Health Informatics

Lecture 7

Samantha Kleinberg
samantha.kleinberg@stevens.edu

Midterm

- Next week!
- In class
- Closed book
- On everything covered up to the midterm

Final project

- **Proposals due March 19 at latest**
  - **Max possible grade on project is A- if you do not submit a proposal. I encourage you to submit early!**
- Proposal content:
  - What will you do? How will you do it? If you need data, do you have it?
  - How will you determine if you were successful? What are the key outcomes of your project?
  - If 2 person team, how labor will be divided.
- 1 page. This is a hard limit.
- **Must use MIMIC-III**
  - No exceptions
  - See also https://github.com/YerevaNN/mimic3-benchmarks

Paper/presentation/etc

- Presentations during final class
- Papers
  - 6-8 pages in conference format (e.g. NIPS, AAAI).

Today’s paper

Back to data challenges

- Observational Data in general
  - Nonstationarity
  - Selection bias
  - Choosing variables
  - Seeing vs doing
- Biomedical Data
  - Biased approximation of truth
  - Sample size
  - Censoring (left, right)
  - Institutional differences
  - Fragmentation of data

Variable selection/granularity

- Problem specific
  - What level in ICD9 hierarchy to use?
  - May use different granularity for different codes w/in an experiment
- Redundancy
  - HbA1C and BG provide information at different scales
  - Rx for different antihypertensives doesn't necessarily provide new info
- Hidden common causes
  - What missing variables may be responsible for observed relationships?

Observation vs randomization

- Nurses' Health study followed ~122,000 nurses every 2 years since the 1970's
- Analysis in the 90's showed women taking HRT after menopause have decreased risk of heart attacks (37% lower death rate, 53% lower risk of CV death)
- HERS trial: RCT showing no effect
- WHI: RCT where heart attacks increase 29% (from 30 to 37 per 10,000 person-years)
- Latest: HRT may be beneficial if it's started early

Doing vs seeing: randomization (more in a few weeks)

- Sever link between causes of intervention and effects (selection bias)
  - E.g. birth control pills and pregnancy
- Isolate cause
  - Single difference between groups, removes confounding
- Blinding
  - Confirmation bias

Sample size

- Big data doesn't guarantee sufficient power!

- Would you rather have 10 patients monitored in great detail or a few datapoints on 10K patients?

Sample size needed depends on:

- Margin of error
  - Do you need to estimate prevalence of a disease to within 2% or is 10% ok?
  - Common default: 5%
- Confidence level
  - Do you need to be 95% certain that your estimate will be within the margin of error, or can you accept an 80% confidence level?
  - Common default: 95%
- Effect size
  - Does the drug increase survival by 25%? Or 2%? Bigger effect means smaller sample required

Sample size/power

- If sample size fixed, can calculate statistical power:
  - how likely is it that you'll see a statistically significant result in population size N with given effect size and significance level

Useful links

- http://powerandsamplesize.com

Missing data/error

- Device malfunction
  - E.g. thermometer moves and now measure room temp
- Network problem
  - Data recorded but not stored
- Data exists… somewhere
- Human error – misrecorded data/omitted data
- Patient factors – omitting data/giving false info

Censoring

- Left: what happened before admission to hospital?
- Right: what's outcome after leaving hospital?

Timing variability

- Note describes chronology after the fact
- Labs not exact
- When do people go to dr? (hint: after symptoms start)
- Interventions recorded after they're done

Using Empiric Semantic Correlation to Interpret Temporal Assertions in Clinical Texts

George Hripcsak, MD, MS, Noémie Elhadad, PhD, Yueh-Hsia Chen, MS, Li Zhou, BMed, PhD, Frances P. Morrison, MD, MPH

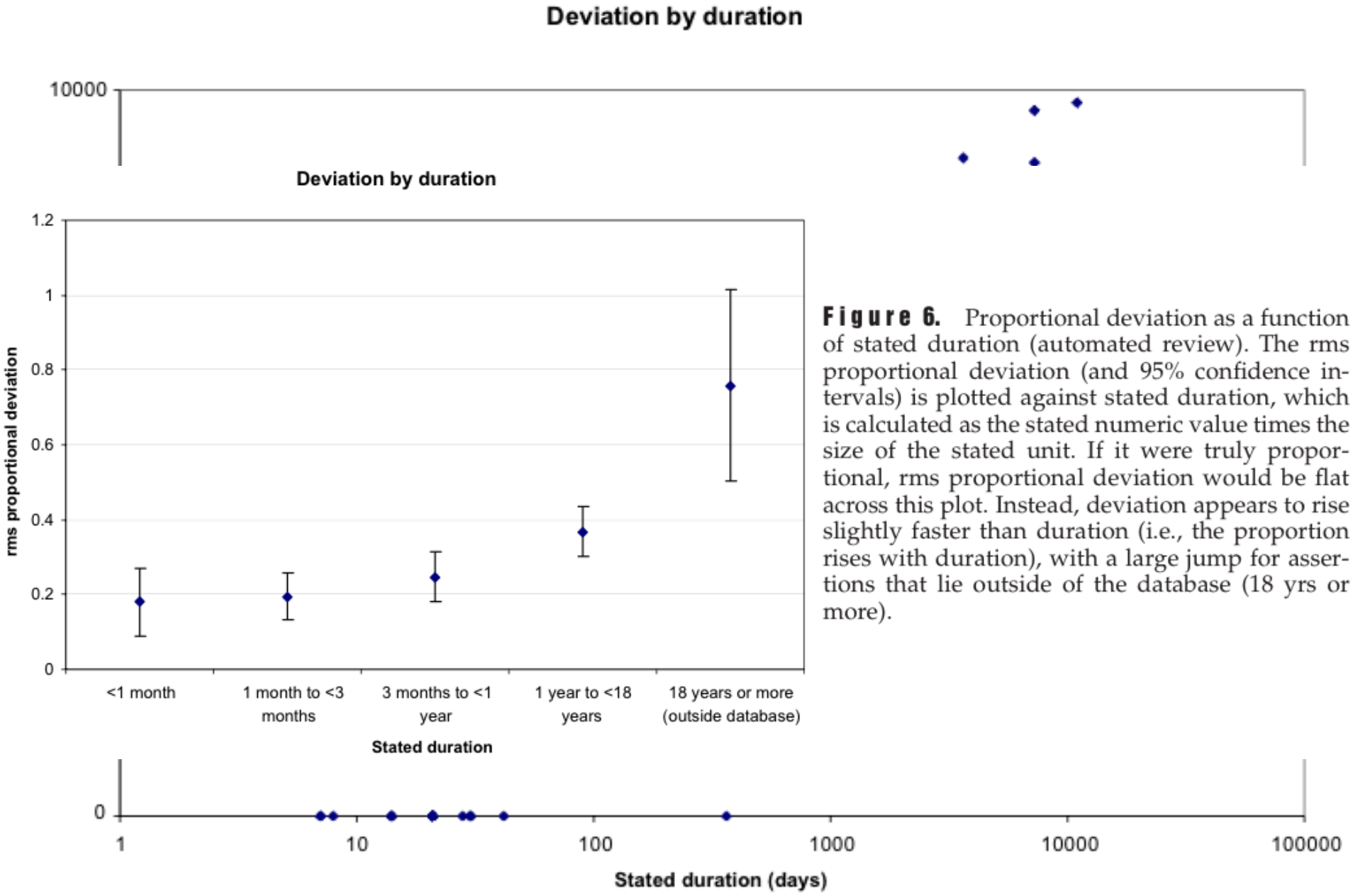


Figure 6. Proportional deviation as a function of stated duration (automated review). The rms proportional deviation (and 95% confidence intervals) is plotted against stated duration, which is calculated as the stated numeric value times the size of the stated unit. If it were truly proportional, rms proportional deviation would be flat across this plot. Instead, deviation appears to rise slightly faster than duration (i.e., the proportion rises with duration), with a large jump for assertions that lie outside of the database (18 yrs or more).

Fragmentation

- Do you go to the same doctors at home and when classes are in session?
- How are records shared?

Controls

- Usually matched to cases, but technical difficulties
  - Is selection criteria structured or unstructured?
  - Is same data available on both?
  - If comparing against “healthy” people, will there be enough data?
- What happens if some cases are actually controls?

Multiple testing

- With many measured variables, we can do many more comparisons in EHR data than in a normal experiment
- Any downsides?

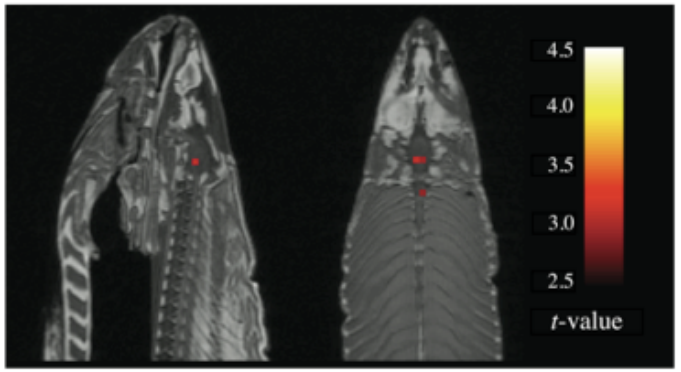
Multiple Testing

METHODS

Subject. One mature Atlantic Salmon (Salmo salar) participated in the fMRI study. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was not alive at the time of scanning.

Task. The task administered to the salmon involved completing an open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

Design. Stimuli were presented in a block design with each photo presented for 10 seconds followed by 12 seconds of rest. A total of 15 photos were displayed. Total scan time was 5.5 minutes.



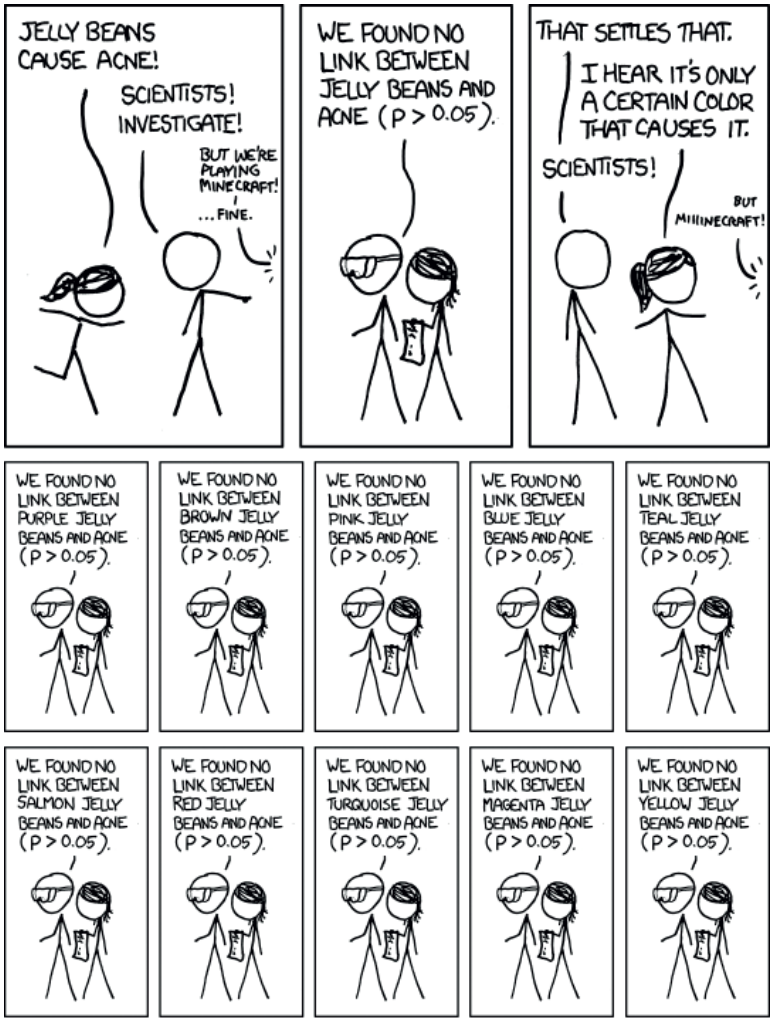
A *t*-contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, p(uncorrected) < 0.001, 3 voxel extent threshold.

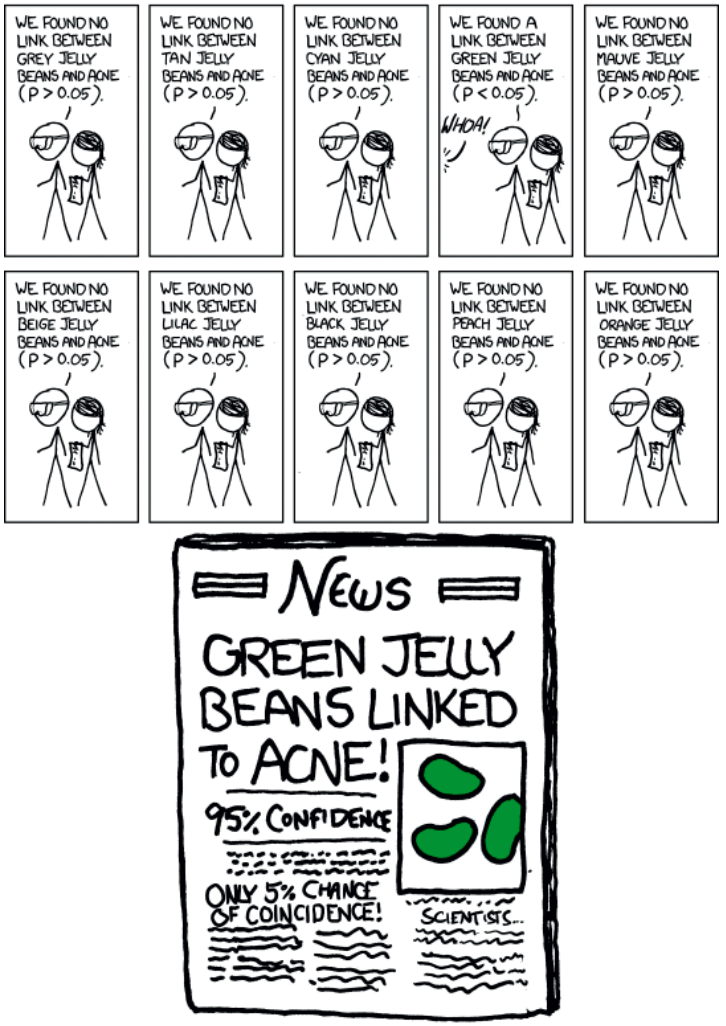
Several active voxels were discovered in a cluster located within the salmon's brain cavity (Figure 1, see above). The size of this cluster was 81 mm^3 with a cluster-level significance of $p = 0.001$. Due to the coarse resolution of the echo-planar image acquisition and the relatively small size of the salmon brain further discrimination between brain regions could not be completed. Out of a search volume of 8064 voxels a total of 16 voxels were significant.

Identical *t*-contrasts controlling the false discovery rate (FDR) and familywise error rate (FWER) were completed. These contrasts indicated no active voxels, even at relaxed statistical thresholds ($p = 0.25$).

Bennett, C. M., Miller, M. B., & Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: An argument for multiple comparisons correction. *NeuroImage*, *47*(1), 125.

Multiple comparisons





http://xkcd.com/882/

Next week

- Midterm!