Increased Efficiency in a NeuroImaging Workflow

Frank M. Skidmore, M.D., Thomas Anthony, M.S.; Joanna Myers; Jon Marstrander, M.S.

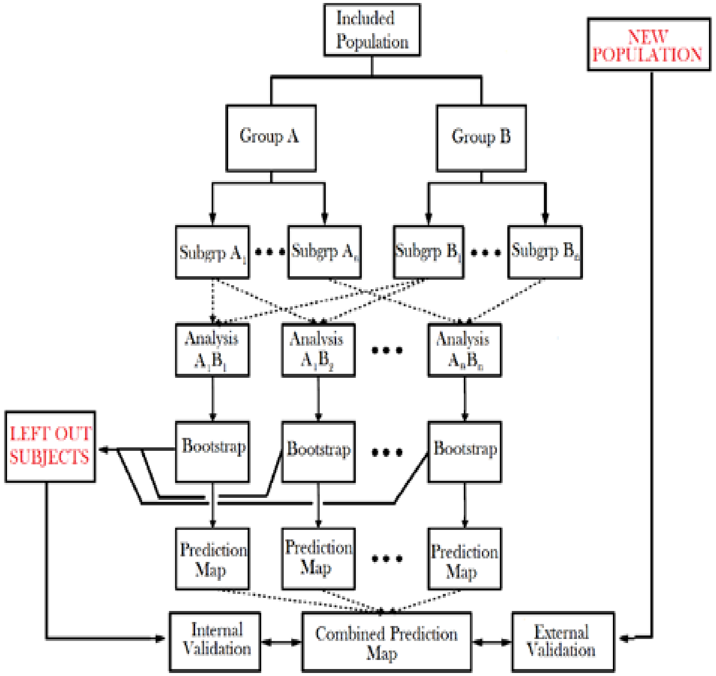
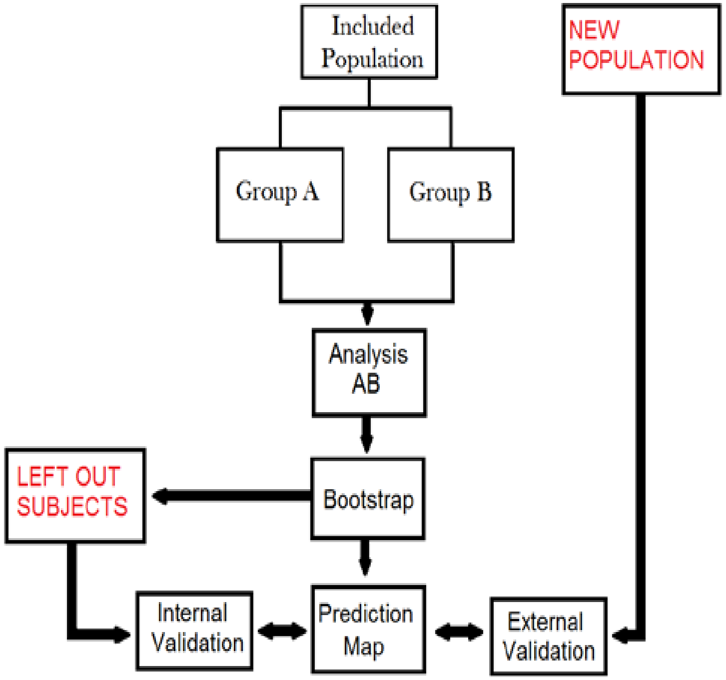
Background – The Standard

- Most brain imaging analyses are focused on generating group-wise results, to prove or disprove a scientific hypothesis
- A key focus is proving or disproving a hypothesis – the subject grouping (either patient based or condition based) is used to support the hypothesis
- A group-wise analysis (rather than a subject specific analysis) is used to support or disprove the hypothesis

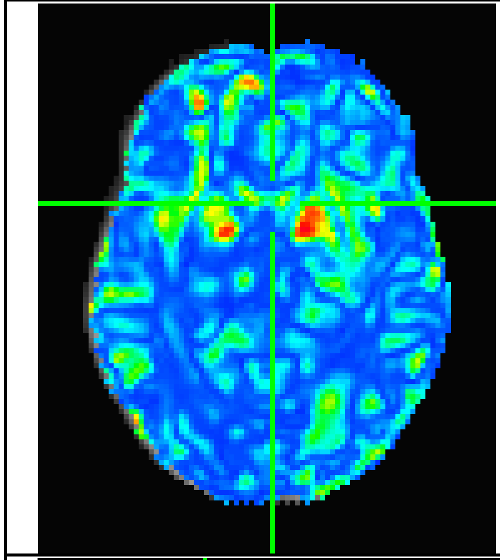
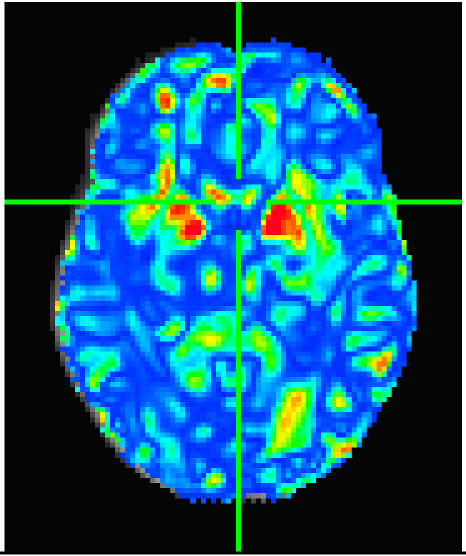
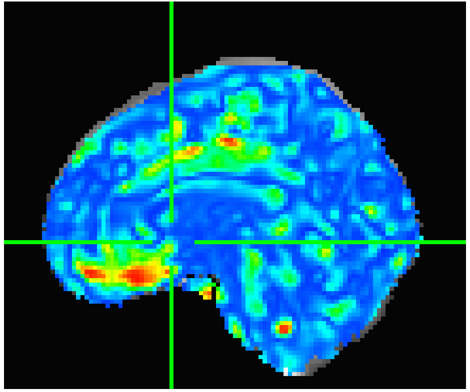
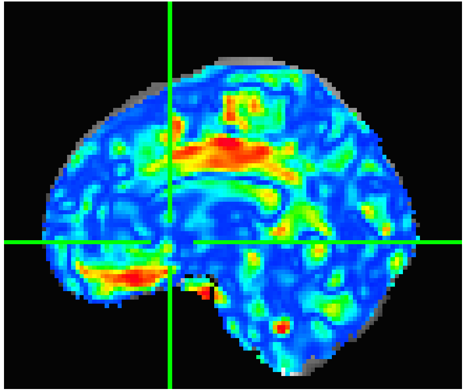
WHAT WE AIM FOR

- Our group is specifically interested in prediction of individual group membership
- For example, we would like to use imaging to diagnosis presence or absence of a clinical condition
- Alternatively, we might wish to use imaging to evaluate if we can predict current or downstream severity of illness

Method – Stratified Bootstrap

Stratified Bootstrap	Simple Bootstrap
	

Example of Mapped Results

Stratified Bootstrap	Simple Bootstrap
	
	

Example of Results

- 20 Subjects with Parkinson Disease
- 22 Healthy Controls
- Diffusion Tensor Fractional Anisotropy (DTI-FA) Map
- Goal: Develop a predictive template that can predict subject class of left-out subjects
- Method: Process DTI-FA Image, Iterative Bootstrap Analysis
  - Lengthy Processing of each image
  - 2,000 to approximately 4,000 t-tests for analysis

Results

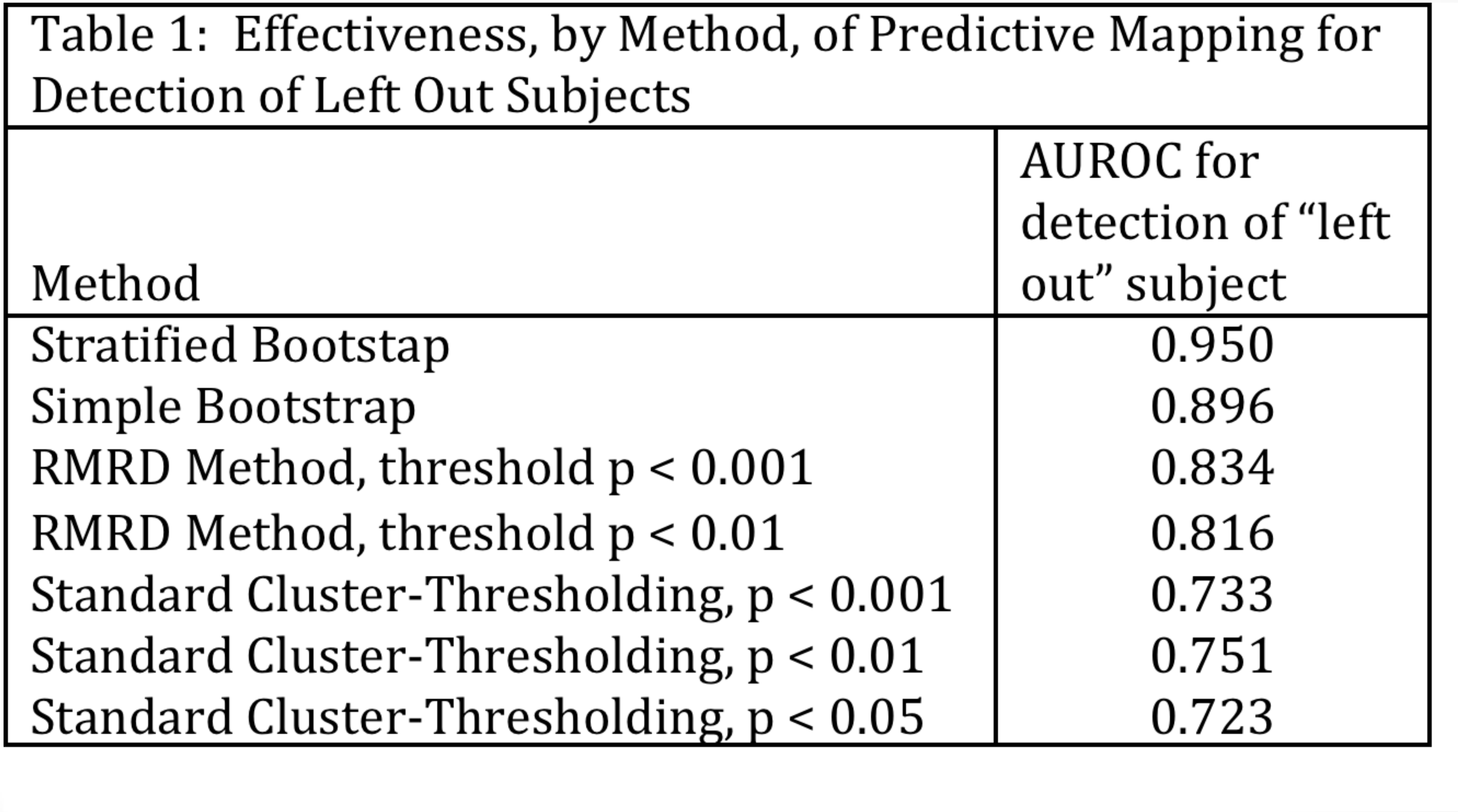
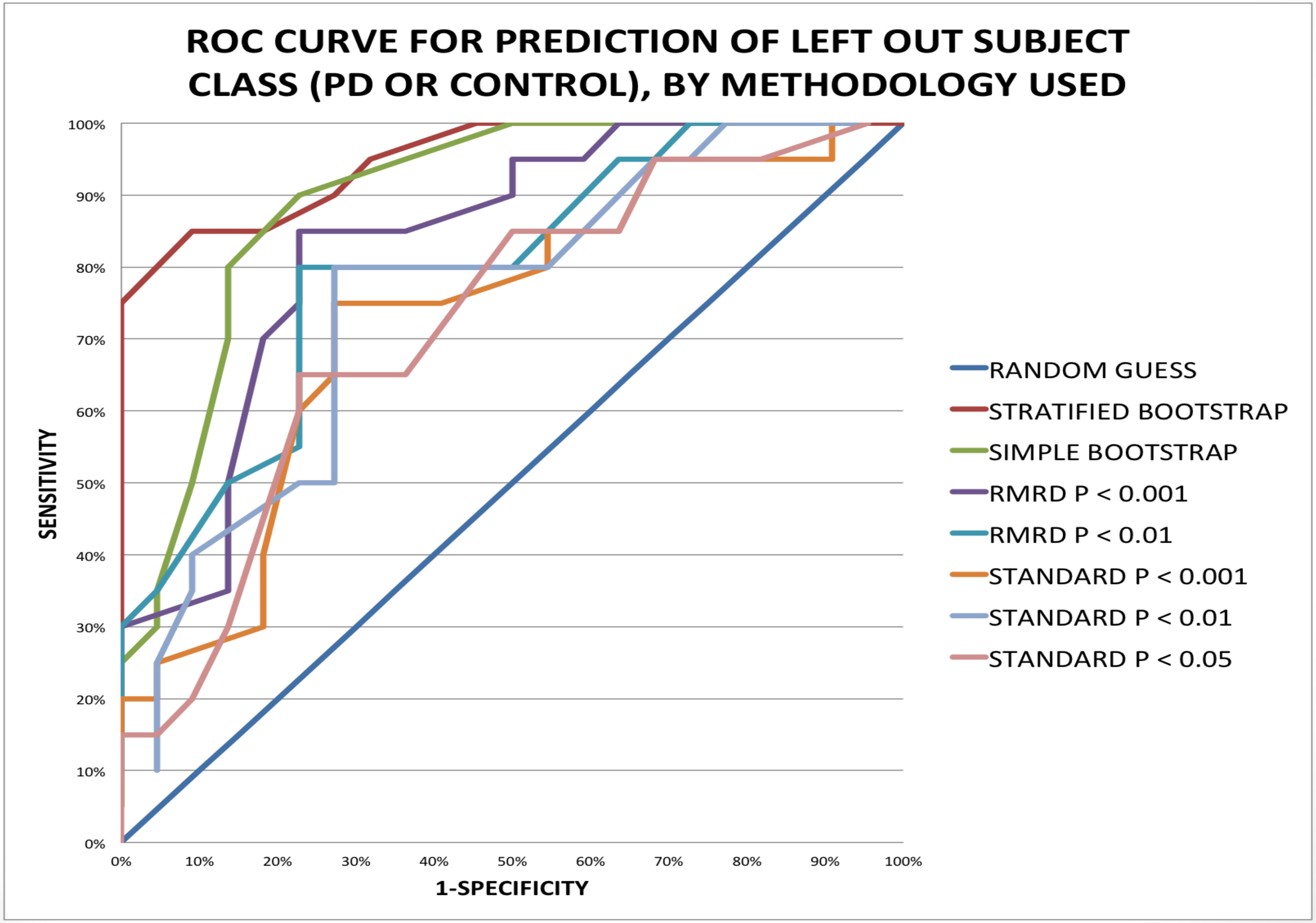


Table 1: Effectiveness, by Method, of Predictive Mapping for Detection of Left Out Subjects

Method	AUROC for detection of “left out” subject
Stratified Bootstap	0.950
Simple Bootstrap	0.896
RMRD Method, threshold $p < 0.001$	0.834
RMRD Method, threshold $p < 0.01$	0.816
Standard Cluster-Thresholding, $p < 0.001$	0.733
Standard Cluster-Thresholding, $p < 0.01$	0.751
Standard Cluster-Thresholding, $p < 0.05$	0.723

Results


ROC CURVE FOR PREDICTION OF LEFT OUT SUBJECT CLASS (PD OR CONTROL), BY METHODOLOGY USED
SENSITIVITY
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
1-SPECIFICITY
RANDOM GUESS
STRATIFIED BOOTSTRAP
SIMPLE BOOTSTRAP
RMRD P < 0.001
RMRD P < 0.01
STANDARD P < 0.001
STANDARD P < 0.01
STANDARD P < 0.05

Analysis Time, Standard

- Processing per subject
  - Data importation: 5-10 minutes per dataset
  - Calculate FA map: 30 minutes
  - Visual Image Quality Analysis: 5-10 minutes
  - Prepare approved subjects: 12 hours
  - Calculate FA map from prepped data: 1 hour
  - Conversion to afni or other processing set: 1 hour
- Bootstrap analysis
  - 2,000 – 4,000 serial t-tests: 6-8 weeks
- Total investigator time for analysis: 1 year for 42 subjects

Computer Processing Estimate

- New sample, with now validated method: 120 subjects, 350 images for analysis
  - 775 – 803 hours of processing time for just the processing step – 32-34 days of computation
  - Conservatively, this would require 100 or more days of processing by a single individual on a high performance single machine
  - Subsequent analysis, on the same machine, would require several weeks for running the algorithm, on a dedicated machine (presuming no power outage)
- On Cheaha, the processing step required < 3 weeks total time
- Pilot of a single bootstrap analysis (2,000 t-tests), required 4 hours using a parallelized processing paradigm across the network, can now plan on rapid analysis in days or less.
- Exploratory analysis is now easier

Summary

- Distributed computing allows large analysis jobs that verge on being impractical in a standard setting to be conducted easily and rapidly
- Dramatic decrease in labor allows mundane processing tasks to be moved to a computer, and reserves more time for creative investigator pursuits
- Significant barriers exist, including requirements to develop knowledge and skills to operate on a “cloud” of processors
- Standard programming tasks often must be altered to operate efficiently on the network
- However, the return on investment is considerable