Do We Have to Deal with Multiple Comparisons in Neuroimaging?

Gang Chen

Scientific and Statistical Computing Core
National Institute of Mental Health
National Institutes of Health, USA

Preview

- **Issues with current correction for multiplicity**
- **Two toy examples**
  - NBA players
  - Kidney cancer
- **Application: region-based analysis (RBA)**
  - Program in AFNI: ***RBA***
- **Other applications**
  - Matrix-based analysis (program in AFNI: ***MBA***)
  - Region-based inter-subject correlation (ISC) analysis
  - Gray matter connectivity analysis
  - Others cases involving multiplicity

Multiplicity in Neuroimaging

- **100,000 spatial units**
- **100,000 models: MUA**
  - Assumption of spatial independence
  - Sharing no information
- **Corrections**
  - Multiplicity + spatial relatedness
  - Problems
    - Heavy penalty: information waste
    - Other issues

Null Hypothesis Significance Testing

- **Straw man H_0: null hypothesis**
  - Witch hunt: Don Quixote's windmills
  - **Type I error** = P(data | H_0) = false positive = p-value
    - Surprise or weirdness of data: **0.05**
    - No effect until shown with small p-value
    - Innocent until proven guilty
  - **Type II error** = P(accept H_0 | H_1) = false negative

	H_0 True	H_0 False
Reject H_0	Type I Error (false positive)	Correct
Fail to Reject H_0	Correct	Type II Error (false negative)

Issues: NHST

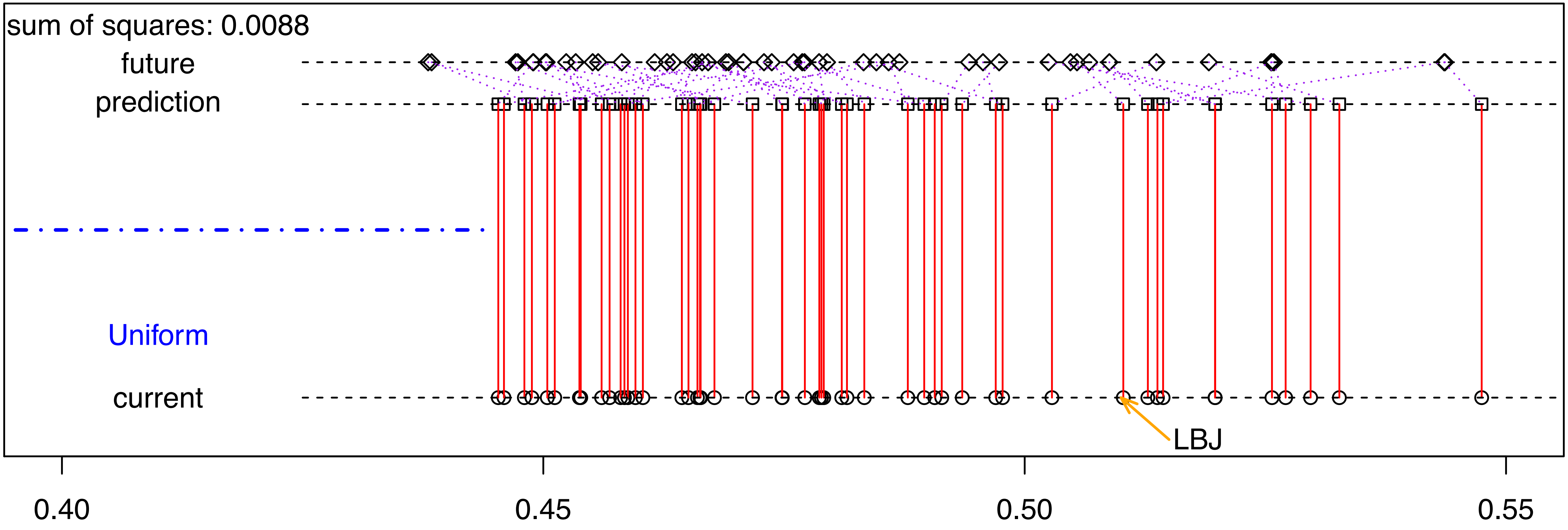
- **Arbitrary dichotomy**
  - Binary or discrete: innocent vs guilty
  - Unrealistic: “activated” vs “not activated”?
- **Vulnerable to misconceptions**
  - p (weirdness | H_o) $\neq$ p (H_o | data)
  - **Absence of evidence ≠ evidence of absence**
- **Vulnerable to data manipulations**
  - **Statistical evidence changes: whole brain, gray matter, region**
- **Inflated effect estimates**
  - **Type M (magnitude) error: biasedness**

Issues: NHST

- Disregarding effect size
- Uncertainty unavailable
  - No standard deviation at voxel or cluster level
- Lack of spatial specificity
  - Locating regions per peak voxel
- Penalizing small regions

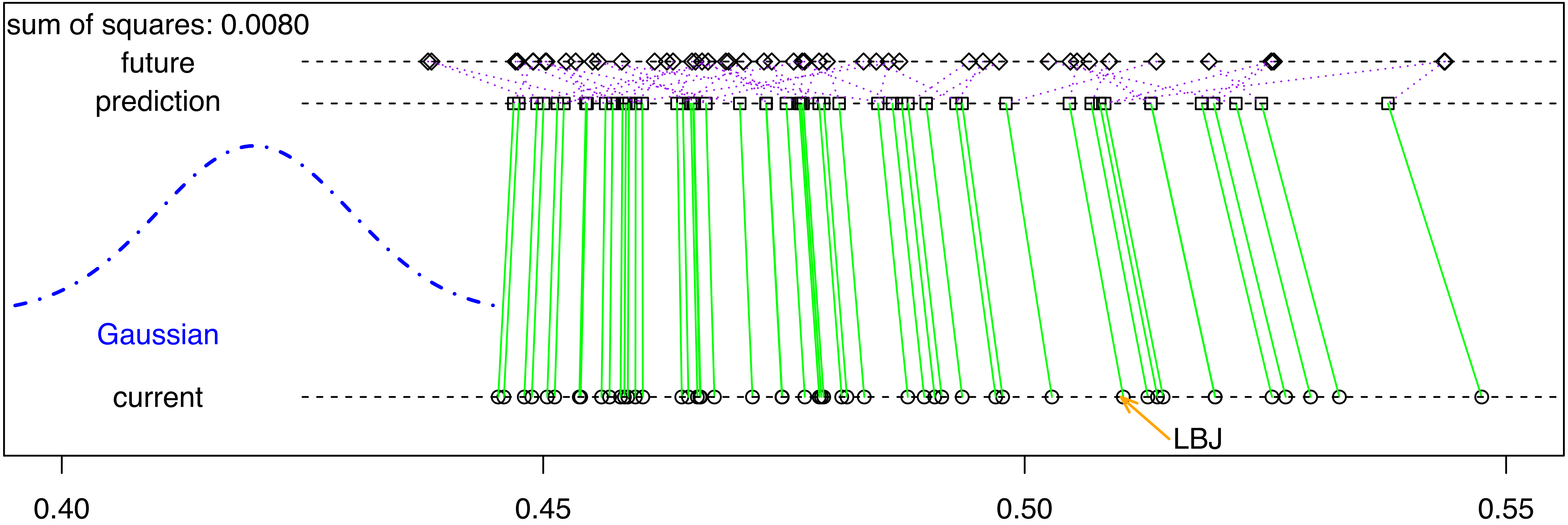
Toy Example 1

- **NBA players**
  - LeBron James field goals percentage: 51%
  - Prediction: performance during next season?
  - One vs. top 50 players: **no pooling** vs complete pooling



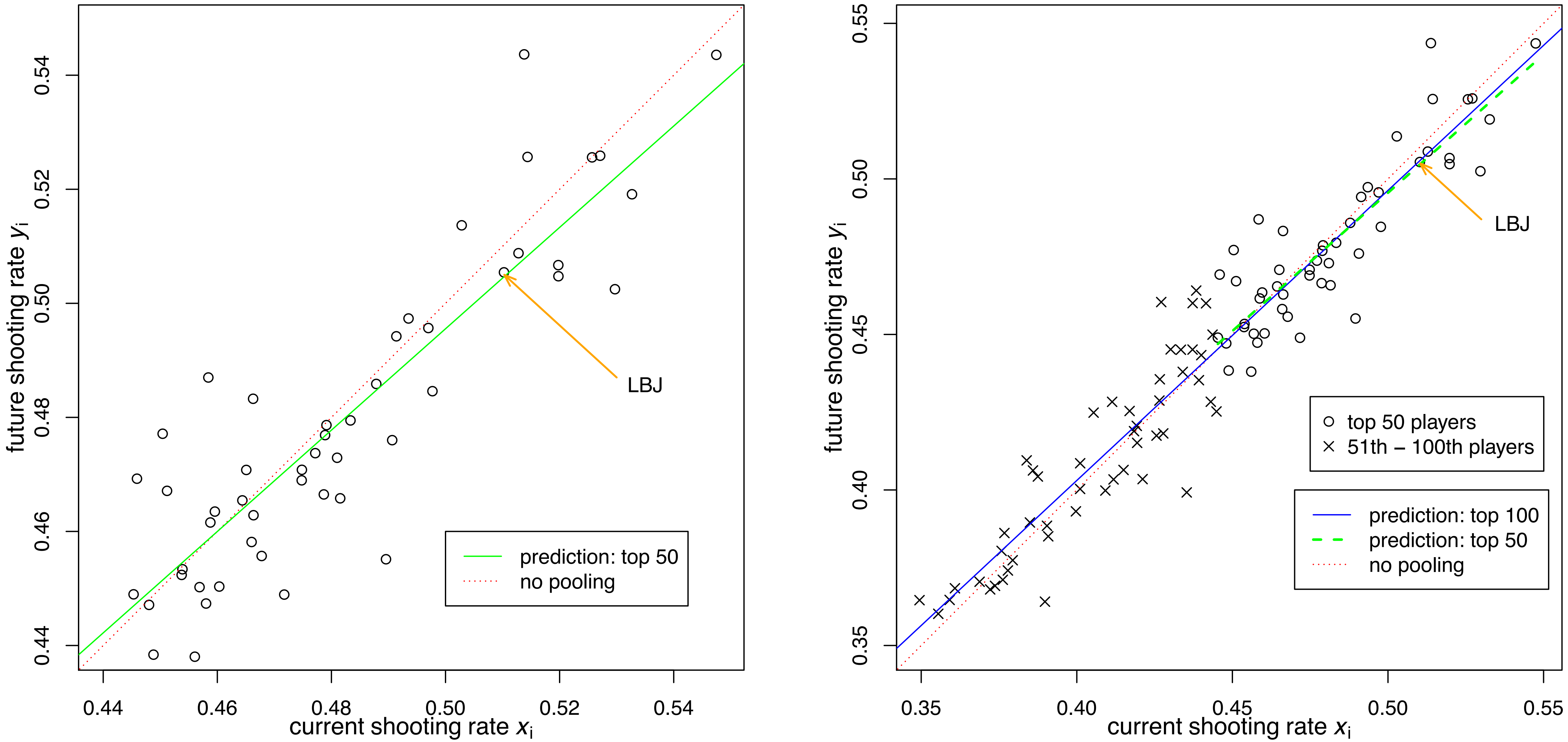
Toy Example 1

- **NBA players**
  - LeBron James field goals percentage during 2019: 51%
  - Prediction: performance during 2020?
  - One vs. top 50 players: **partial pooling** (regression to the mean)



Toy Example 1

- **Top 50 vs. 100 NBA players: adaptivity**

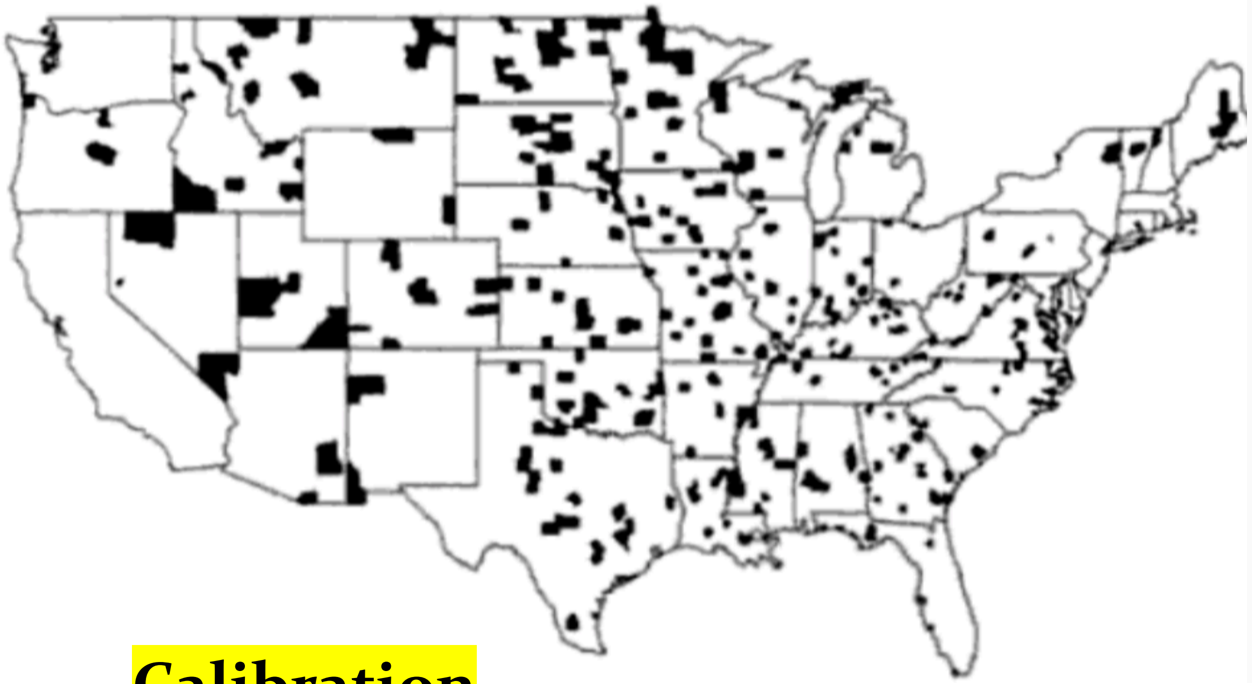


Toy Example 2

- **Kidney cancer distribution among U. S. counties**

Highest rate

lowest rate



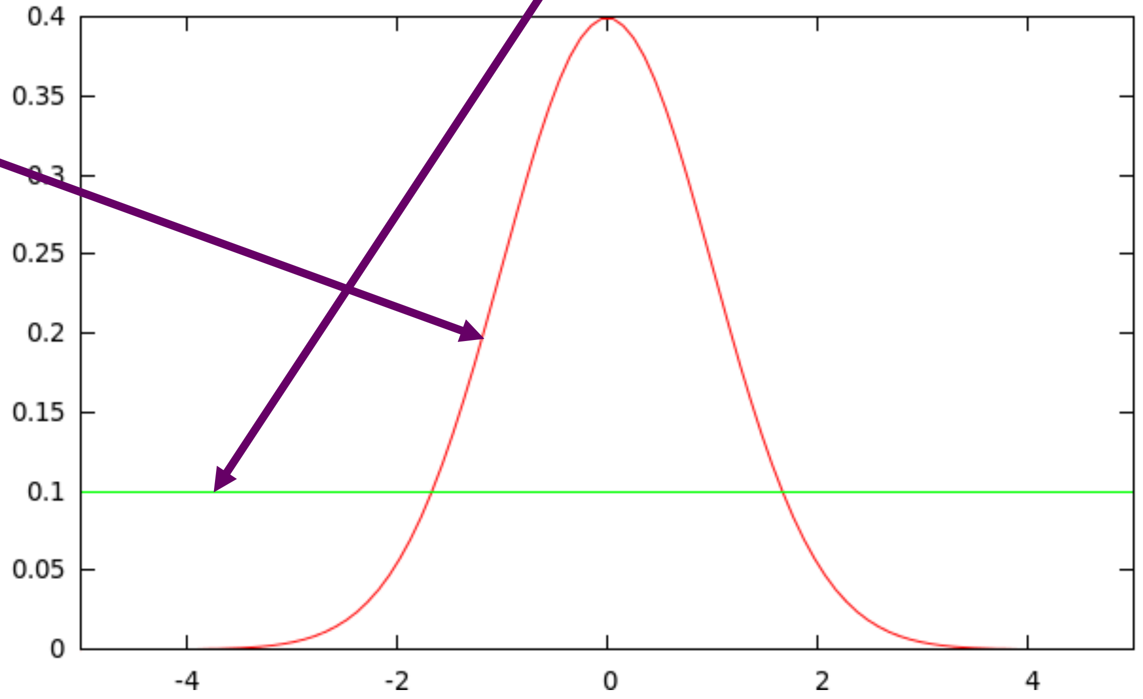
Calibration

Morals from kidney cancer data

- **Multiplicity problem: > 3000 counties!**
  - Divide *p*-value by number of counties?
  - Borrow idea from neuroimaging: leverage geographical relatedness?
- **What can we learn from the example? Food for thought**
  - Care about strawman H_0 (zero kidney rate), false positives, *p*-value?
  - Trust individual county-wise estimates? Unbiased! **BLUE**
    - **Incorrect sign errors** (type S): some counties really have higher kidney cancer rate than others?
    - **Incorrect magnitude** (type M): some counties really have higher/lower cancer rate?
  - Would correction for multiplicity help at all?
    - Useless in controlling for type S and M errors
- **How can we do better?**
  - Information share: across spatial elements
  - **Research hypothesis: P (effect > 0 | data)**

What do we know about spatial elements?

- **Element-wise modeling**
  - Pretend full ignorance: fully trust the data
  - Uniform distribution: each element equally likely to have any value in (- ∞, +∞)
  - Similar for variances: variances can be negative in ANOVA
- **One crucial prior for spatial elements**
  - Reasonable to assume Gaussian distribution?
  - Gaussian assumption adopted everywhere!
    - Subjects, residuals across TRs
  - How can Gaussian assumption help?
    - Loosely constraining elements
    - No full trust for individual estimates
    - Information sharing: shrinkage or partial pooling
    - Controlling type S and M errors



Short summary: what we intend to achieve

- **Abandon strawman and *p*-value**
  - Directly focus on research interest: $P\,(\text{effect} > 0 \mid \text{data})$ vs. $P\,(\text{data} \mid \text{effect} = 0)$
- **Build one model**
  - Incorporate all elements into a multilevel or hierarchical structure
  - Loosely constrain elements: leverage **prior knowledge**
  - Achieve higher modeling efficiency: **no more multiplicity**!
  - Validate the model by comparing with potential competitors
  - Be conservative on effect estimates by controlling type S and M errors: **biased?**
  - Always be mindful of uncertainties: strength of evidence (no proof)
  - Less vulnerable to data manipulations: whole brain, gray matter, regions, ...
- **Avoid dichotomous decisions**
  - Report full results if possible
  - Highlight instead of hide based on gradient of evidence

Application: region-based analysis

- **Dataset**
  - Subjects: n = 124 children; resting-state data (Xiao et al., 2019)
  - Individual subjects: seed-based correlation for each subject
    - 3D correlation between seed and whole brain ("functional connectivity")
  - Explanatory variable (behavior data): Theory of Mind Index x_i

- **Voxel-wise group analysis: GLMs**
  - Focus: association between x and seed-based correlation (z-score)
  - Pretense: voxels unrelated - equal likelihood within $(-\infty, \infty)$
  - Information waste!
  - GLMs: mass univariate - multiplicity

  m = 100,000 voxels → 100,000 models

Uniform distribution: total freedom - each parameter on its own

$$
\begin{aligned}
\text{1st voxel: } & \boldsymbol{y}_1 = a_1 + b_1\boldsymbol{x} + \boldsymbol{\epsilon}_1 \\
\text{2nd voxel: } & \boldsymbol{y}_2 = a_2 + b_2\boldsymbol{x} + \boldsymbol{\epsilon}_2 \\
& \cdots \\
m\text{th voxel: } & \boldsymbol{y}_m = a_m + b_m\boldsymbol{x} + \boldsymbol{\epsilon}_m
\end{aligned}
$$

Xiao et al., 2019. Neuroimage 184:707-716

GLMs: dealing with multiplicity!

- **Voxel-based analysis: GLMs**
  - Penalty time for pretense: multiple testing (m = 100,000), magic 0.05
  - Show time for various correction methods
    - Voxel-wise p, FWE, FDR, spatial smoothness, clusters, …
    - Simulations, random field theory, permutations, …
    - How would dataset turn out under GLM? 4 lucky clusters manage to survive

voxel p	cluster threshold	surviving ROIs	ROIs
0.001	28	2	R PCC, PCC/PrC
0.005	66	4	R PCC, PCC/PrC., L IPL, L TPJ
0.01	106	4	R PCC, PCC/PrC., L IPL, L TPJ
0.05	467	4	R PCC, PCC/PrC., L IPL, L TPJ

Switching from voxels to ROIs: still GLMs

- **Region-wise analysis : GLMs**
  - Focus: association between and seed-based correlation (z-score)
  - Pretense: ROIs unrelated
  - GLMs: mass univariate

    m = 21 ROIs $\longrightarrow$

    21 models
  - Penalty time for pretense:

    multiple testing – what to do?
    - Bonferroni? Unbearable
    - What else?

Uniform distribution: total freedom - each parameter on its own.

$$
\begin{aligned}
\text{1st ROI: } \boldsymbol{y}_1 &= a_1 + b_1\boldsymbol{x} + \boldsymbol{\epsilon}_1 \\
\text{2nd ROI: } \boldsymbol{y}_2 &= a_2 + b_2\boldsymbol{x} + \boldsymbol{\epsilon}_2 \\
&\cdots \\
m\text{th ROI: } \boldsymbol{y}_m &= a_m + b_m\boldsymbol{x} + \boldsymbol{\epsilon}_m
\end{aligned}
$$

Switching from GLMs to LME

- **Region-wise analysis : Linear Mixed-Effects (LME) model**
  - One model integrates all regions
  - ROIs loosely constrained instead of being unrelated
    - Gaussian distribution: Is it far-fetched or subjective?
    - Similar to cross-subject variability
  - Goal: effect of interest- $a + \alpha_j$, $b + \beta_j$
  - Differentiation: fixed vs. random
    - Fixed: **epistemic** uncertainty
    - Random: **aleatoric** uncertainty
    - Julius Caesar: Alea iacta est. January 10, 49 BC
  - What can we get out of LME?
    - Conventional framework
    - Estimates for fixed effects
    - Variances for random effects
  - ***Dead end!***

New components

idiosyncratic effect of *i*th subject

Overall effect: shared by all ROIs and subjects

Unique effect of *j*th ROI

$$z_{ij} = \boxed{a + bx_i} + \boxed{\pi_i} + \boxed{\alpha_j + \beta_j x_i} + \epsilon_{ij}$$

$$\pi_i \overset{iid}{\sim} \mathcal{N}(0, \tau^2), (\alpha_j, \beta_j)^T \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\lambda})$$

$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), i = 1, 2, ..., n, \ j = 1, 2, ..., m$$

Switching from GLMs to BML

- **Region-wise analysis : Bayesian multilevel (BML) model**
  - One model integrates all regions: basically same as LME
  - ROIs loosely constrained instead of being unrelated
    - Gaussian distribution: Is it far-fetched or subjective?
    - Similar to cross-subject variability
  - Goal: effect of interest $b + \beta_j$
  - No more differentiation: fixed vs. random
    - All parameters: **aleatoric**
  - Same model as LME plus priors
    - Markov chain Monte Carlo (MCMC)
    - Inferences via posterior distribution
  - ***Ka-ching!***

New components

Overall effect: shared by all ROIs and subjects

Idiosyncratic effect by *i*th subject

Unique effect by *j*th ROI

$$z_{ij} = \boxed{a + bx_i} + \boxed{\pi_i} + \boxed{\alpha_j + \beta_j x_i} + \epsilon_{ij}$$

$$\pi_i \overset{iid}{\sim} \mathcal{N}(0, \tau^2), (\alpha_j, \beta_j)^T \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\lambda})$$

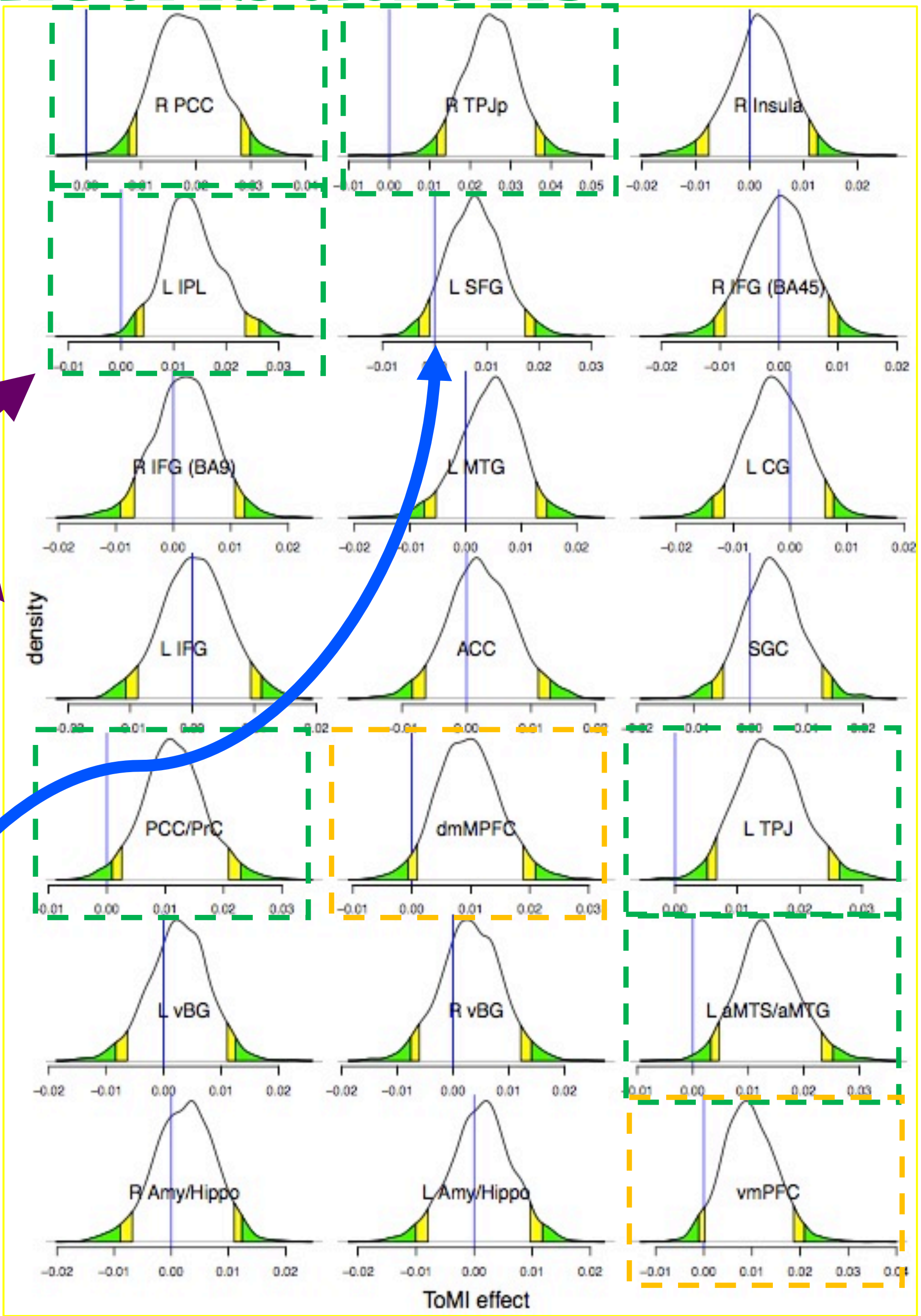
$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), i = 1, 2, ..., n, \; j = 1, 2, ..., m$$

Chen, et al, 2019. Handling Multiplicity in Neuroimaging through Bayesian Lenses with Multilevel Modeling. Neuroinformatics.

Inferences from BML: full distributions

- **Region-based BML: 21 ROIs**
- **Full report with richer information: posterior distributions for each ROI**
  - No dichotomization
  - No results hiding
  - No discrimination against small regions
  - No ambiguities about spatial specificity
  - No inconvenient interpretation of confidence interval
  - Evidence for each ROI: ***P*** **(effect > 0 | data)**

Highlight, not hide



- **9 ROIs with strong evidence of effect compared to**
  - Region-wise GLM with Bonferroni correction
  - Voxel-wise GLM at cluster level: 2 clusters

How about Left SFG?

Inferences from BML: uncertainty

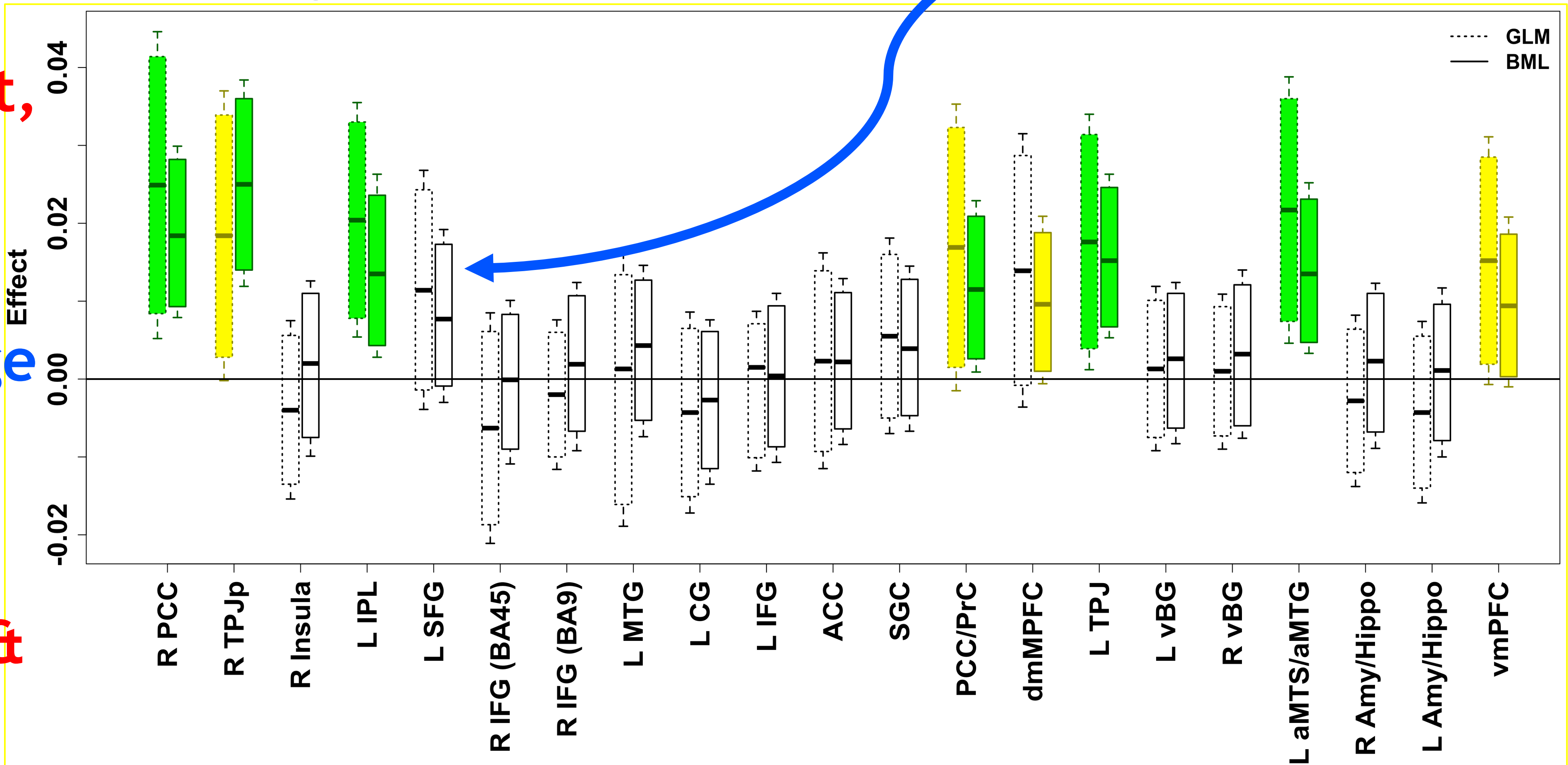
- ROI-based BML: 21 ROIs
- Full report with bar graph uncertainty intervals
  - Nothing hidden under sea level
- 8 ROIs with strong evidence for effect of interest

How about Left SFG?

Highlight, not hide

Shrinkage / partial pooling

Type M & S errors



BML: model validations

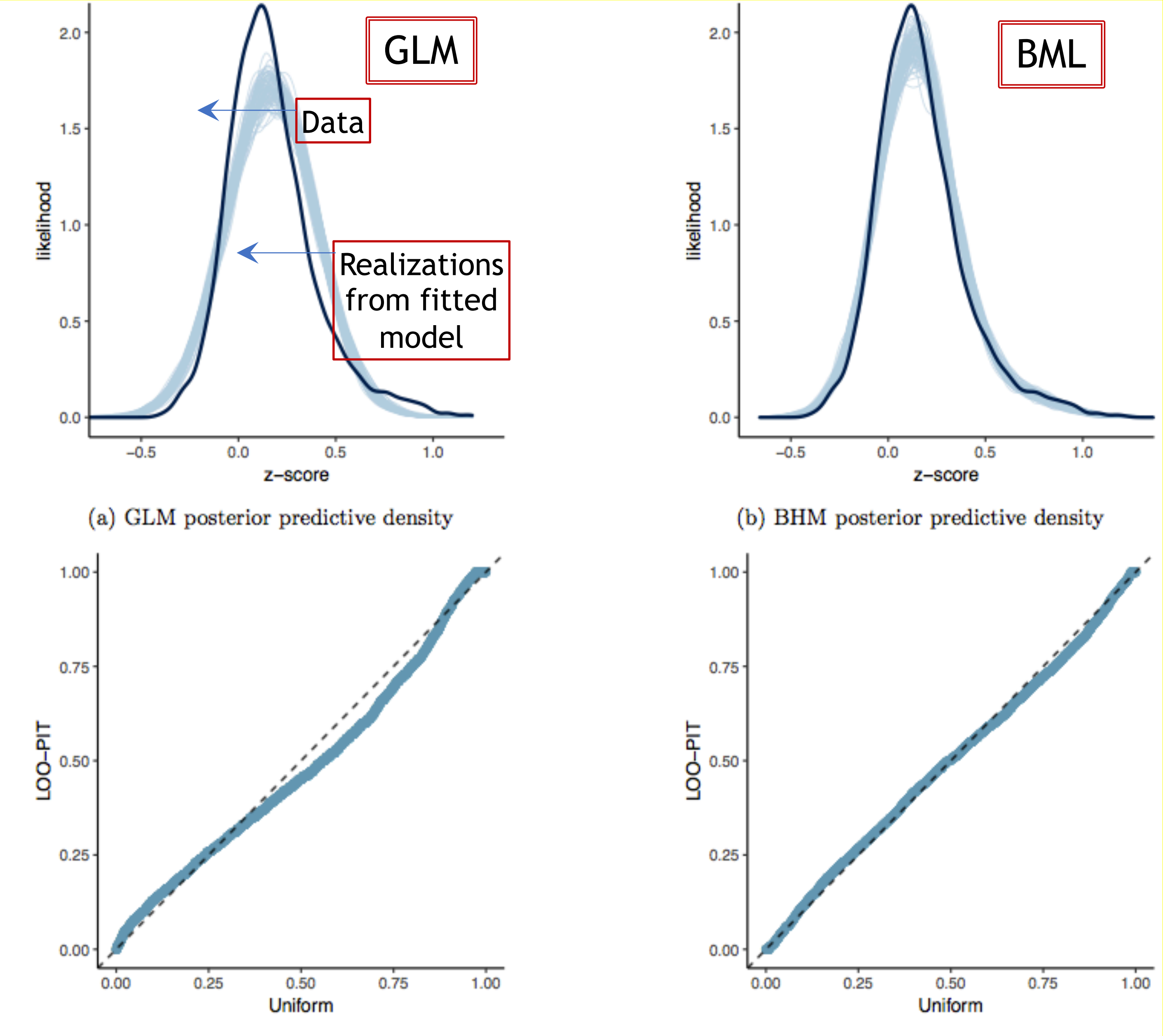
- **Cross-validation**
  - Leave-one-out information criterion (LOOIC)

Cross-validation

	LOOIC	SE
GLM	-300.39	98.25
BML	-2247.06	86.42
GLM - BML	1946.67	96.35

  - Posterior predictive checking

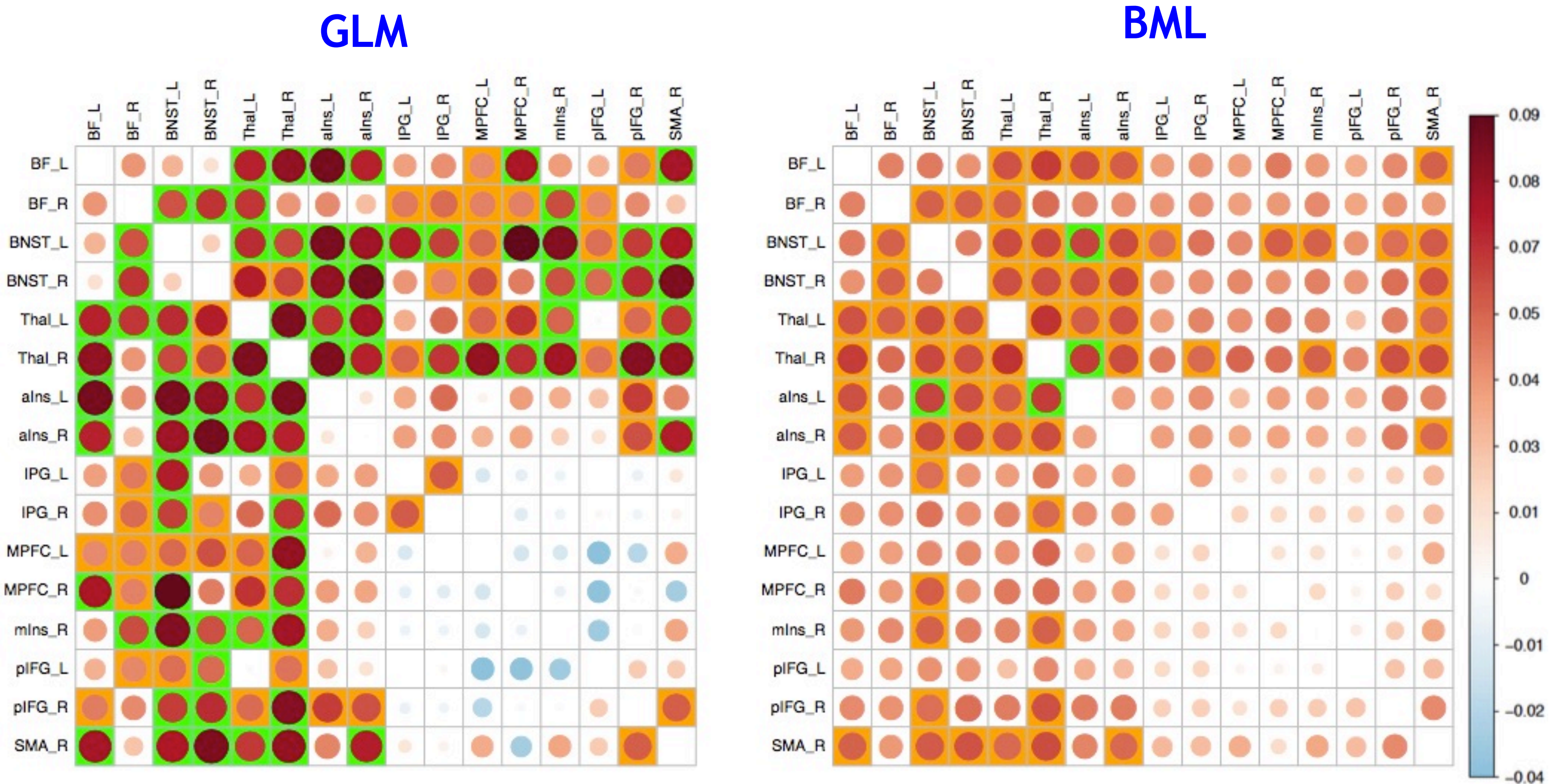
- **Effects of BML**
  - Regularizing ROIs: don't fully trust individual ROI data
  - Sacrificing fit at each ROI; achieving better overall fit



(a) GLM posterior predictive density
(b) BHM posterior predictive density
(c) GLM cross-validation: Q-Q plot (uniform)
(d) BHM cross-validation: Q-Q plot (uniform)

Other applications

- **Matrix-based analysis**
  - 63 RPs identified by GLMs with *p* of 0.05
    - none survived after correction with NBS via permutations
  - 33 RPs with strong evidence under BML



Summary

- **Issues with current correction for multiplicity**
- **Two toy examples**
  - NBA players
  - Kidney cancer
- **Application: region-based analysis (RBA)**
  - Program in AFNI: *RBA*
- **Other applications**
  - Matrix-based analysis (program in AFNI: *MBA*)
  - Region-based inter-subject correlation (ISC) analysis
  - Gray matter connectivity analysis
  - Others cases involving multiplicity

Keep Kidney Cancer in Mind!

- Kidney cancer distribution among counties

Highest rate

lowest rate

Calibration

Acknowledgements

- **Paul-Christian Bürkner** (Department of Psychology, University of Münster)
- **Andrew Gelman** (Columbia University), **Stan Development Team, R Foundation**

- **Yaqiong Xiao, Elizabeth Redcay, Tracy Riggins, Fengji Geng**
- **Luiz Pessoa, Joshua Kinnison** (Depart of Psychology, University of Maryland)

- **Zhihao Li** (School of Psychology and Sociology, Shenzhen University, China)
  Lijun Yin (Department of Psychology, Sun Yat-sen University, China)

- **Emily Finn, Daniel Handwerker** (SFIM/NIMH, National Institutes of Health)

- **Paul A. Taylor, Daniel R. Glen, Justin K. Rajendra, Richard C. Reynolds, Robert W. Cox** (SSCC/NIMH, National Institutes of Health)