

The perfect data science experience (versus real life)

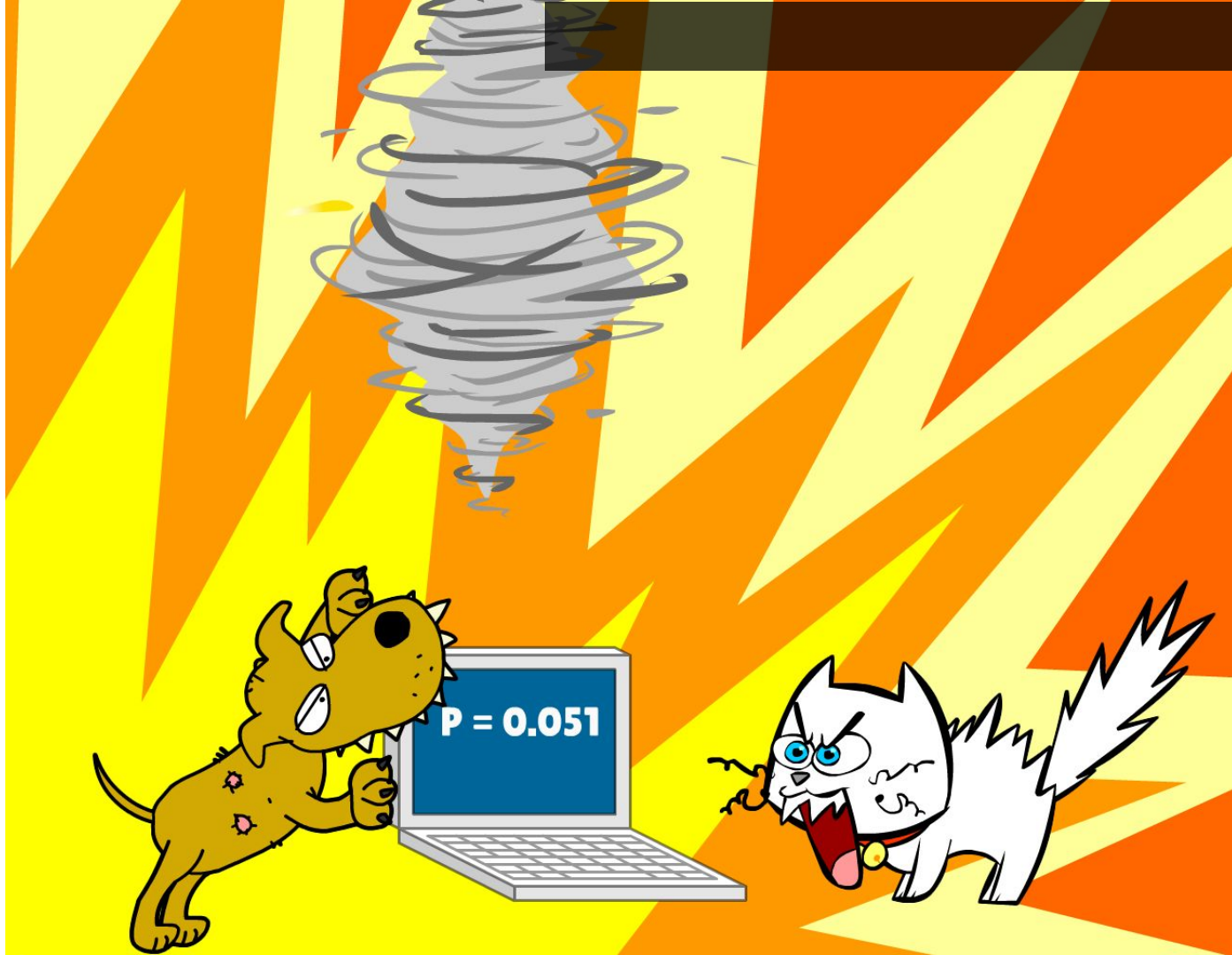
Brian Caffo, Jeff Leek, Roger Peng

@bcaffo

www.bcaffo.com



- Clearly defined hypotheses of interest, specified a priori
- Experimental design available
 - Randomization used across a treatment of interest
 - Stratification on nuisance variables
- Random sample from a population of interest
- Data directly able to interrogate hypotheses
- Dataset creation/merging goes smoothly
- No missing data or dropout
- Analysis is robust without need for advanced modeling
- Conclusions are clear
 - Parsimonious knowledge gained via the experiment
- Decision is obvious given the data

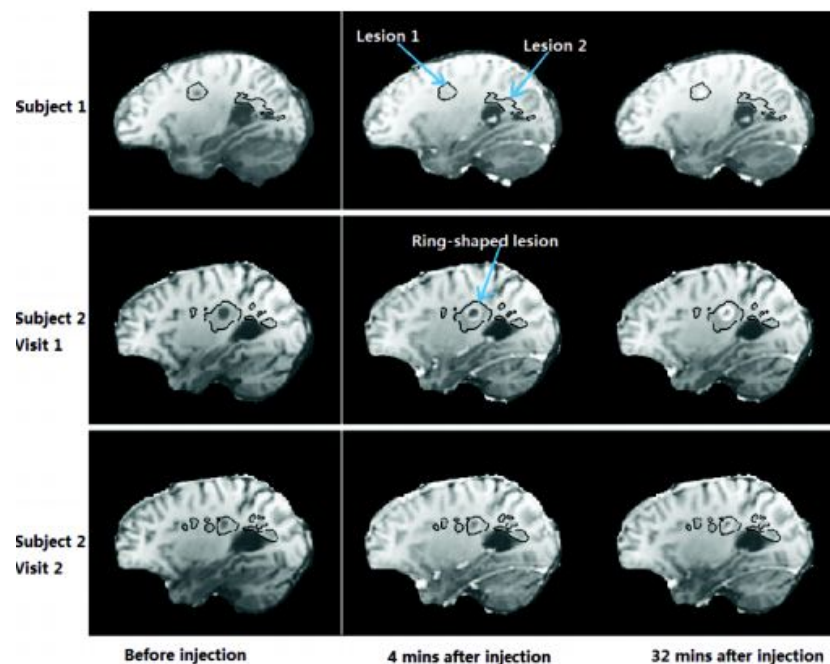


- Data is needed to inform hypotheses and interrogate them
- Multiple comparisons are an issue
- Experimental design options are limited. Or, data is observational
 - Randomization is not available
 - Data is retrospective
- The population being studied isn't the population of interest
- The data don't have the exact measurements that you need to evaluate the hypotheses
- Dataset is problematic
 - Merging is problematic with multiple matches, no matches
 - Data entry errors
- Missing data
- Advanced modeling is required, similarly advanced computing needed to fit the model, issues with robustness and bugs
- Conclusions are indeterminant
- Decision is not substantially further informed by the data

Soft Null Hypotheses: A Case Study of Image Enhancement Detection in Brain Lesions

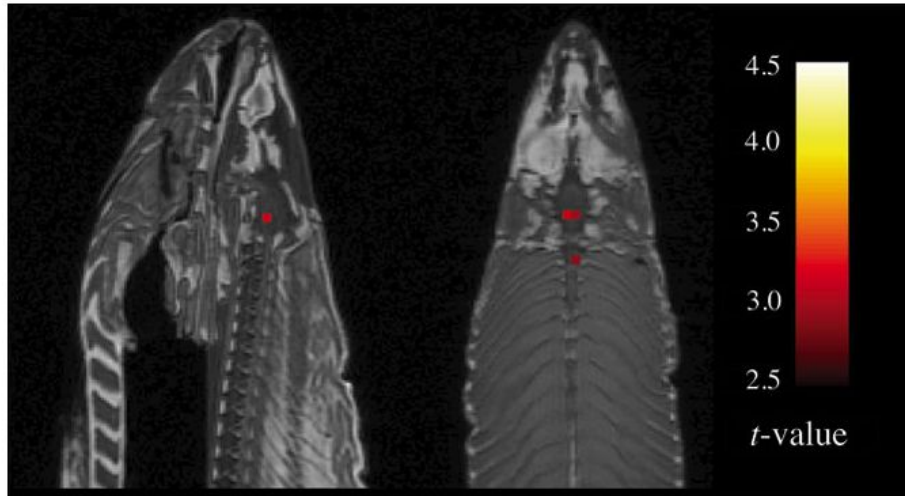
Haochang Shou, Russell T. Shinohara, Han Liu, Daniel S. Reich, Ciprian M. Crainiceanu

(Submitted on 24 Jun 2013)



Motivating example

Brain activation



<http://www.wired.com/2009/09/fmrisalmon/>



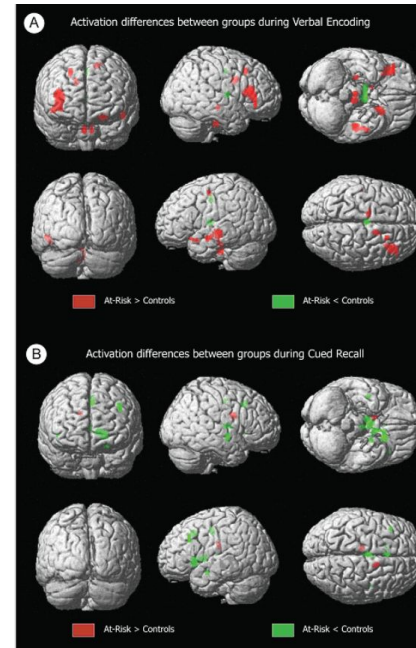
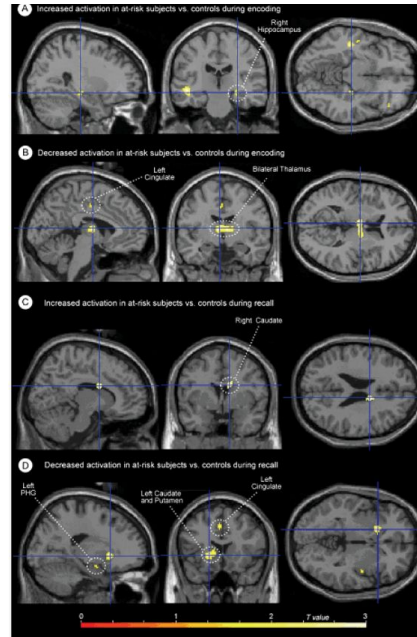
Questions and Answers About the WHI Postmenopausal Hormone Therapy Trials

Questions

- [Can you summarize the results of the estrogen-plus-progestin and estrogen-alone studies?](#)
- [What advice can you give to women about taking estrogen-alone and estrogen-plus-progestin therapy?](#)
- [Are there alternatives for postmenopausal hormone therapy?](#)
- [How can women learn more about WHI and hormone therapy?](#)

Familial risk for Alzheimer's disease alters fMRI activation patterns

Susan Spear Bassett,¹ David M. Yousem,² Catherine Cristinzio,¹ Ivana Kusevic,¹ Michael A. Yassa,¹ Brian S. Caffo³ and Scott L. Zeger³



- We'll dig through the various facets of the perfect data science experience versus real life
- We'll give practical recommendations for the messy situations that arise in real life