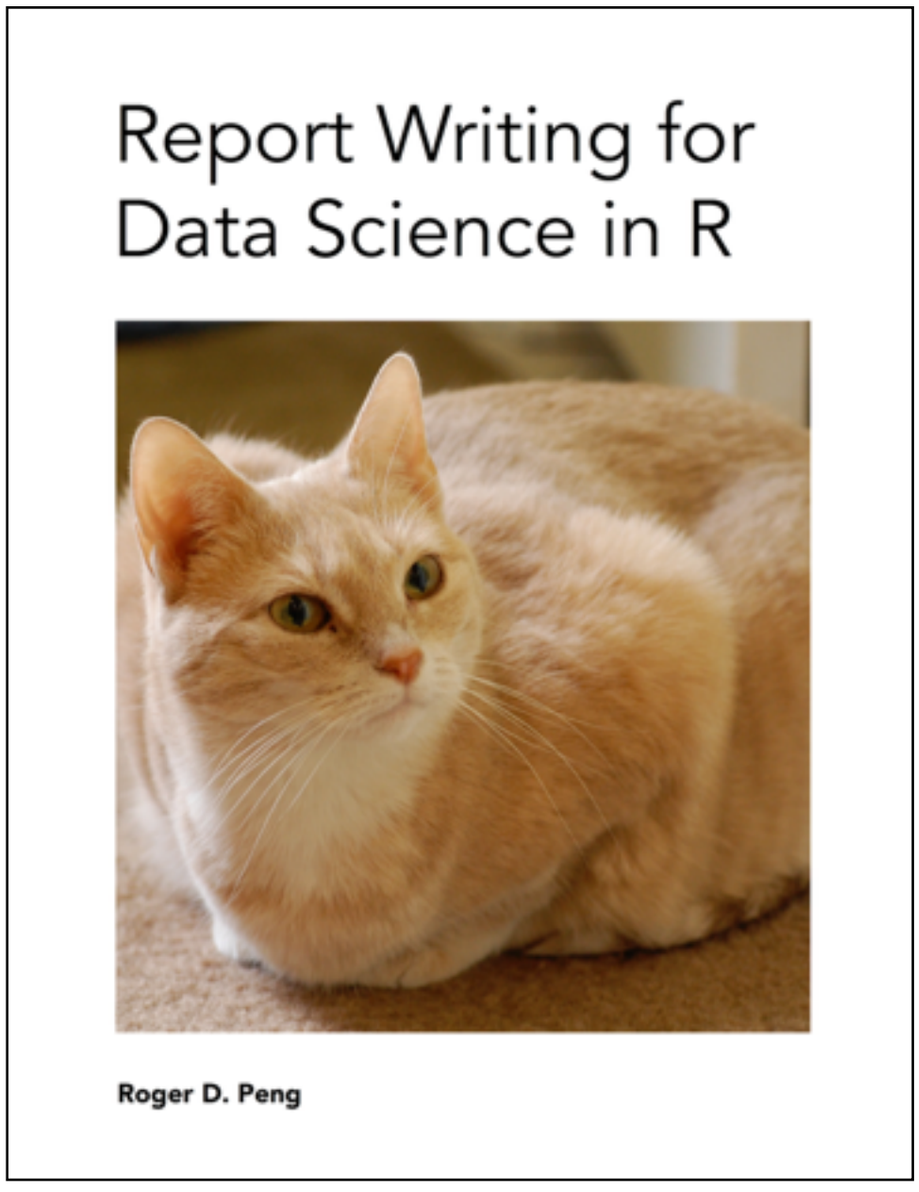
JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Reproducible Research: Medication or Prevention?

Roger D. Peng
@rdpeng, @simplystats, simplystatistics.org

UW SISBID
July 2015

About Me



leanpub.com/reportwriting

simplystats

simplystatistics.org

@simplystats

Johns Hopkins Data Science Specialization

The Johns Hopkins Data Science Specialization

Powered by Tumblr
Colorfall Theme by Paul Mackenzie

1. The Data Scientist's Toolbox - Get yourself set up.
2. R programming - Learn to code.
3. Getting and Cleaning Data - You need data. Get some.
4. Exploratory Data Analysis - What's that in my data?
5. Reproducible Research - Did you do what you think you did?
6. Statistical Inference - You don't have infinite money. Try sampling.
7. Regression Models - The duct tape of data science.
8. Practical Machine Learning - Predict the future with data. Easy.
9. Developing Data Products - There better be an app for that data.

http://jhudatascience.org

Johns Hopkins Data Science Specialization

- Data science toolbox
- Probability / math stat, statistical inference
- Getting + cleaning data
- R programming
- Regression modeling / machine learning
- **Reproducible research tools**
- Exploratory data analysis
- Data products
- Capstone project with industry partners

Incorporating Statistical Expertise Into Software (1991)

"Throughout American or even global industry, there is much advocacy of statistical process control and of understanding processes. **Statisticians have a process they espouse but do not know anything about**. It is the process of putting together many tiny pieces, the process called data analysis, and is not really understood."

Daryl Pregibon, NRC Report 1991

http://goo.gl/vsBUKn

http://youtu.be/8h96LgVpUrI

Parable #1

ARTICLES

nature medicine

Genomic signatures to guide the use of chemotherapeutics

Anil Potti[1,2], Holly K Dressman[1,3], Andrea Bild[1,3], Richard F Riedel[1,2], Gina Chan[4], Robyn Sayer[4], Janiel Cragun[4], Hope Cottrill[4], Michael J Kelley[2], Rebecca Petersen[5], David Harpole[5], Jeffrey Marks[5], Andrew Berchuck[1,6], Geoffrey S Ginsburg[1,2], Phillip Febbo[1–3], Johnathan Lancaster[4] & Joseph R Nevins[1–3]

What Happened?

Clinical medicine (tightly controlled)

+

Genomics (wild west)

=

Wild West

Parable #2

Sciencexpress Report

Genetic Signatures of Exceptional Longevity in Humans

Paola Sebastiani,[1]* Nadia Solovieff,[1] Annibale Puca,[2] Stephen W. Hartley,[1] Efthymia Melista,[3] Stacy Andersen,[4] Daniel A. Dworkis,[3] Jemma B. Wilk,[5] Richard H. Myers,[5] Martin H. Steinberg,[6] Monty Montano,[3] Clinton T. Baldwin,[6,7] Thomas T. Perls[4]*

[1]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA. [2]IRCCS Multimedica, Milano, Italy; Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, Segrate, 20122, Italy. [3]Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA. [4]Section of Geriatrics, Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA 02118, USA. [5]Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA. [6]Departments of Medicine and Pediatrics, Boston University School of Medicine and Boston Medical Center, Boston, MA 02118, USA. [7]Center for Human Genetics, Boston University School of Medicine, Boston, MA 02118, USA.

Parable #2

RETRACTION

Post date 22 July 2011

After online publication of our Report "Genetic signatures of exceptional longevity in humans" (*1*), we discovered that technical errors in the Illumina 610 array and an inadequate quality control protocol introduced false-positive single-nucleotide polymorphisms (SNPs) in our findings. An independent laboratory subsequently performed stringent quality control measures, ambiguous SNPs were then removed, and resultant genotype data were validated using an independent platform. We then reanalyzed the reduced data set using the same methodology as in the published paper. We feel the main scientific findings remain supported by the available data: (i) A model consisting of multiple specific SNPs accurately differentiates between centenarians and controls; (ii) genetic profiles cluster into specific signatures; and (iii) signatures are associated with ages of onset of specific age-related diseases and subjects with the oldest ages. However, the specific details of the new analysis change substantially from those originally published online to the point of becoming a new report. Therefore, we retract the original manuscript and will pursue alternative publication of the new findings.

PAOLA SEBASTIANI,[1]* NADIA SOLOVIEFF,[1] ANNIBALE PUCA,[2] STEPHEN W. HARTLEY,[1] EFTHYMIA MELISTA,[3] STACY ANDERSEN,[4] DANIEL A. DWORKIS,[3] JEMMA B. WILK,[5] RICHARD H. MYERS,[5] MARTIN H. STEINBERG,[6] MONTY MONTANO,[3] CLINTON T. BALDWIN,[6,7] THOMAS T. PERLS[4]*

Parable #2

Serious flaws revealed in "longevity genes" study

BY DANIEL MACARTHUR 07.07.10 | 8:30 PM | PERMALINK

When an article was published in *Science* last week reporting that DNA samples from exceptionally long-lived individuals differed detectably from those of normal individuals, it got plenty of positive attention from the mainstream media. However, the buzz from experts was rapid and telling: my colleagues in the statistical genetics community weren't excited about the results, but immediately, profoundly skeptical.

http://goo.gl/5Ti4YA

Parable #2

Reproducibility!

What needs to happen next? For a start, **the authors should release the raw intensity data for their genotyping experiments**, which would allow independent investigators to spot obvious problems. Doing so immediately on a public database would go a long way towards showing they're not trying to cover up any methodological flaws. Ideally, they should also validate their putative associated SNPs using an independent platform and release those raw data as well

More broadly, **this is an important lesson for the increasing number of investigators wandering into the GWAS arena**: they need to be aware that the genotype data they're working with aren't just clean, digital data points, but best-guess estimates (typically very reliable, but sometimes badly flawed) based on an noisy fluorescent intensity signal. There's a reason why researchers working on GWAS spend so much of their time on a regimented series of upstream "data cleaning" steps and careful downstream validation of new associations – it's all too easy for noisy data to introduce bias that produces a false association signal. So, kids, don't end up in Newsweek for all the wrong reasons: talk to someone who really knows what they're doing when it comes to GWAS data.

What Happened?

What's Next for Reproducibility?

- Reproducibility is critical for communicating a data analysis

- One cannot sufficiently describe an analysis in journal pages or supplementary materials

- General consensus about its importance

- No credible plan (yet) for how to implement such a requirement (hint: this is what's next)

Reproducible Research at *Biostatistics*

Biostatistics (2012), **13**, 1, *pp*. 166–178
doi:10.1093/biostatistics/kxr013
Advance Access publication on June 17, 2011

R

Significance analysis and statistical dissection of variably methylated regions

ANDREW E. JAFFE

Departments of Epidemiology and Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

ANDREW P. FEINBERG

Center for Epigenetics, Johns Hopkins University, Baltimore, MD 21205, USA

RAFAEL A. IRIZARRY, JEFFREY T. LEEK*

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
jleek@jhsph.edu

Biostatistics (2009), **10**, 3, *pp*. 409–423
doi:10.1093/biostatistics/kxp010
Advance Access publication on April 17, 2009

R

Air pollution and health in Scotland: a multicity study

DUNCAN LEE*, CLAIRE FERGUSON

Department of Statistics, University of Glasgow, Glasgow, G12 8QQ UK
duncan@stats.gla.ac.uk

RICHARD MITCHELL

Public Health and Health Policy, University of Glasgow, Glasgow, G12 8QQ UK

Biostatistics (2009), **10**, 4, *pp*. 756–772
doi:10.1093/biostatistics/kxp029
Advance Access publication on July 27, 2009

C

Second-order estimating equations for the analysis of clustered current status data

RICHARD J. COOK*, DAVID TOLUSSO

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1
rjcook@uwaterloo.ca

What Problem Does Reproducibility Solve?

- What we get
  - Transparency / Improved information transfer
  - Data availability
  - Software / Methods
- What we do NOT get
  - Validity / Correctness of the analysis

What Problem Does Reproducibility Solve?

- An analysis can be reproducible and still be wrong
- We want to know “can we trust this analysis?”
- Does requiring reproducibility deter bad analysis?
- While reproducibility is a key part of science, what people really want is *correct* or *well-done* research

Limitations of Reproducibility

- The premise of reproducible research is that with data/code available, people can check each other and the whole system is self-correcting

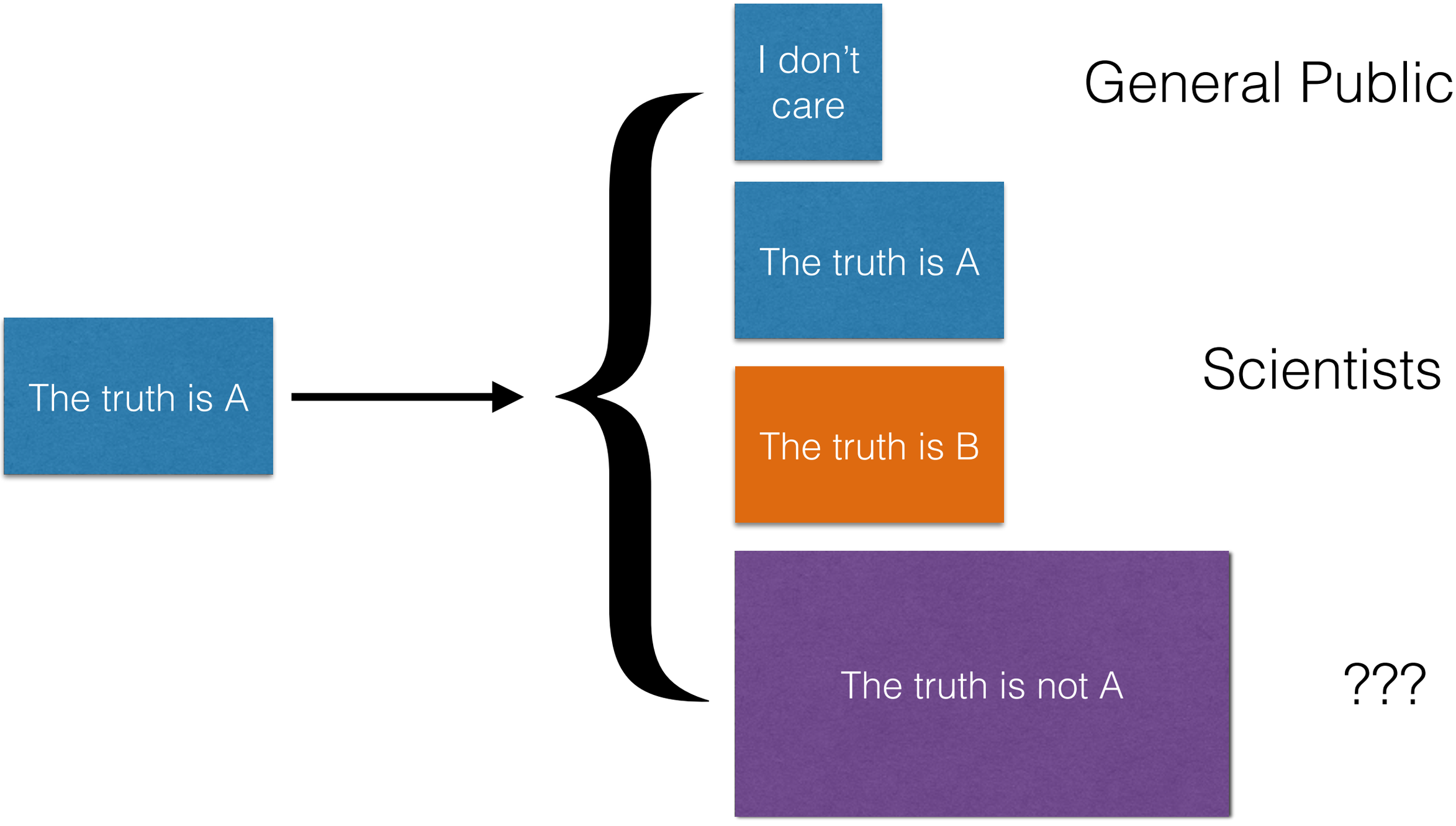
- Addresses the most “downstream” aspect of the research process – post-publication

- Assumes everyone is capable of doing analysis and wants to achieve the same goals (i.e. scientific discovery, search for truth)

Who Reproduces Research?

- Someone needs to *do* something
  - Re-run the analysis; check results match
  - Check the code for bugs/errors
  - Try alternate approaches; check sensitivity
- The need for someone to do something is inherited from traditional notion of replication
- Who is “someone” and what are their goals?

Who Reproduces Research?



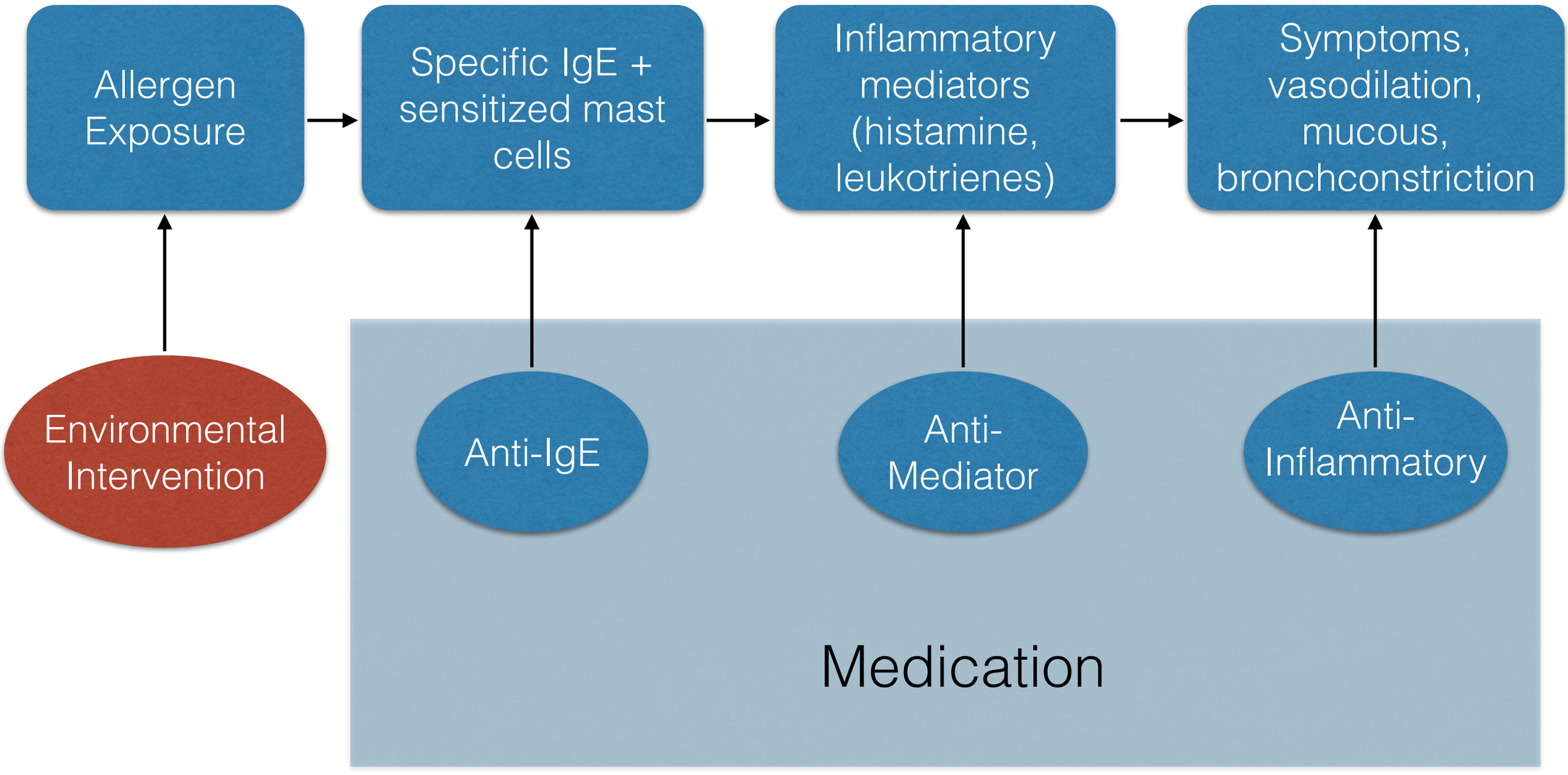
Primary Prevention

- Once bad research is published, it can be a long road to rectify (Duke episode is a prime example)

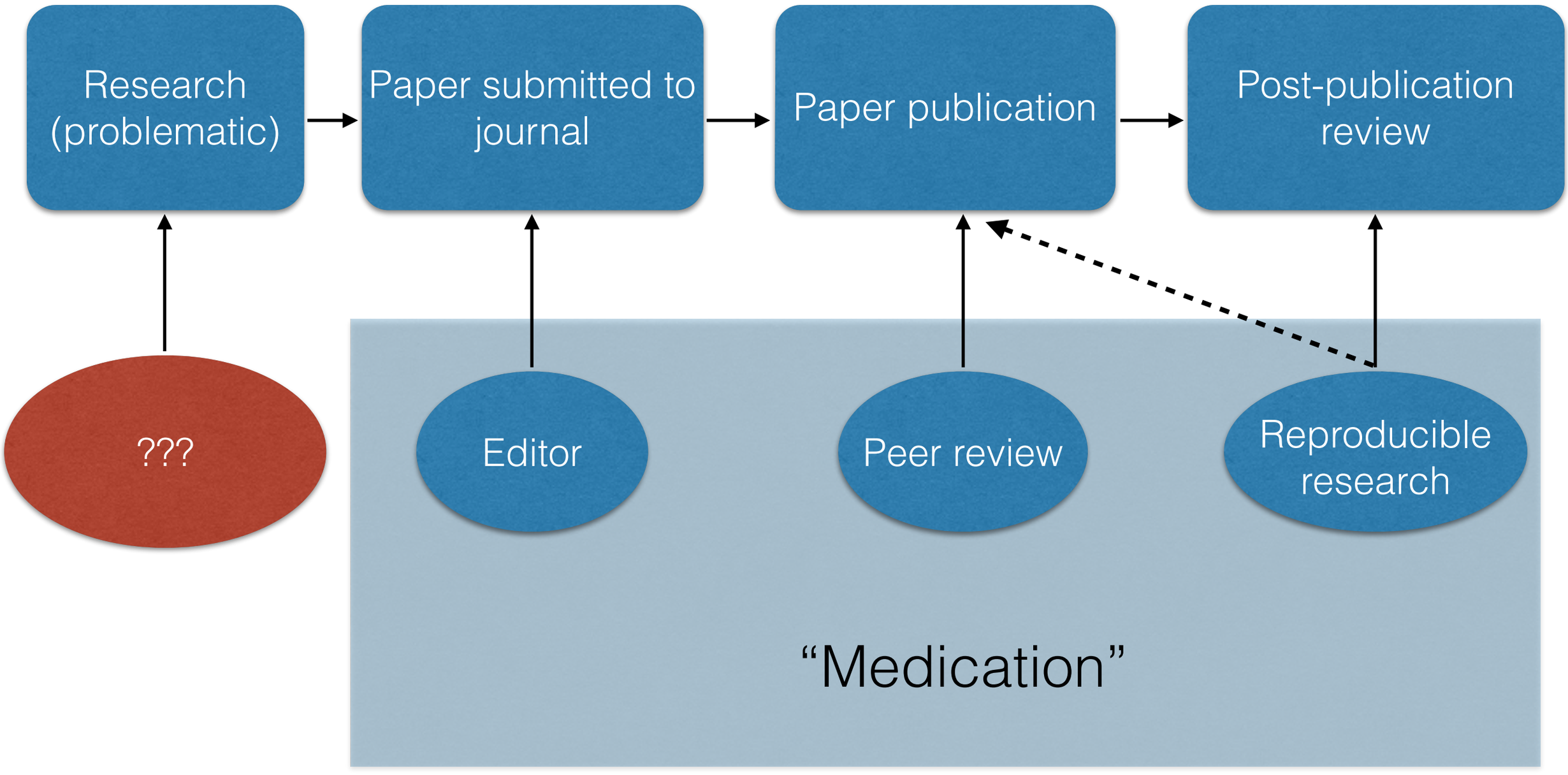
- Even reproducible research can be difficult to untangle unless examined by knowledgeable people (e.g. Baggerly and Coombes)

- How do we prevent shoddy / fraudulent research from appearing *in the first place*?

An Analogy from Asthma



Scientific Dissemination



“Evidence-based Data Analysis”

- Most data analyses involve stringing together many different tools and methods (“data science”)

- Some methods are standard for a given field, others are often applied ad hoc

- We should apply thoroughly studied, mutually agreed upon methods to analyze data whenever possible

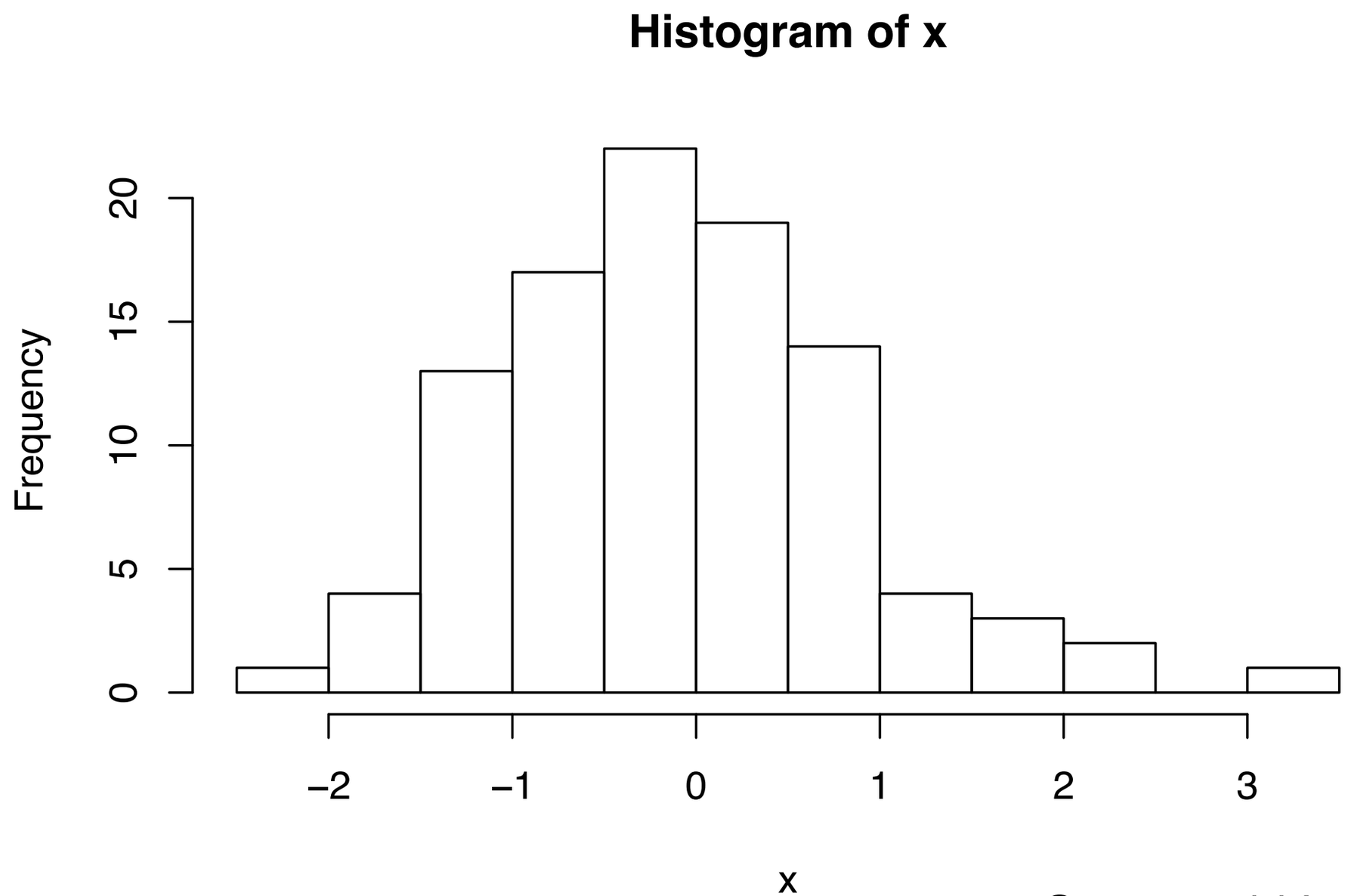
- There should be evidence to justify the use of a given analysis method

- Many methods are used “off-label”; often okay, but….

“Evidence-based Data Analysis”

- Create analytic pipelines from evidence-based components
- Create a standard by which we can judge deviations
- *Deterministic Statistical Machine* (http://goo.gl/Qvlhuv)
- Don’t mess with evidence-based pipeline (“transparent box” analysis)
- Reduce the “researcher degrees of freedom”
- Analogous to a pre-specified clinical trial protocol

Evidence-based Histogram




```
hist(x)
```

Sturges HA (1926), *JASA*
Scott DW (1979), *Biometrika*

Evidence-based Regression

- How do you solve the matrix equation $X^TX\beta = X^Ty$?
- Textbooks tell us $\beta = (X^TX)^{-1}X^Ty$ but we *never actually compute the solution this way*
- Matrix inversion is unstable; we know that because numerical analysts always yell at us
- R uses the QR decomposition; other decompositions like LU can be used too

Case Study: Estimating Acute Effects of Ambient Air Pollution Exposure

- Acute/short-term effects typically estimated via panel studies or time series studies

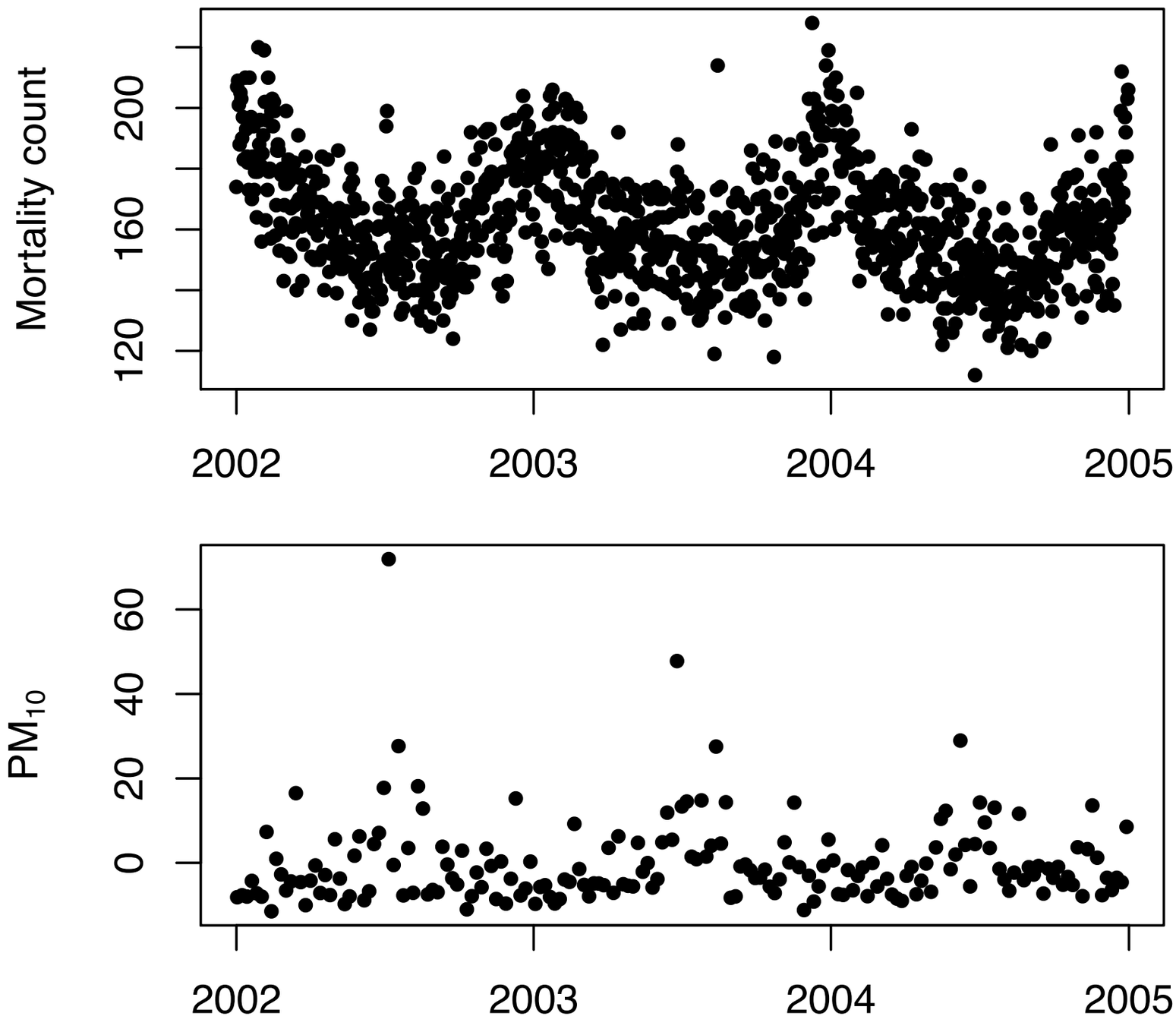
- Work originated in late 1970s early 1980s

- Key question: “Are short-term changes in pollution associated with short-term changes in a population health outcome?”

- Studies usually conducted at community level

- Long history of statistical research investigating proper methods of analysis

New York Data



Case Study: Estimating Acute Effects of Ambient Air Pollution Exposure

- Can we encode everything that we have found in statistical/epidemiological research into a single package?

- Time series studies do not have a huge range of variation; typically involves similar types of data and similar questions

- We can create a deterministic statistical machine for this area?

DSM Modules for Time Series Studies of Air Pollution and Health

1. Check for outliers, high leverage, overdispersion
2. Fill in missing data? NO!
3. Model selection: Estimate degrees of freedom to adjust for unmeasured confounders
   - Other aspects of model not as critical
4. Multiple lag analysis
5. Sensitivity analysis wrt
   - Unmeasured confounder adjustment
   - Influential points

Dominici, McDermott, Hastie (2004) *JASA*; Peng, Dominici, Louis (2006) *JRSS-A*

Where’s the Curation?

MUSIC

tracks

PANDORA
internet radio

Spotify™

rdio™

Curation for Data Analysis

- Provide packages that encode data analysis pipelines for given problems, technologies, questions

- Curated by experts knowledgeable in the various areas (i.e. statisticians)

- Documentation given supporting each module in the pipeline

- Changes introduced after passing relevant benchmarks and unit tests

- We already have the tools and much of the knowledge

Cochrane Summaries

COCHRANE SUMMARIES
Independent high-quality evidence for health care decision making

English | 简体中文 | Hrvatski | Français | Português | Español

to Cochrane.org
to The Cochrane Library

Enter words like "aspirin for headaches" or "vaccines for influenza"

Search

Browse health topics | New and updated

A product of The Cochrane Collaboration

How to use this site

If you are looking for information you can trust,
to help you make choices about health care, this site is for you.

---

http://summaries.cochrane.org

Cochrane Summaries

COCHRANE SUMMARIES
Trusted evidence. Informed decisions. Better health.

English | 简体中文 | 繁體中文
Hrvatski | Français | Português | Español

to Cochrane.org
to The Cochrane Library

ibuprofen for headaches

Search

Browse health topics | Retain current filters | Reset

A product of The Cochrane Collaboration

How to use this site

Search results: 6

Current search: ibuprofen for headaches | Dentistry & oral health

Sort by: Best match | Results per page: 10 | Subscribe to this search More... | ShareThis | +1 | Like 0

Filter by

Health topics

- Child health (1)
- Dentistry & oral health
  - Antibiotic therapy (1)
  - Craniofacial anomalies (2)
  - Oral pain (3)
- Rheumatology (1)

Cochrane group topics

- Child Health (1)
- Complementary Medicine (1)
- Nursing Care (1)
- Oral Health (6)
- Pre-hospital and Emergency Care (1)

Ibuprofen versus paracetamol (acetaminophen) for pain relief after surgical removal of lower wisdom teeth *12 Dec 2013*

Review question This review, carried out by the Cochrane Oral Health Group, seeks to compare the effectiveness of two commonly used painkillers, paracetamol and ibuprofen and the combination of both in a single tablet in the relief of pain following surgical removal of lower wisdom teeth. Background Worldwide the number of surgical operations ...

Does giving children painkillers such as paracetamol and ibuprofen before dental treatment help reduce pain after the treatment? *12 Sep 2012*

Dental pain is common after dental procedures and can be linked with increase in fear of dental treatment, avoidance of dental treatment and other associated problems. Reducing pain is important, particularly in children and adolescents. One way of managing this might be to give painkillers before treatment so that they can start to work right away. ...

Antibiotic use for severe toothache (irreversible pulpitis) *18 Dec 2013*

Review question Are oral antibiotics effective and safe for treating pain in irreversible pulpitis (inflammation of the nerve inside the tooth/nerve damage)?

Cochrane Summaries

Ibuprofen versus paracetamol (acetaminophen) for pain relief after surgical removal of lower wisdom teeth

Bailey E, Worthington HV, van Wijk A, Yates JM, Coulthard P, Afzal Z

Published Online: 12 December 2013

Key results

Ibuprofen is more effective than paracetamol at all doses studied in this review. On limited evidence, the combination of ibuprofen and paracetamol appeared to be no more effective than the single drugs when measured two hours after surgery. However, again on limited evidence, it was found to be more effective than the drugs taken singly when measured at six hours after surgery. Participants taking the combined drug also had a smaller chance of requiring rescue medication.

Quality of the evidence

All of the results (outcomes) comparing ibuprofen to paracetamol are of high quality. This means that further research is very unlikely to change our confidence in the estimates of the effect.

http://goo.gl/2Lz3ax

FDA Drug “Curation”

- Before a company can market a drug to the public, it must go through rigorous approval process
- Phase I, II, and III clinical trials
- Goal is to weight health benefits vs. health risks
- Not perfect by any means but a useful model

Evidence-based Analysis for Data Science

Raw Data

Cleaning / Validation

Pre-processing

Exploratory data analysis

Statistical model development

Sensitivity analysis

Finalize and report results

Data Science

We Need to Avoid…

Lots of data

+

Not sure what to do with data

=

Whatever works or is easiest!

Summary

- Reproducibility is important, but likely would not have prevented recent notorious examples

- Reproducibility focuses on the most “downstream” aspect of the research dissemination process

- Evidence-based data analysis would provide standardized, best practices for given scientific areas and questions

- With development of personalized medicine, poor data analysis has the potential to seriously harm people

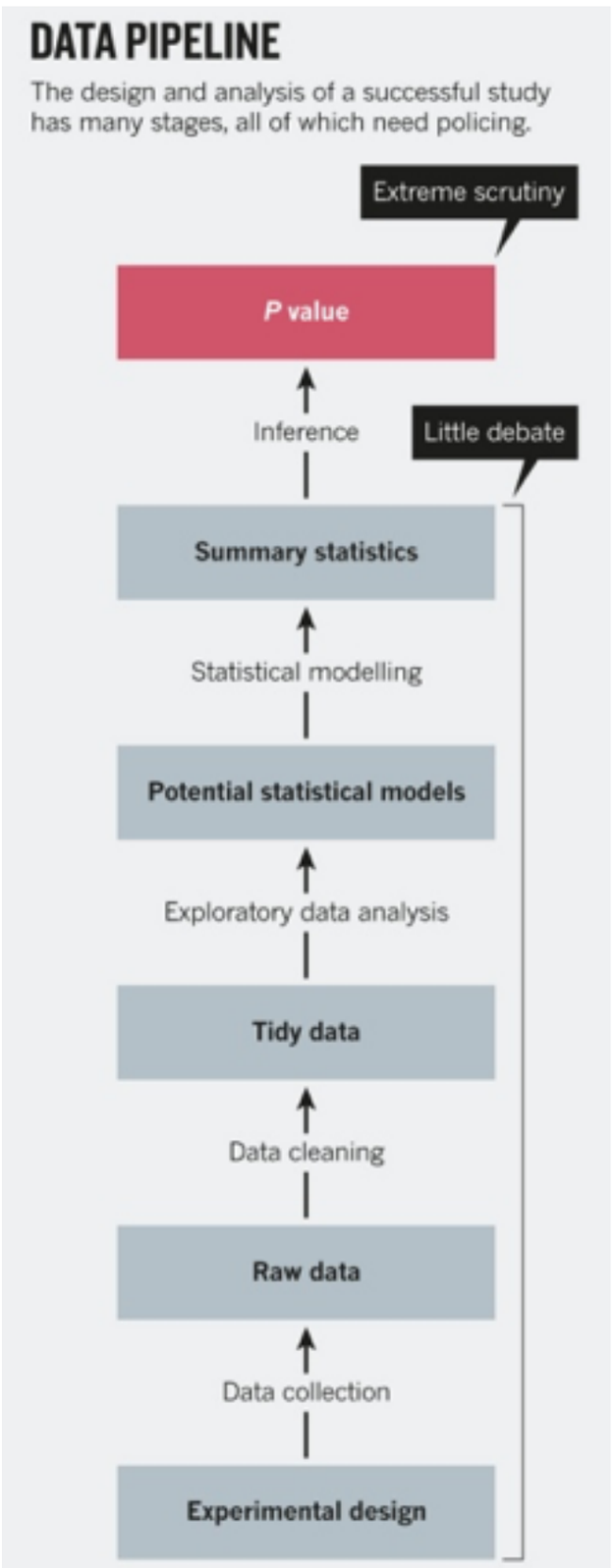
- More effort should be put into improving the quality of “upstream” aspects of scientific research

NATURE | COMMENT

Statistics: *P* values are just the tip of the iceberg

Nature, April 2015

Jeffrey T. Leek & Roger D. Peng



Opinion: Reproducible research can still be wrong: Adopting a prevention approach

PNAS, February 2015

Jeffrey T. Leek[a,1] and Roger D. Peng[b]

PERSPECTIVE

STATISTICS

What is the question?

Science, March 2015

Jeffery T. Leek, Roger D. Peng