Population-Level Analysis in AFNI

Gang Chen

Scientific and Statistical Computing Core
National Institute of Mental Health
National Institutes of Health
U.S.A.

Polls: statistics knowledge

- priors: t-tests, regression, correlation, general linear model (GLM)

- AN(C)OVA: between- and within-individual factors

- linear mixed-effects (hierarchical, multilevel) modeling

- meta-analysis

- Bayesian modeling

- programming languages: R, Matlab, SPSS, SAS

- statistics is hard
    - ⋆ science: assessing variability/uncertainty
        - popular face: a single p-value
    - ⋆ art: extracting information
        - 1-sample t-test with outliers?

# Big picture 1: theory vs statistics
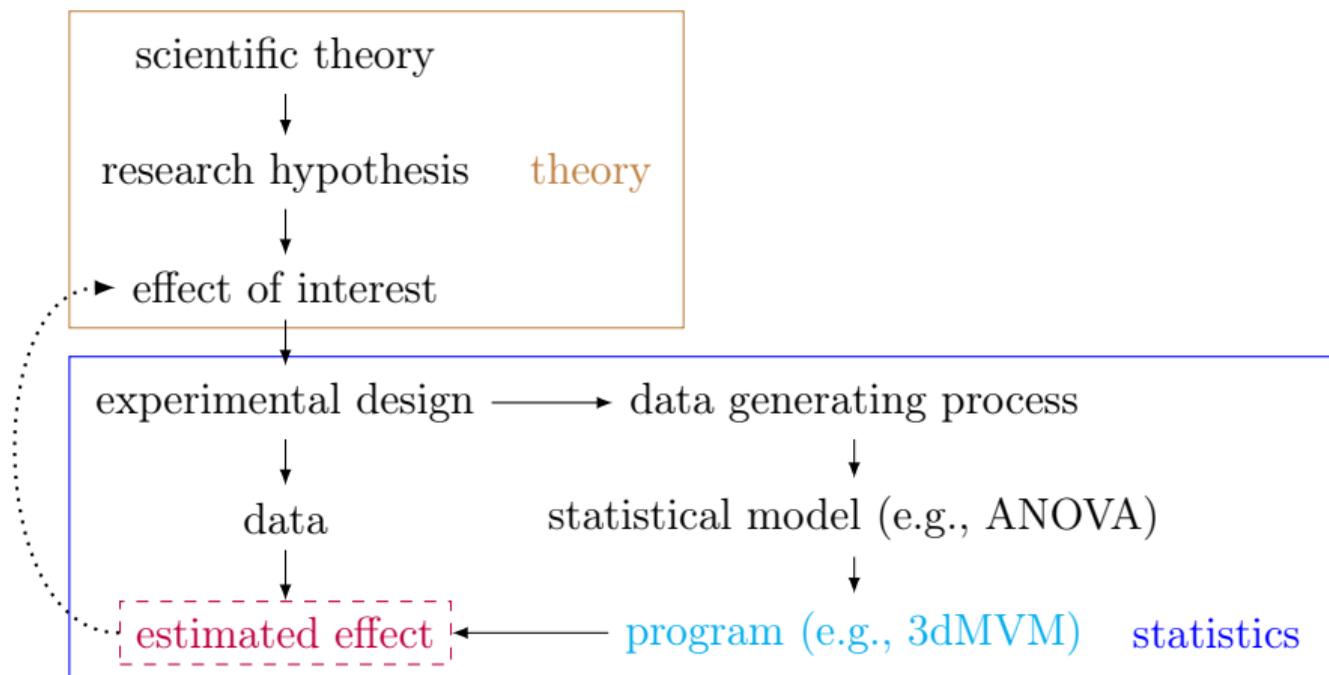
- See the forest for the trees
  - ⋆ investigation = decision based on a single p-value ("significance")?
  - ⋆ how to select variables and build a model?

Big picture 2: experiment hierarchy
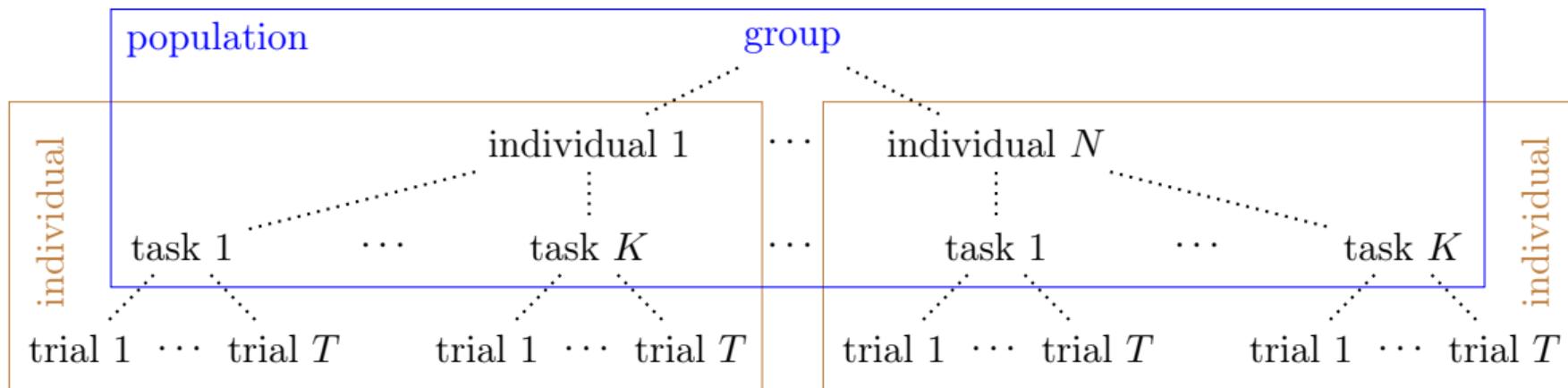
- splitting into two-stage modeling
  - ⋆ ✓ reducing conceptual & modeling complexity
  - ⋆ ✓ reducing computational hurdle
  - ⋆ ✗ no free lunch: neglecting the role of trials
    - no cross-trial variability assumption
    - uncertainty of effect estimation at the individual level
    - importance of trial sample size

Big picture 3: spatial hierarchy
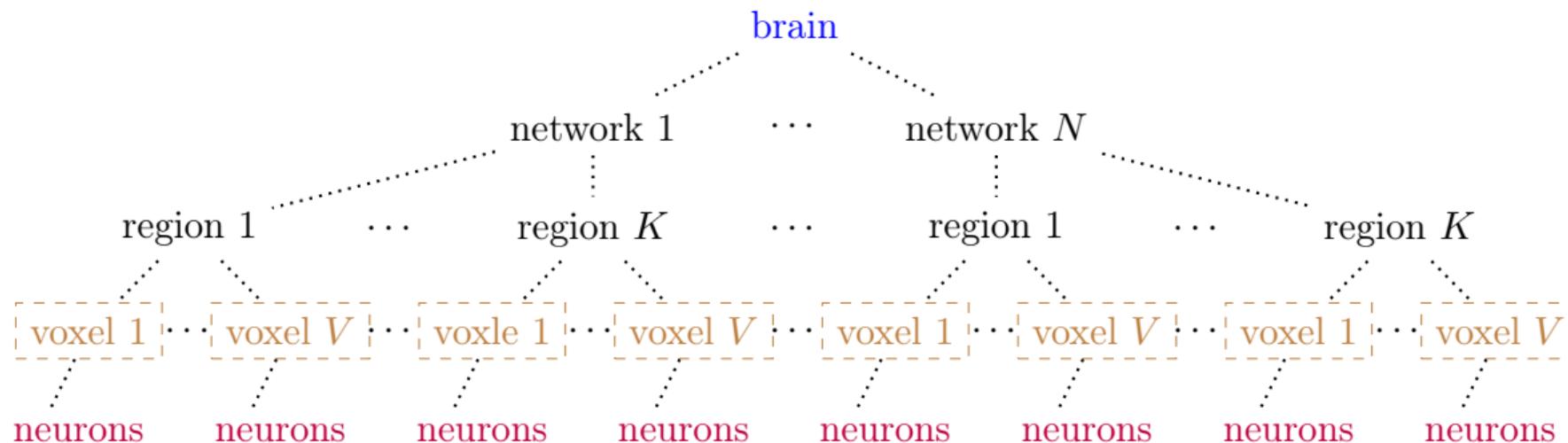
- approximation/idealization: networks, regions, voxels
- information integrity: a single integrative model
- practicality: as many models as number of voxels
  - ⋆ ✓ reducing conceptual & modeling complexity
  - ⋆ ✓ reducing computational hurdle
  - ⋆ ✗ penalty for multiplicity: type I & II
  - ⋆ ✗ estimation errors due to lack of regularization: type M & S (vs type I & II)

Why individual followed by population?

- Integrative modeling
  - ⋆ ideal: one model that incorporates all hierarchical levels
  - ⋆ reducing information loss
  - ⋆ impractical: unwieldy models and prohibitive computation cost

- Two-stage methodology
  - ⋆ individual level
    - time series regression: 3dDeconvolve
    - GLS: accounting for residual temporal correlation using ARMA(1,1) with 3dREMLfit
  - ⋆ Population level
    - response variable: individual-level effect estimates (β values) and their uncertainty info
    - predictors: experiment factors, covariates (sex, age, ...)

- Generalizability from samples to hypothetical groups/conditions
  - ⋆ two types of samples: trials and participants
  - ⋆ two types of generalization: trials → task condition; individuals → population
  - ⋆ prior assumption: cross-individual/trial variability $\sim \mathcal{N}(0, \sigma^2)$
  - ⋆ cross-trial variability: usually not properly modeled

Terminology

- Factors (discrete/categorical variables): within- vs between-individuals
  - ⋆ between-individuals (patient vs control, sex): independence
  - ⋆ within-individual (task conditions): relatedness, variance-covariance
- Quantitative (continuous) variables: within- vs between-individuals
  - ⋆ between-individuals: age (cross-sectional), brain volume
  - ⋆ within-individual: age (longitudinal), rating across conditions
- Covariates: all explanatory variables, quantitative variables, variables of no interest
- Fixed- vs random-effects
  - ⋆ fixed: constant; population level (condition, group); effects of interest
  - ⋆ random: varying; lower levels (e.g., participants, trials); exchangeable, generalizeable
- Model structure
  - ⋆ Student's t, GLM, AN(C)OVA, hierarchical models
- R notations
  - ⋆ population level: $A * B = 1 + A + B + A : B$; lower levels: (1|Subj), (1|A:Subj)
- Decision vs effect estimation
  - ⋆ estimation: accuracy vs precision

Program list: population level

| Spatial Unit	Program	Model
voxel, node, ROI massively univariate	3dttest++	t-tests, GLM
	3dMEMA	3dttest++ analog with $\beta + t$ as input
	3dMVM	GLM, AN(C)OVA
	3dLMEr	LME: hierarchical modeling
	3dMSS	multilevel smoothing splines: nonlinearity
	3dISC	inter-subject correlation: naturalistic data
	3dICC	intra-class correlation: reliability
ROI	RBA	region-based analysis: Bayesian modeling
	PTA	profile trajectory analysis: nonlinearity
	TRR	test-retest reliability

Others: 3dANOVA, 3dANOVA2, 3dANOVA3, 3dRegAna, 3dLME

Student's t-test

- One-sample: 1 group with 1 effect
  - ⋆ data: $y_i, \; i = 1, 2, ..., n$

  individual: ϵ_i

  task: μ ········· response: y_i
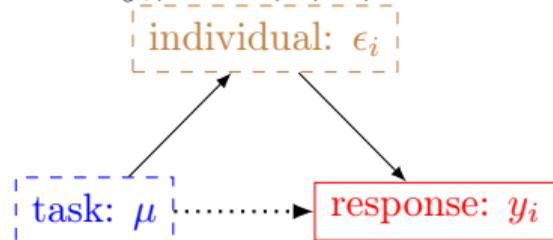
  - ⋆ model: $y_i = \mu + \epsilon_i, \; \epsilon_i \sim \mathcal{N}(0, \sigma^2)$
  - ⋆ OLS/ML estimation
    $\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i, \; \widehat{\sigma} = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \widehat{\mu})^2$
    $t(n-1)$-statistic
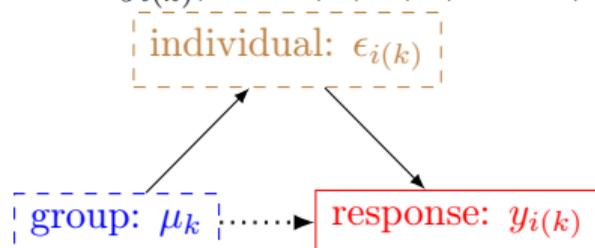 - ⋆ program in AFNI: 3dttest++

- Paired: 1 group with 2 conditions
  - ⋆ data: $(y_{i1}, y_{i2}), \; i = 1, 2, ..., n$
  - ⋆ reducing to one-sample: $y_{i1} - y_{i2}$
  - ⋆ program: 3dttest++ -paired

- Two-sample: 2 groups with 1 effect
  - ⋆ data: $y_{i(k)}, \; i = 1, 2, ..., n; \; k = 1, 2$

  individual: $\epsilon_{i(k)}$

  group: μ_k ······ response: $y_{i(k)}$

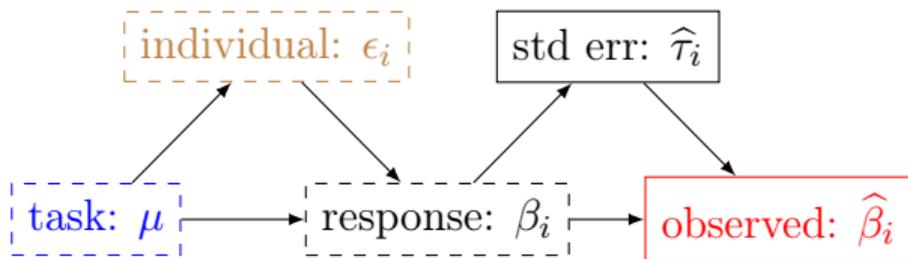 - ⋆ $y_{i(k)} = \mu_k + \epsilon_{i(k)}, \; \epsilon_{i(k)} \sim \mathcal{N}(0, \sigma^2)$
  - ⋆ special case: 1-way between-individual ANOVA; univariate GLM

- Handling missing voxel values: -zskip
  - ⋆ excluding 0 values at the voxel level
  - ⋆ z-stat instead of t: unequal # individuals across voxels

Accounting for estimation uncertainty

individual: ϵ_i std err: $\widehat{\tau}_i$

task: μ → response: β_i → observed: $\widehat{\beta}_i$

- Methodology
  - ⋆ data: $(\widehat{\beta}_i,\ t_i),\ i = 1, 2, ..., n;\ \widehat{\tau}_i = \widehat{\beta}_i/t_i$
  - ⋆ model:
    $\widehat{\beta}_i = \beta_i + \pi_i,\ \beta_i = \mu + \epsilon_i,$
    $\pi_i \sim \mathcal{N}(0,\ \widehat{\tau}_i^2),\ \epsilon_i \sim \mathcal{N}(0,\ \sigma^2)$
  - ⋆ same methodology as meta-analysis
  - ⋆ program in AFNI: 3dMEMA
  - ⋆ missing data at voxel level: -missing_data 0

- example
 - ⋆ 2 groups: patient, control
  - ⋆ input data: estimate, t-statistic

```
3dMEMA -prefix output          \
       -jobs 16                \
       -groups pat ctr         \
       -missing_data 0         \
       -set   pat              \
       p1  p1_B+tlrc  p1_T+tlrc \
       p2  p2_B+tlrc  p2_T+tlrc \
       ...
       -set   ctr              \
       c1  c1_B+tlrc  c1_T+tlrc \
       c2  c2_B+tlrc  c2_T+tlrc \
       ...
```

Chen, G., Saad, Z.S., Nath, A.R., Beauchamp, M.S., Cox, R.W., 2012. FMRI group analysis combining effect estimates and their variances. NeuroImage 60, 747–765.

GLM

- all between-individual variables: ≥ 1 groups; ≥ 0 quantitative variables
- AN(C)OVA without within-individual variables (factors, quantitative variables)
- Data at each spatial unit: $(y_i, x_{i1}, ...), \ i = 1, 2, ..., n$
- Model: $y_i = \alpha + \beta_1 x_{i1} + ... + \epsilon_i$
- Effects of interest: $\alpha, \ \beta_1, ...$
- When x_i is quantitative
  - ⋆ centering: not needed for β_i; crucial for some effects (e.g., α)
  - ⋆ linearity assumption: too strong?
- How to decide what covariates to include? More later
- Special GLMs
  - ⋆ regresion
  - ⋆ one-, two-sample t-tests
  - ⋆ AN(C)OVA w/o within-individual variables
- Programs: 3dttest++ -covariates, 3dMVM

ANOVA with within-individual factors
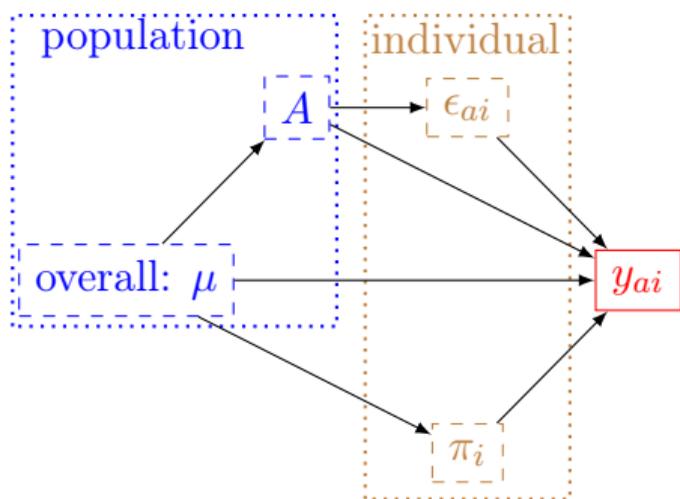
- 1-way within-individual
  ANOVA $y_{ai} = \mu + A + \pi_i + \epsilon_{ai}$
  population: μ, A; lower: π_i, ϵ_{ai}
- extension of paired t

- 2-way within-individual ANOVA:
  $$y_{abi} = \mu + A + B + A:B + \pi_i + \alpha_{ai} + \beta_{bi} + \epsilon_{abi}$$

- Traditional ANOVA
  - ⋆ economical: sums of squares
  - ⋆ 3dANOVA, 3dANOVA2, 3dANOVA3
  - ⋆ separate for each number of factors
  - ⋆ quantitative covariates: ✗

- Univariate GLM: unwieldy & problematic

- Multivariate GLM
  - ⋆ more flexible than univariate GLM
  - ⋆ missing data: ✗
  - ⋆ within-individual quantitative covariates: ✗

Chen, G., Adleman, N.E., Saad, Z.S., Leibenluft, E., Cox, R.W., 2014. Applications of multivariate modeling to neuroimaging group analysis: A comprehensive alternative to univariate general linear model. NeuroImage 99, 571–588.

AN(C)OVA approach: multivariate GLM - 3dMVM

- Example: 3 factors
  - ⋆ 1 between-individuals - group (Grp): patient & control
  - ⋆ 2 within-individual - intensity (Int): hi & lo; condition (Cond): Pos, Neg, Neu
  - ⋆ $2 \times 2 \times 3$ mixed ANOVA

```
3dMVM -prefix    Output  -jobs 16                                              \
      -bsVars  'Grp'  -wsVars    'Int*Cond'                                     \
      -num_glt  4                                                              \
      -gltlabel 1 Pat_Pos      -gltCode 1 'Grp : 1*Pat Cond : 1*Pos'           \
      -gltLabel 2 Ctl_Pos-Neg  -gltCode 2 'Grp : 1*Ctl Cond : 1*Pos -1*Neg'    \
      -gltlabel 3 Grp_Pos-Neg  -gltCode 3 'Grp : 1*Ctl -1*Pat Cond : 1*Pos -1*Neg' \
      -gltlabel 4 hi-lo.P-N    -gltCode 4 'Int : 1*hi  -1*lo  Cond : 1*Pos -1*Neg' \
      -dataTable                                                               \
      Subj        Grp         Int         Cond        ImputFile                \
      s1          ctl         hi          Pos         $1_Pos.nii               \
      s1          ctl         hi          Neg         S1_Neg.nii               \
      s1          ctl         hi          Neu         $1_Neu.nii               \
      s2          Pat         lo          Pos         $2_Pos.nii               \
      s2          Pat         lo          Neg         s2_Neg.nii               \
      S2          Pat         lo          Neu         s2_Neu.nii               \
      ...
```

AN(C)OVA: hierarchical modeling - 3dLMEr

$$y_{abi} = \mu + A + B + A:B + \delta_i + \alpha_{ai} + \beta_{bi} + \epsilon_{abi}$$

```
3dLMEr -prefix   Output  -jobs 16                                          \
       -model    'Int*Cond*Grp+(1|Subj)+(1|Int:Subj)+(1|Cond:Subj)' \
       -gltCode Pat_Pos     'Grp : 1*Pat Cond : 1*Pos'                     \
       -gltCode Ctl_Pos-Neg 'Grp : 1*Ctl Cond : 1*Pos -1*Neg'             \
       -gltCode Grp_Pos-Neg 'Grp : 1*Ctl -1*Pat Cond : 1*Pos -1*Neg' \
       -gltCode hi-lo.P-N   'Int : 1*hi  -1*lo  Cond : 1*Pos -1*Neg' \
       -dataTable                                                          \
       Subj       Grp          Int          Cond       ImputFile           \
       s1         ctl          hi           Pos        $1_Pos.nii          \
       s1         ctl          hi           Neg        S1_Neg.nii          \
       s1         ctl          hi           Neu        $1_Neu.nii          \
       s2         Pat          lo           Pos        $2_Pos.nii          \
       s2         Pat          lo           Neg        s2_Neg.nii          \
       S2         Pat          lo           Neu        s2_Neu.nii          \
       ...
```

blog post: https://tinyurl.com/4cx47ew9

Chen, G., Saad, Z.S., Britton, J.C., Pine, D.S., Cox, R.W., 2013. Linear mixed-effects modeling approach to FMRI group analysis. NeuroImage 73, 176–190.

Inter-subject correlation analysis

- Naturalistic scanning
  - ⋆ task-related FMRI: remote from real life
  - ⋆ movie watching, speech/music listening

- Data structure
  - ⋆ correlations: r_{ij}
  - ⋆ n individuals $\to \frac{1}{2}n(n-1)$ ISC pairs
  - ⋆ intricate correlation structure
  - ⋆ how to disentangle the 'messy' structure?

- Previous methods
  - ⋆ random shuffling time series: ?
  - ⋆ one individual vs sum of all others: ?
  - ⋆ permutations, bootstrapping

$n = 5$ individuals: 10 ISC pairs

	z_{21}	z_{31}	z_{41}	z_{51}	z_{32}	z_{42}	z_{52}	z_{43}	z_{53}	z_{54}
z_{21}	1	ρ	ρ	ρ	ρ	ρ	ρ	0	0	0
z_{31}	ρ	1	ρ	ρ	ρ	0	0	ρ	ρ	0
z_{41}	ρ	ρ	1	ρ	0	ρ	0	ρ	0	ρ
z_{51}	ρ	ρ	ρ	1	0	0	ρ	0	ρ	ρ
z_{32}	ρ	ρ	0	0	1	ρ	ρ	ρ	ρ	0
z_{42}	ρ	0	ρ	0	ρ	1	ρ	ρ	0	ρ
z_{52}	ρ	0	0	ρ	ρ	ρ	1	0	ρ	ρ
z_{43}	0	ρ	ρ	0	ρ	ρ	0	1	ρ	ρ
z_{53}	0	ρ	0	ρ	ρ	0	ρ	ρ	1	ρ
z_{54}	0	0	ρ	ρ	0	ρ	ρ	ρ	ρ	1

Chen, G., Shin, Y.-W., Taylor, P.A., Glen, D.R., Reynolds, R.C., Israel, R.B., Cox, R.W., 2016. Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. NeuroImage 142, 248–259.

Inter-subject correlation analysis


```
3dISC  -prefix ISC    -jobs 16                        \
       -model 'grp+(1|Subj1)+(1|Subj2)'              \
       -gltCode ave      '1 0 -0.5'                   \
       -gltCode G11      '1 1 0'                       \
       -gltCode G12      '1 0 1'                       \
       -gltCode G22      '1 -1 -1'                     \
       -gltCode G11vG22  '0 2 1'                       \
       -gltCode G11vG12  '0 1 -2'                      \
       -gltCode G12vG22  '0 1 2'                       \
       -gltCode ave-G12  '0 0 -1.5'                    \
       -dataTable                                      \
       Subj1 Subj2   grp      InputFile               \
       s1    s2      G11      s1_2+tlrc               \
       s1    s3      G11      s1_3+tlrc               \
       ...
       s1    s24     G12      s1_25+tlr              \
       s1    s25     G12      s1_26+tlr              \
       ...
       s25   s24     G22      s25_26+tlr             \
       s25   s25     G22      s25_27+tlr
```
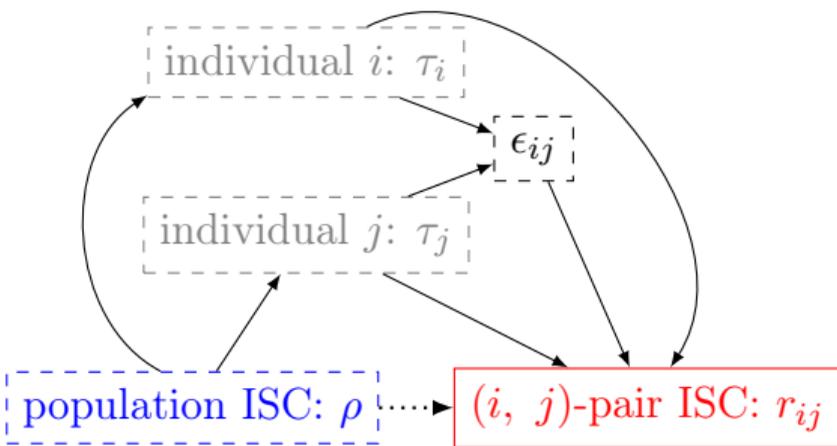
- Modeling approach
  - ★ n individuals → $\frac{1}{2}n(n-1)$ ISC pairs
  - ★ each pair: $r_{ij}, \ i \neq j$
  - ★ model: $z_{ij} = \mu + \tau_i + \tau_j + \epsilon_{ij}$
  - ★ program: 3dISC (variation of 3dLMEr)

Chen, G., Taylor, P.A., Shin, Y.-W., Reynolds, R.C., Cox, R.W., 2017. Untangling the relatedness among correlations, Part II: Inter-subject correlation group analysis through linear mixed-effects modeling. NeuroImage 147, 825–840.

Capturing nonlinearity

- Handling a quantitative variable (e.g. age)
    - ⋆ linearity by default in common practice
    - ⋆ alternative: polynomials

- Linearity and polynomials
    - ⋆ + Simple and economical
    - ⋆ + Fitness: $(k-1)$th order for k points
    - ⋆ − Difficult to guess the order
    - ⋆ − Maladaptive for whole brain
    - ⋆ − Assess difference between 2 curves?
    - ⋆ − Non-locality, instability

- Extending polynomial fitting
    - ⋆ Fitting with splines
    - ⋆ Similar to HDR estimation at individual level
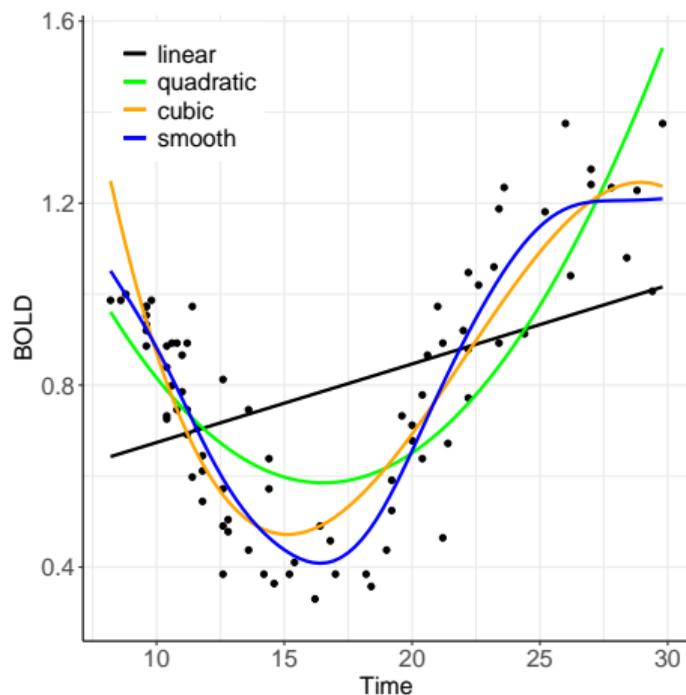
Fitting nonlinearity with splines

- Thin plate splines
  - ⋆ incremental nonlinearity
  - ⋆ penalty against roughness: quadratic+
  - ⋆ unique: baseline and linearity
  - ⋆ programs in AFNI
    - voxel/node-wise: 3dMSS
    - region-level: PTA
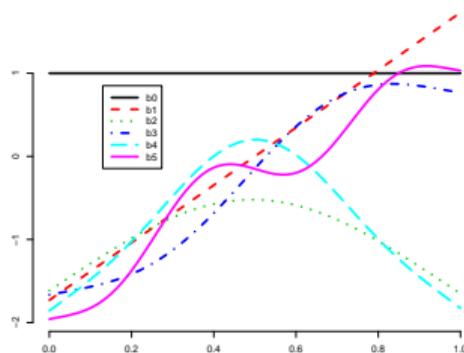
demo script

```
3dMSS -prefix MSS -jobs 16                        \
      -mrr 's(age,k=10)+s(age,k=10,by=grp)' \
      -qVars 'age'                                \
      -prediction @pred.txt                       \
      -dataTable  @data.txt
```

separation of linearity:

```
-mrr 'grp*age+s(age,k=10,m=c(2,0))+
      s(age,k=10,by=grp,m=c(2,0))'
```

data.txt   ### pred.txt

Subj	grp	age	InputFile		label	grp	age
S1	-1	17	S1.nii		p.t1	-1	11
S2	1	12	S2.nii		p.t2	-1	11.25
...					...		

Chen, G., Nash, T.A., Cole, K.M., Kohn, P.D., Wei, S.-M., Gregory, M.D., Eisenberg, D.P., Cox, R.W., Berman, K.F., Shane Kippenhan, J., 2021. Beyond linearity in neuroimaging: Capturing nonlinear relationships with application to longitudinal studies. NeuroImage 233, 117891.

Nonlinear modeling: profile tracking analysis

- Layer fMRI data
  - ★ 16 individuals
  - ★ 11 layers: BOLD/VASO per individual
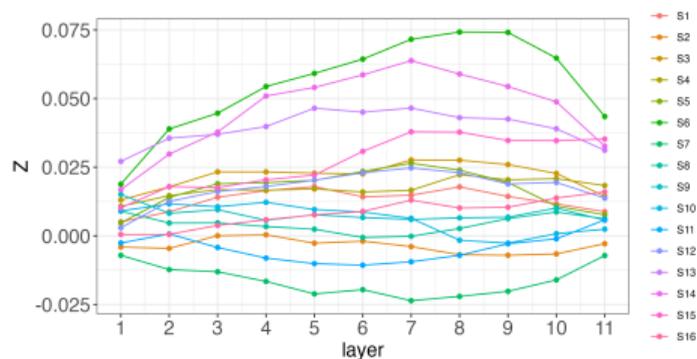
input data

```
Subj layer Z
S1   1 0.00511335
S1   2 0.00875758
S1   3 0.0140701
...
```

prediction

```
label layer
L1    1
L2    2
...
```
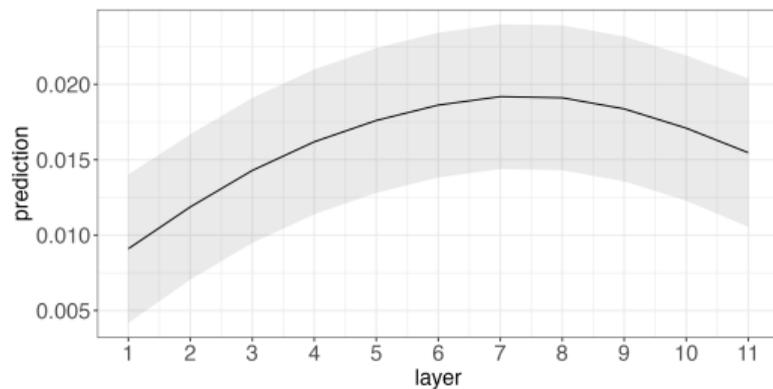


```
PTA -prefix output  -Y Z                        \
    -input input.txt                            \
    -model 's(layer,k=10)+s(Subj,bs="re")'      \
    -vt Subj 's(Subj)'                          \
    -prediction pred.txt
```

Blog post: https://shorturl.at/hwFK9

Role of sample sizes
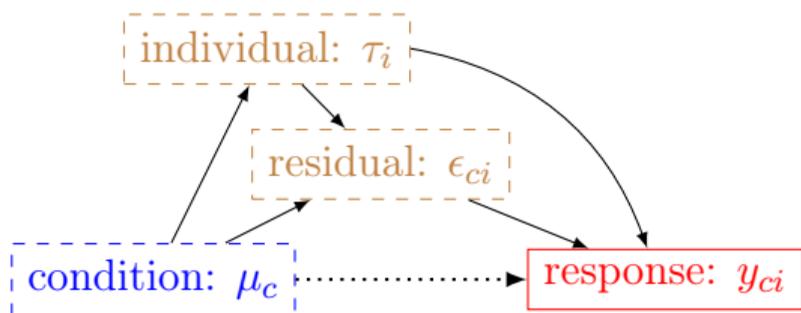
- Participant sample size N
    - ★ data at condition level: $y_{ci}, \; c = 1, 2, ...$
    - ★ model: $y_{ci} = \mu_c + \tau_i + \epsilon_{ci}$
    - ★ efficiency $\sim \sqrt{N}$
    - ★ trial sample size does not matter?

Role of sample sizes

individual: τ_{ci}

residual: ϵ_{cit}

condition: μ_c $\cdots\cdots\cdots\blacktriangleright$ response: y_{cit}
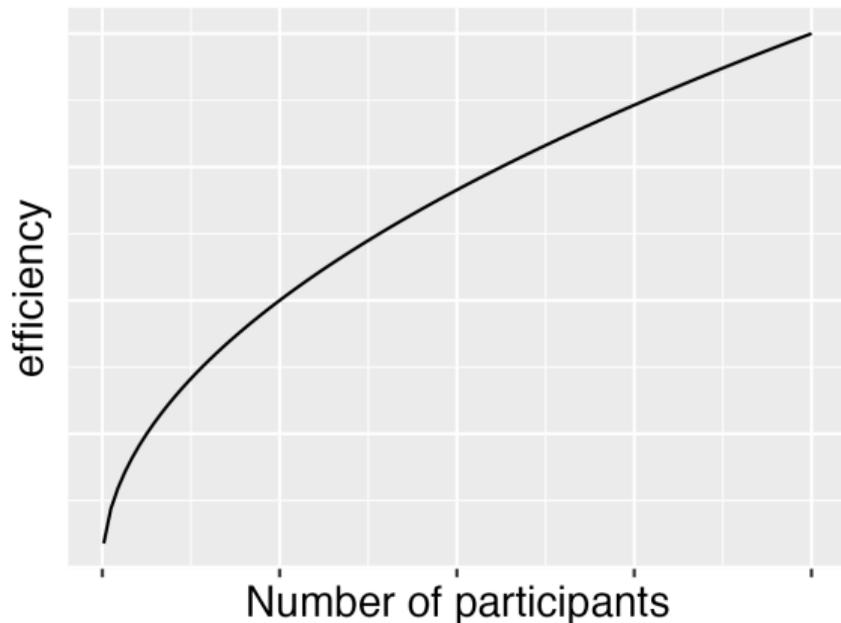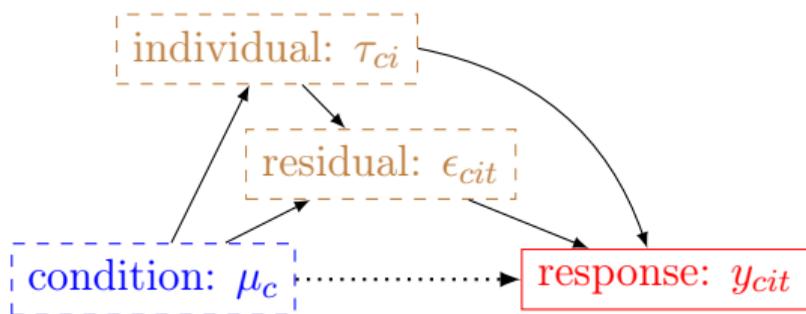
- Trial sample size T
  - ⋆ data at trial level: $y_{cit}, \; c = 1, 2, \ldots$
  - ⋆ model: $y_{cit} = \mu_c + \pi_{ci} + \epsilon_{cit}$
    $\pi_{ci} \sim \mathcal{N}(0, \; \sigma_\pi^2), \; \epsilon_{cit} \sim \mathcal{N}(0, \; \sigma_\epsilon^2)$
  - ⋆ efficiency hyperbolically related to S, T
  - ⋆ $R_v = \sigma_\epsilon / \sigma_\pi \gg 1$: large cross-trial var.
  - ⋆ T almost as important as S
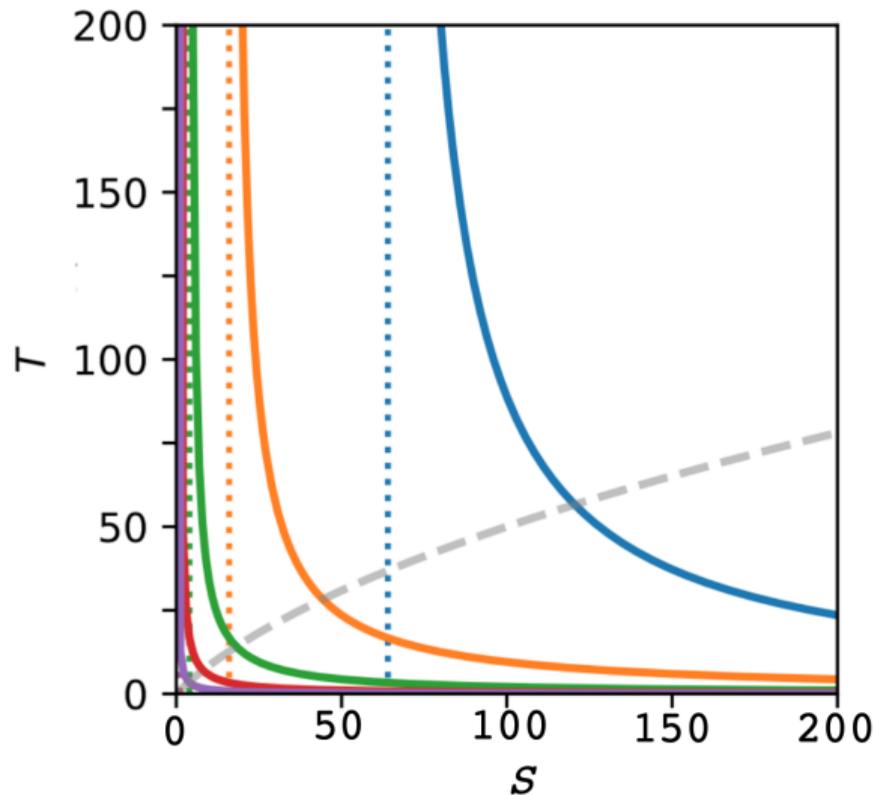
Chen, G, Taylor, PA, Haller, SP, Kircanski, K, Stoddard, J, Pine, DS, Leibenluft, E, Brotman, MA, Cox, RW, 2018. Intraclass correlation: Improved modeling approaches and applications for neuroimaging. Human Brain Mapping 39, 1187–1206.

Simple size considerations

- Difficulty in estimating sample sizes
  - ⋆ effect sizes usually not reported
  - ⋆ results dichotomized at peak voxels
  - ⋆ region-specific: substantial variability across regions
  - ⋆ current power analysis analysis tools
    - solely focusing on participants
    - pacifiers?

- Suggestions
  - ⋆ gather information from literature
  - ⋆ balance trial and participant samples
    - hyperbolic relationship: leveraging between the two in both efficiency and financial cost
  - ⋆ Interactions
    - 2-way interactions: at least a few times more samples than main effects (> 100)
    - 3-way interactions: challenging (> 1000)

Covariate selection

- Statistical modeling
  - ⋆ One model for all effects?
    - step-up/down, statistical metrics (p-values, R^2, information criteria)
  - ⋆ Two goals
    - prediction: forecasting future responses
    - inference: estimating the impact of a predictor on response → causal effects

- Data structure for each adult participant
  - ⋆ response variable: short-term memory (STM)
  - ⋆ predictor: <u>voxel-level</u> gray matter density (GMD)
  - ⋆ 5 covariates
    - 2 between-individual factors: sex, APOE genotype
    - 3 quantitative variables: age, weight, intracranial volume (ICV)

- Questions
  - ⋆ OK to switch predictor and response variable?
  - ⋆ OK to include all covariates?
  - ⋆ are all estimated effects interpretable?
  - ⋆ could more variables have been collected: height, sleep data?

Directed Acyclic Graph (DAG)

- Express prior knowledge or hypothesized relations among variables with graphs
  - ★ nodes: variables; arrows: directional influence
  - ★ directed acyclic graph (DAG): a common language of graphical representation
  - ★ jargon: causal path, front/back door, minimally sufficient set, ...
- 3 basic types

(A) confounder/fork (noncausal & open)

(B) collider (noncausal & closed)

(C) mediator/pipe (causal & open)
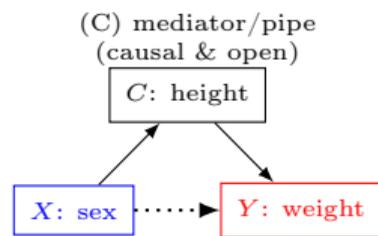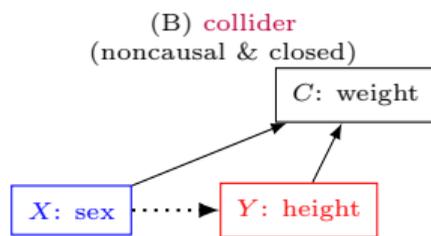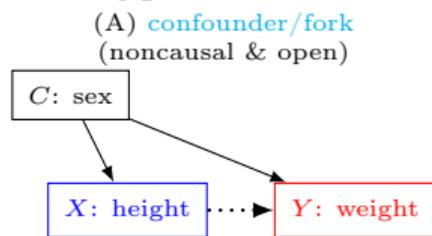
- 4 auxiliary types: covariate influences either predictor or response, but not both

(A) child/descendant of predictor
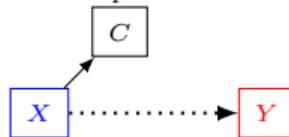
(B) child/descendant of response

(C) parent/ancestor of predictor
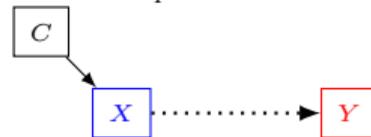
(D) parent/ancestor of response
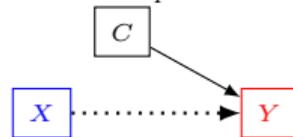
- Role of statistical metrics
  - ★ correlation of 0.8 btw C and Y (or X); multicollinearity

Quiz

age/site relative to sex/task & BOLD?

head size relative to sex & BOLD
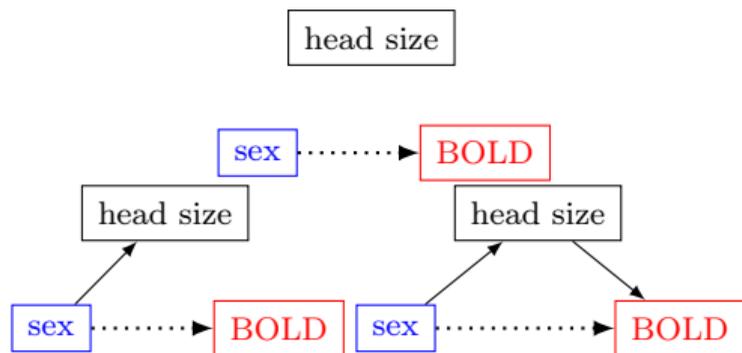
slow drift relative to task & BOLD

head motion relative to task & BOLD
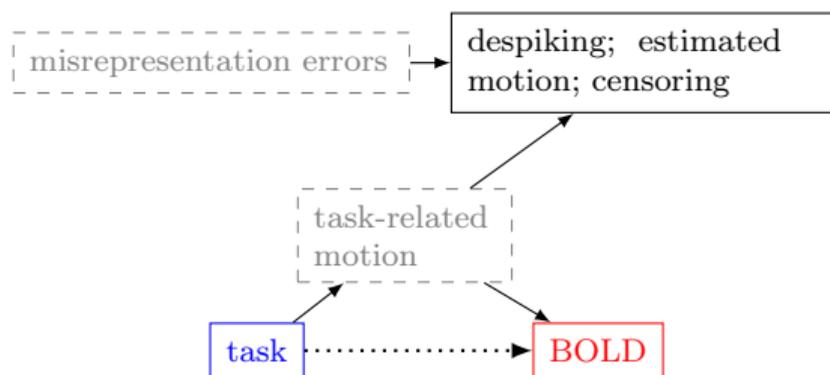
Censoring: data points or participants?

Revisiting motivating example

- Data structure for each adult participant
  - ⋆ Response variable: short-term memory (STM)
  - ⋆ Predictor: voxel-level gray matter density (GMD)
  - ⋆ 5 covariates
    - 2 between-individual factors: sex, APOE genotype
    - 3 quantitative variables: age, weight, intracranial volume (ICV)
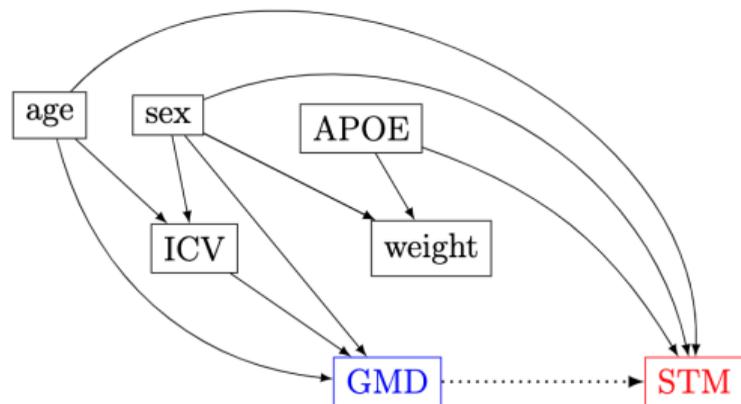
- Addressing four questions
  - ⋆ switch predictor and response variable?
  - ⋆ include all covariates?
  - ⋆ are all estimated effects interpretable?
  - ⋆ could more variables have been collected? height, sleep data?

Summary: variable selection

- DAGs for model selection
  - ⋆ confounder: ✓; collider: ✗; mediator: ⚠
  - ⋆ ancestors/descendants: only condition on ancestors of response
- Suggestions
  - ⋆ drawing DAGs
    - experiment planning & modeling
    - all (including latent) variables
  - ⋆ modeling
    - each effect may require a separate model
    - centering, interactions, nonlinearity
  - ⋆ reporting
    - state effects of interest
    - present DAGs when necessary: transparency
    - avoid listing all estimated effects from a model (table 2 fallacy)
    - avoiding dichotomization: highlight-but-not-hide
  - ⋆ motion/physiological contamination: be cautious

BOLD response: standard approach

- Canonical: shape-fixed HRF
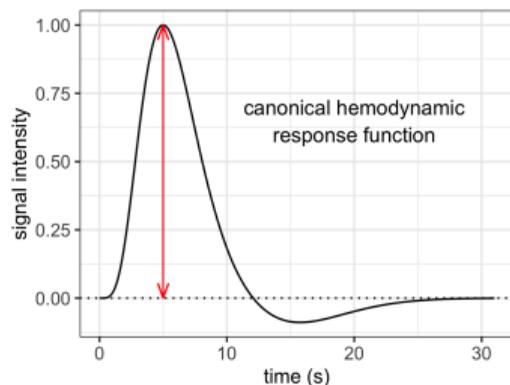
- $h(t) = 5.7t^5 e^{-t}/\Gamma(6) - 0.95t^{15} e^{-t}/\Gamma(16)$
- 2 phases: overshoot & undershoot
- overshoot peaks @ 5s
- overshoot / overall duration: 12 / 32s
- undershoot depth: 9% of peak; no initial dip

- Benefit in modeling: widely adopted
  - complexity reduction: 1D → 0D (peak height)
  - simplicity: one β per response/condition

- Empirical BOLD response profile
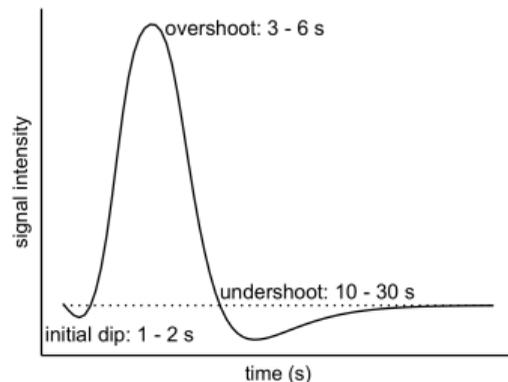
- 3 phases: initial dip, overshoot & undershoot
- large variability (eg Handwerker et al 2004)

- Issues with canonical HRF
  - seeing what one wanted to see
  - inflexible: maladaptive to shape variations
  - lost details: peak location, undershoot, ...
  - info loss: inaccuracies & distortion

BOLD response: other approaches

- Adjusting canonical HRF
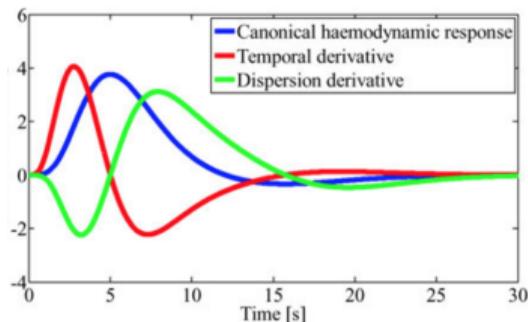
  - ⋆ 1 or 2 more bases: time derivative, dispersion
  - ⋆ increased adaptivity: peak location & width
  - ⋆ improved model fit
  - ⋆ auxiliary info abandoned or rarely utilized
  - ⋆ variability ignored: undershoot, initial dip

- Estimating HDRs at individual level
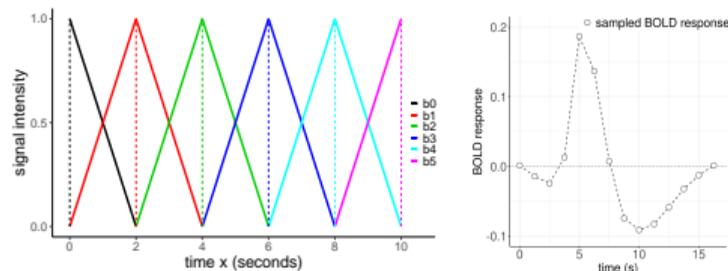
  - ⋆ piece-wise linear splines: tents/sticks, FIR
    3dDeconvolve -stim_times 1 stim.1D
    'TENT(2,16,8)'
  - ⋆ estimated HDR: at sampled data points
  - ⋆ shape info: sampled HDR vs 0D (scalar)
  - ⋆ more accurate: data-driven
  - ⋆ weaker assumption: pure morphology vs peak
  - ⋆ challenging for trial-level modeling
  - ⋆ complication: dealing with HDR samples
  - ⋆ sporadically adopted in neuroimaging

Estimating HDRs: population level

individual-level HDR — smooth splines — smooth fitting
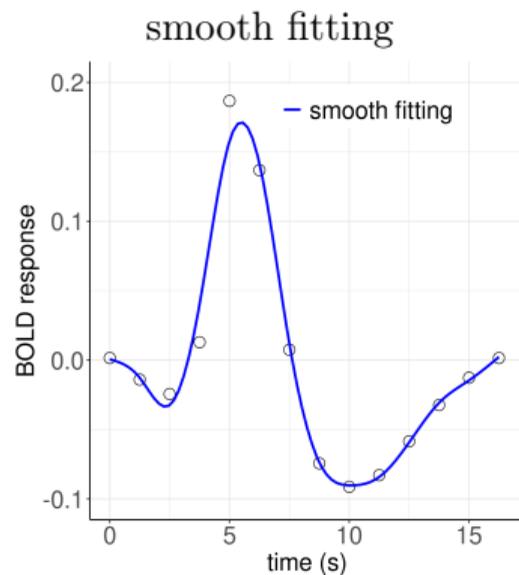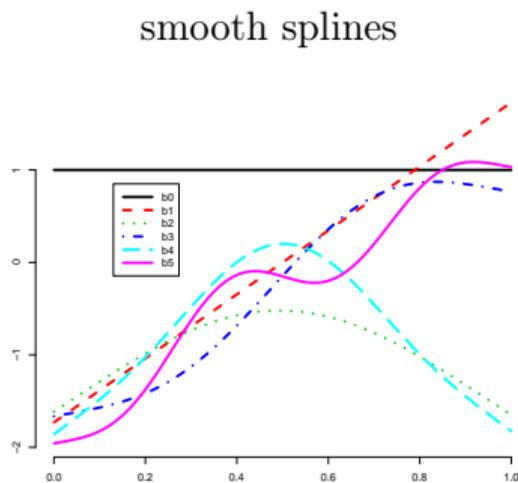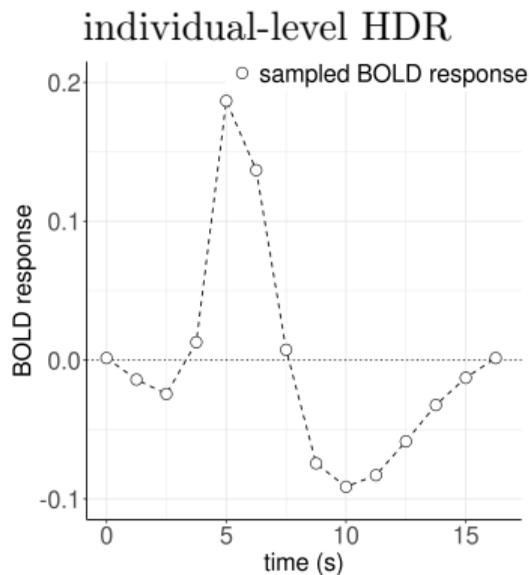
- Implementation: 3dMSS in AFNI

Estimating HDRs: applied to real data
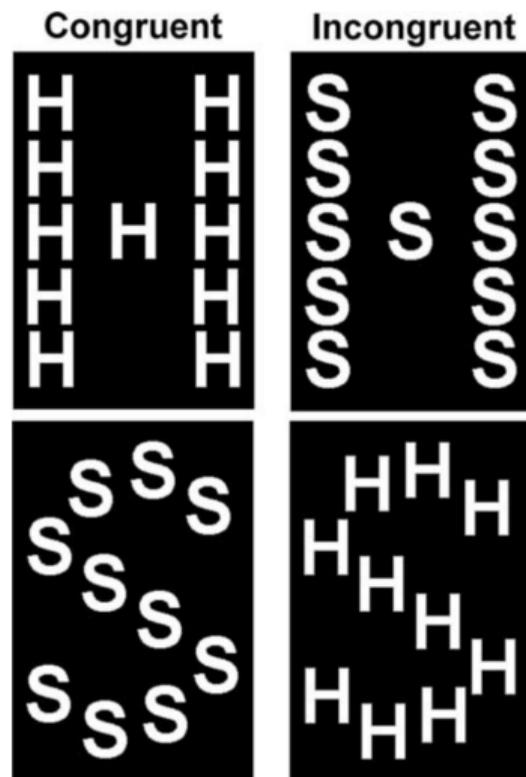
- Experiment: 2×2 fMRI
  - ⋆ 2 conditions: attention – congruent & incongruent
  - ⋆ 2 groups: 44 HV & 43 BP
  - ⋆ 288 trials per condition; TR: 1.25 s

- Individual: 2 approaches
  - ⋆ canonical HRF: 1 scalar per condition
  - ⋆ estimated HDRs: sampled at 14 time points -
    stim_times ... tent(-2.5,13.75,14)

- Population: 2 approaches
  - ⋆ canonical HRF: 2×2 ANOVA – 3dMVM
  - ⋆ estimated HDRs: smooth splines – 3dMSS

Model comparisons: overall

Much higher sensitivity by estimated HDRs
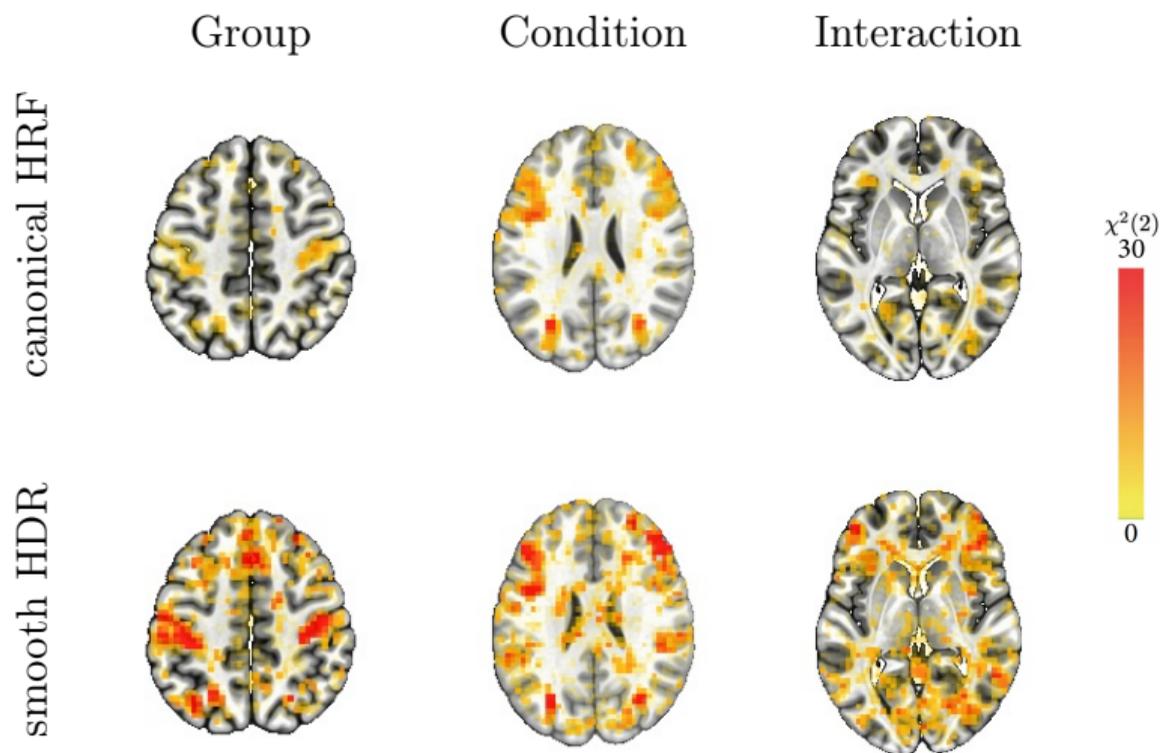
Zooming in: weaker evidence with canonical HRF
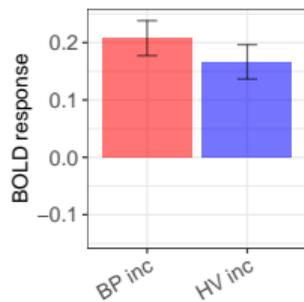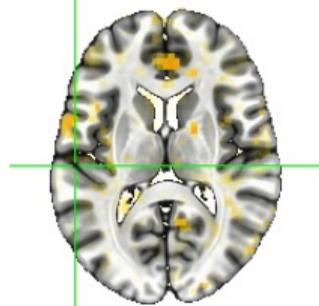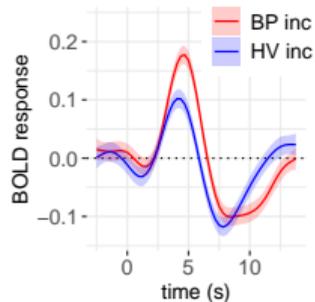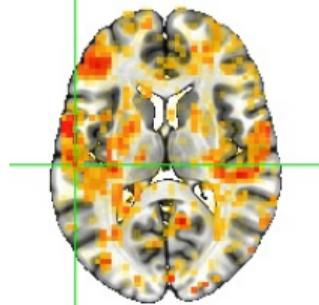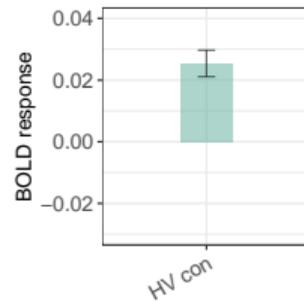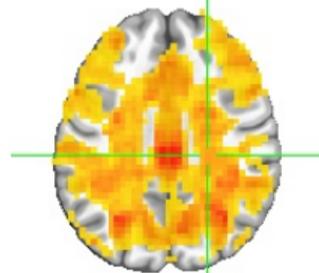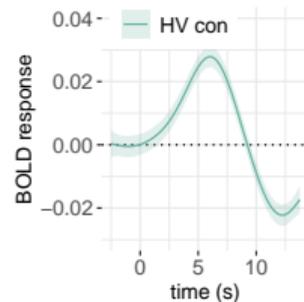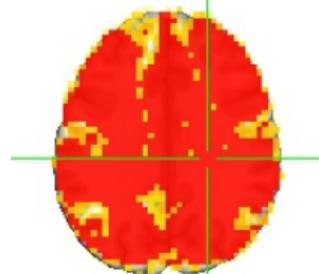
HDR estimation: inter-individual variability

different peak & peak/nadir ratio

HDR: cross-region variability
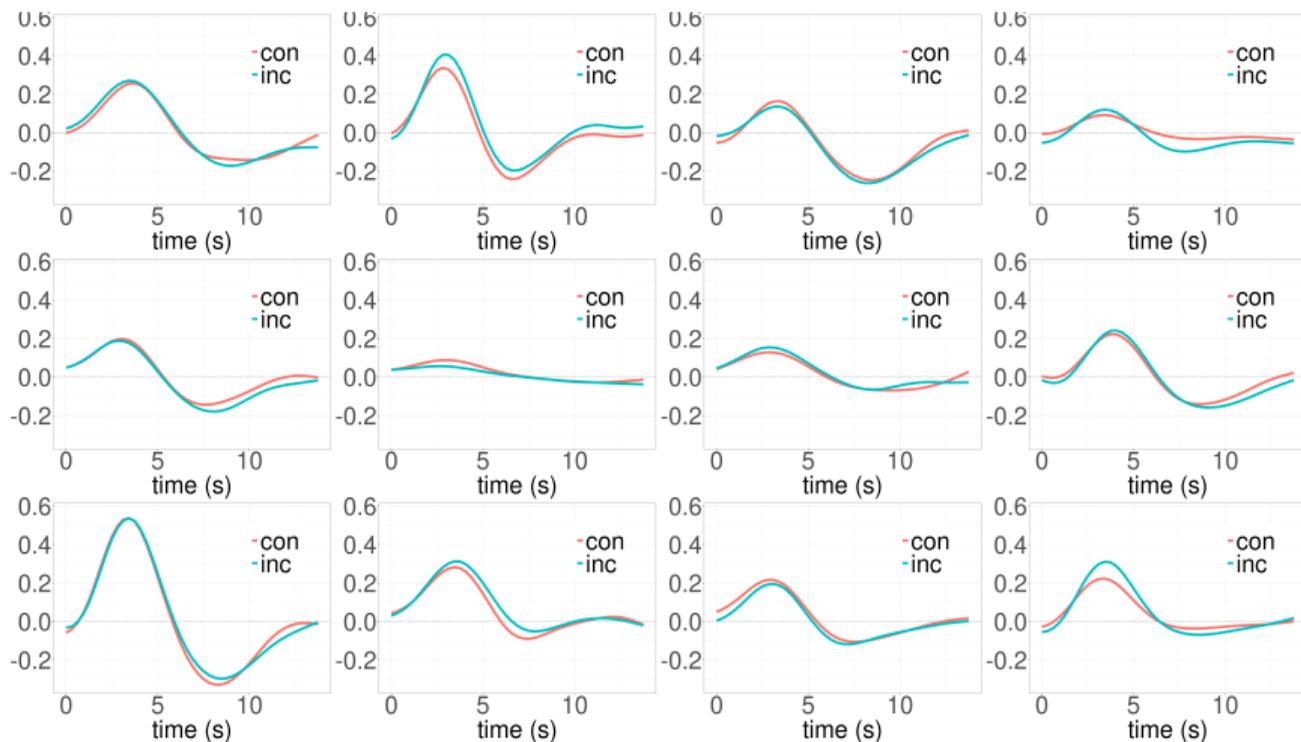
canonical HRF lacking adaptivity
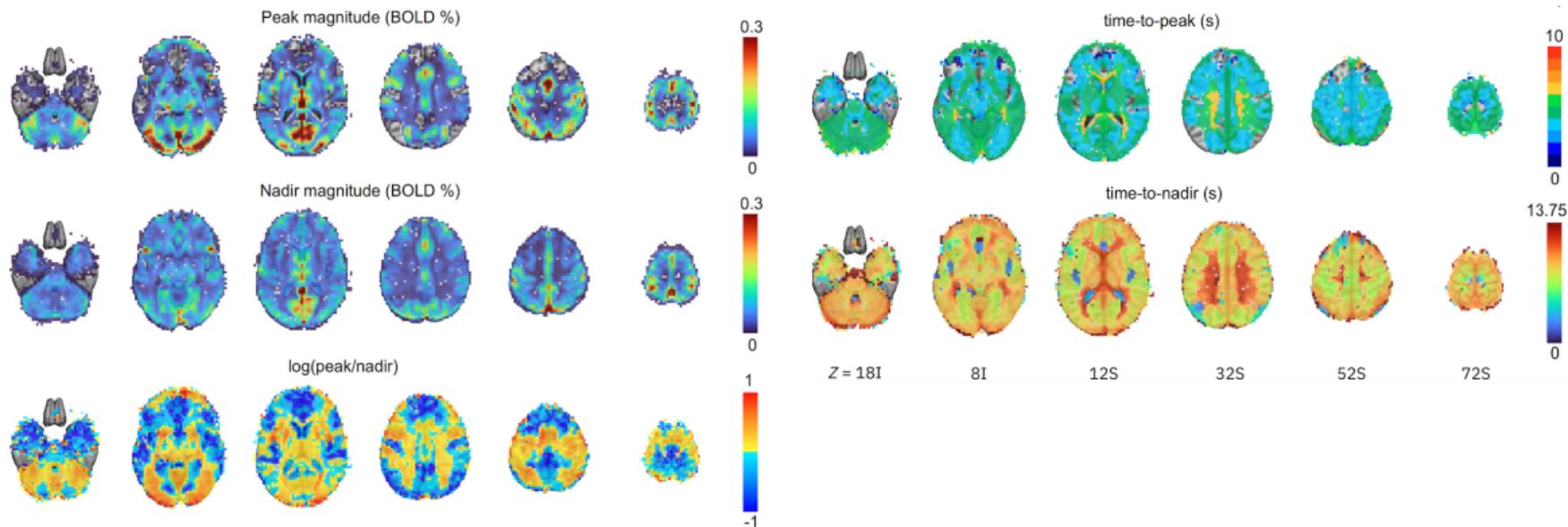
peak, nadir & their ratio

peak & nadir locations

Summary: capturing HDR profiles

```
3dMSS -prefix output -jobs 16                          \
        -lme 's(TR,k=10)+s(TR,k=10,by=group)'          \
        -ranEff 'list(Subj=~1)'                        \
        -qVars 'TR,group'                              \
        -prediction @HDR.table                         \
        -dataTable  @data.table
```

data.table

```
Subj group TR   InputFile
 s1    1    0   s1.Inc.b0.nii
 s1    1    1   s1.Inc.b1.nii
 s1    1    2   s1.Inc.b2.nii
...
```

HDR.table

```
label   group    TR
p.t1       1    0.00
p.t2       1    0.25
   ...
c.t1      -1    0.00
c.t2      -1    0.25
...
```

- Shape variability
  - ⋆ large: across regions
  - ⋆ subtle: groups/tasks/individuals

- accuracy/precision/sensitivity
  - ⋆ canonical HRF < esimated HDR
  - ⋆ visualization: scalar vs full HDR

- Program: 3dMSS in AFNI

- Blog post:
  https://shorturl.at/ahmp2

Chen, G., Taylor, P.A., Reynolds, R.C., Leibenluft, E., Pine, D.S., Brotman, M.A., Pagliaccio, D., Haller, S.P., 2023. BOLD response is more than just magnitude: improving detection sensitivity through capturing hemodynamic profiles. NeuroImage 277, 120224.

Multiplicity - NARPS: what to conclude?

- long history of emphasizing stringency: controlling false positives
  - ⋆ salmon (2010), cluster failure (2016), NARPS (2020)
  - ⋆ current result reporting: permutations, voxel-level p of 0.001 (Monte Carlo simulations, random field theory)

- NARPS: Neuroimaging Analysis Replication and Prediction Study
  - ⋆ ∼ 70 teams analyzing same dataset
  - ⋆ requirement: reporting dichotomized decisions on 9 hypotheses

- 2 conclusions: half full, half empty?
  - ⋆ reproducibility crisis: "sizeable variation"
  - ⋆ largely consistent results: "significant consensus"

- major points
  - ⋆ can individual studies be decisive/conclusive?
  - ⋆ does result reporting have to be dichotomized?
  - ⋆ OTOH, the intention of imposing strict multiple testing adjustment is to improve reproducibility, but the emphasis on rigor → information loss & reproducibility crisis?
  - ⋆ scientific investigation: decision-making vs quantification?

# Info integrity

- Voxelwise visualization
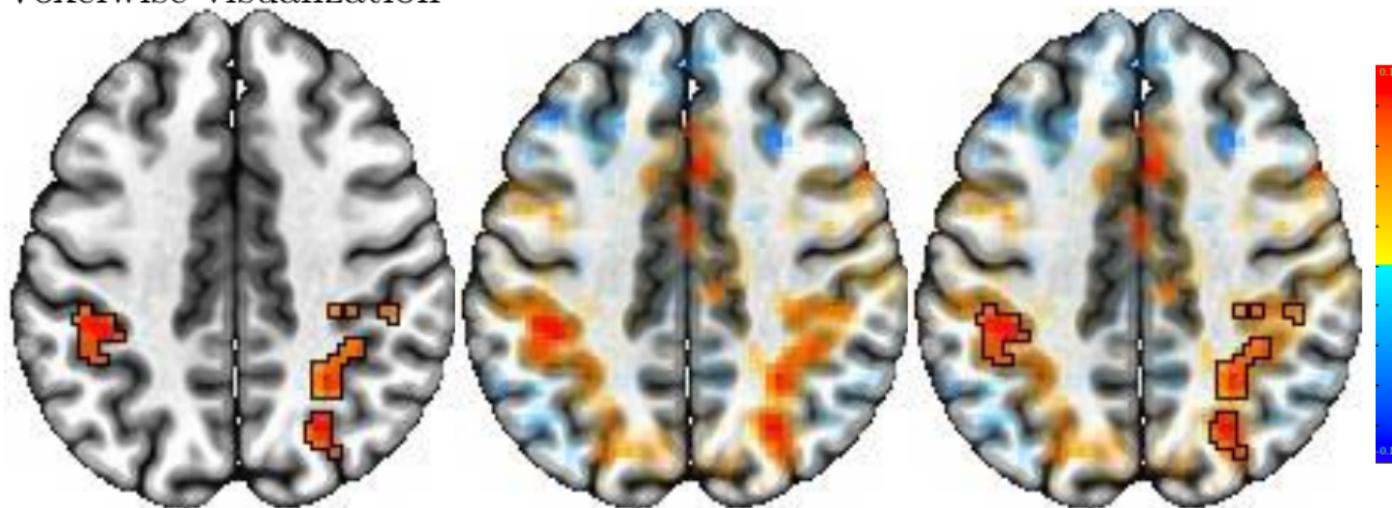
- Tabulation: problematic
  - ⋆ selection bias: artificial dichotomization (cf *t*-test example)
  - ⋆ huge info reduction/loss: cluster ⇒ peak voxel ⇒ region
  - ⋆ unsuitable for meta-analysis

- Optimal meta-analysis: full reporting & data sharing

Multiplicity

- Definition: simultaneous inferences (key: simultaneity!)
  - ⋆ M1: inferences in a study (e.g., voxels, regions, correlation matrix)
  - ⋆ M2: searching a parameter space (e.g., processing/modeling options/pipelines)
  - ⋆ M3: all similar studies (e.g., unipolar depression, ABCD, UK Biobank)
  - ⋆ M4: all studies in a journal, a field (e.g., neuroimaging) or entire history

- Multiple testing adjustment: classical solutions
  - ⋆ focus on overall statistical evidence, but not on effect magnitude, uncertainty, voxel-level statistics
  - ⋆ Penalize/dilute statistical evidence
    - FWE: adjusting for overall false positive rate: leveraging local (not global) spatial relatedness based on infinite results of pure noise; cluster
    - FDR: controlling for overall false discovery rate with current result; voxels (no spatial consideration)

- Issues
  - ⋆ heavy penalty: losing the sight of the whole hierarchical structure
  - ⋆ FWE: discrimination against small regions
  - ⋆ dichotomization
 - ⋆ disconnection with anatomy

Multiplicity: conventional approach

- NARPS: 87 ROIs, 47 individuals (equal range group); effect: gain
- Massively univariate analysis
  - ⋆ Strong prior: uniform distribution; only local (not global) relationship
  - ⋆ More reasonable prior: normal distribution
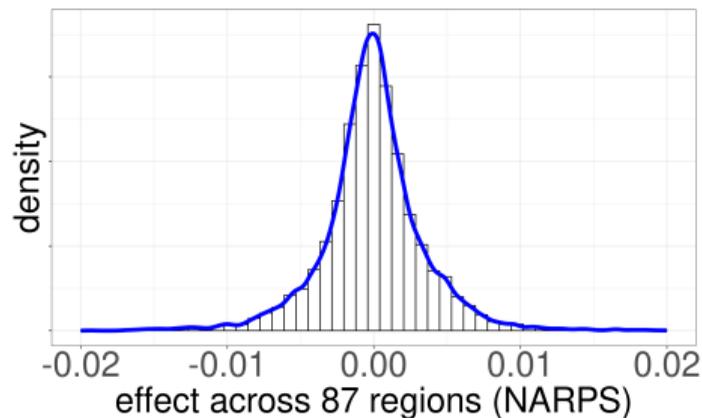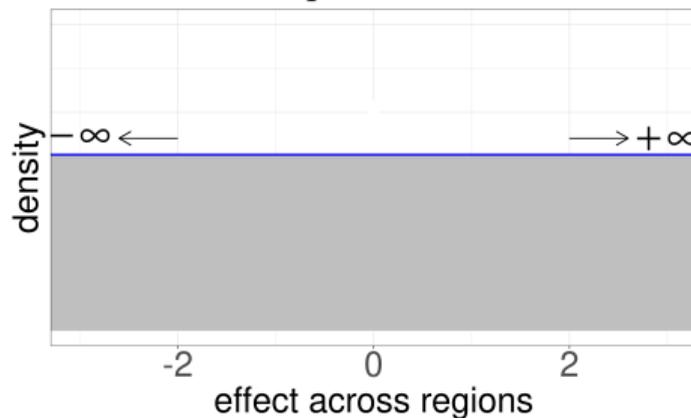
- Consequence of an unreasonable prior (uniform)
  - ⋆ Model quality: compromised
  - ⋆ Penalty from multiple testing

Handling multiplicity through incorporating spatial hierarchy
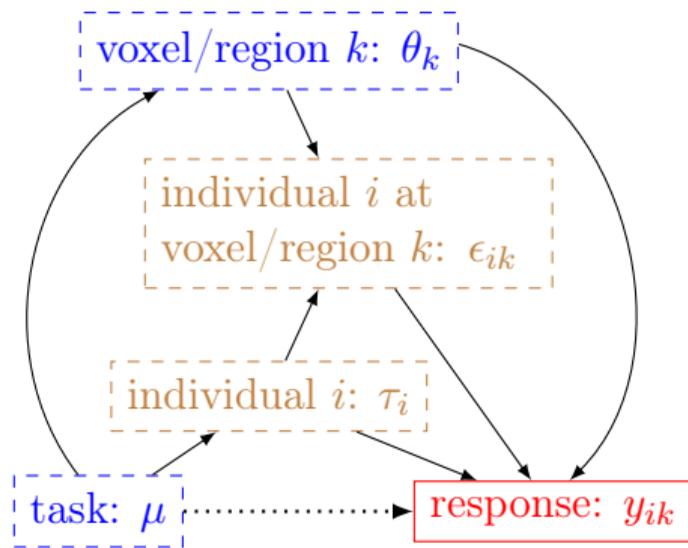
- Mass univariate (1-sample t):

$$y_{i1} = \mu + \epsilon_{i1}$$
$$y_{i2} = \mu + \epsilon_{i2}$$
...
$$y_{iK} = \mu + \epsilon_{iK}$$

  - ⋆ K voxels/regions → K models → multiplicity
  - ⋆ FWE only compensates local, but not global, relatedness → information loss

- Hierarchical: $y_{ik} = \mu + \tau_i + \theta_k + \epsilon_{ik}$
  - ⋆ one single model
  - ⋆ info shared/calibrated/regularized across space
  - ⋆ same mechanism as one-sample t-test process
  - ⋆ program for region-based analysis: RBA

Chen, G., Xiao, Y., Taylor, P.A., Rajendra, J.K., Riggins, T., Geng, F., Redcay, E., Cox, R.W., 2019. Handling Multiplicity in Neuroimaging through Bayesian Lenses with Multilevel Modeling. Neuroinformatics 17, 515–545.

Chen, G., Taylor, P.A., Cox, R.W., Pessoa, L., 2020. Fighting or embracing multiplicity in neuroimaging? Neighborhood leverage versus global calibration. NeuroImage 206, 116320.

Multiplicity: hierarchical modeling

- Data at population level
  - ⋆ 124 individuals; explanatory variable: behavior measure
  - ⋆ effect of interest: association

- Conventional mass univariate analysis
  - ⋆ 2 clusters survived FWE adjustment based on voxel-level p of 0.001

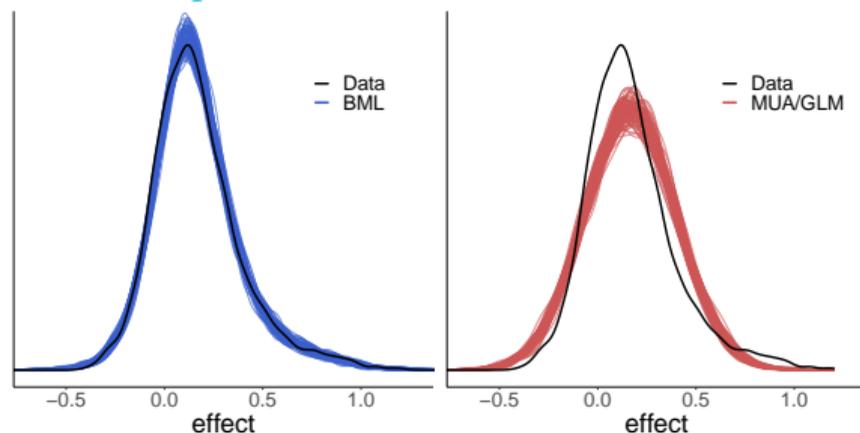- Hierarchical modeling
  - ⋆ 21 regions
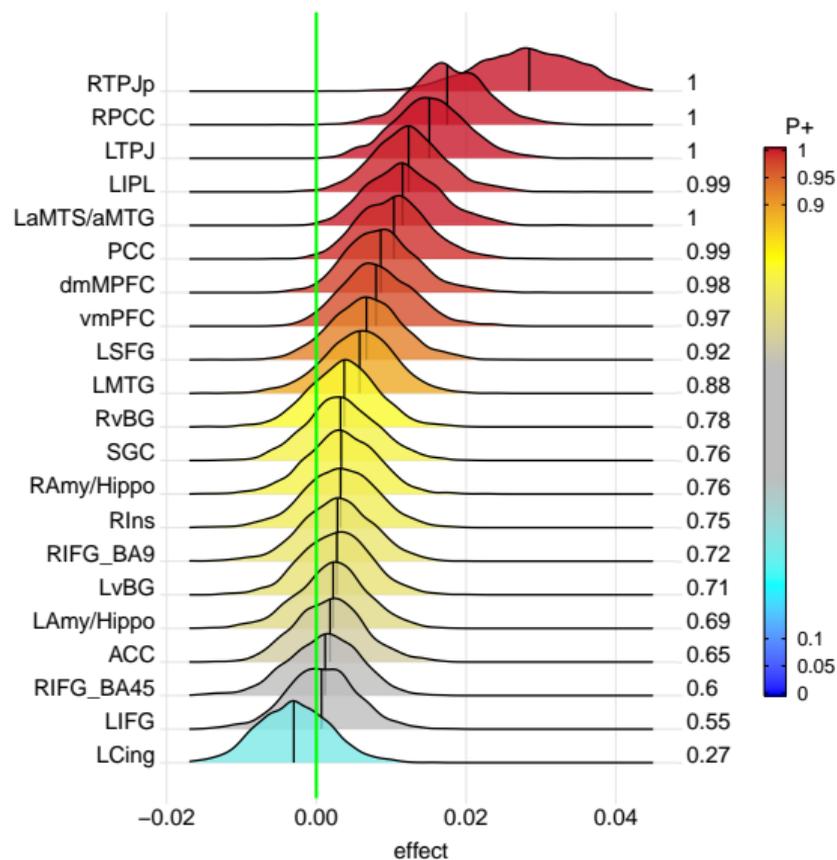
Hierarchical modeling: model quality

- Data at population level
  - ★ 124 individuals; explanatory variable: behavior measure
  - ★ effect of interest: association

- Hierarchical modeling:

Leave-one-out cross-validation

	LOOIC	SE
GLM	-300	98
RBA	-2247	86
GLM vs RBA	1947	96

Posterior predictive checks

Take-home messages

- Experimental design: sample sizes
  - ⋆ participants
  - ⋆ trial numbers: almost equally important as participants
  - ⋆ higher-order interactions requires much larger sample sizes
    - differences of differences: much smaller effect magnitude
    - for example, 2-way interactions requires a few times samples than main effects
  - ⋆ avoiding tasks with substantial head motion

- Experimental design: variable selection
  - ⋆ draw DAGs → collect data for relevant covariates
  - ⋆ multicollinearity: not a modeling issue - no statistical solution
    - human errors, experimental design

- Modeling
  - ⋆ quality control: cautious with task-related motion
  - ⋆ canonical HRF vs estimation of HDRs
  - ⋆ voxel- vs region-level
  - ⋆ draw DAGs: confounder, collider, mediator, parents/children, ...
  - ⋆ different effects may require separate models
  - ⋆ interactions, nonlinearity, centering, data hierarchy, ...

Take-home messages (cont.)

- Model caveats
  - ⋆ a model is just a data machine
  - ⋆ understanding the caveats of a model enables us to interpret its results with caution

- Result reporting
  - ⋆ science: quantification vs decision
  - ⋆ show effect magnitude and uncertainty
  - ⋆ avoid strict thresholding and adopt soft thresholding
    - don't let statistics fully dictate the process
    - OK to set p-value of 0.01 with a cluster size of 20 voxels
    - importance of domain knowledge: literature; anatomical structure
    - highlight, but don't hide
  - ⋆ supporting material
    - Taylor, P.A., Reynolds, R.C., Calhoun, V., Gonzalez-Castillo, J., Handwerker, D.A., Bandettini, P.A., Mejia, A.F., Chen, G., 2023. Highlight results, don't hide them: Enhance interpretation, reduce biases and improve reproducibility. NeuroImage 274, 120138.
    - Chen, G., Taylor, P.A., Stoddard, J., Cox, R.W., Bandettini, P.A., Pessoa, L., 2022. Sources of information waste in neuroimaging: Mishandling structures, thinking dichotomously, and over-reducing data. Aperture Neuro 2.
  - ⋆ promote transparency and open science: require collective effort!