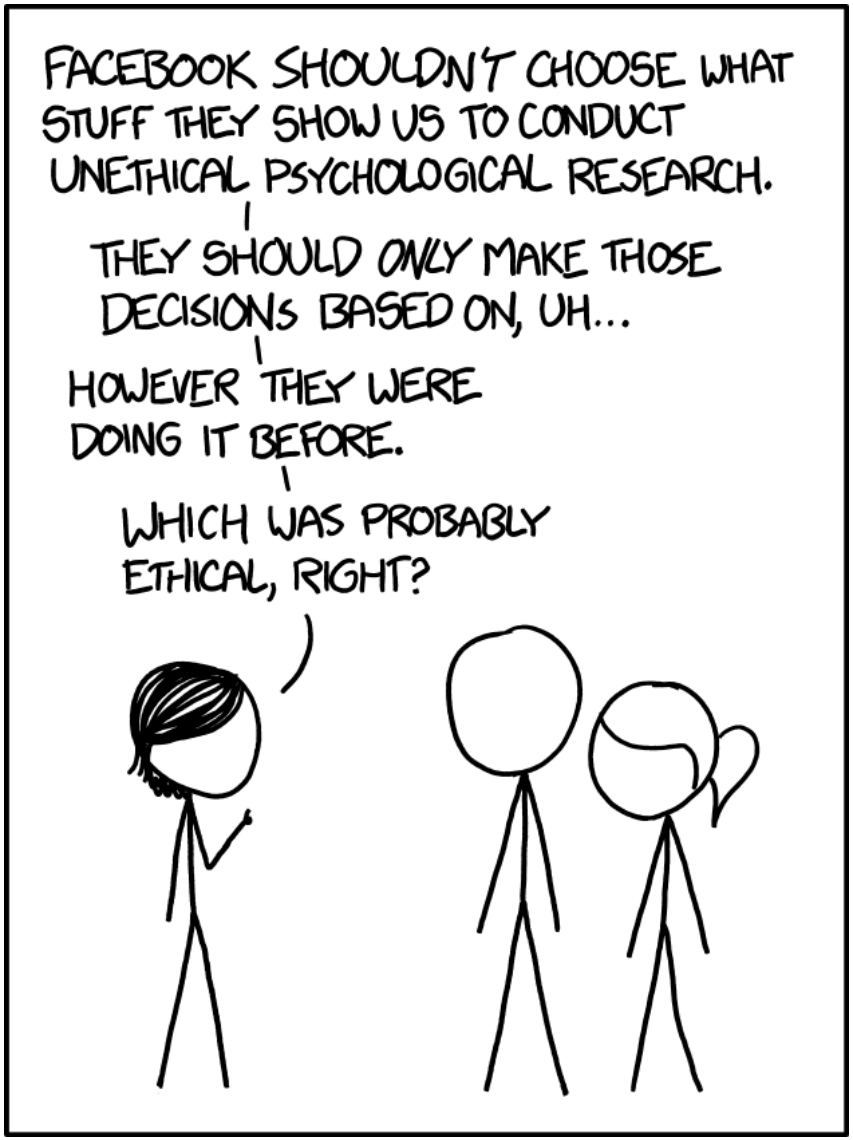
Ethics in Data Science

Author: Nicholas G Reich

Research Ethics[1]



[1] https://xkcd.com/1390/

Professional ethics require vigilance, common sense

Important to have an open, honest orientation

- Work to be conscious of your and others' biases.
- Be willing to stand up for what you believe is right.
- Ask for help when you need it or if you feel unsure.
- Admit mistakes when you make them.
- Be open to learn.
- Create open, transparent, tested workflows.

Responsibility to stakeholders[2]

Know and understand your obligations

- **Are you legally bound to your employer or client?** What terms and conditions apply?
- **Do you have a responsibility to the general public or people whose data you have at your fingertips?** Should you keep datasets on your laptop? etc...
- **Are there legal responsibilities that you need to be aware of?** Sometimes, there are legal terms around data sharing and/or confidentiality.
- **Are there responsibilities, binding or not, that you have to your professional community?** Make your data and code public, to the extent possible.

[2] Adapted from *Modern Data Science with R*

Present results with humility

Overconfidence is a common error when interpreting data/models

- Scrutinize your results from every angle. What would a critic say?
- Use language like "the models says..." to reinforce that outputs come from a model with assumptions.
- Present all evidence honestly, including that which doesn't totally fit the storyline, especially if the story fits your pre-conceived notion. Why are outliers outlying?

Conflicts of interest (COI)

COIs arise frequently and naturally

- You have a responsibility to identify COIs. e.g.

  ... I have been paid for my work by Company X.

  ... I have collaborated with colleague Y on a large group project in the last two years. I do not feel that this precludes me from providing an unbiased review of the manuscript.

- They can become an issue if not identified early.
- Often, different fields will have policies and standards in places to ensure reporting and compliance.

Acknowledge mistakes

From: Nicholas Reich <nick@schoolph.umass.edu>
Subject: correction needed in PLOS NTD article
Date: March 31, 2017 at 10:31 PM
To: corrections@plos.org
Cc: Justin Lessler <justin@jhu.edu>

Dear PLOS Editors -

We write to you in regards to our paper published June 15, 2016 in PLOS NTD entitled "Challenges in Real-Time Prediction of Infectious Disease: A Case Study of Dengue in Thailand":
http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004761

Recently, we discovered an unintentional coding error that led to pervasive minor inaccuracies in the manuscript. Specifically, in processing and merging multiple real-time data files from the Thai Ministry of Public Health, duplicate case records of dengue were created when a single record should have been in place. This led to over-counting of cases in many periods of time, an error that impacted our forecast models and our evaluation of the forecasts themselves.

We have reviewed guidelines for "Corrections, Retractions, Republications and Version Control" on the ICMJE website that was referenced on the PLOS NTD corrections and retractions page:
http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/corrections-and-version-control.html

Based on our thorough re-analysis of the corrected data, these errors do not change the overall interpretation, significance, or conclusion of the manuscript, although minor changes are necessary throughout the article, including in tables and figures.

We are seeking guidance on the appropriate way to prepare these updated results and a revised manuscript so that the editorial team can assess the extent of corrections needed and decide on the best way forward.

Best regards,
Nick Reich and Justin Lessler

--
Nicholas G. Reich, Ph.D.
Assistant Professor
Department of Biostatistics and Epidemiology
School of Public Health and Health Sciences
University of Massachusetts, Amherst

Ethical principles for data science[3]

Summary of key points

- Do your work well by your own standards and by the standards of your profession.
- Recognize the parties to whom you have a special professional obligation.
- Report results and methods honestly Respect your responsibility to identify and report flaws and shortcomings in your work.
- Acknowledge possible conflicts of interest at appropriate times.

[3] Adapted from *Modern Data Science with R*