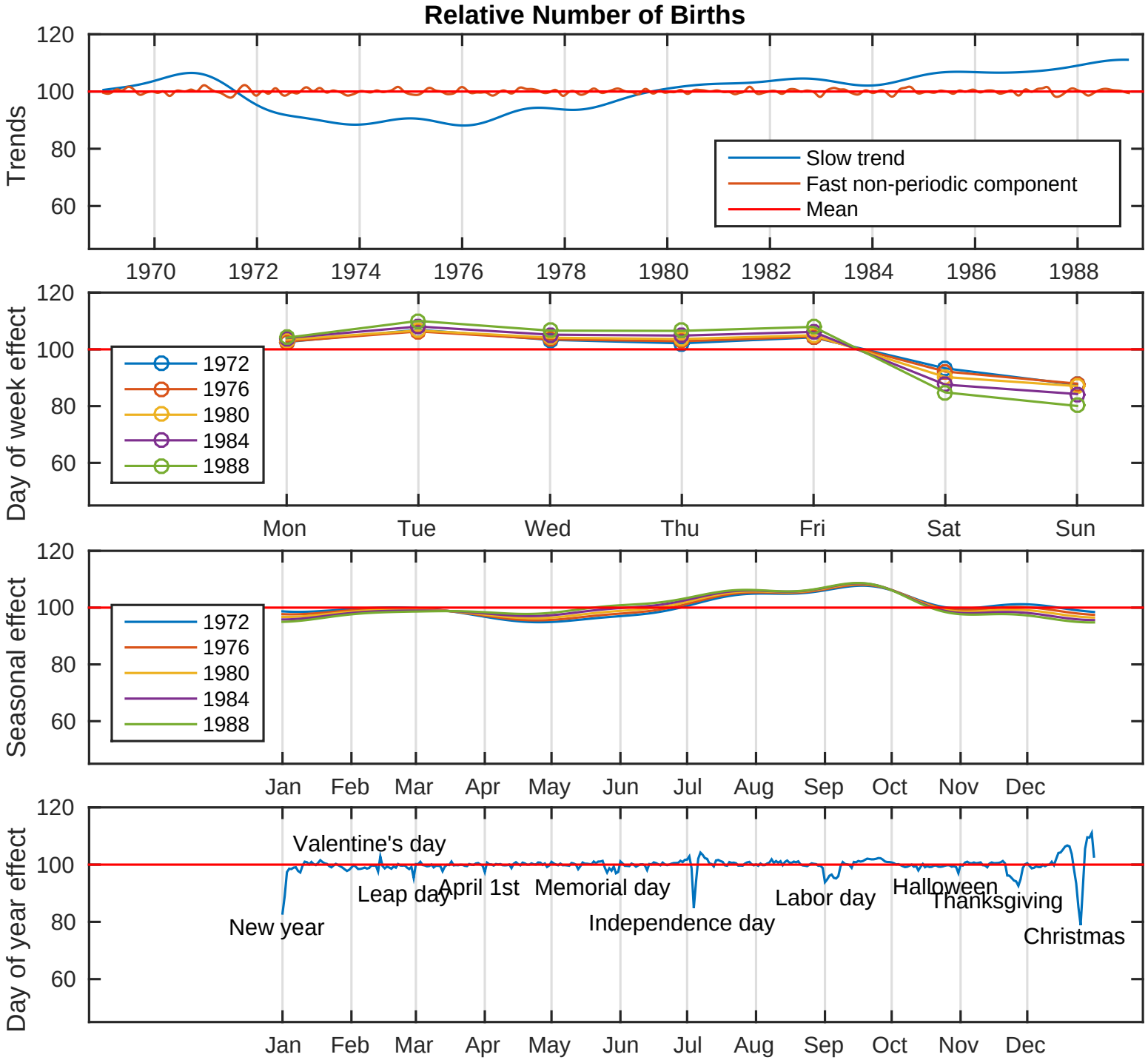
The published graphs show data from 30 days in the year

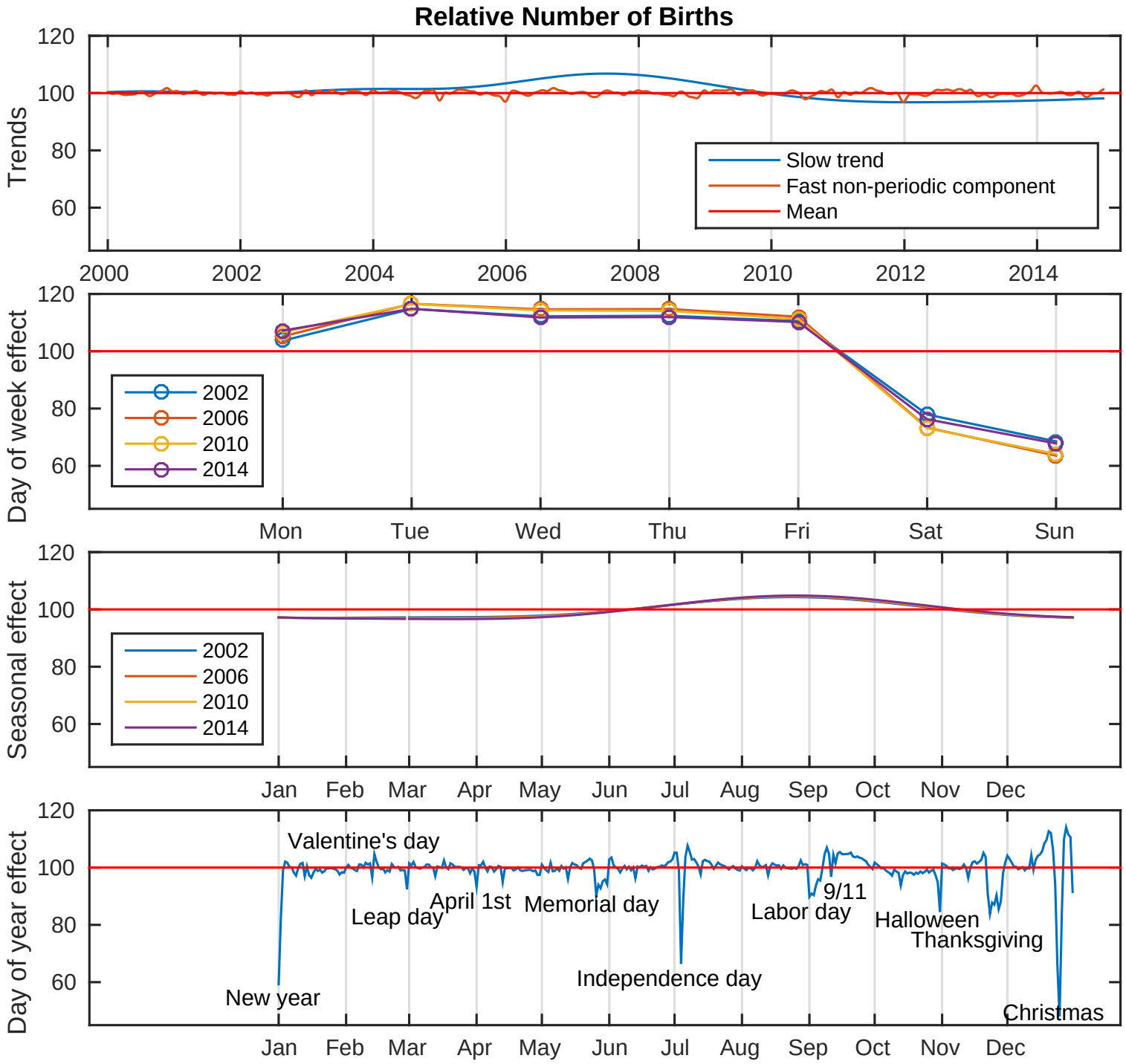


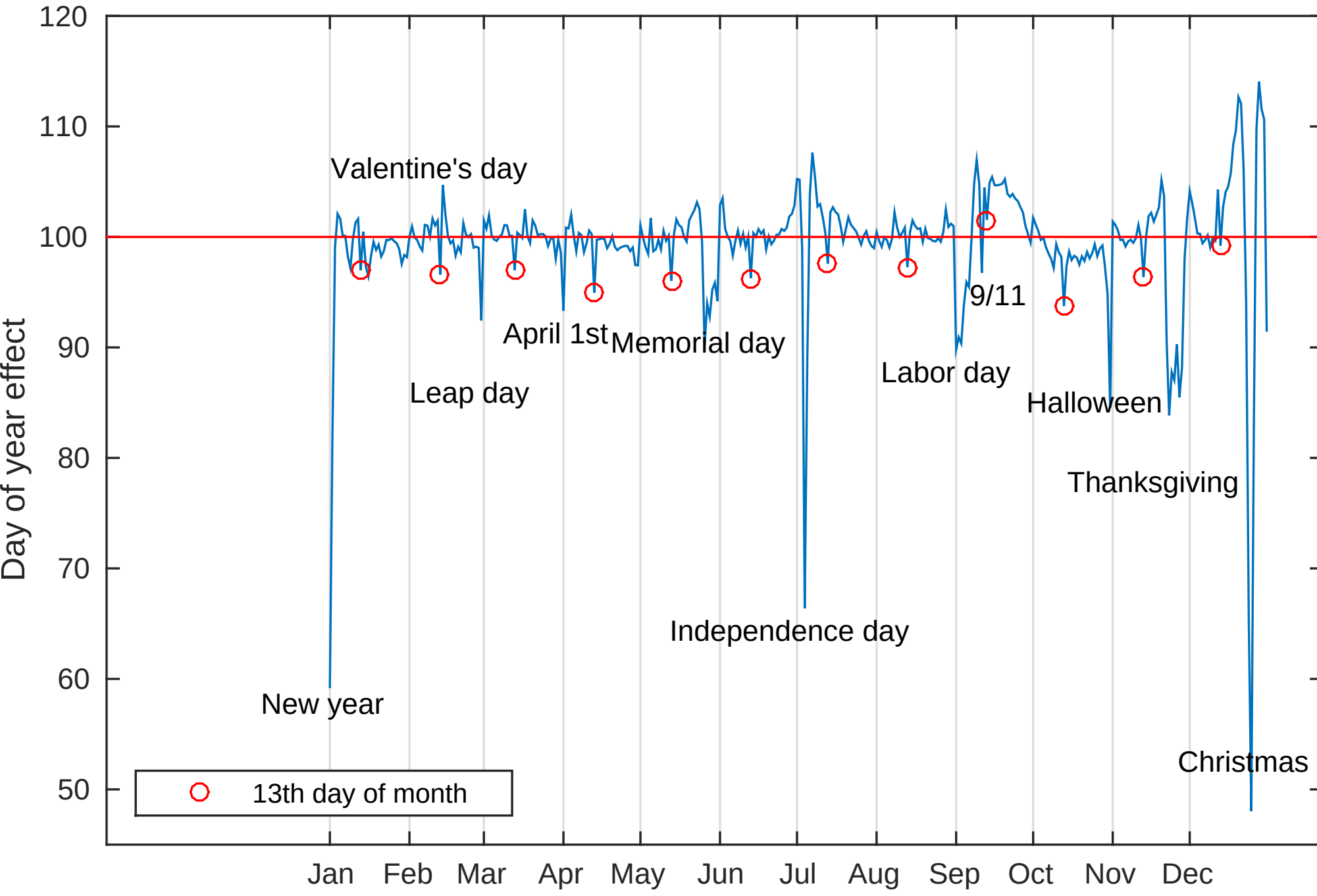



Births by Day of Year
Births
Smoothed
Mean
210000
200000
190000
180000
170000
160000
Births
12 / 30
2 / 14
2 / 29*4
10 / 31
7 / 4
12 / 24
1 / 1
12 / 25
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Day
Source: National Vital Statistics System natality data, as provided by Google BigQuery. Graph by Chris Mulligan (chmullig.com)


Which Birth Dates Are Most Common?
DAY JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
BIRTHDAY RANK
Less common
More common


Relative Number of Births
Trends
120
100
80
60
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988
Slow trend
Fast non-periodic component
Mean
Day of week effect
Mon
Tue
Wed
Thu
Fri
Sat
Sun
1972
1976
1980
1984
1988
Seasonal effect
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Day of year effect
Valentine's day
New year
Leap day
April 1st
Memorial day
Independence day
Labor day
Halloween
Thanksgiving
Christmas


Relative Number of Births
Trends
Slow trend
Fast non-periodic component
Mean
Day of week effect
2002
2006
2010
2014
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Seasonal effect
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Day of year effect
Valentine's day
Leap day
April 1st
Memorial day
Independence day
Labor day
9/11
Halloween
Thanksgiving
New year
Christmas


Day of year effect
120
110
100
90
80
70
60
50
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Valentine's day
Leap day
April 1st
Memorial day
Independence day
Labor day
9/11
Halloween
Thanksgiving
Christmas
New year
13th day of month

The blessing of dimensionality

- We learned by looking at 366 questions at once!
- Consider the alternative . . .

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

P
X
©
R
sa
D
p

Kristina M. Durante[1], Ashley Rae[1], and Vladas Griskevicius[2]

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

Abstract

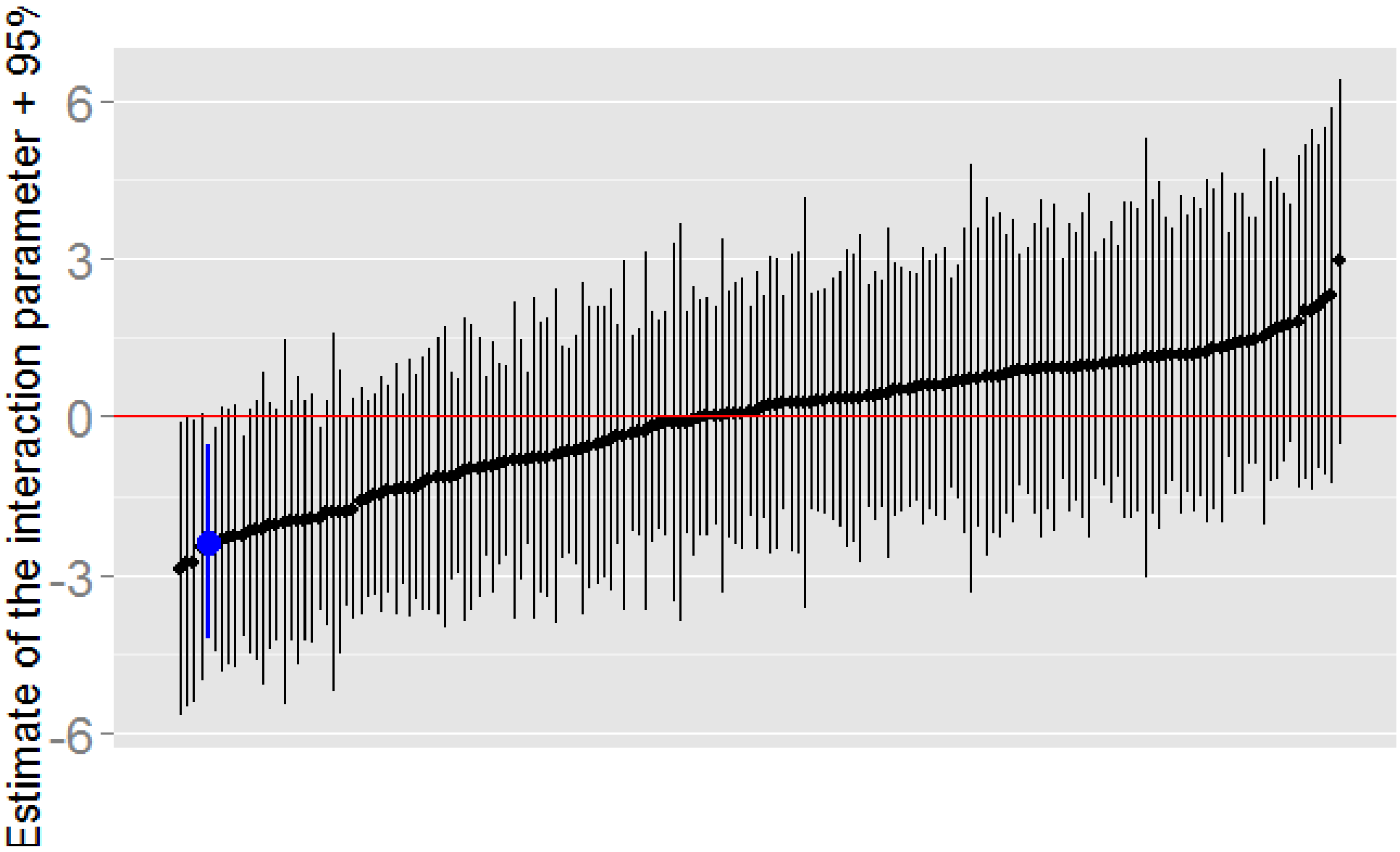
Each month, many women experience an ovulatory cycle that regulates fertility. Although re
cycle influences women's mating preferences, we proposed that it might also change wome
views. Building on theory suggesting that political and religious orientation are linked to repr
how fertility influenced women's politics, religiosity, and voting in the 2012 U.S. presidentia
with large and diverse samples, ovulation had drastically different effects on single women a
relationships. Ovulation led single women to become more liberal, less religious, and more
Obama. In contrast, ovulation led women in committed relationships to become more con
and more likely to vote for Mitt Romney. In addition, ovulation-induced changes in polit
women's voting behavior. Overall, the ovulatory cycle not only influences women's politics
differently for single women than for women in relationships.

Choices!

1. Exclusion criteria based on cycle length (3 options)
2. Exclusion criteria based on "How sure are you?" response (2)
3. Cycle day assessment (3)
4. Fertility assessment (4)
5. Relationship status assessment (3)

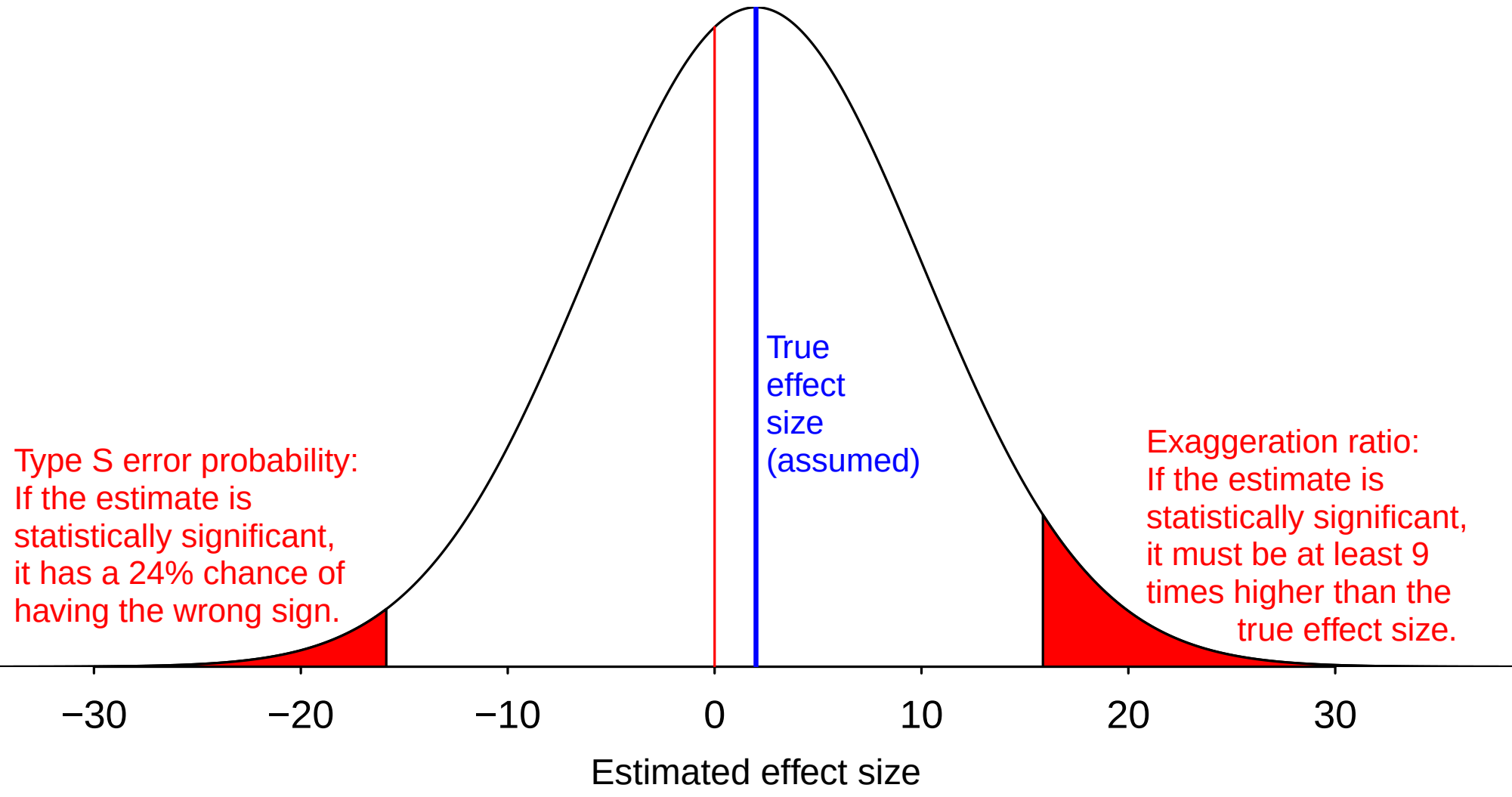
168 possibilities (after excluding some contradictory combinations)

Living in the multiverse




DEAD
ARRIVAL

This is what "power = 0.06" looks like. Get used to it.

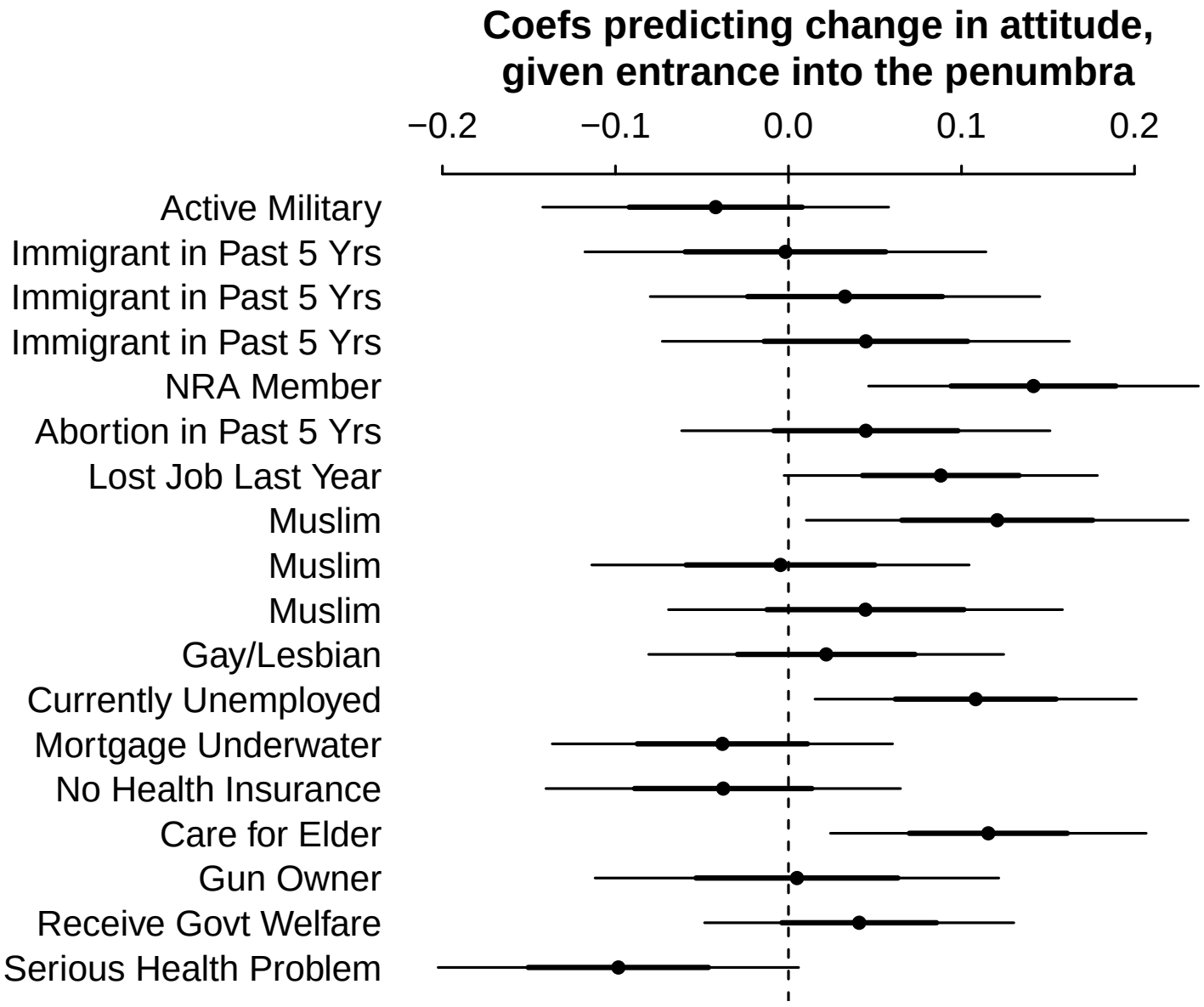


David Weisburd
with Anthony Petrosino and Gail Mason

Design Sensitivity in Criminal Justice Experiments

It is commonly assumed that increasing the size of a sample provides the most straightforward method for increasing the statistical power of a research design and thus avoiding the possibility that an investigation is biased toward a finding of no difference or no effect (e.g., see

sanctions. Contrary to conventional wisdom advocating large sample designs, little relationship is found in practice between sample size and statistical power. Difficulty in maintaining the integrity of treatments and the homogeneity of samples or treatments employed offsets the design advantages of larger investigations.



- Small but nonzero effects
- No effects with cross-predictions, reverse-time predictions

Policy!

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.

My new favorite example

My new favorite example

- Model:

My new favorite example

- Model:
  - $y|\theta \sim \mathrm{N}(\theta, 1)$

My new favorite example

- Model:
  - $y|\theta \sim \mathrm{N}(\theta, 1)$
  - $p(\theta) \propto 1$

My new favorite example

- Model:
  - $y|\theta \sim \mathrm{N}(\theta, 1)$
  - $p(\theta) \propto 1$
- Data:

My new favorite example

- Model:
  - $y|\theta \sim \mathrm{N}(\theta, 1)$
  - $p(\theta) \propto 1$
- Data:
  - $y = 1$

My new favorite example

- Model:
  - $y|\theta \sim \mathrm{N}(\theta, 1)$
  - $p(\theta) \propto 1$
- Data:
  - $y = 1$
- Inference:

My new favorite example

- Model:
  - $y|\theta \sim \mathrm{N}(\theta, 1)$
  - $p(\theta) \propto 1$
- Data:
  - $y = 1$
- Inference:
  - $\theta|y \sim \mathrm{N}(y, 1)$

My new favorite example

- Model:
  - $y|\theta \sim \mathrm{N}(\theta, 1)$
  - $p(\theta) \propto 1$
- Data:
  - $y = 1$
- Inference:
  - $\theta|y \sim \mathrm{N}(y, 1)$
  - $\Pr(\theta > 0|y) = .84$

My new favorite example

- Model:
  - $y|\theta \sim \mathrm{N}(\theta, 1)$
  - $p(\theta) \propto 1$
- Data:
  - $y = 1$
- Inference:
  - $\theta|y \sim \mathrm{N}(y, 1)$
  - $\Pr(\theta > 0|y) = .84$
- Wanna bet??

Where to go next?

- Scale-free modeling
- Weakly informative priors
- Prior information wipes out the multiple comparisons problem
- Computational stability and inferential stability; the folk theorem of statistical computing
- Implications for "big data"