Predicting Treatment Efficacy via quantitative MRI: A Bayesian Joint Model

Timothy D. Johnson

Department of Biostatistics
University of Michigan

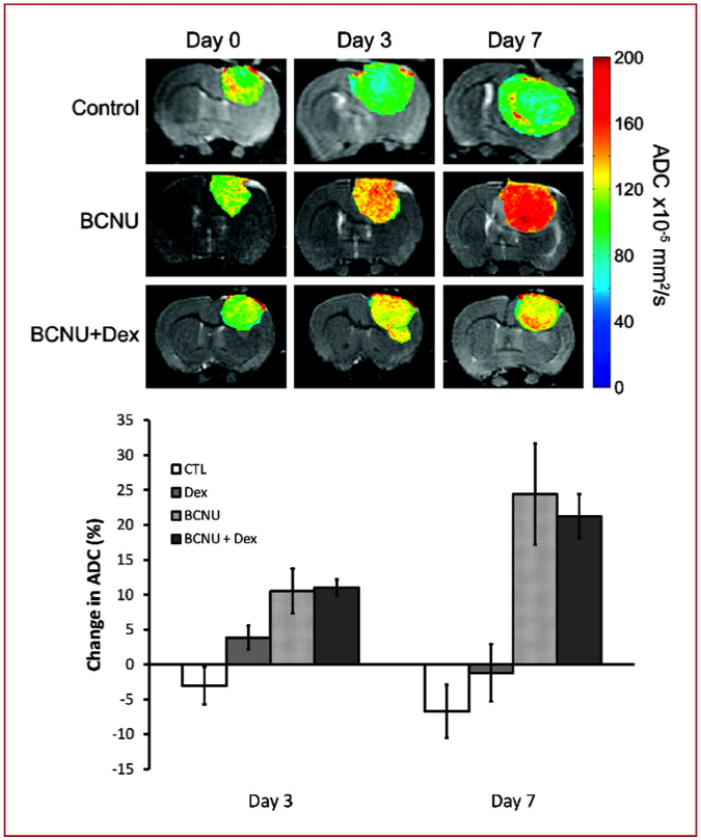
Outline

Use qMRI to predict treatment response early

- High-grade Gliomas
  - 1 year median survival after diagnosis
- treatment lasts $\approx$ 3 months
- another 2 months before radiological response measured
  - Complete Response — no visible sign of tumor
  - Partial Response — $>$ 25% volume reduction
  - Stable disease — $<$ 25% reduction and $<$ 25% volume increase
  - Progressive disease — $>$ 25% volume increase
- Second line therapies may then be given (usually too late to have any effect)
- Goal: predict response within 2-3 weeks of treatment initiation

qMRI Biomarkers for Treatment Response

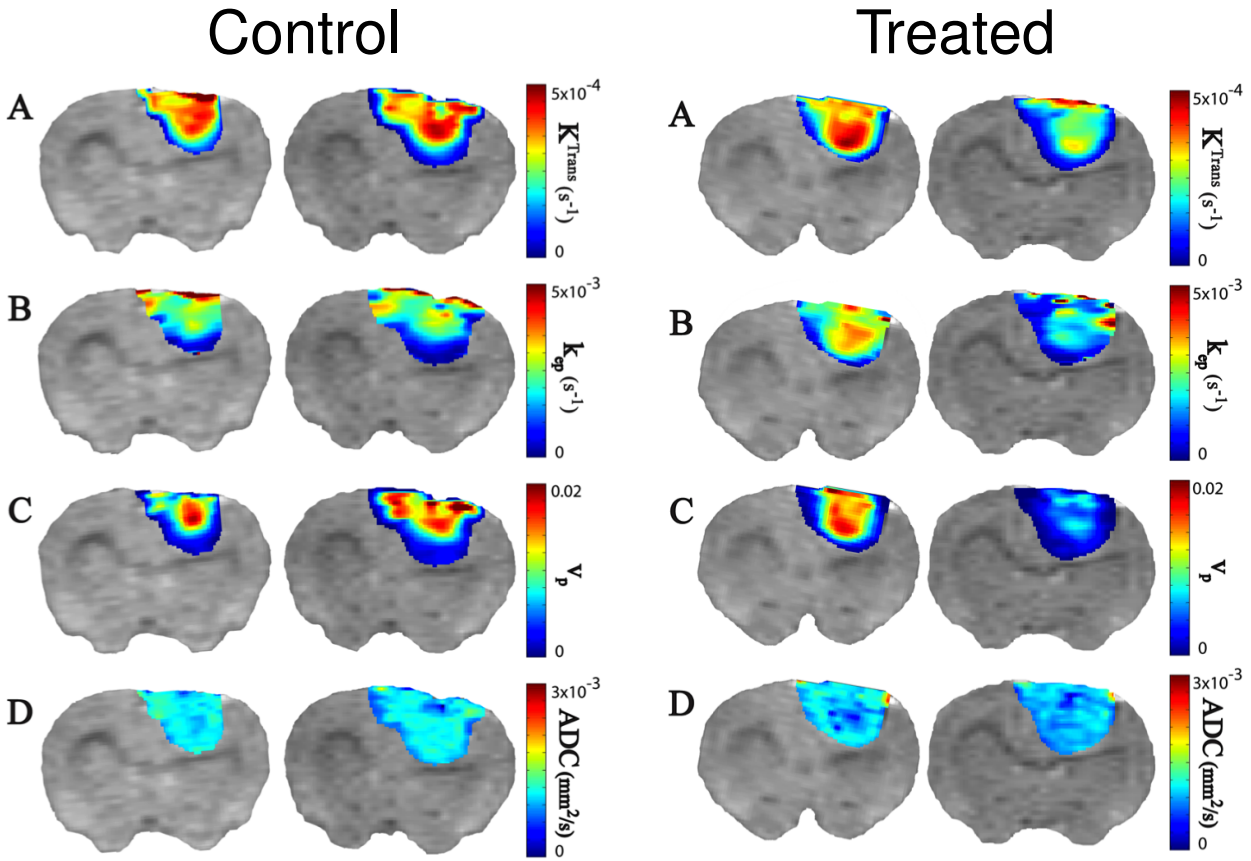
- Diffusion — measure of Brownian motion of water molecules
- Apparent Diffusion Coefficient (ADC)
  - Magnitude of the diffusion tensor



- high cellular density = low diffusion
- Cytotoxic drugs/Radiation kill cells which then lyse
  - low cellular density = high diffusion

qMRI Biomarkers for Treatment Response

- Perfusion — measure of blood flow or blood volume



  - low blood volume = less nutrients = retarded growth

Early Human Trial

Hypothesis

quantitative MRI can predict treatment efficacy early

Early Results

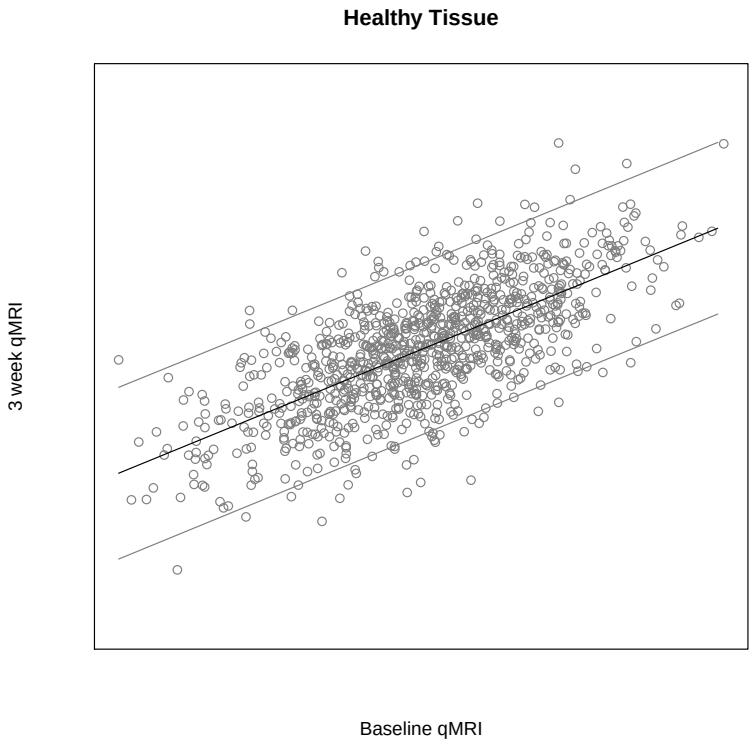
Human study (Glioma tumors) appeared futile

- no significance change in mean ADC due to treatment
- mean ADC could not predict outcome (radiological response)

- Colleagues did not give up —
  - An entire program project grant was funded based on early animal models
  - They noted that regions of tumors had large changes in ADC
  - Noticed changes in the tails of the tumor histogram

New summary statistic

- Moffat et al. (2005) developed a new summary statistic
  - functional diffusion map — FDM (and FPM)
  - group means significantly different
    - (SD + PR + CR) vs. PD

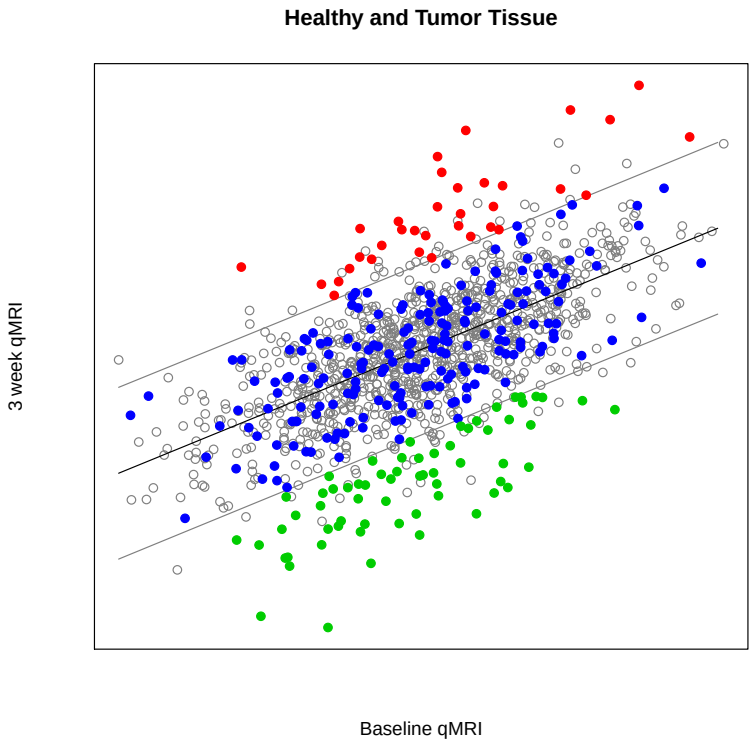


New summary statistic

- Moffat et al. (2005) developed a new summary statistic
  - functional diffusion map — FDM (and FPM)
  - group means significantly different
    - (SD + PR + CR) vs. PD

New summary statistic

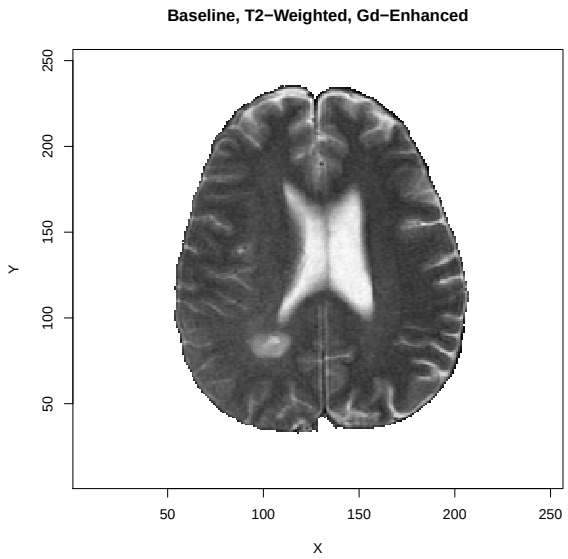
- Moffat et al. (2005) developed a new summary statistic
  - functional diffusion map — FDM (and FPM)
  - group means significantly different
    - (SD + PR + CR) vs. PD

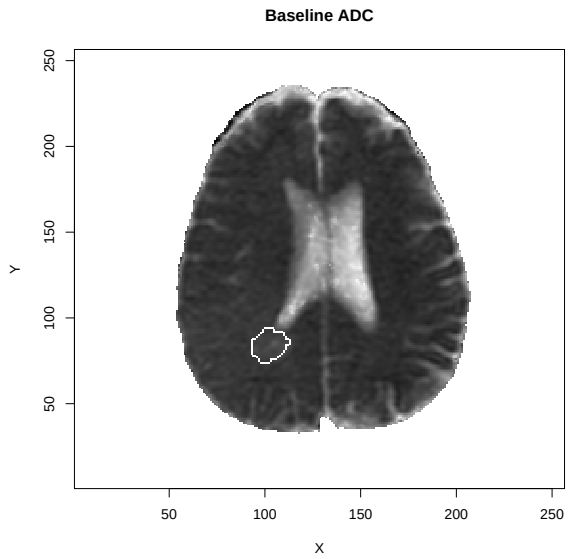


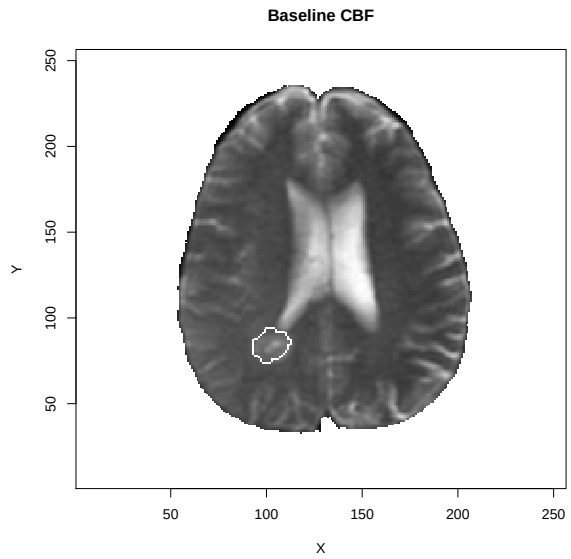
New summary statistic

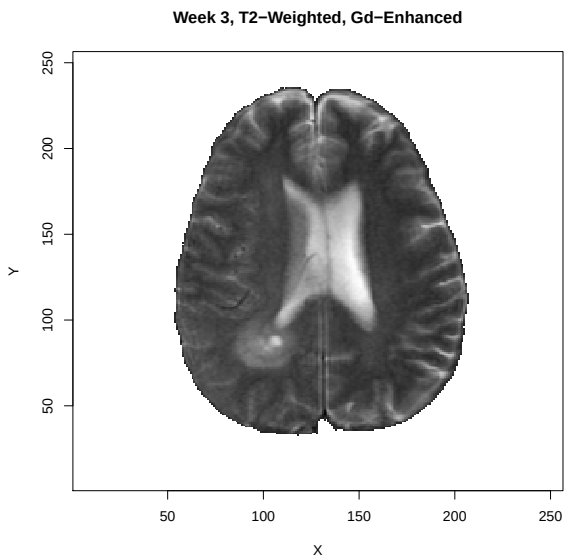
- Moffat et al. (2005) developed a new summary statistic
  - functional diffusion map — FDM (and FPM)
  - group means significantly different
    - (SD + PR + CR) vs. PD
- I was still skeptical
  - showing a difference in means does not imply predictive power
- After obtaining the data
  - tried using FDM and FPM to predict one-year survival status
    - leave-one-out CV: 63% correct classification (Logistic classifier)
  - I had to try harder
    - a large chunk of my salary comes from the P01!

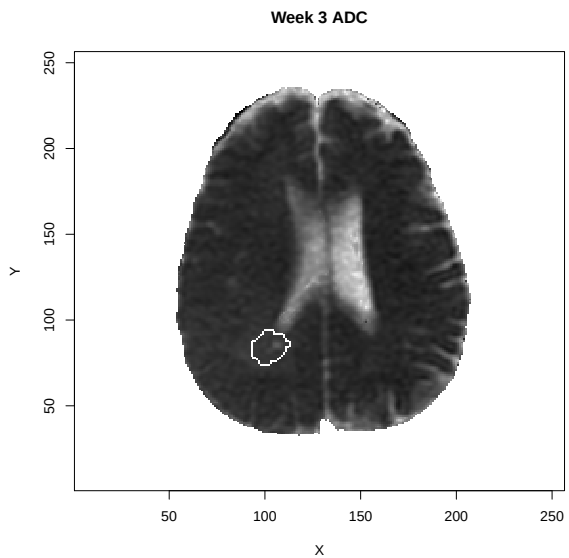
Sample Images

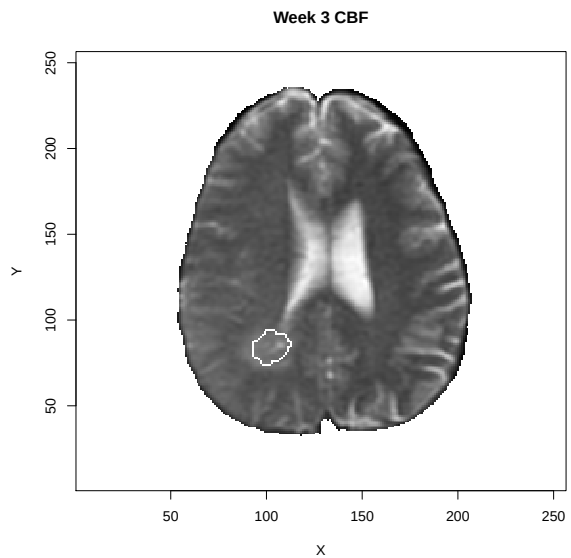












Two-Stage Joint Model

Stage I: Multivariate spatio-temporal pairwise difference prior [1]

- $\mathcal{Y}$ will denote the set of all images over all M subjects
- Ω_1 denotes the stage I parameters
- Summary statistics derived in stage I denoted by $\mathcal{X}$
  - functionals of Ω_1: $\mathcal{X} = F(\Omega_1)$

[1] Besag (1993), *Towards Bayesian Image Analysis*, Journal of Applied Statistics (20) 107–119.

Two-Stage Joint Model

Stage I: Multivariate spatio-temporal pairwise difference prior [1]

- $\mathcal{Y}$ will denote the set of all images over all M subjects
- Ω_1 denotes the stage I parameters
- Summary statistics derived in stage I denoted by $\mathcal{X}$
  - functionals of Ω_1: $\mathcal{X} = F(\Omega_1)$

Sampling distribution

$$[\mathbf{Y}_i \mid \boldsymbol{\mu}_i, \Sigma] \sim N(\boldsymbol{\mu}_i, \Sigma), \quad \forall \text{ tumor voxels } i$$

[1] Besag (1993), *Towards Bayesian Image Analysis*, Journal of Applied Statistics (20) 107–119.

Two-Stage Joint Model

Stage I: Multivariate spatio-temporal pairwise difference prior [1]

- $\mathcal{Y}$ will denote the set of all images over all M subjects
- Ω_1 denotes the stage I parameters
- Summary statistics derived in stage I denoted by $\mathcal{X}$
  - functionals of Ω_1: $\mathcal{X} = F(\Omega_1)$

Sampling distribution

$$[\mathbf{Y}_i \mid \boldsymbol{\mu}_i, \Sigma] \sim N(\boldsymbol{\mu}_i, \Sigma), \quad \forall \text{ tumor voxels } i$$

Prior distribution (pairwise-difference prior)

$$\pi(\boldsymbol{\mu}) \quad \propto \quad \exp\left[-\sum_{i \sim j}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^{\mathrm{T}} \Psi^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\right]$$

[1] Besag (1993), *Towards Bayesian Image Analysis*, Journal of Applied Statistics (20) 107–119.

Two-Stage Joint Model

Stage II: Generalized non-linear model [2]

- $\mathbb{Z}$ will denote the M-vector of 1-year survival statuses
- Probit link, MARS[3] basis
- Ω_2 denotes the stage II parameters
- Stages linked via summary statistics

[2] Holmes and Denison (2003), *Classification with Bayesian MARS*, Machine Learning (50) 159–173.

[3] Friedman (1991), *Multivariate adaptive regression splines*, The Annals of Statistics (19) 1–61.

Two-Stage Joint Model

Stage II: Generalized non-linear model [2]

- $\mathbb{Z}$ will denote the M-vector of 1-year survival statuses
- Probit link, MARS[3] basis
- Ω_2 denotes the stage II parameters
- Stages linked via summary statistics

GNLM-BMARS

$$\Pr(Z_j = 1 \mid \mathbf{X}_j, \mathbf{\Omega}_2) = \Phi(\eta_j), \qquad \eta_j = \sum_{k=0}^{K} \beta_k B_k(\mathbf{X}_j),$$

$$B_k(\mathbf{X}_j) = \begin{cases} 1, & k = 0, \\ \prod_{\ell=1}^{L_k} [s_{\ell k}(X_{jw_{\ell k}} - t_{\ell k})]_+, & k = 1, 2, \ldots, K \end{cases}$$

[2] Holmes and Denison (2003), *Classification with Bayesian MARS*, Machine Learning (50) 159–173.

[3] Friedman (1991), *Multivariate adaptive regression splines*, The Annals of Statistics (19) 1–61.

Two-Stage Joint Model

Stage II: Generalized non-linear model [2]

- $\mathbb{Z}$ will denote the M-vector of 1-year survival statuses
- Probit link, MARS[3] basis
- Ω_2 denotes the stage II parameters
- Stages linked via summary statistics

GNLM-BMARS

$$\Pr(Z_j = 1 \mid \mathbf{X}_j, \mathbf{\Omega}_2) = \Phi(\eta_j), \qquad \eta_j = \sum_{k=0}^{K} \beta_k B_k(\mathbf{X}_j),$$

$$B_k(\mathbf{X}_j) = \begin{cases} 1, & k = 0, \\ \prod_{\ell=1}^{L_k} [s_{\ell k}(X_{jw_{\ell k}} - t_{\ell k})]_+, & k = 1, 2, \ldots, K \end{cases}$$

Posterior factorization

$$\pi(\Omega_1, \Omega_2 \mid \mathcal{Y}, \mathbb{Z}) = \pi(\Omega_2 \mid \mathbb{Z}, F(\Omega_1)) \times \pi(\Omega_1 \mid \mathcal{Y})$$

[2] Holmes and Denison (2003), *Classification with Bayesian MARS*, Machine Learning (50) 159–173.

[3] Friedman (1991), *Multivariate adaptive regression splines*, The Annals of Statistics (19) 1–61.

Prediction

- Ultimately interested in predicting a new patient's survival status given his/her imaging data

Posterior Predictive Expectation

$$\mathrm{E}(Z_{\text{new}} \mid \mathcal{Y}_{\text{new}}, \mathbf{Z}, \mathcal{Y}) = \int \pi(Z_{\text{new}} = 1 \mid \mathcal{Y}_{\text{new}}, \Omega)\pi(\Omega \mid \mathcal{Y}, \mathbf{Z})d\Omega$$

  - $\Omega = \{\Omega_1, \Omega_2\}$

- We will use cross-validation to assess model

Prediction of tumor response in the contralateral hemisphere

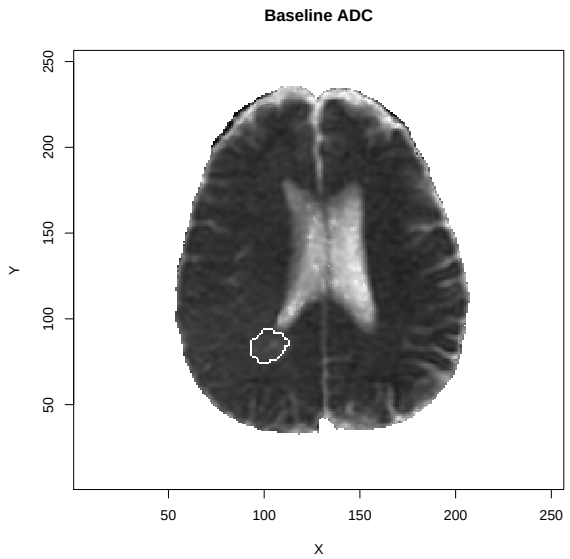
- Would like to compare tumor response under treatment vs. under no treatment
  - impossible

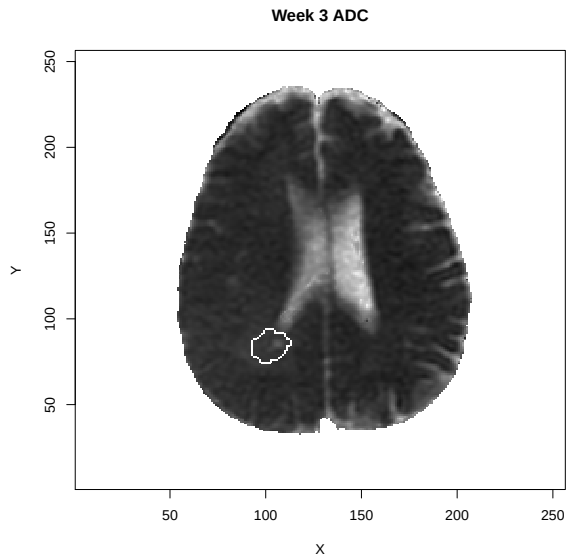
Prediction of tumor response in the contralateral hemisphere

- Would like to compare tumor response under treatment vs. under no treatment
  - impossible
- Next best: compare tumor response under treatment vs. tumor response as though it responds similar to healthy tissue in the contralateral hemisphere of the brain
  - Since this is not observed, we predict it

Prediction of tumor response in the contralateral hemisphere

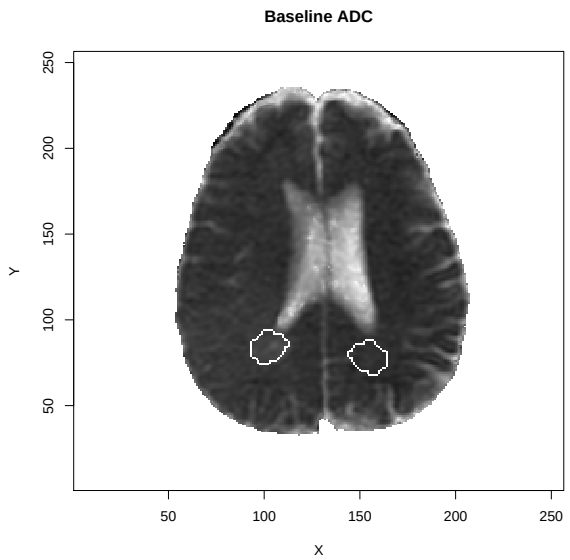
- Would like to compare tumor response under treatment vs. under no treatment
  - impossible
- Next best: compare tumor response under treatment vs. tumor response as though it responds similar to healthy tissue in the contralateral hemisphere of the brain
  - Since this is not observed, we predict it

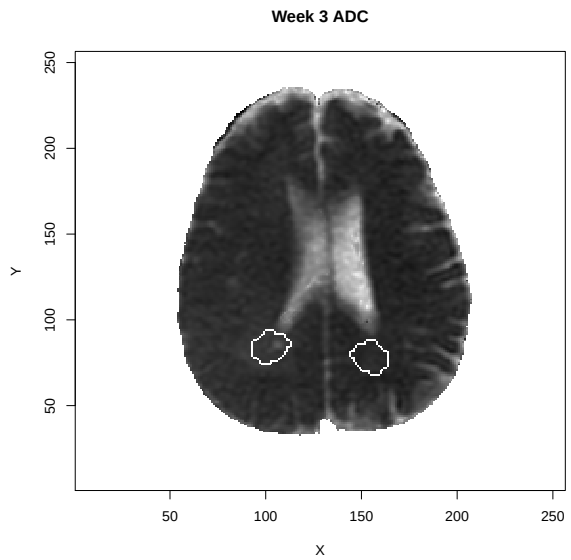




Prediction of tumor response in the contralateral hemisphere

- Would like to compare tumor response under treatment vs. under no treatment
  - impossible
- Next best: compare tumor response under treatment vs. tumor response as though it responds similar to healthy tissue in the contralateral hemisphere of the brain
  - Since this is not observed, we predict it





Summary Statistics

- Kullback-Leibler Divergence between
  - estimated change in tumor means (FDM or FPM) and predicted change in contralateral hemisphere
- Conditional diffusion (perfusion) statistic:
  - conditional distribution given spatial information
  - prop. of week 3 tumor voxel means > 0.975 (diffusion) or < 0.025 (perfusion) quantile of the conditional predictive mean distr. in contralateral hemisphere

Algorithm Highlights

- Latent variable representation[4]
  - transforms probit model into a (marginally) equivalent linear model
- RJMCMC[5]
  - number of MARS basis is unknown and random
  - integrate regression coefficients out of joint likelihood
- Importance sampling for c.v. [6]
  - only run algo. once with full data
- run algorithm for 100K iterations, burnin of 50K
  - oversample stage II 10:1
  - roughly 20 hours on a 3.0GHz Mac Xserve server
  - vast majority of computation spent in stage I

---

[4] Albert and Chib (1993), *Bayesian analysis of binary and polychotomous response data*, JASA (88) 669–679.

[5] Green (1995), *Reversible jump Markov chain Monte Carlo computation and Bayesian model determination* , Biometrika (82) 711–732.

[6] Gelfand, Dey, Chang (1992), *Model determination using preditive distributions with implementation via sample-based methods*, Bayesian Statistics 4, 147–167.

Comparison with Simpler Models

$$\text{If } \Pr(Z_j = 1 \mid Z_{-j}, \mathcal{Y}) > 0.5, \text{ then predict } Z_j = 1$$

- Using all summary statistics

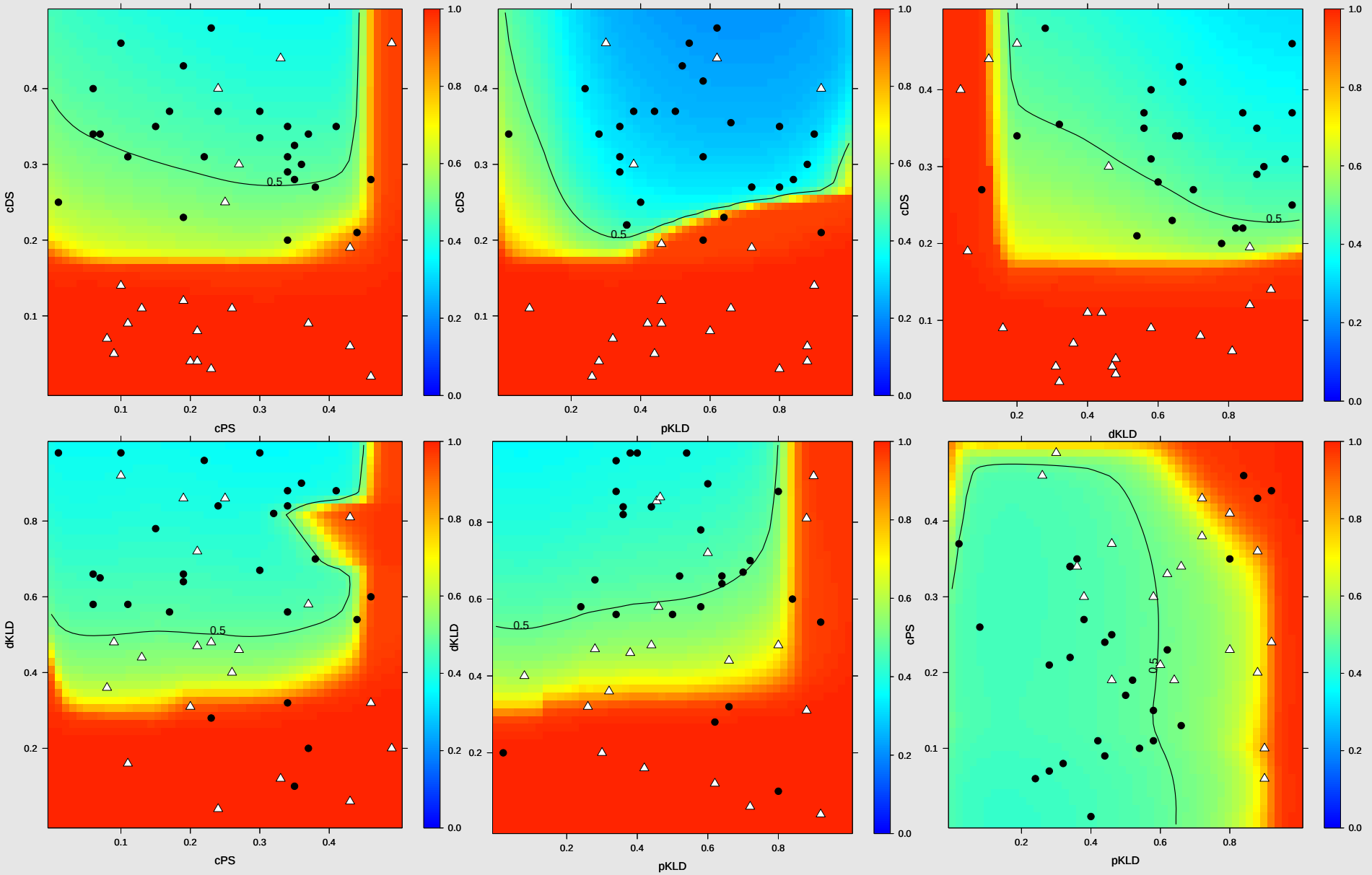
Model	[1]CCR_{cv}
Bayesian joint model	0.79
Separate models (stage I + GLM)	0.62
fDM/fPM + GLM	0.63

[1] Correct cross-validated classification rate.

- Only using the Kullback-Leibler statistics

Model	CCR_{cv}
Bayesian joint model	0.72
Single model (Obs. data + GNLM)	0.64

Marginal Decision Boundaries: $\Pr(Z_{new} = 1 \mid \mathbf{Z}, \mathcal{Y}, \mathbf{X}_{new,i,j}) = 0.5$



Remarks

- Manuscript to appear:
  - Wu and Johnson (2011), *Predicting treatment efficacy via Quantitative MRI: a Bayesian joint model*, JRSSC (in press).
    - currently available at `http://www.bepress.com/umichbiostat/paper86`
- Accounting for spatial correlation and complex decision boundary increases prediction rates over simpler models
- Summary statistics may not be ideal—more work is needed with collaborators to define better summaries
  - currently reducing a large amount of data to a few summary values
  - perhaps a larger vector would afford better prediction
- Currently small trials under way to determine if qMRI can be used in other tumors
  - breast cancer
  - prostate cancer bone metastases
  - sarcomas

Acknowledgements

- Jincao Wu, NCI
- Brian D. Ross, Department of Radiology
- Charles R. Meyer, Department of Radiology
- Thomas L. Chenevert, Department of Radiology
- Craig J. Galbán, Department of Radiology
- Christina Tsien, Department of Radiation Oncology
- Daniel A. Hamstra, Department of Radiation Oncology
- The Center for Molecular Imaging, UofM