Estimation Target Is Relevant

Brian Caffo, Jeff Leek, Roger Peng

@bcaffo

www.bcaffo.com

“If you can't be with the one you love, love the one you're with”

Even less inspiring

http://quoteinvestigator.com/wp-content/uploads/2013/04/muttjeff01.jpg

- It’s very common to not be able to directly measure what is most germane to your investigation
  - BMI as a measure of body fat percentage
  - GDP as a measure of economic well being
  - Food frequency questionnaires for calorie consumption
  - Likely voters versus actual voters
  - Web hits versus actual web traffic
  - BOLD signal versus brain activation

- Good surrogate outcomes:
  - Unbiased
  - Known variance around the desired outcome

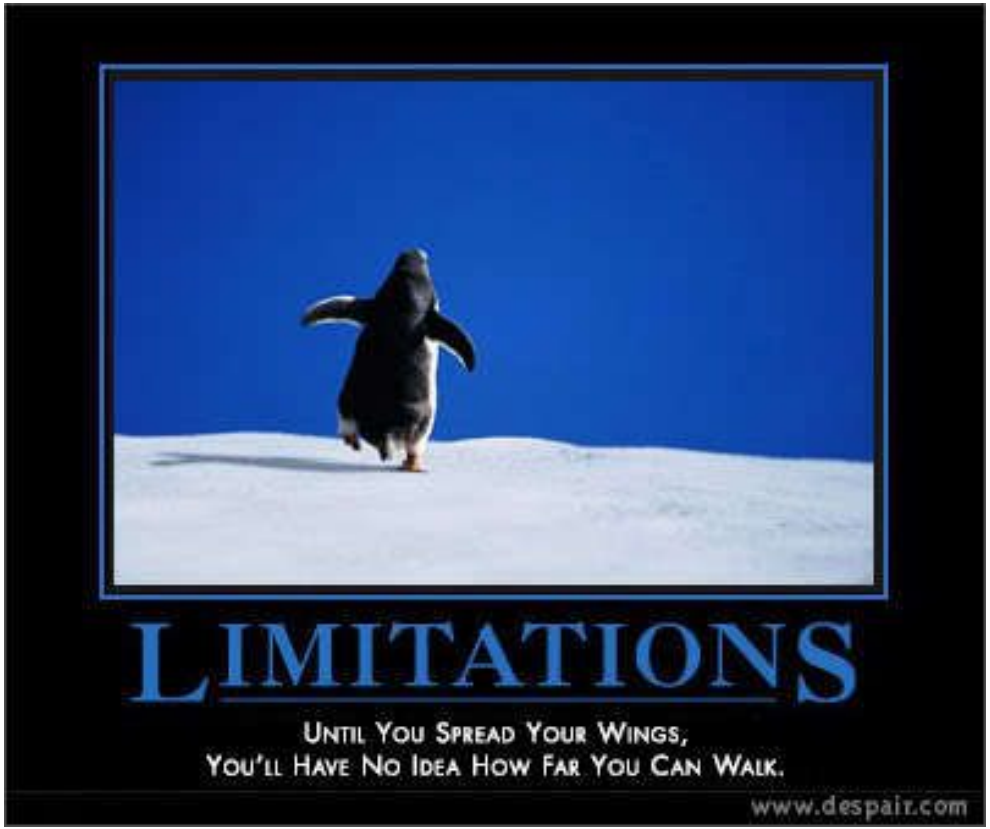
Build models of the truth on a subset where the true outcome is known

- Use the discovered relationships to inform the cases where it’s not
- Example fitbit data - run experiments where you know actual steps

http://cliparts.co/walking-cartoon

If there's no way to validate your surrogate

- Use skepticism of results
- Use sensitivity analysis for bias and variability of the surrogate



Prior to running your analysis, think about how you interpret results

If the surrogate is so unknown and unreliable an estimate of the desired outcome that no results would change your decision or state of knowledge, then why run the analyses at all?