

Percent Change and Power Calculation

UCLA Advanced NeuroImaging
Summer School, 2007

Outline

- Calculating %-change
 - How to do it
 - What featquery does
- Power analysis
 - Why you should use power calculations
 - How to carry out power calculations

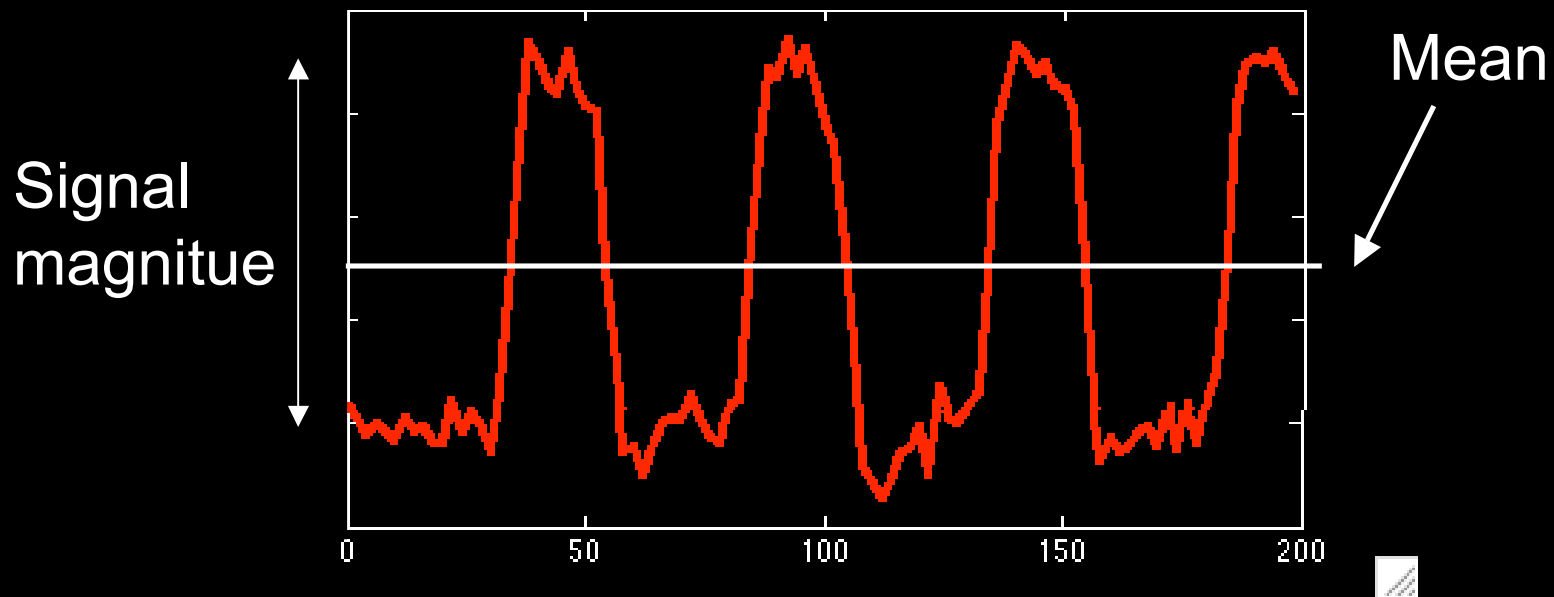
Why %-Change?

- As it is, parameter estimates do not reflect a specific unit
- T-stats are okay (they are unitless)
- What if we want to tell other people how large our activation was?
 - Convert to %-change

%-change

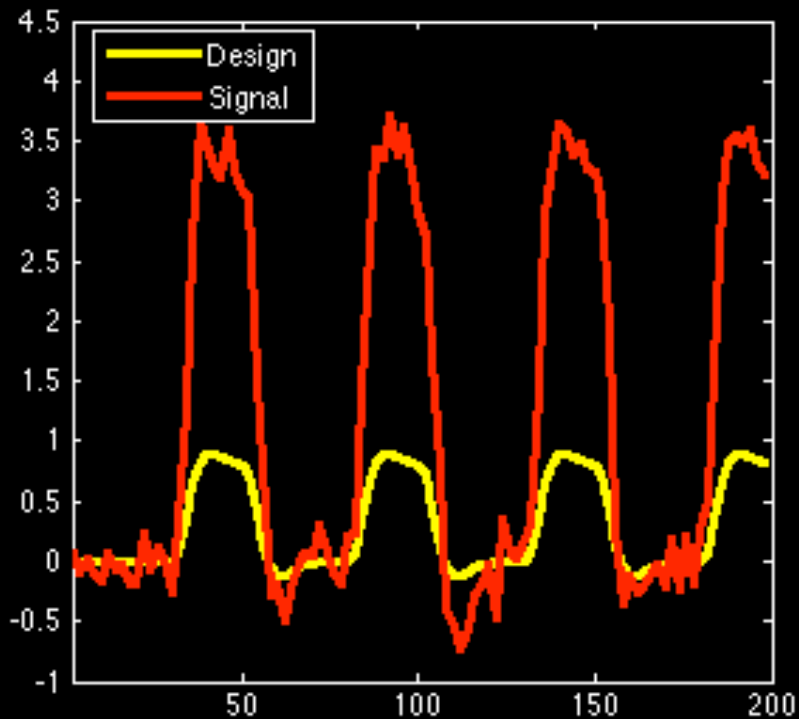
- How big is the signal magnitude relative to baseline?

$$\% \text{-change} = \frac{\textit{sig. magnitude}}{\textit{mean}} \times 100$$

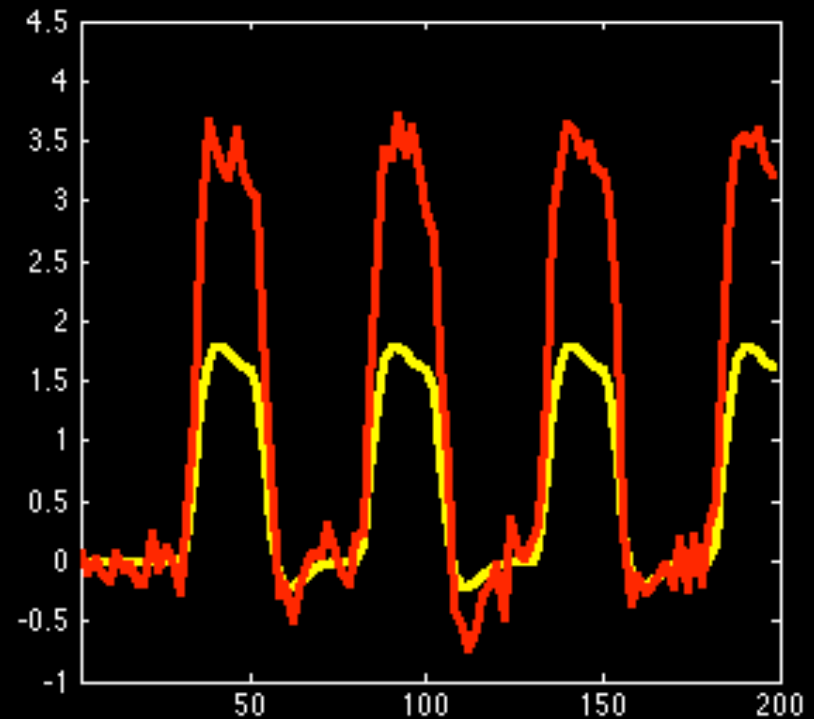


Block Design

- How to get the signal magnitude from parameter estimates...



p.e.=4, EV height=1
1x4=signal magnitude



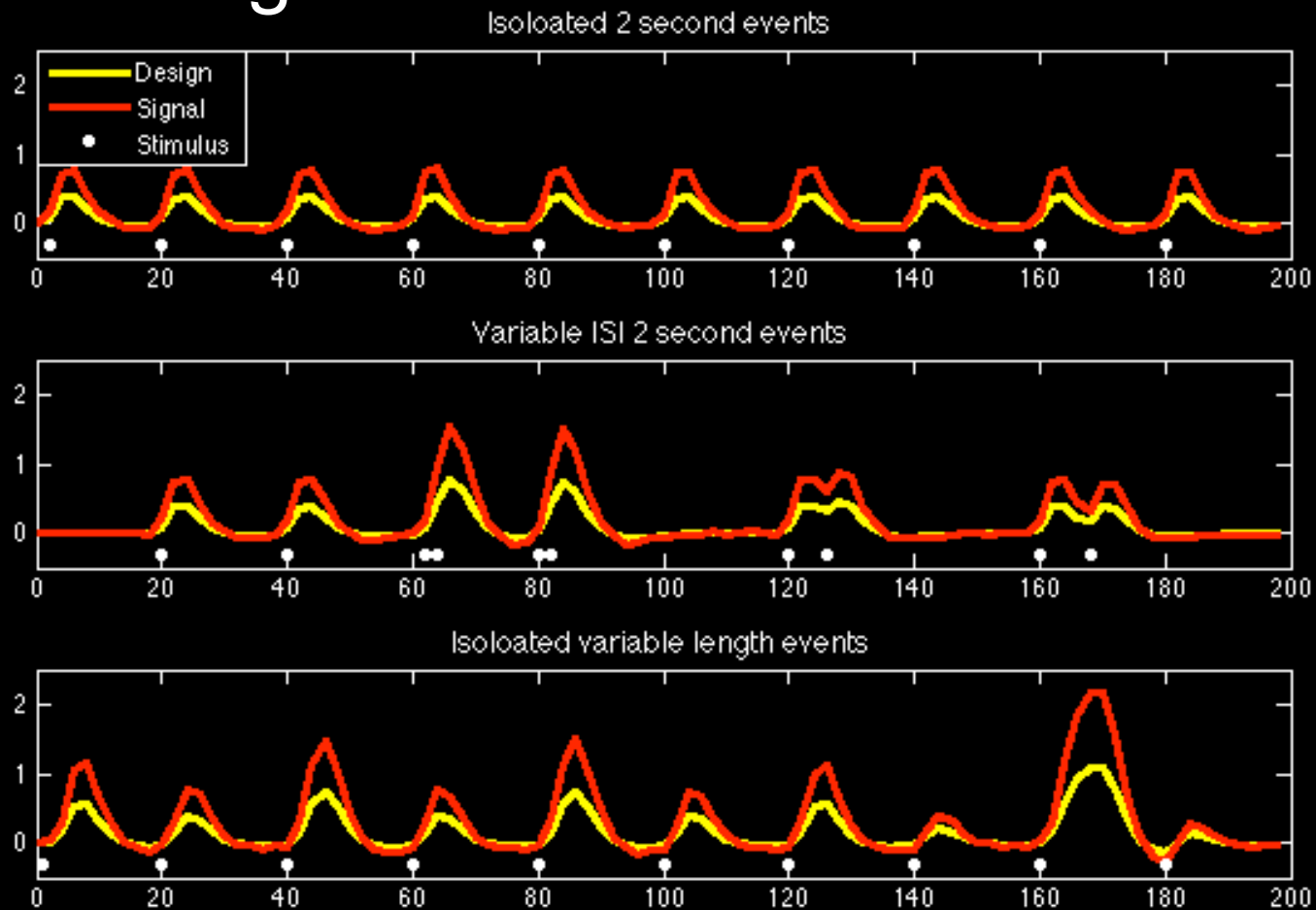
p.e.=2, EV height=2
2x2=signal magnitude

Block Design

- To make life easier, set min/max range of EV's=1!
 - In FSL the Grand Mean Scaling sets mean $\sim 100^2$ in all voxels
 - PE/100=%-change! (roughly)
 - To be completely accurate you should divide by the true mean

Event Related Design

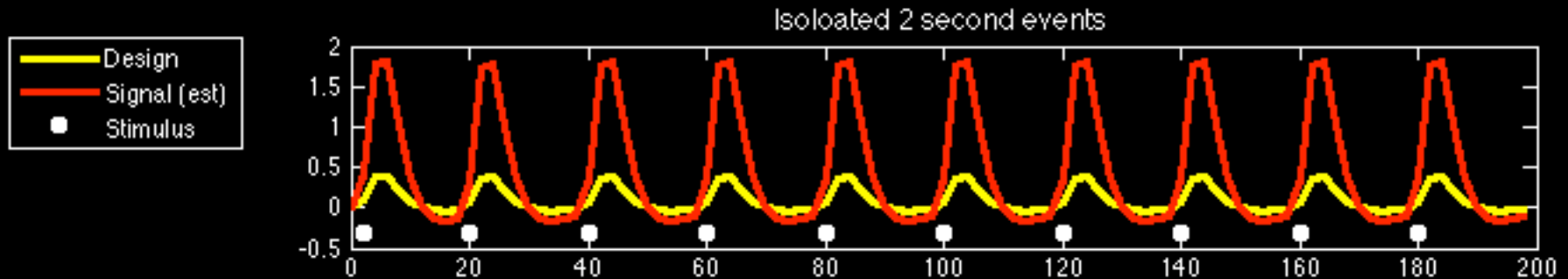
- Tricky! Proximity and length of events affects height



Event Related Design

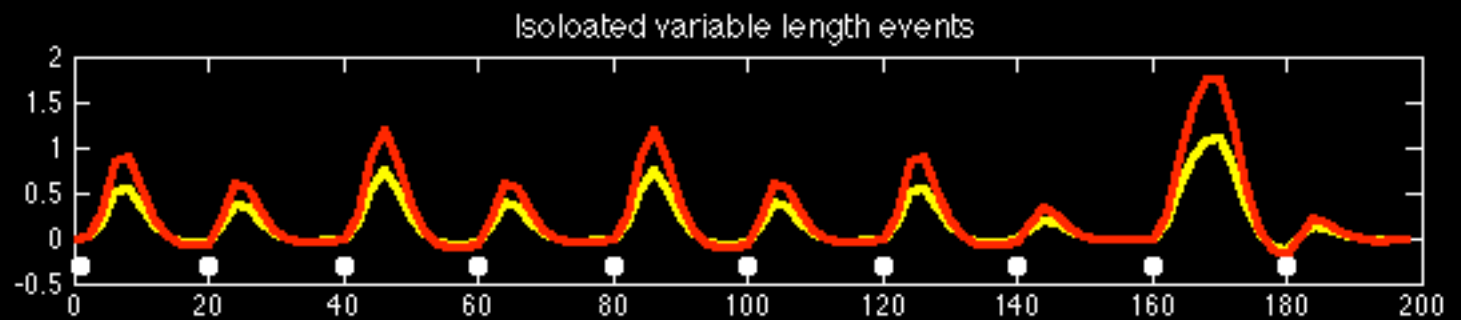
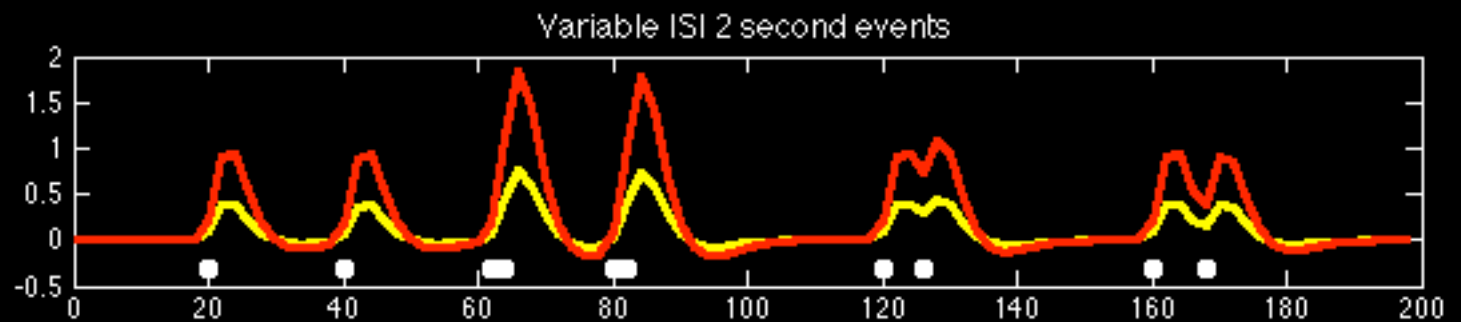
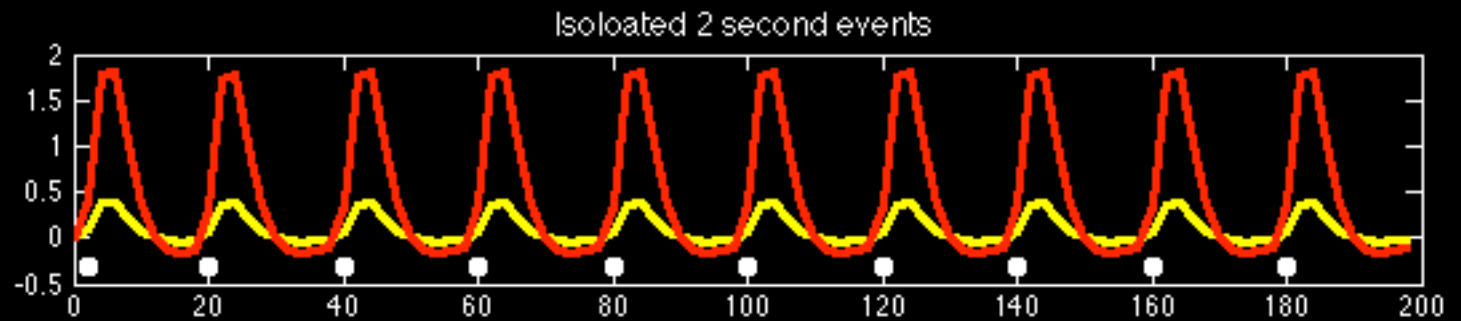
- How about using min/max range?
 - Is it interpretable?
 - I tell you I found a 2% change, calculated using the min/max range. Can you interpret this with your own design?

ER design

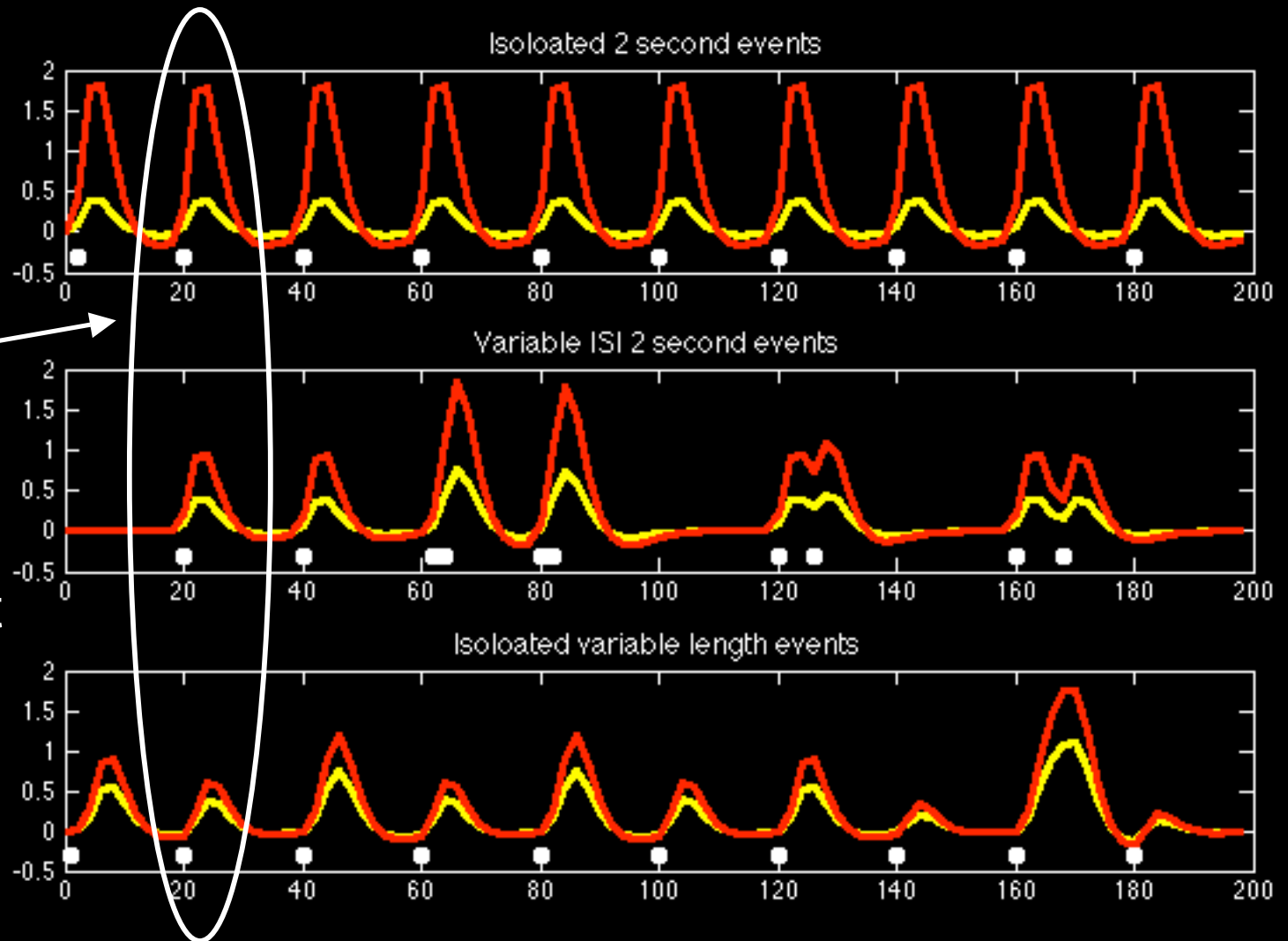


- Assume mean=100
- %-change=(PE)(Min/Max range)
- $2=PE*(0.5) \Rightarrow PE=4$

ER design



ER design



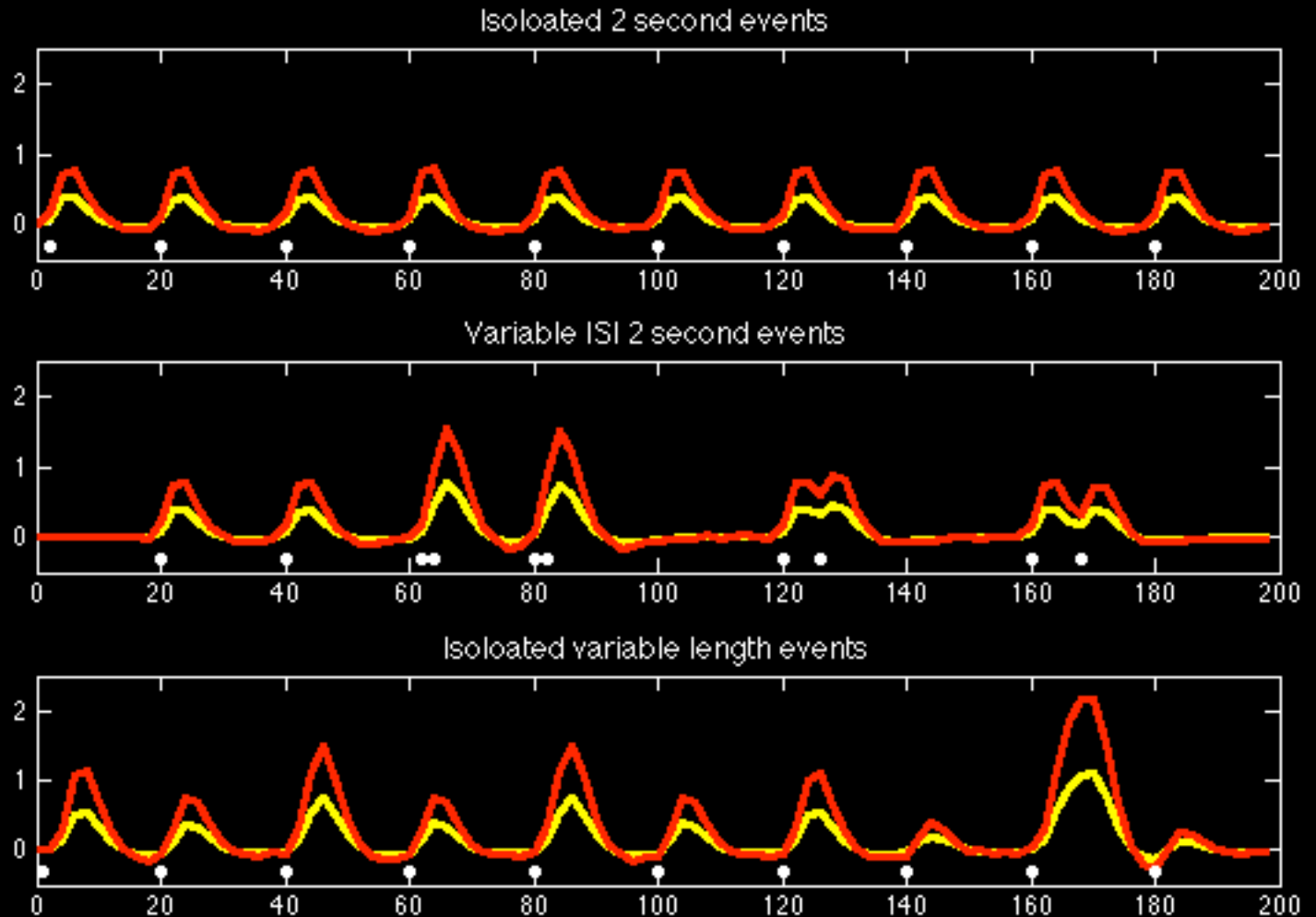
All designs have an isolated 2 second event **BUT** the estimated signal is very different!

ER Design

- Min/max range doesn't work as well for event related design
 - Doesn't translate well to other designs.
 - *warning* Featquery uses min/max range
- Instead of min/max range, choose something specific (isolated 2 second event) and make sure to report what you used!

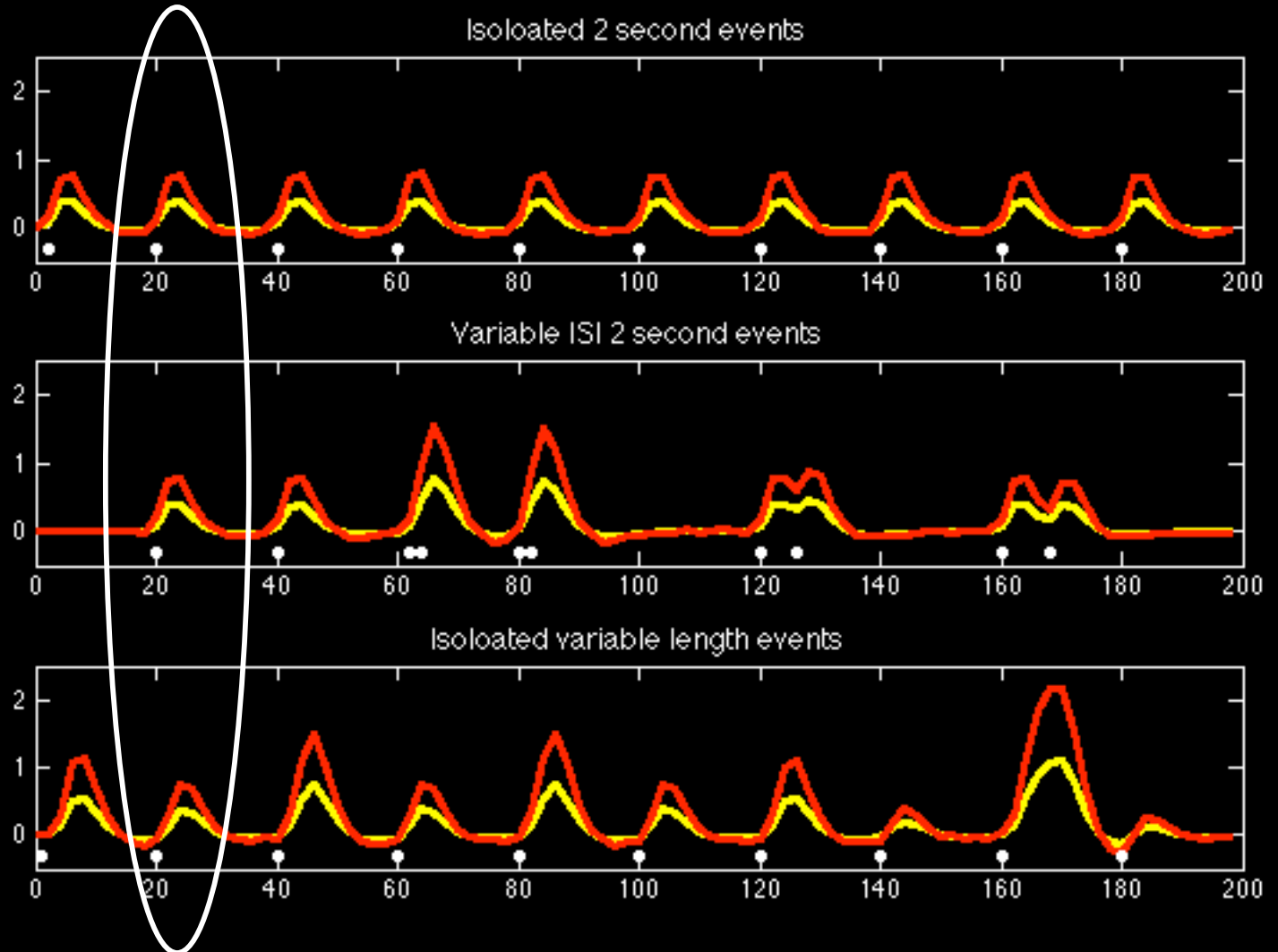
ER Design

- 1% change based on height of isolated 2-s event



ER Design

- 1% change based on height of isolated 2-s event
- Now we can compare across designs!



A note about contrasts

- What does the contrast $[1 \ 1 \ -1 \ -1]$ mean?
 - $(\beta_1 + \beta_2) - (\beta_3 + \beta_4)$
 - beta's are mean activations for 4 levels of subjects performing some task (beginners, some training, medium training, experts)
 - Test beginners/some training - med/experts
 - $\beta_1 = 2, \beta_2 = 2.5, \beta_3 = 5, \beta_4 = 5.2$
 - $\beta_1 + \beta_2 = 4.5$ and $\beta_3 + \beta_4 = 10.2$
 - Difference is twice the mean difference!

A note about contrasts

- Although $[1 \ 1 \ -1 \ -1]$ implies a difference that is twice the original scale, our test statistic is okay
- Since we're interested in preserving scale, use contrast that give us means
 - Positive parts sum to 1
 - Negative parts sum to -1

Rules to get *almost* %-change in FSL

- Baseline is already $\sim 100^2$
- EV is constructed appropriately
 - Boxcar height = 1
 - ER design: Specified event has height=1
 - I like to ignore the post stimulus undershoot
- Construct contrast that follow the rules
 - Positive parts sum to 1
 - Negative parts sum to -1
- $PE * 100 \sim \%$ -change

What if you didn't follow the rules

- Calculate the height of regressor as it was
 - Height of block
 - Height of specified typed of event
- Calculate contrast fix
 - Number you'd have to divide contrast by to fix it

- $\% \text{ change} = \frac{PE(\text{EV height})}{(\text{contrast fix})(\text{baseline})} \times 100$

Example

- Study with subjects that were beginners, some training, medium training, experts
- Level 1, isolated 2 second event (using the gamma HRF) has height=0.2917
- Level 2, I used the contrast [1 1 -1 -1]
- COPE=500 (from contrast estimate)
- $\% \text{ change} = \frac{(500)(0.2917)}{(2)(100^2)} \times 100 = 0.729$

Featquery

- Featquery calculates %-change for ROI's
- Uses min/max range of 'effective regressor height'
 - Roughly speaking effective regressor corrects violation of the contrast rule and correlation between regressors
- Probably works fine for block design
- Very bad to run separately on first level runs for ER studies!
 - Will use a different scale factor for each subject
- Min/Max range isn't best for ER design

What to do instead of featquery?

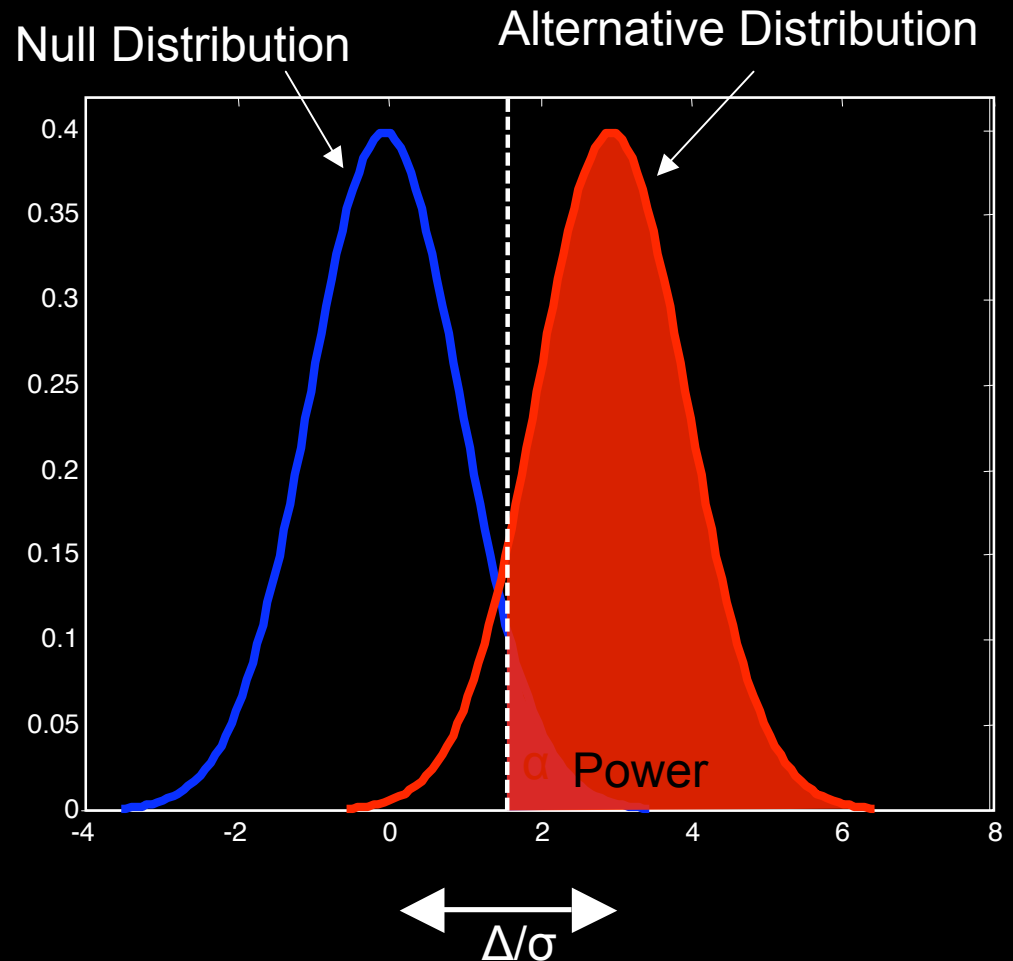
- There's a writeup with some code that uses `avwstats` and `avwmaths` located here
 - http://mumford.bol.ucla.edu/perchange_guide.pdf
- Some isolated event heights are in a table in this document
 - Isolated event height should be calculated from a design with *very* small TR for best resolution

Power Analysis-Why?

- To answer the question....
 - How many subjects do I need for my study?
 - How many runs per subject should I collect?
- To create thorough grant applications that will make reviewers happy!
 - Don't waste money on underpowered studies OR collecting data on more subjects than you need

What is Power?

- Power: The probability of rejecting H_0 when H_A is true
- Specify your null distribution
 - Mean=0, variance= σ^2
- Specify the effect size (Δ), which leads to alternative distribution
- Specify the false positive rate, α



fMRI Power

- Interpretation
 - 1100 total voxels
 - 100 voxels have $\beta=\Delta$
 - A test with 50% power on average will detect 50 of these voxels with true activation
 - 1000 voxels have $\beta=0$
 - $\alpha=5\%$ implies on average 50 null voxels will have false positives

| | | Truth (unobserved) | |
|------------------------|-----------------|------------------------------------------|-------------------------------|
| | | H_0 True | H_0 False |
| Test Result (observed) | Reject H_0 | Type I Error α 50 | Power 50 |
| | Accept H_0 | Correct 950 | Type II Error 50 |
| | | 1000 | 100 |

Necessary information

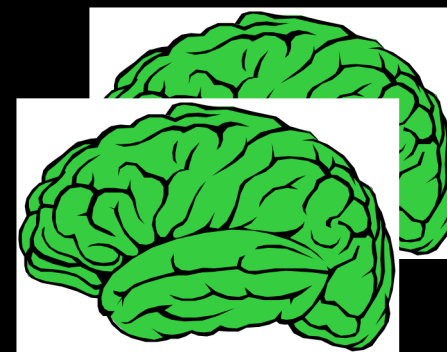
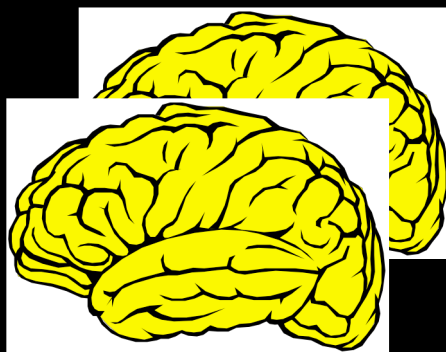
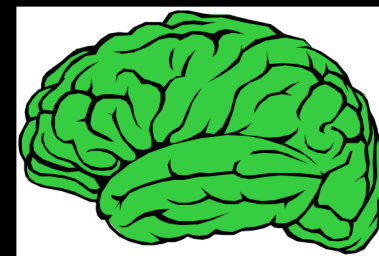
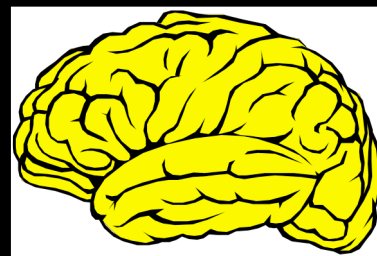
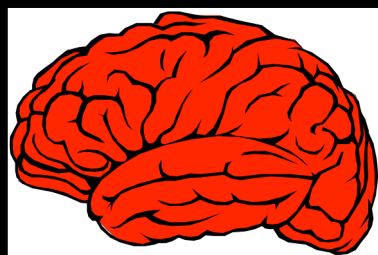
- **N**: Number of Subjects
 - Adjusted to achieve sufficient power
- **α** : The size of the test you'd like to use
 - Commonly set to 0.05 (5% false positive rate)
- **Δ** : The size of the effect you're interested in detecting
 - Based on intuition or similar studies
- **σ^2** : The variance of Δ
 - Has a complicated structure with very little intuition
 - Depends on many things ...

Why is it so difficult for group fMRI?

Temporal autocorr.

$$\text{Cov}(Y) = \sigma_w^2 V$$

Time



Subject 1

Subject 2



Subject N

Between subject variability, σ_B^2

Level 1

$$Y_k = X_k \beta_k + \varepsilon_k$$

The diagram illustrates the Level 1 model equation $Y_k = X_k \beta_k + \varepsilon_k$. On the left, Y_k is represented as a vertical vector of four red time series plots. This is followed by an equals sign. In the center, X_k is shown as a matrix with four columns: a gray column with a red vertical line, a white column with a red wavy line, a gray column with a red wavy line, and a white column with a red wavy line. To the right of X_k is another equals sign, followed by a vertical vector of four parameters: β_{k0} , β_{k1} , β_{k2} , and β_{k3} . This is followed by a plus sign and a vertical vector of four red error term plots, labeled ε_k .

- Y_k : T_k -vector timeseries for subject k
- X_k : $T_k \times p$ design matrix
- β_k : p -vector of parameters
- ε_k : T_k -vector error term, $\text{Cov}(\varepsilon_k) = \sigma_k^2 V_k$

Level 2

$$\hat{\beta}_{cont} = X_g \beta_g + \varepsilon_g$$

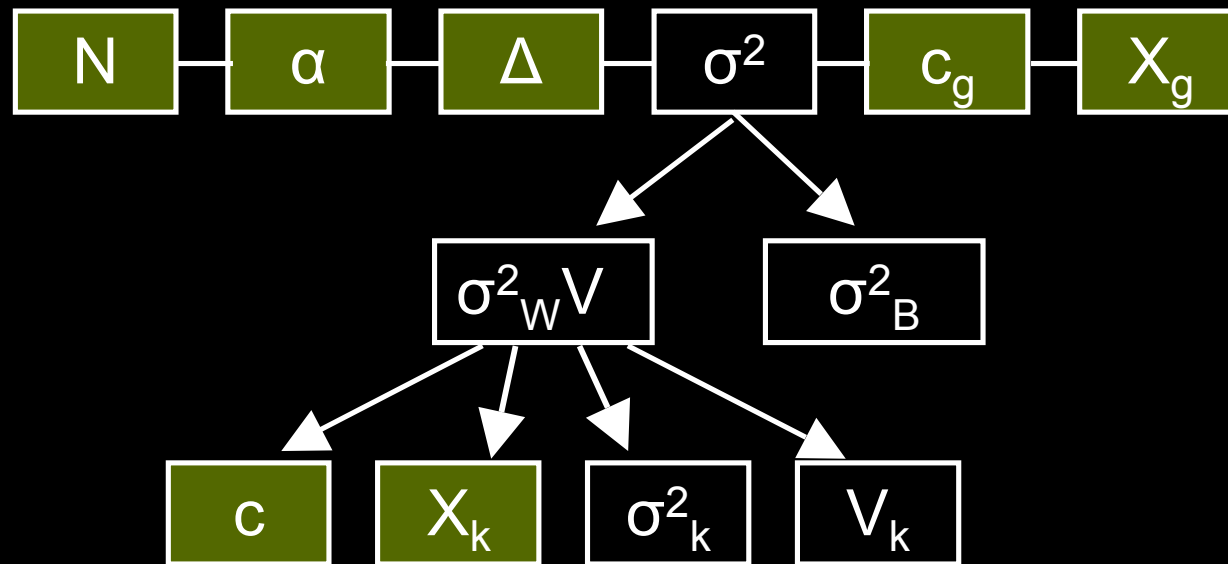
- $c\hat{\beta}_k$
- X_g : $N \times p_g$ design matrix
- β_g : p_g -vector of parameters
- ε_g : N -vector error term

$$- \text{Cov}(\varepsilon_g) = V_g = \underbrace{\text{diag}\{c(X_k^T V_k^{-1} X_k)^{-1} \sigma_k^2 c^T\}}_{\text{Within subject variability}} + \underbrace{\sigma_B^2 I_N}_{\text{Between subject variability}}$$

Within subject variability Between subject variability

Alternative distribution

- For a specific $H_A: c_g \beta_g = \Delta$
- t is distributed $T_{n-pg, ncp}$
 - $ncp = \Delta / c_g (X_g^T V_g^{-1} X_g) c_g^T$
- What do we need?



Estimating Parameters

- How to you estimate parameters for a future study?
 - Look at other people's study results for similar studies
 - Look at your own similar studies
- Average parameter estimates over ROIs of interest

Estimating $\sigma_k^2 V_k$

- Use covariance estimate supplied by software
 - FSL uses a unstructured covariance that can have up to 13 or more parameters
 - SPM2 uses a 2 parameter AR(1) approximation
- Fit AR(1)+WN model to software estimate
 - Only 3 parameters
 - Summarize parameters over ROI's
 - Simple parameterization is easy to report and bridges covariance estimates across software
 - Assume covariance is same over all subjects (necessary to extrapolate to larger populations)

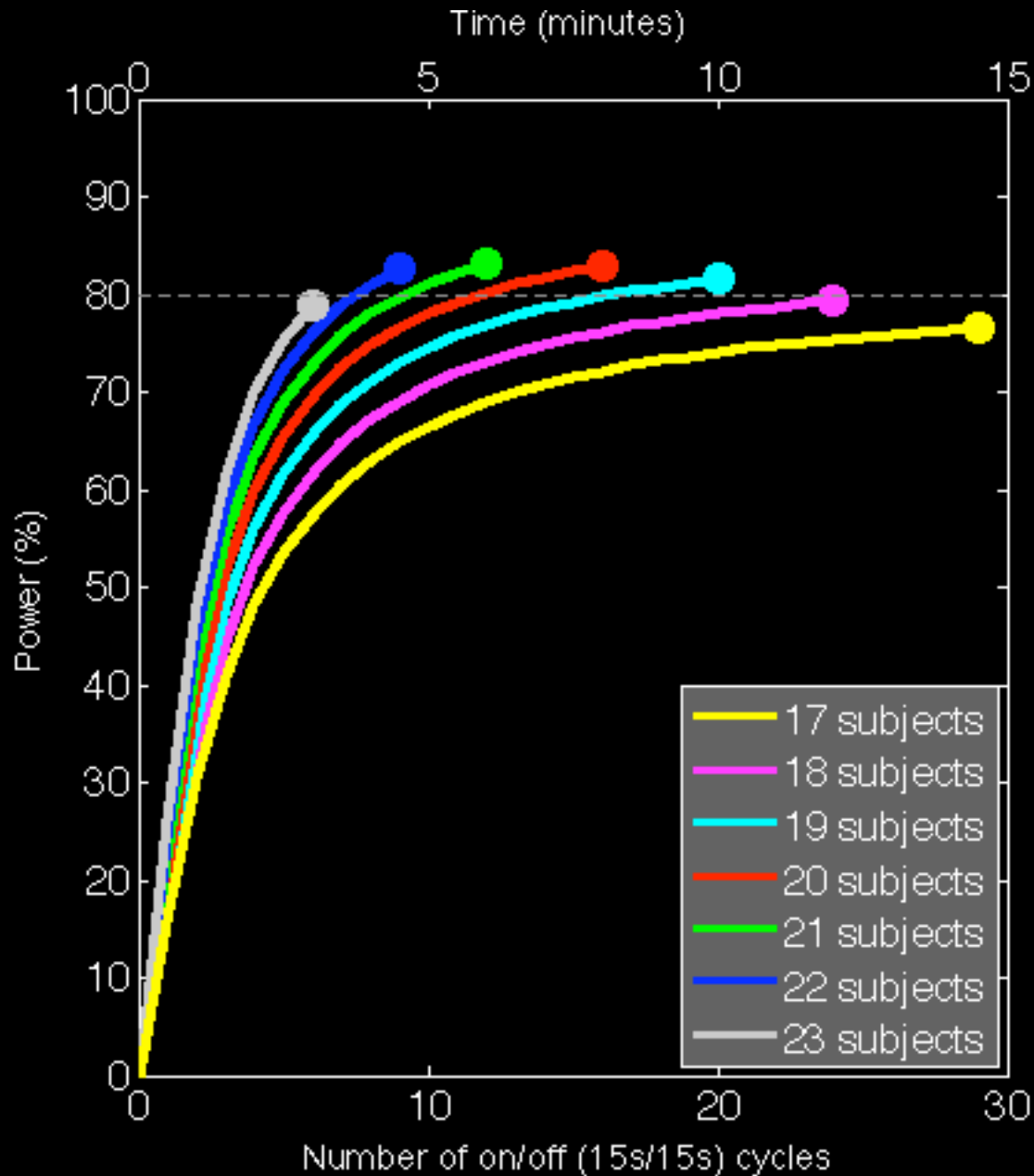
Estimating σ^2_B

- SPM2 doesn't estimate, but we still need to approximate it for the power calculation
 - You can see my paper for the details
- FSL estimates and saves this parameter in an image file
 - Average over the ROI

Model

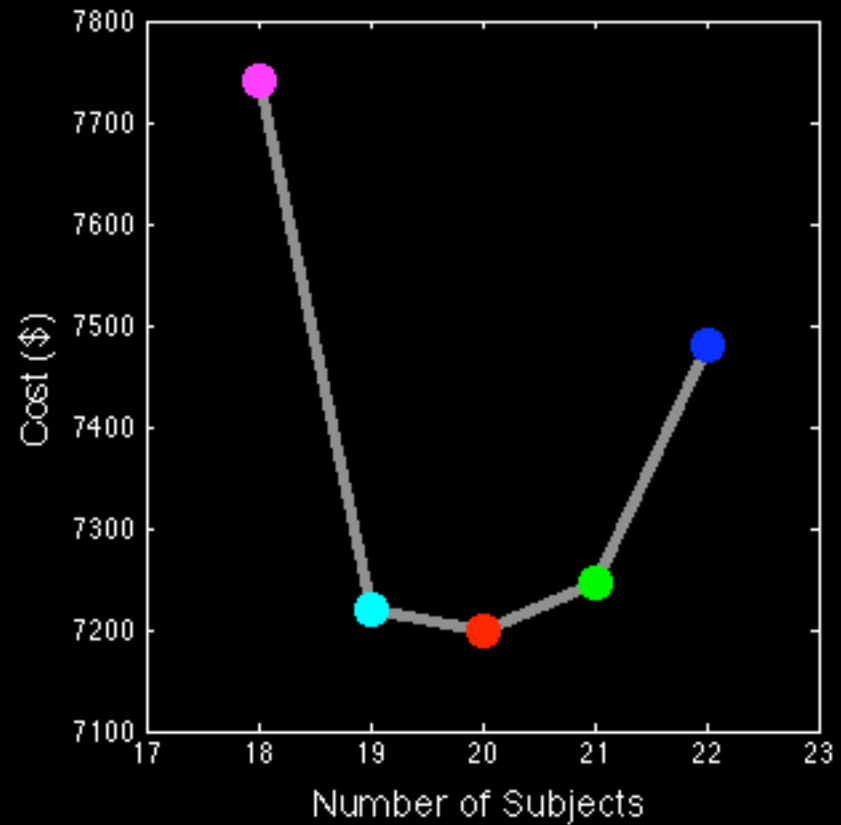
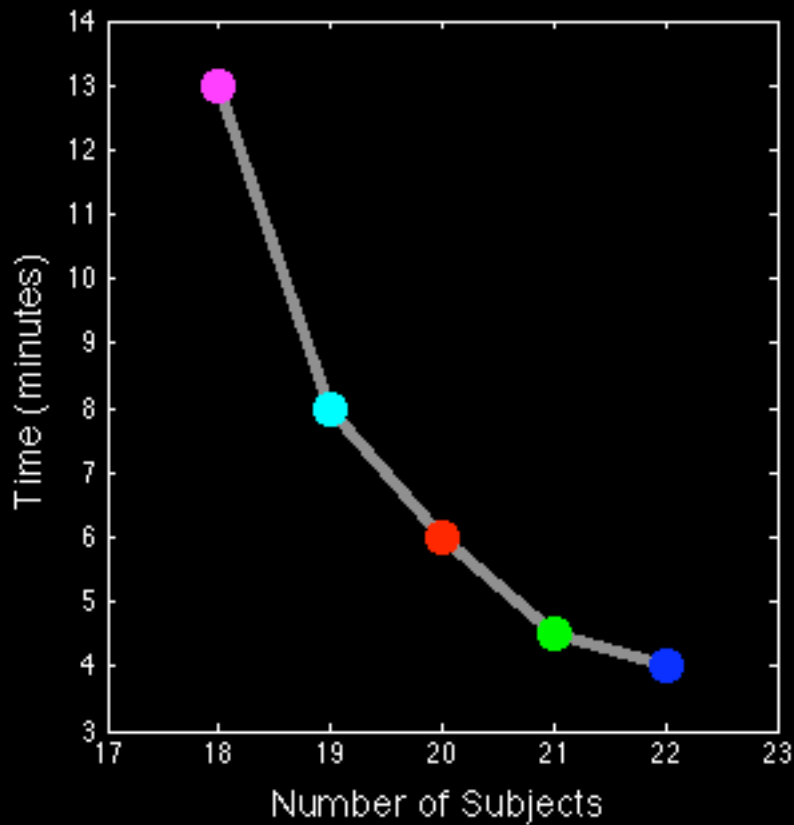
- Block design 15s on 15s off
- TR=3s
- Hrf: Gamma, sd=3
- Parameters estimated from Block study
 - FIAC single subject data
 - Read 3 little pigs
 - Same/different speaker, same/different sentence
 - Looked at blocks with same sentence same speaker

Power as a function of run length and



More importantly....cost!

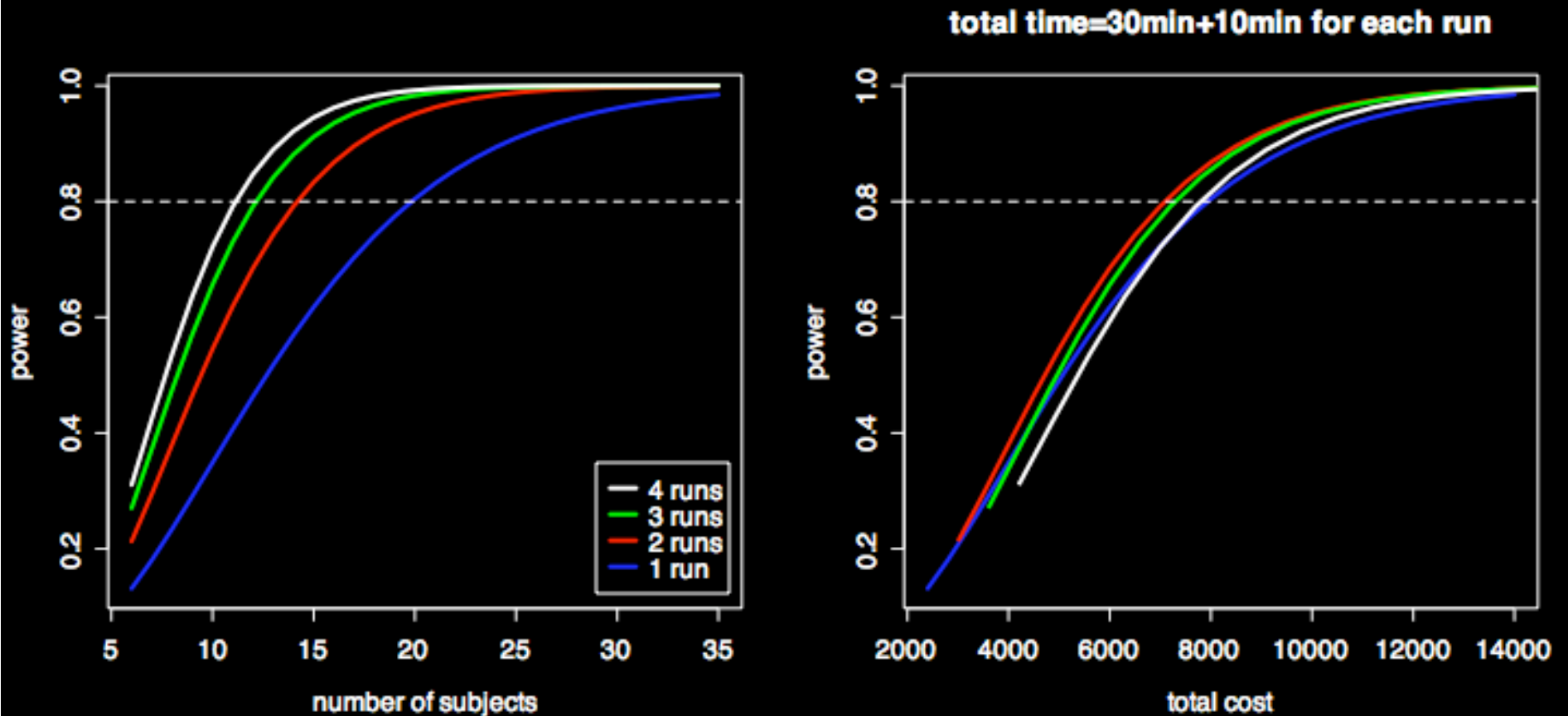
- Cost to achieve 80% power
- Cost=\$300 per subject+\$10 per each extra minute



How Many Runs?

- Can also expand to a 3 level model and study impact of adding runs
- Example
 - ER study
 - Study used 3 runs per subject
 - Estimate between run variability
 - Assume within subject variability is the same across subjects
 - Assume study design is same across subjects

How many runs?

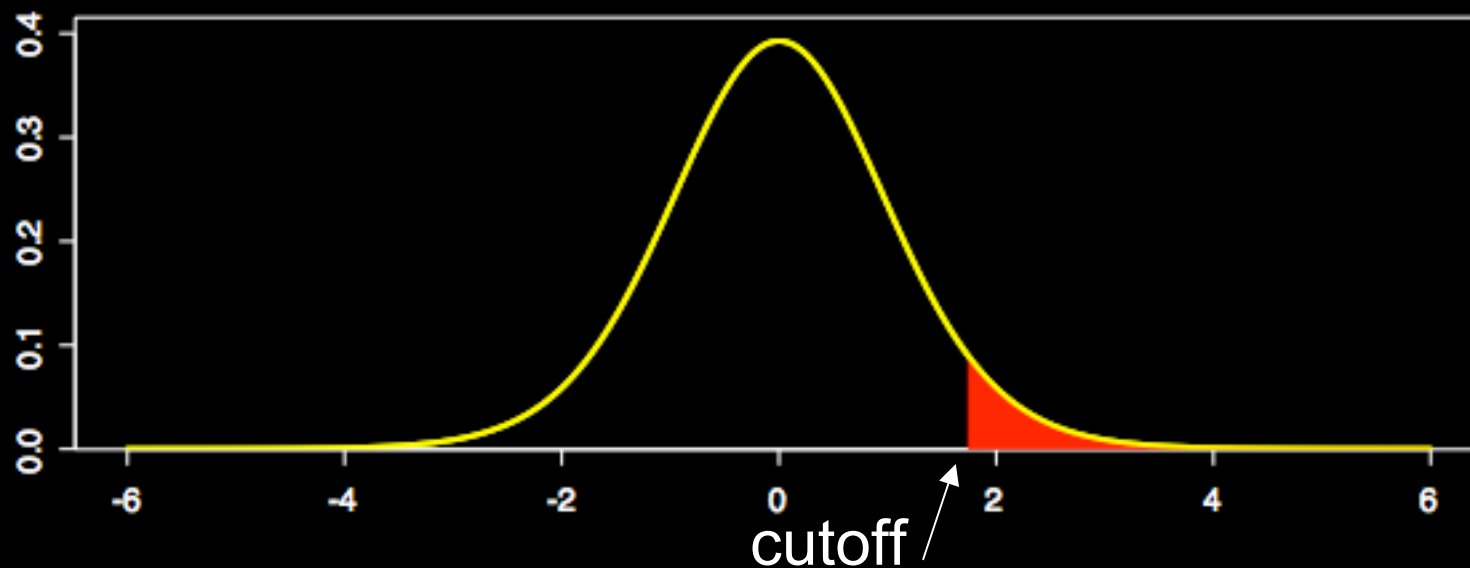


How can you calculate power?

- Only 1 tool that I know of
 - Fmripower! (created by me)
 - Beta version at fmripower.org
 - ROI based power analysis
 - Can apply to old FSL analyses
 - Runs in Matlab
 - Current version only allows user to specify different #'s of subjects
 - Assumes # of runs for future study will be the same
 - Assumes between subject variability is same across subjects
 - Doesn't control for multiple comparisons

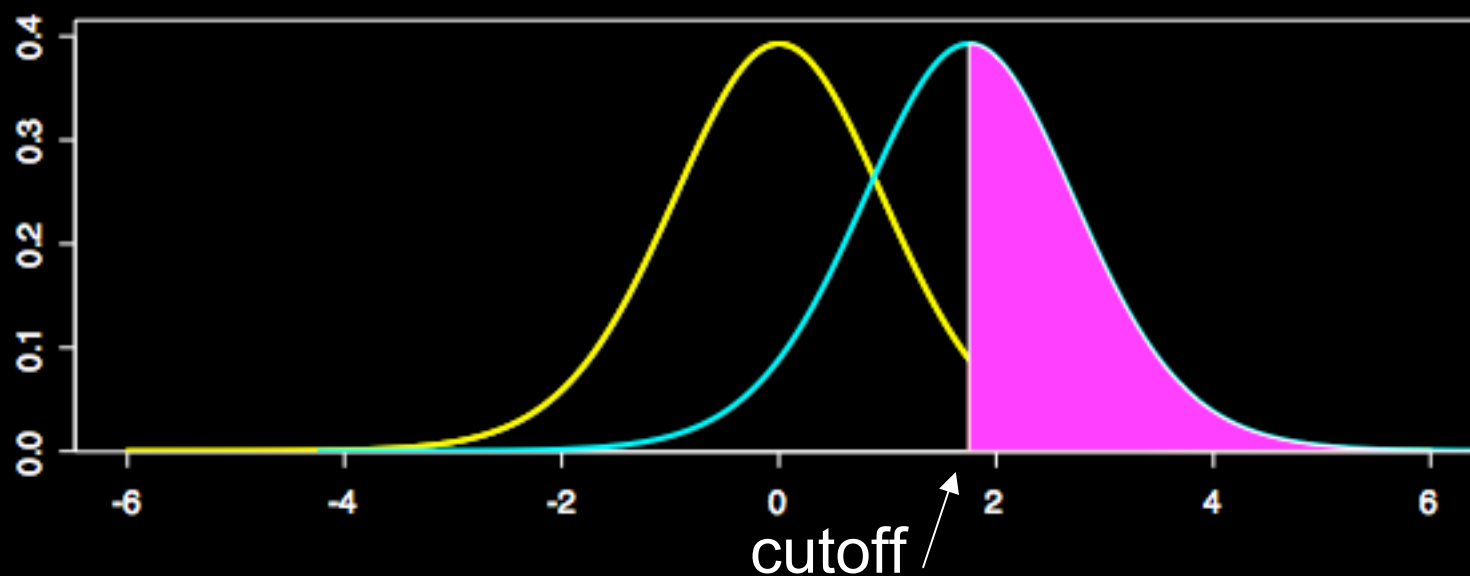
Can I use power post hoc?

- Please don't! It doesn't make sense
 - Power is a function of alpha
 - If you rejected your null, post hoc power is always less than 50%
 - See *The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis* Hoenig et al, Amer. Stat. 2001
- Try percent change threshold (see Tom Nichols website)
 - Based on confidence intervals



Can I use power post hoc?

- Please don't! It doesn't make sense
 - Power is a function of alpha
 - If you rejected your null, post hoc power is always less than 50%
 - See *The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis* Hoenig et al, Amer. Stat. 2001
- Try percent change threshold (see Tom Nichols website)
 - Based on confidence intervals



fmripower

fmRipower

Set .gfeat options

.gfeat directory
/rack/revlearn/group/posneg/3rdlev/9_tp2_post_corr-tp1_post_i

Select lower level cope of interest
cope1_feat

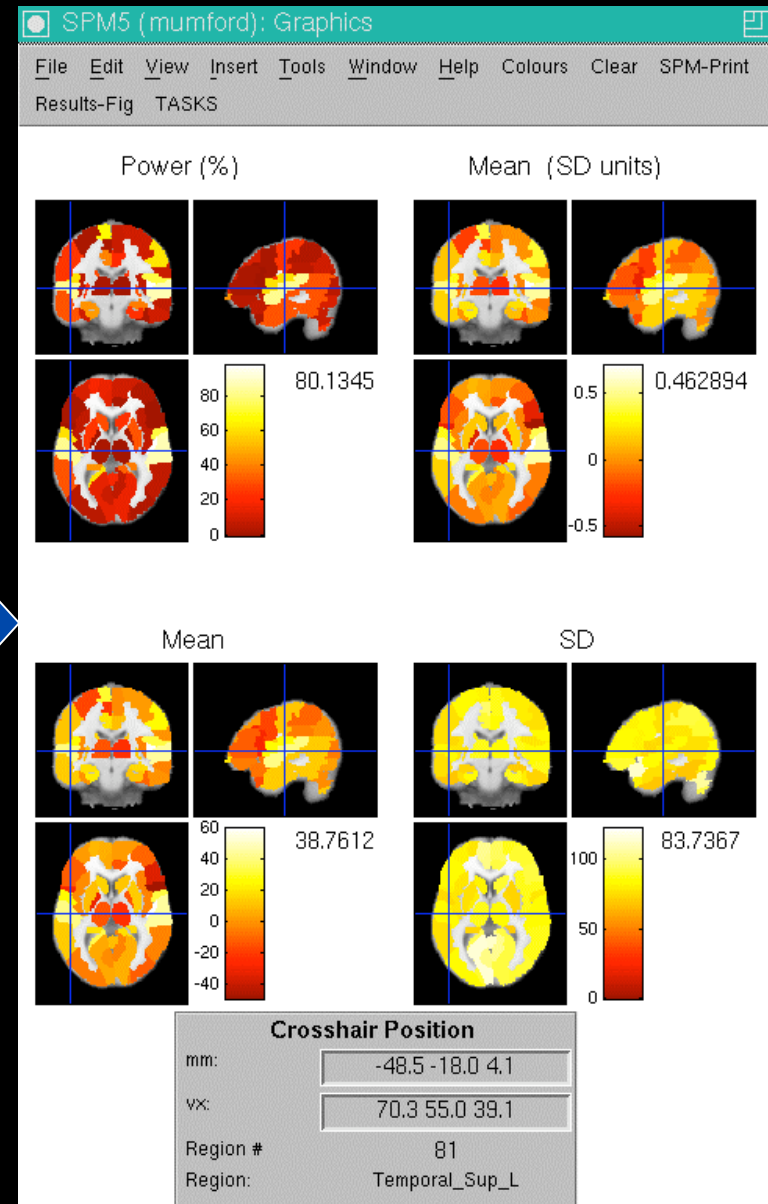
Select top level cope of interest
cope1.nii.gz

Power calculation options

group design matrix
/rn/group/posneg/3rdlev/9_tp2_post_corr-tp1_post_incorr/Des

ROI mask
/raid/home/mumford/MatlabCode/New_gui/fmripower_gui/aa_2t

Type I error rate
.05



Conclusion

- Percent change should be calculated in a manner that is interpretable
 - Min/max range does not lead to interpretable %-change
 - Interpretable %-change is very handy for power calculations
- Power calculations can save you time and money
 - Scan enough subjects to see an effect, don't waste time on an underpowered study
 - Don't keep subjects in the scanner too long if you don't need to