

Trial design in the presence of non-exchangeable subpopulations

Brian P. Hobbs, PhD

Cancer Biostatistics Section Head in The Taussig Cancer Institute
Associate Staff, Department of Quantitative Health Sciences in The
Lerner Research Institute



April 2018

Methodology Research Overview

- ① Methodology and applications: cancer imaging and radiotherapy
 - radiomic-derived immune phenotypes
 - texture analysis
 - perfusion markers
 - radiation dosimetrics
- ② Methodology: precision medicine
 - personalized treatment selection
 - characterizing local benefit
 - comparing biomarker-guided strategies
- ③ Methodology: trial design
 - Basket Trials
 - Multi-source Adaptive Designs
 - Platform Design
 - Safety & efficacy designs based on Total Toxicity Burden

Limitations with current translational paradigm

Translation paradigm was not formulated for precision medicine

- population-average effects with the extent of patient heterogeneity with respect to tumor micro-environment, tumor ecology, and immune-environment

Recent failures elucidate intrinsic limitations with:

- non-informative phase II trials that ignore prognostic heterogeneity
- population-average treatment comparisons that enroll heterogeneous subpopulations
- surrogate endpoint selection (pseudo-progression with immunotherapies)
- absence of projections of phase III success with mediation models

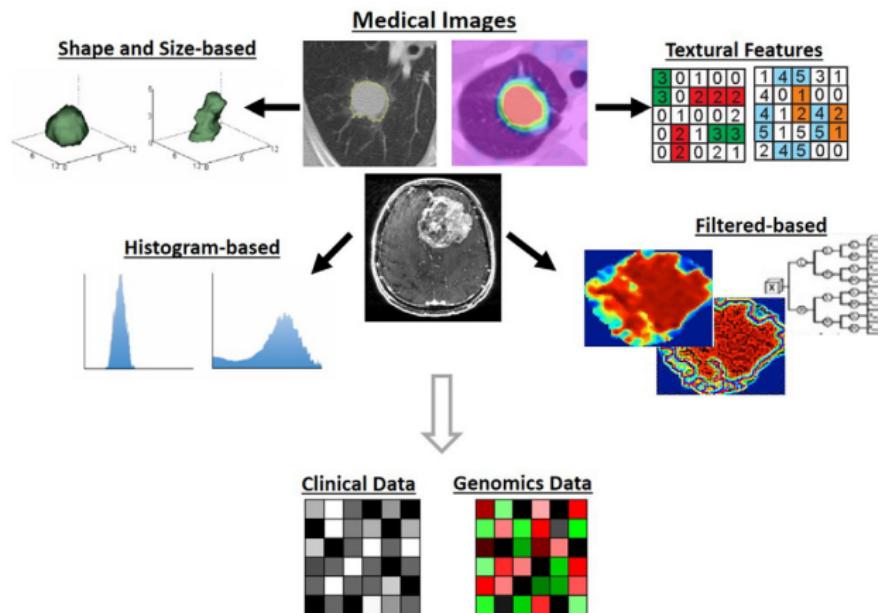
Methodological Solutions

Data science may address contemporary issues through:

- ① trial designs that estimate exchangeable effects for pre-specified subtypes based on known sources of heterogeneity
- ② characterizing and elucidating unknown sources of heterogeneity (immune-phenotypes)
- ③ single-arm trials that use prognostic models to assess counterfactuals (predictive probability)
 - comparative projections within sub-populations in advance of confirmatory trials elucidating eligibility criteria as well as during accrual with adaptively designs that delineate biomarker domains predictive of response
- ④ understanding treatment selection and surrogate endpoint selection for immunotherapies

Methodology Cancer Radiomics

High-throughput extraction of quantitative image features



Example Texture features

Table 1: Commonly used GLCM-based features.

Statistic	Formula
Correlation	$\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p_{ij}}{\sigma_i \sigma_j}$
Energy	$\sum_{i,j} p(i,j)^2$
Contrast	$\sum_{i,j} i - j ^2 p(i,j)$
Entropy	$\sum_{i,j} p(i,j) \log(p(i,j))$
Homogeneity	$\sum_{i,j} \frac{p(i,j)}{1 + i - j ^2}$

Notes:

p_{ij} = $(ij)^{th}$ entry in normalized GLCM

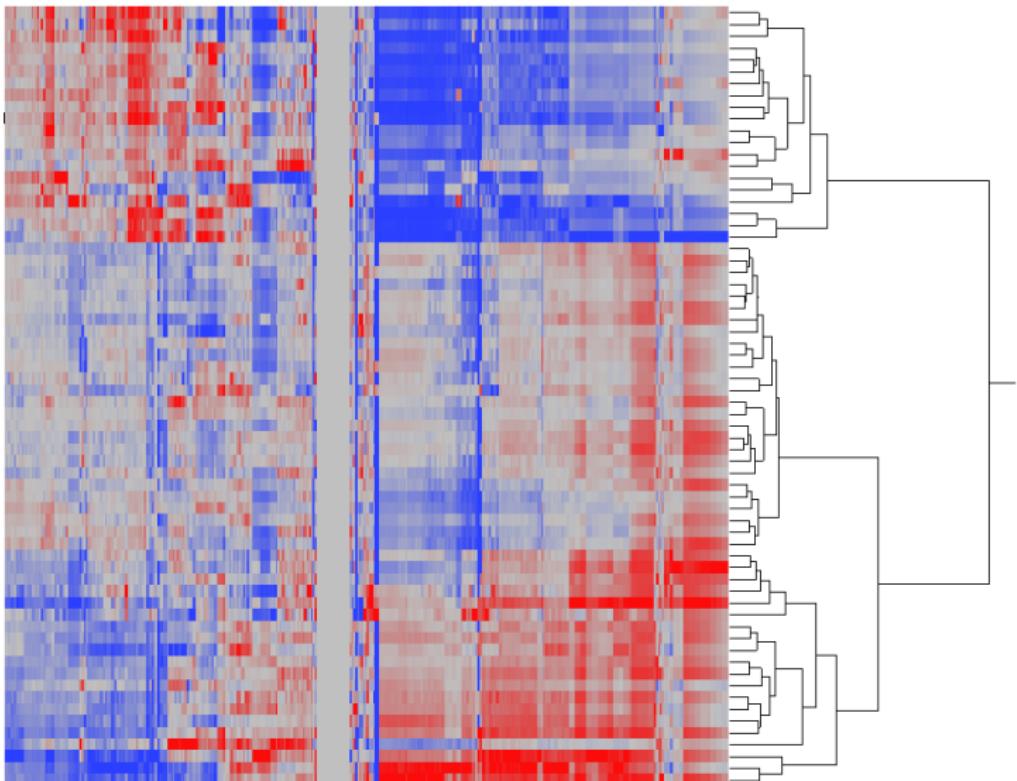
μ_i = mean of row i in normalized GLCM

μ_j = mean of column j in normalized GLCM

σ_i = standard deviation of row i in normalized GLCM

σ_j = standard deviation of column j in normalized GLCM

515 Radiomics NSCLC Features produced by IBEX



Cancer Radiomics

Current analytical frameworks inherently limited:

- Reductive mapping of the complex patterns
 - Redundant sets of highly correlated features
 - Univariate/regression models applied
 - Variable selection with each application
 - Limited Reproducibility
-
- **Absent:** multivariate inferential modeling strategies that captures the complex patterns of interest
 - **derive from foundational probability and statistics**

Beyond “Radiomics-based” Feature Summaries

Spatial Bayesian Models of GLCM

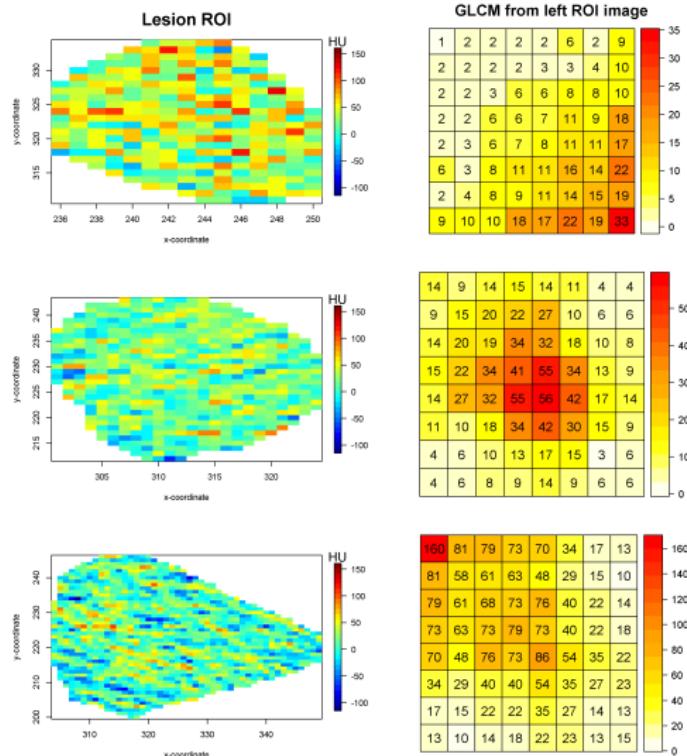
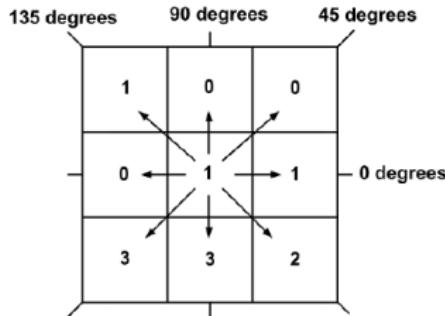


Figure 1: Illustrative patterns of GLCM in adrenal lesions from CT imaging for 3 representative subjects. Pixel-level ROIs on the left, GLCMs on the right. The color



How $\#(1,0)$ is calculated

Gray Level	0	1	2	3
0	$\#(0,0)$	$\#(0,1)$	$\#(0,2)$	$\#(0,3)$
1	$\#(1,0)$	$\#(1,1)$	$\#(1,2)$	$\#(1,3)$
2	$\#(2,0)$	$\#(2,1)$	$\#(2,2)$	$\#(2,3)$
3	$\#(3,0)$	$\#(3,1)$	$\#(3,2)$	$\#(3,3)$

The general form of GLCM

- A matrix defined over an image to be the distribution of co-occurring gray-scale pixel values at the given distance and angle, also known as spatial dependence matrix
- The $(i,j)^{th}$ entry of GLCM represents how often a pixel with gray level i occurs either horizontally, vertically or diagonally to adjacent pixels with gray level j
- Mapping the irregular ROI to regular grid

Supervised Learning: Bayesian Predictive Modeling of GLCM objects

Spatial Bayesian Models of GLCM

Bayesian Hierarchical model and Posterior inference

Denote $\mathbf{y}_i = (y_i(s_1), y_i(s_2), \dots, y_i(s_n))'$, values of normalized GLCM counts for the i^{th} patient; $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$, the p -dimensional coefficient vector; \mathbf{x}_i the $n \times p$ dimensional patient-level design matrix.

The observed normalized GLCM for patient i is assumed to follow Gaussian Distribution:

$$\mathbf{y}_i | \boldsymbol{\beta}, \boldsymbol{\eta}_i, \tau_\epsilon \sim MVN(\mathbf{x}_i \boldsymbol{\beta} + \boldsymbol{\eta}_i, \tau_\epsilon^{-1} \mathbf{I})$$

where $\boldsymbol{\eta}_i = (\eta_i(s_1), \eta_i(s_2), \dots, \eta_i(s_n))'$ defines the spatial random field capturing the spatial dependencies among the GLCM cells.

We further define a Gaussian Markov Random Field by specifying a set of full conditionals distributions for $\boldsymbol{\eta}_i$:

$$\begin{aligned} \eta_i(s_k) | \eta_i(s_1), \dots, \eta_i(s_{k-1}), \eta_i(s_{k+1}), \dots, \eta_i(s_n) \\ \sim N\left(\sum_{d \neq k} \frac{W_{kd} \eta_i(s_d)}{B_k}, \frac{1}{\tau_\eta B_k}\right) \end{aligned}$$

where W_{kd} indicating whether lattice elements k and d are adjacent neighbors and $B_k = D_k + q$, where D_k denotes the number of neighbors for lattice k , and $q > 0$ a diagonal offset term ensures the properness of the joint distribution of $\boldsymbol{\eta}_i$.

Hierarchical model specification is complete with prior distributions for the parameters:

$$\begin{aligned} \boldsymbol{\beta} &\sim MVN(\boldsymbol{\beta} | \mathbf{0}, 1000 * \mathbf{I}) \\ \tau_\epsilon &\sim \text{Gamma}(\tau_\epsilon | 0.001, 0.001) \\ \tau_\eta &\sim \text{Gamma}(\tau_\eta | 0.1, 0.1) \\ q &\sim \text{Gamma}(q | 1, 100) \end{aligned}$$

Spatial Bayesian Models of GLCM

Denote the observed GLCM of a new, heretofore unclassified object by \mathbf{y}_{N+1} with potential covariates \mathbf{x}_{N+1} . Additionally, let $\mathbf{c} = \{c_1, \dots, c_h\}$ denote the set of all possible classes to which object \mathbf{y}_{N+1} could be assigned. The classification probability for class c_k is:

$$\frac{P(c = c_k \mid \mathbf{y}_{N+1}, \mathbf{y}, \mathbf{x}_{N+1}, \mathbf{x}) =}{\sum_{l=1}^h P(\mathbf{y}_{N+1} \mid \mathbf{y}, \mathbf{x}_{N+1}, \mathbf{x}, c = c_l) P(c = c_l)}$$

where $P(\mathbf{y}_{N+1} \mid \mathbf{y}, \mathbf{x}_{N+1}, \mathbf{x}, c = c_k) =$

$$\int P(\mathbf{y}_{N+1} \mid \boldsymbol{\eta}_{N+1}^{c_k}, \mathbf{y}^{c_k}, \mathbf{x}_{N+1}, \mathbf{x}^{c_k}, \boldsymbol{\beta}^{c_k}, \tau_{\epsilon}^{c_k}, \tau_{\eta}^{c_k}, q^{c_k})$$

$$P(\boldsymbol{\beta}^{c_k}, \boldsymbol{\eta}_{N+1}^{c_k}, \tau_{\epsilon}^{c_k}, \tau_{\eta}^{c_k}, q^{c_k} \mid \mathbf{y}^{c_k}, \mathbf{x}^{c_k})$$

$$d(\boldsymbol{\beta}^{c_k}, \boldsymbol{\eta}_{N+1}^{c_k}, \tau_{\epsilon}^{c_k}, \tau_{\eta}^{c_k}, q^{c_k})$$

- Li, Guindani, Ng, Hobbs (2017) Classification of adrenal lesions through spatial Bayesian modeling of GLCM, *IEEE ISBI Proceedings*
- Li, Guindani, Ng, Hobbs (2017) Spatial Bayesian modeling of GLCM with application to malignant lesion characterization, *submitted*

**Retrospective Interrogation of $n = 379$ CT scans of
Adrenal Lesions with pathology verification
(January 2001 to January 2010)**

Case study: Adrenal Lesions

Predictive accuracy of single-scan binary classifiers

- BSGC = Bayesian Spatial Gaussian Classifier (trained on GLCM object)
- versus regression and machine learning techniques trained on radiomics features

Table 2: **Case study results.** Classification accuracy obtained from prediction under LOOCV for non-contrast (NC) and delay (DL) scans, separately

	BSGC	Logistic Regression	SVM	ANN	Random Forest
NC	0.809	0.514	0.505	0.524	0.562
DL	0.771	0.571	0.467	0.538	0.495

Simulation Study

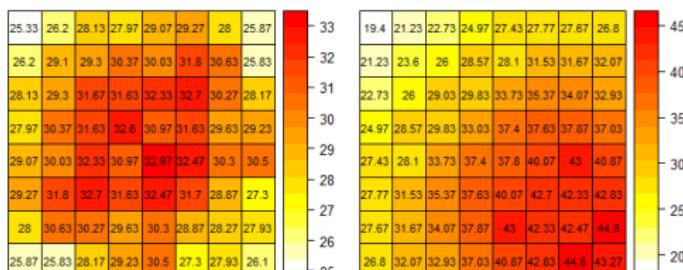
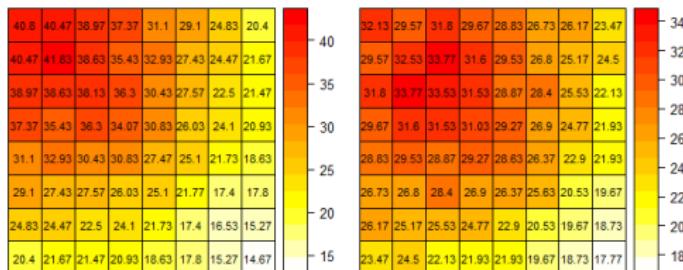
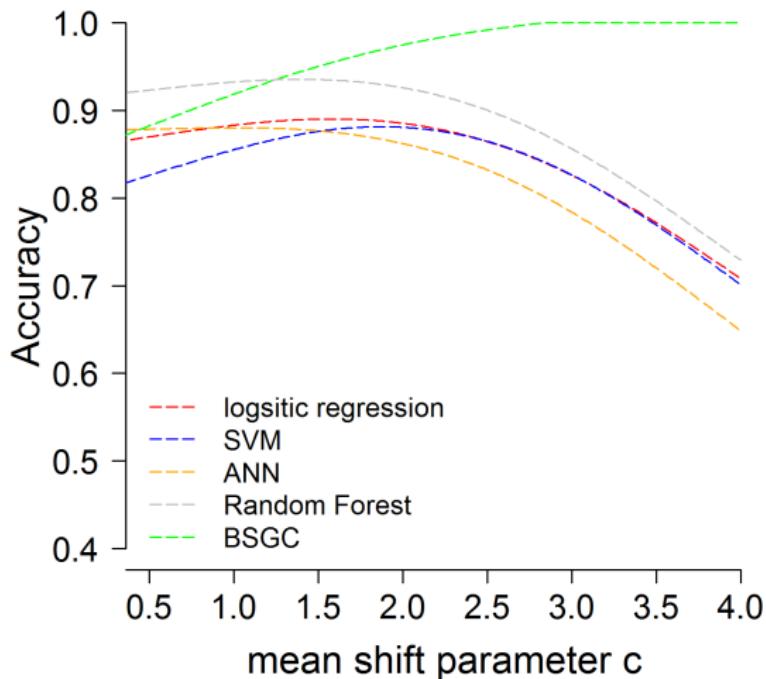


Figure 2: Simulation scenarios for comparing GLCM-based classifiers. Each figure depicts the assumed mean rate surface used to simulate GLCM-derived patterns with different choices of c under $s = 10$. The upper-left is generated with $c = 0$, the upper-right with $c = 10$.

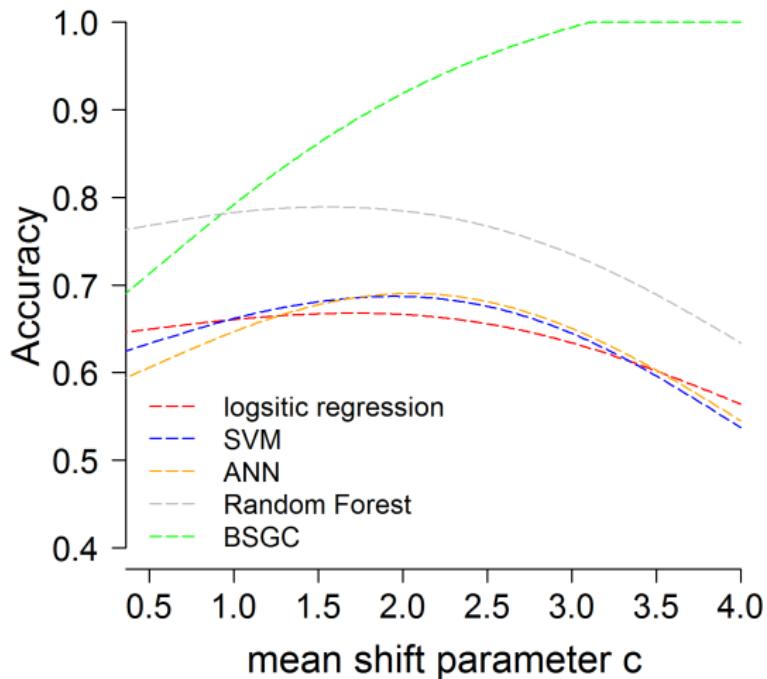
Simulation Study

Simulation Results $s=5$



Simulation Study

Simulation Results $s=12$



Unsupervised Learning: Bayesian Nonparametric Clustering of GLCM objects

Hierarchical Rounded Gaussian Spatial Dirichlet Process model

Dirichlet Process

- A flexible, nonparametric prior over an infinite number of classes as well as the parameters for those classes.
 - Release parametric assumption
 - Clustering
- The Dirichlet Process (DP) is a distribution over distributions.
 - $G \sim DP(\alpha, G_0)$, where α is the concentration parameter, G_0 is the base measure.
 - G is a distribution drawn from the DP.
- **Conjugacy of DP:**
$$G|\eta_1, \eta_2, \dots, \eta_n \sim DP\left(\alpha + n, \frac{\alpha}{\alpha+n} G_0 + \frac{n}{\alpha+n} \frac{\sum_{i=1}^n \delta_{\eta_i}}{n}\right).$$
- **Discreteness of DP:** samples from a DP are discrete with probability 1, $G = \sum_{l=1}^{\infty} w_l \delta_{\theta_l}$, where w_l is the weight and δ_{θ_l} is a Dirac delta at θ_l and $\theta_l \sim G_0$.

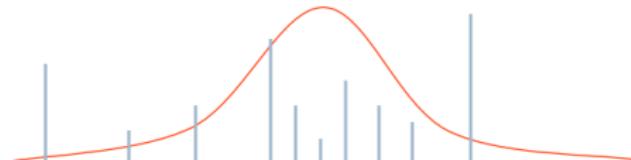
More on Dirichlet process

Dirichlet Process

- ▶ Consider Gaussian G_0



- ▶ $G \sim DP(\alpha, G_0)$

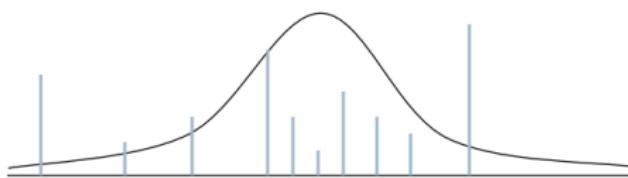


1

More on Dirichlet process

Dirichlet Process

- ▶ $G \sim DP(\alpha, G_0)$



- ▶ G_0 is continuous, so the probability that any two samples are equal is precisely zero.
- ▶ However, G is a discrete distribution, made up of a countably infinite number of point masses [Blackwell]
 - ▶ Therefore, there is always a non-zero probability of two samples colliding

2

Rounded Kernel

- Denote $\mathbf{z}_t = (z_t(s_1), z_t(s_2), \dots, z_t(s_n))'$ as the counts vector from GLCM of the t th patient

Rounded Kernel

- Denote $\mathbf{z}_t = (z_t(s_1), z_t(s_2), \dots, z_t(s_n))'$ as the counts vector from GLCM of the t th patient
- $h(\cdot)$ is the mapping function such that $\mathbf{z}_t = h(\mathbf{y}_t)$, where \mathbf{y}_t is the latent continuous vector associated with observed counts vector \mathbf{z}_t

Rounded Kernel

- Denote $\mathbf{z}_t = (z_t(s_1), z_t(s_2), \dots, z_t(s_n))'$ as the counts vector from GLCM of the t th patient
- $h(\cdot)$ is the mapping function such that $\mathbf{z}_t = h(\mathbf{y}_t)$, where \mathbf{y}_t is the latent continuous vector associated with observed counts vector \mathbf{z}_t
- $P(z_t(s_i) = k)$ can be expressed as the probability that $y_t(s_i)$ lies in the interval $(a_k, a_{k+1}]$, i.e., we have $\underline{z_t(s_i) = k}$ if and only if $\underline{y_t(s_i) \in (a_k, a_{k+1}]}$

Rounded Kernel

- Denote $\mathbf{z}_t = (z_t(s_1), z_t(s_2), \dots, z_t(s_n))'$ as the counts vector from GLCM of the t th patient
- $h(\cdot)$ is the mapping function such that $\mathbf{z}_t = h(\mathbf{y}_t)$, where \mathbf{y}_t is the latent continuous vector associated with observed counts vector \mathbf{z}_t
- $P(z_t(s_i) = k)$ can be expressed as the probability that $y_t(s_i)$ lies in the interval $(a_k, a_{k+1}]$, i.e., we have $z_t(s_i) = k$ if and only if $y_t(s_i) \in (a_k, a_{k+1}]$
- $P(Z(s_1) = k_1, \dots, Z(s_n) = k_n) = \int_{A_k} f(\mathbf{y}(\mathbf{s})) d(\mathbf{y}(\mathbf{s}))$, such that $A_k = \{\mathbf{y}(\mathbf{s}) : a_{k_1} \leq y(s_1) < a_{k_1} + 1, \dots, a_{k_n} \leq y(s_n) < a_{k_n} + 1\}$ defines a disjoint partition of the sample space. The cut-point parameter a_k is chosen to be $a_0 = -\infty$ and $a_k = k - 1$ for $k \in 0, 1, 2, \dots$

Spatial Dirichlet Process model

$$h^{-1}(z_t(s_i)) = y_t(s_i) = \mathbf{x}_t \boldsymbol{\beta} + \gamma_t \theta_t(s_i) + \epsilon_t(s_i)$$

- $y_t(s_i)$, is the foregoing latent continuous variable for subject t at spatial location s_i , generated from observed count variable $z_t(s_i)$.
- \mathbf{x}_t subject level covariates, $\boldsymbol{\beta}$ is the fixed effect parameters, γ_t is scaling covariate to adjust for imaging size.
- $\epsilon_t(s_i)$ is the random noise.
- $\boldsymbol{\theta}_t = (\theta_t(s_1), \theta_t(s_2), \dots, \theta_t(s_n))'$ is the spatial random field for subject t , induced by the Spatial Dirichlet process prior

$$\boldsymbol{\theta}_t \mid G^{(n)} \stackrel{iid}{\sim} G^{(n)}, \quad t = 1, 2, \dots, T$$

$$G^{(n)} \mid v, \sigma^2, q \sim DP(v G_0^{(n)})$$

$$G_0^{(n)}(\cdot \mid \sigma^2, q) = \underline{N_n(\cdot \mid \mathbf{0}, (\mathbf{D} - \rho * \mathbf{W})^{-1} \sigma^2)}.$$

Full Hierarchical model

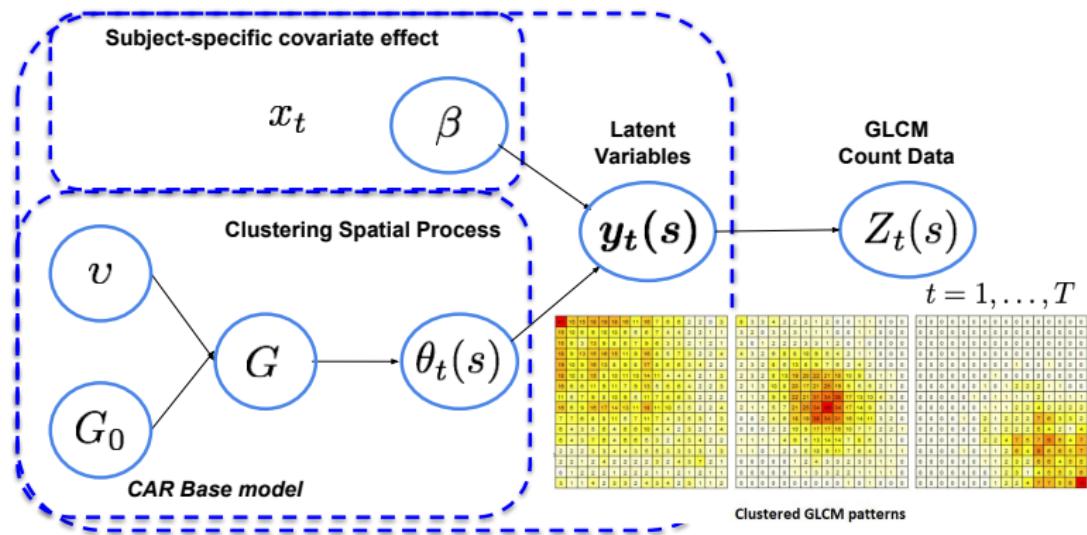
Conditional on the latent \mathbf{y}_t 's, the sampling distribution can be expressed as

$$P(\mathbf{Z}|\mathbf{y}) \propto \prod_{t=1}^T \prod_{i=1}^n I \{ a_{z_t(s_i)} \leq y_t(s_i) < a_{z_t(s_i)+1} \},$$

Then, the hierarchical model on the latent vector \mathbf{y}_t , our hierarchical model can be summarized as follows:

$$\begin{aligned} \mathbf{y}_t | \beta, \theta_t, \tau^2 &\sim MVN(\mathbf{y}_t | \mathbf{x}_t \beta + \gamma_t \theta_t, \tau^2 \mathbf{I}) & \beta &\sim MVN(\beta | \beta_0, \Sigma_\beta) \\ \theta_t | G^{(n)} &\stackrel{iid}{\sim} G^{(n)} & \tau^2 &\sim IG(a_\tau, b_\tau) \\ G^{(n)} | v, \sigma^2, \rho &\sim DP(v G_0^{(n)}) & \sigma^2 &\sim IG(a_\sigma, b_\sigma) \\ G_0^{(n)}(\cdot | \sigma^2, \rho) &\equiv MVN(\cdot | \mathbf{0}, (\mathbf{D} - \rho \mathbf{W})^{-1} \sigma^2) & v &\sim Gamma(a_v, b_v) \\ & & \rho &\sim Uniform(0, 1) \end{aligned}$$

Graphical representation of the proposed HRGSDP model



Following the discreteness property of DP :

$$E(\mathbf{y}(\mathbf{s}) \mid G^{(n)}, \tau^2, \boldsymbol{\beta}) = \underline{\mathbf{x}\boldsymbol{\beta} + \gamma \sum_{l=1}^{\infty} w_l \theta_l(\mathbf{s})}$$

$$var(\mathbf{y}(\mathbf{s}) \mid G^{(n)}, \tau^2, \boldsymbol{\beta}) = \underline{\gamma^2 [\sum_{l=1}^{\infty} w_l \theta_l^2(\mathbf{s}) - (\sum_{l=1}^{\infty} w_l \theta_l(\mathbf{s}))^2] + \tau^2 \mathbf{I}}$$

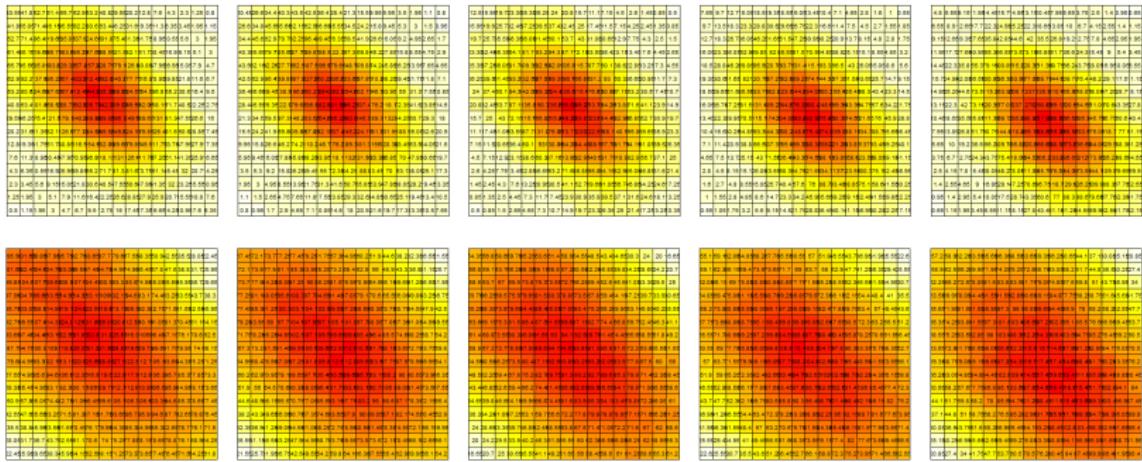
$$cov(y(s_i), y(s_j) \mid G^{(n)}, \tau^2, \boldsymbol{\beta}) = \underline{\gamma^2 [\sum_{l=1}^{\infty} w_l \theta_l(s_i) \theta_l(s_j) - (\sum_{l=1}^{\infty} w_l \theta_l(s_i)) (\sum_{l=1}^{\infty} w_l \theta_l(s_j))]}$$

- The random spatial surface $\{\mathbf{y}(\mathbf{s}) : \mathbf{s} \in D\}$ generated from $G^{(n)}$ has nonconstant variance and is **nonstationary**.
- $f(\mathbf{y}(\mathbf{s}) \mid G^{(n)}, \tau^2, \boldsymbol{\beta})$ can be almost surely expressed as infinite mixture expression $\sum_{l=1}^{\infty} w_l N_n(\mathbf{y}(\mathbf{s}) \mid \mathbf{x}\boldsymbol{\beta} + \gamma \boldsymbol{\theta}_l, \tau^2 \mathbf{I})$ and is **non-Gaussian**.

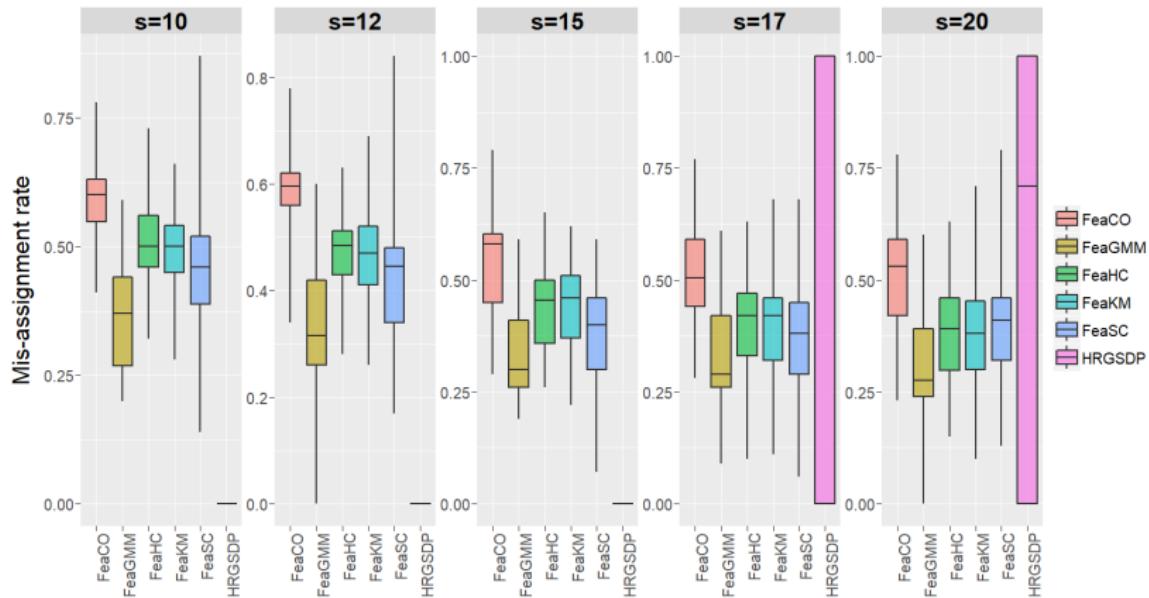
Simulation strategies

- 1 Generating points over 16×16 grided surface with a bivariate normal distribution with $\mu = (2 + c, 14 - c)'$ and
$$\Sigma = s * \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$
- 2 For a given s , we generated GLCMs as a mixture of five components represented by $c \in \{5, 5.5, 6, 6.5, 7\}$.
- 3 We also evaluated performance by varying $s \in \{10, 12, 15, 17, 20\}$, representing the extent to which the noise covers the true signal in the underlying spatial patterns intrinsic to each class.

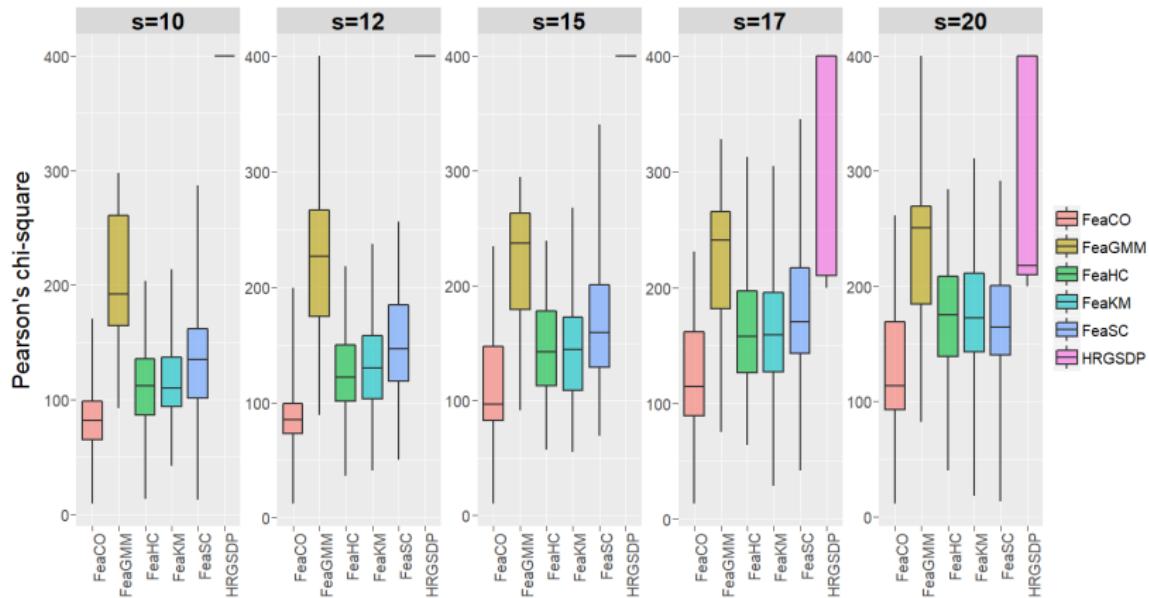
Simulated mean count surfaces for the five components



Each column depicts the mean count surface of the five spatial patterns with different choices of $c \in \{5, 5.5, 6, 6.5, 7\}$, the upper figure is generated under $s = 10$, the lower figure is generated under $s = 20$



Simulation results. Boxplot of mis-assignment rate for all methods. 'FeaCO' stands for feature based [consensus clustering](#), 'FeaGMM' stands for feature based [Gaussian mixture model](#), 'FeaHC' stands for feature based [hierarchical clustering](#), 'FeaKM' stands for feature based [k-means clustering](#), 'FeaSC' stands for feature based [spectral clustering](#), 'HRGSDP' stands for our proposed [Hierarchical Rounded Gaussian Spatial Dirichlet Process](#) approach. $s \in \{10, 12, 15, 17, 20\}$, representing the extent to which the noise covers the true signal in the underlying spatial patterns.



Simulation results. Boxplot of Pearson's χ^2 for all methods. 'FeaCO' stands for feature based **consensus clustering**, 'FeaGMM' stands for feature based **Gaussian mixture model**, 'FeaHC' stands for feature based **hierarchical clustering**, 'FeaKM' stands for feature based **k-means clustering**, 'FeaSC' stands for feature based **spectral clustering**, 'HRGSDP' stands for our proposed **Hierarchical Rounded Gaussian Spatial Dirichlet Process** approach. $s \in \{10, 12, 15, 17, 20\}$, representing the extent to which the noise covers the true signal in the underlying spatial patterns.

Case study: Adrenal Lesions

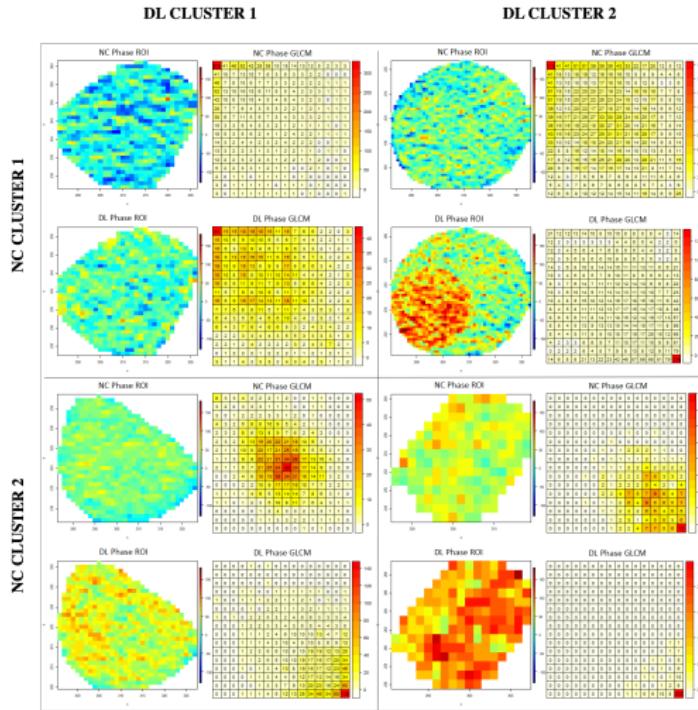
Unsupervised clustering of GLCM objects: Multi-scan associations ($ACC = 0.781$)

Table 1: **Adrenal case study results:** Cross Table for clusters of non-contrast and delay images. The number in the brackets in each cell represents the number of true malignant lesions in that specific subgroup.

		Non-contrast scan	
		cluster 1	cluster 2
Delay scan	cluster 1	59(2)	52(20)
	cluster 2	5(2)	94(72)

- Li, Guindani, Ng, Hobbs (2017) A hierarchical rounded Gaussian spatial Dirichlet process model for GLCM. *in process*

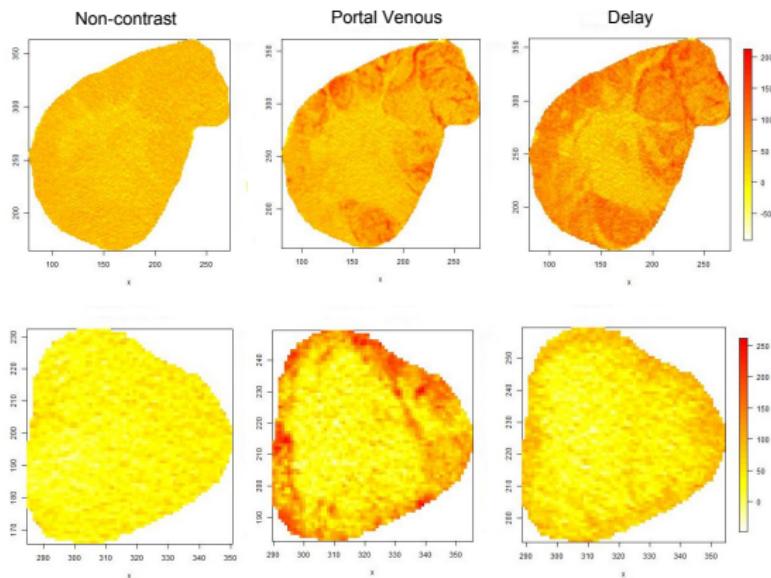
Example of derived subtypes of adrenal lesion



Bayesian Unsupervised Networks:

Integrating Scans and delay time

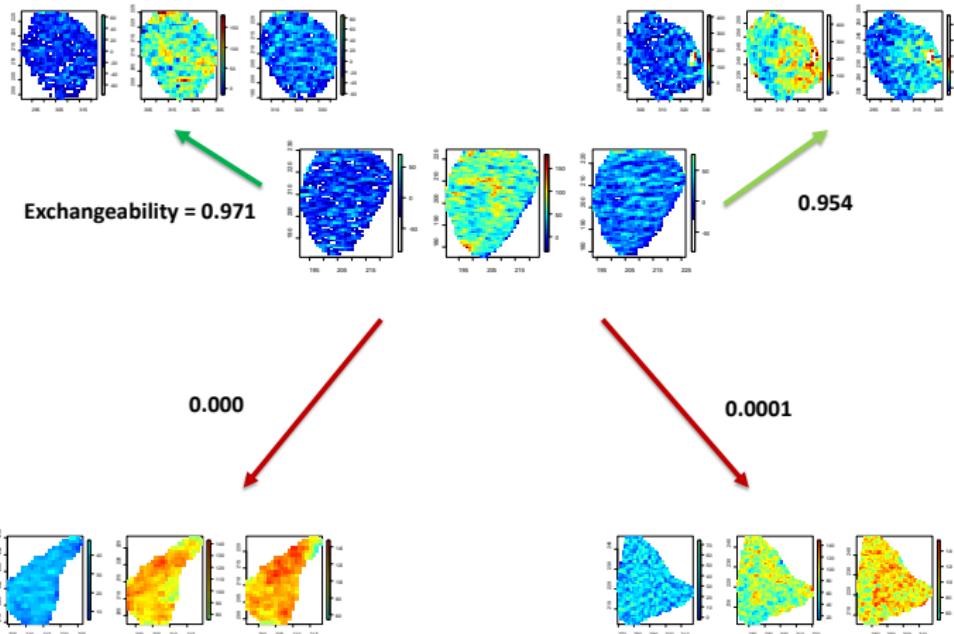
Models of “Washout” for Detection of Adrenal Cancer



Distributions of HU density over three successive scans within two ROIs: one containing a malignant (top) the other a benign (bottom) adrenal lesion. In contrast to the benign lesion, the malignant tissues exhibit relatively high HU density in the noncontrast scan as well as the relative absence of “washout” between portal venous and delay scans.

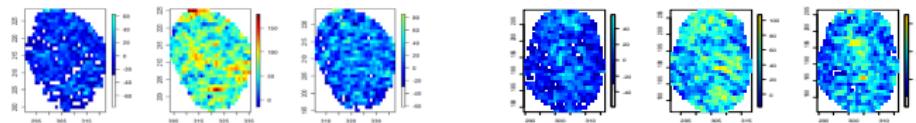
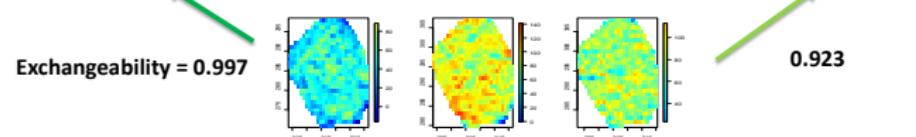
Bayesian Networks of dynamic CT

Bayesian “Similarity” Networks of Enhancement Patterns

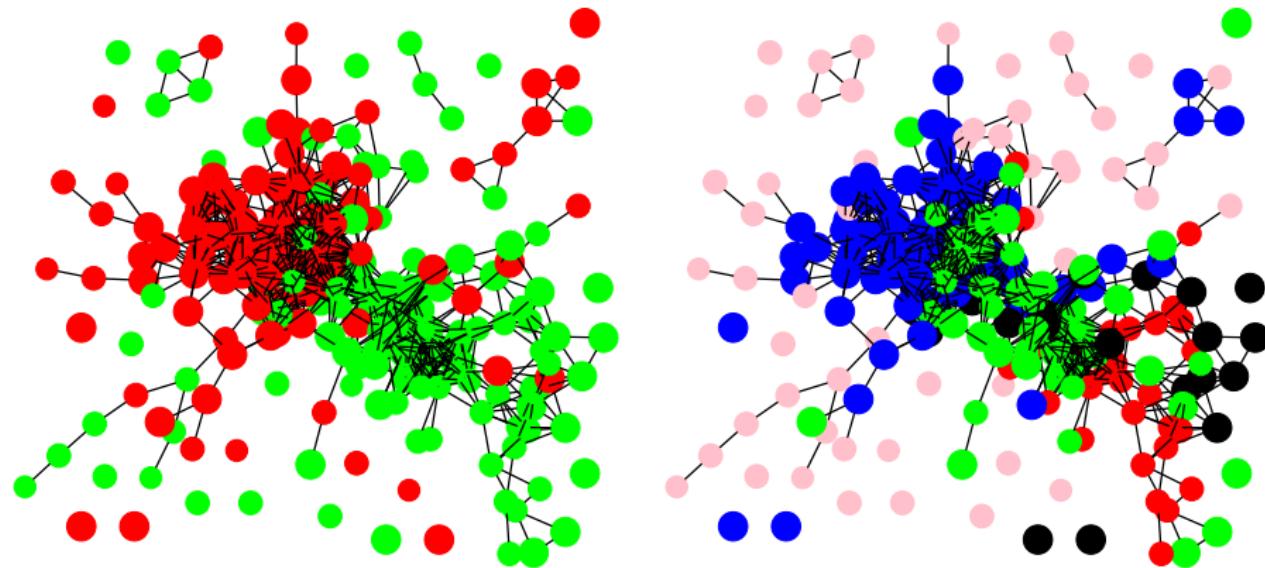


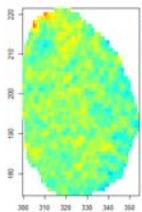
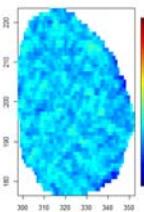
Bayesian Networks of dynamic CT

Bayesian “Similarity” Networks of Enhancement Patterns

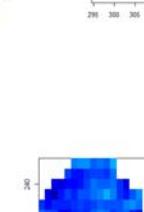
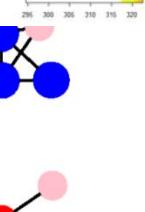
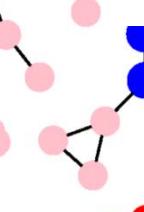
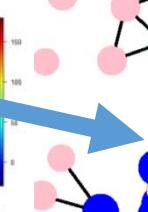
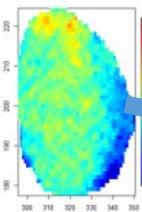


Partial Exchangeability Network of 222 Adrenal Lesions (Hobbs, Chen, Ng; in process)

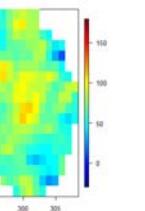
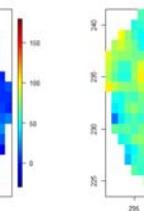
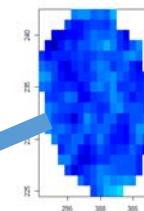




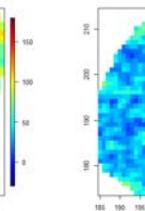
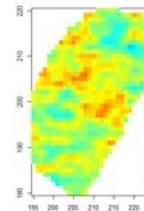
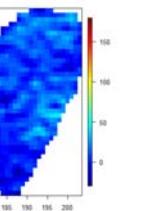
Cluster 5



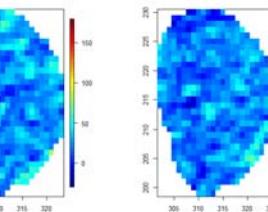
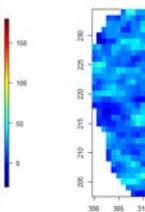
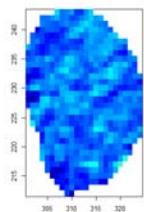
Cluster 2



Cluster 3

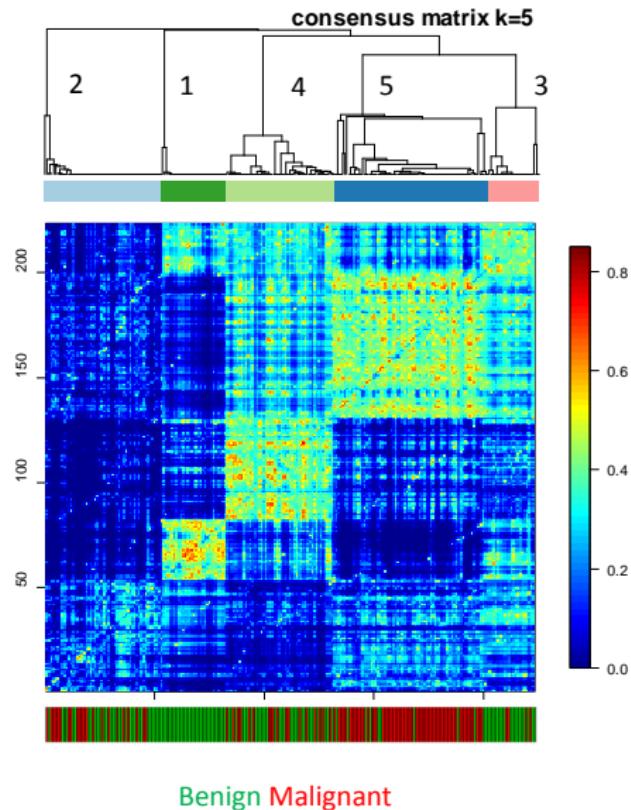


Cluster 4



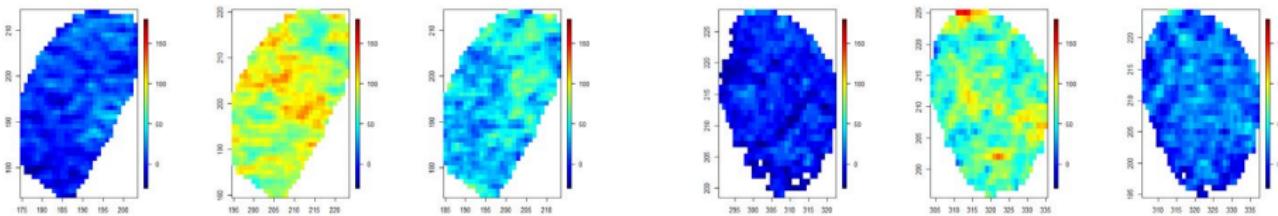
Bayesian Networks of dynamic CT

- 29 of 29 Benign Patients in Cluster 1
-
- 31 of 53 Benign Patients in Cluster 2
-
- 19 of 23 Benign Patients in Cluster 3
-
- 31 of 49 Benign Patients in Cluster 4
-
- 10 of 69 Benign Patients in Cluster 5

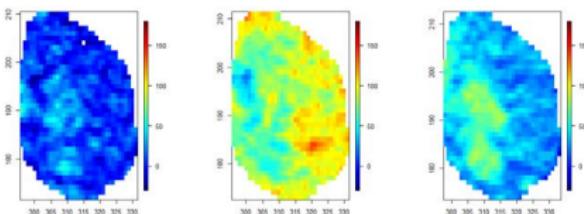


Bayesian Networks of dynamic CT

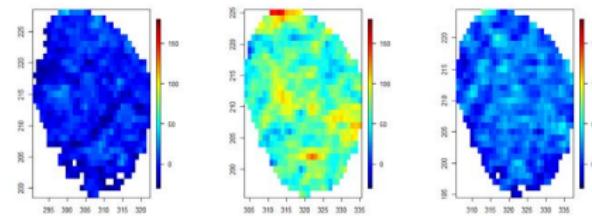
Cluster 1



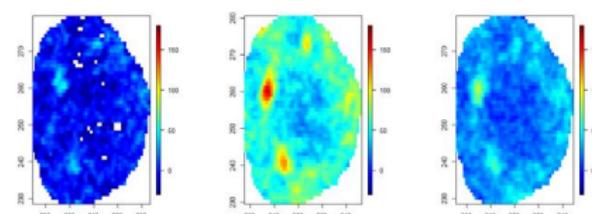
Delay 863



Delay 830



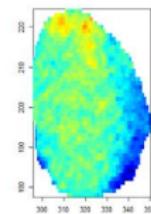
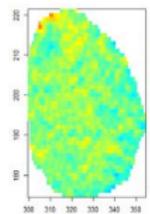
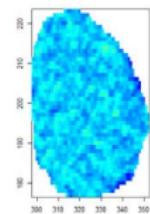
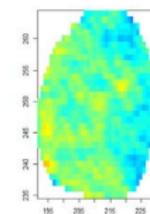
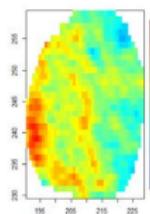
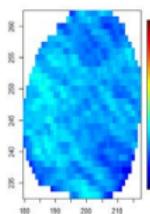
Delay 796



Delay 166

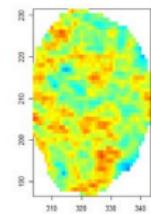
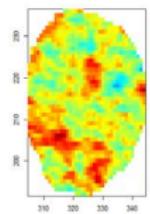
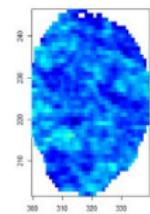
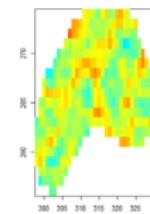
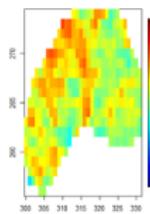
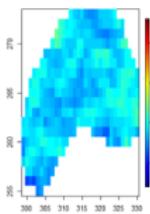
Bayesian Networks of dynamic CT

Cluster 5



Delay 136

Delay 104



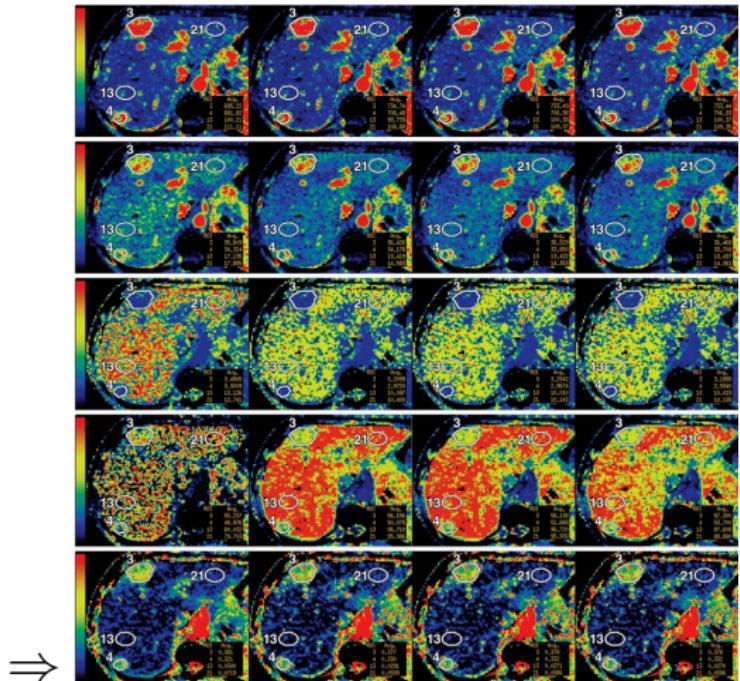
Delay 75

Delay 104

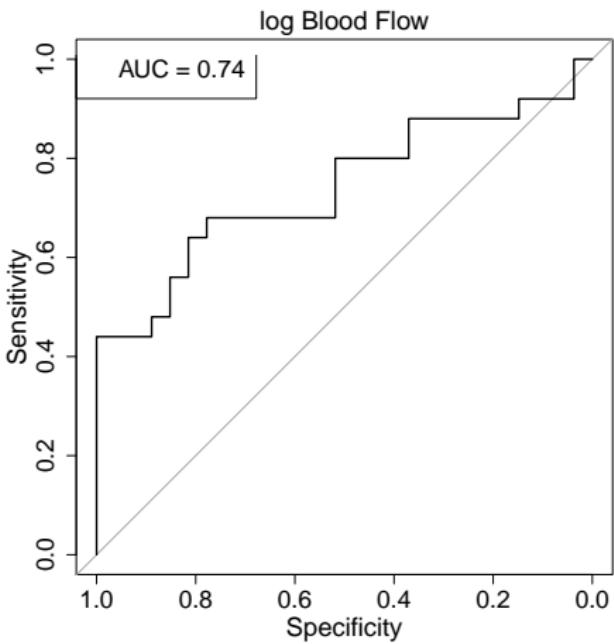
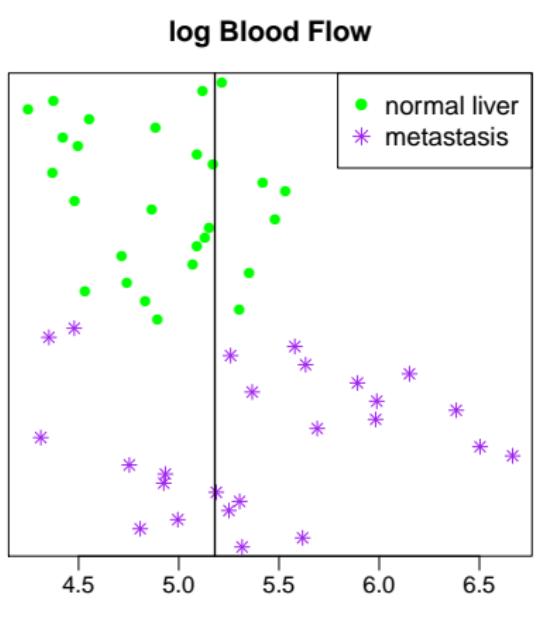
Hepatic Metastatic Detection using CT perfusion

Diagnostic tool enabling quantitative evaluation

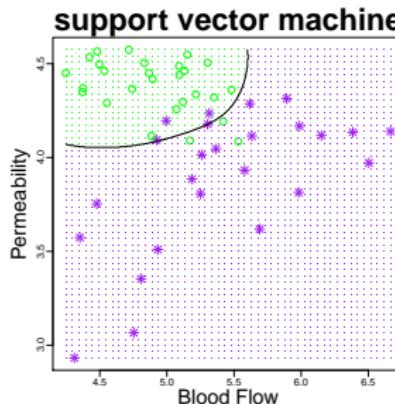
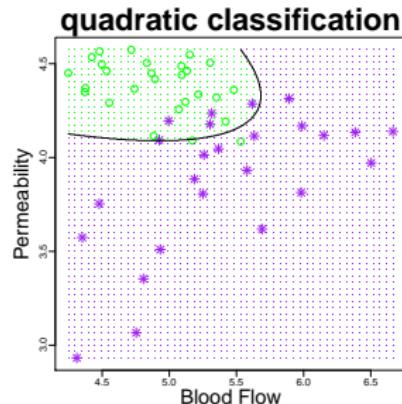
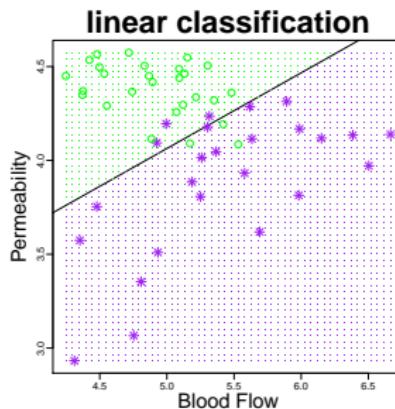
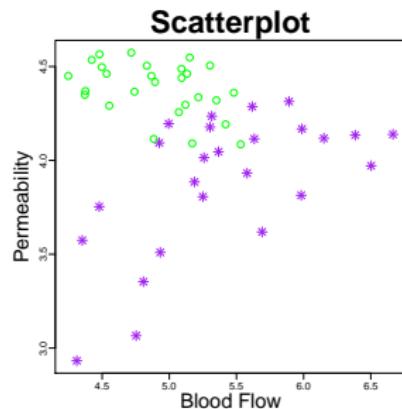
Unenhanced axial CT scan \Rightarrow 5 perfusion characteristics over 4 scans



Current practice in diagnostic radiology

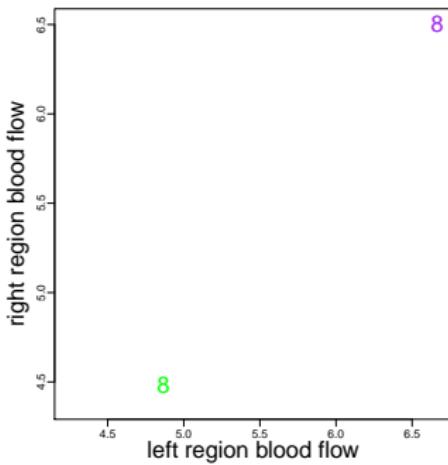
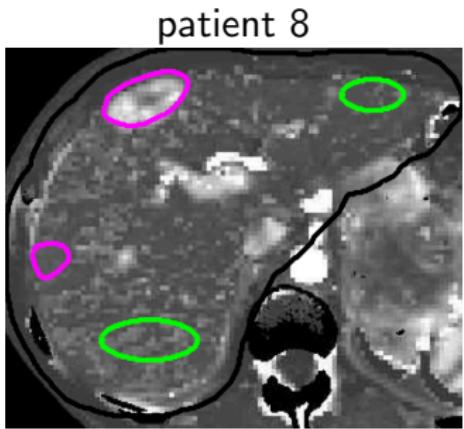


Approaches that leverage between feature dependence

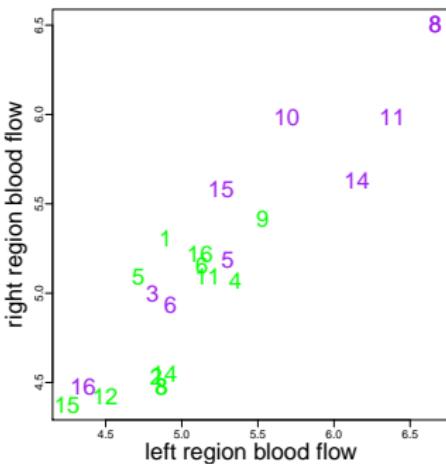
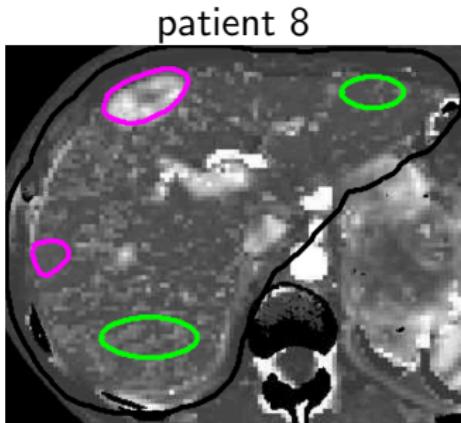


Class assignments are evaluated independently

Inter-Region Correlation



Inter-Region Correlation



Sample Correlation for neighboring regions of the same tissue type

	BF	BV	MTT	PS	HAF
	0.88	0.86	0.72	0.71	0.90

Simultaneous Classification for ROI level inference

	region 1	region 2	region 3	region 4
1	tumor	tumor	tumor	tumor
2	normal	tumor	tumor	tumor
3	tumor	normal	tumor	tumor
⋮	⋮	⋮	⋮	⋮
6	normal	normal	tumor	tumor
7	tumor	normal	normal	tumor
⋮	⋮	⋮	⋮	⋮
16	normal	normal	normal	normal

Wang Y, Hobbs, BP, Hu J, Ng C, Do K, "Predictive Classification of Correlated Targets with Application to Detection of Metastatic Cancer using Functional CT Imaging," *Biometrics*, 2015.

Model

Notation

- z : tissue type; m : number of features
- i : patient index, $i = 1, \dots, N$; j : region index, $j = 1, \dots, n_i$
- y_{ij} : measurement for the patient i region j
- s_{ij} : region location

Distribution:

- $\text{vec}(\mathbf{Y}_i^z) \sim N(\mathbf{1}_{n_i^z} \otimes \boldsymbol{\mu}_z, \boldsymbol{\Psi}_i^z \otimes \boldsymbol{\Sigma}_z)$
- Intra-region:
 - Mean: $\boldsymbol{\mu}_z = \mathbb{E}[y_{ij}|Z = z]$
 - Covariance: $\boldsymbol{\Sigma}_z = \text{cov}(y_{ij}|Z = z)$
- Inter-region:
 - Independence across patients
 - Independence across tissue classes
 - Within the same patient and the same tissue class

$$\text{cov}(y_{ij}, y_{ij'}) = \psi(s_{ij}, s_{ij'}) \boldsymbol{\Sigma}_z$$

Model

Notation

- z : tissue type; m : number of features
- i : patient index, $i = 1, \dots, N$; j : region index, $j = 1, \dots, n_i$
- y_{ij} : measurement for the patient i region j
- s_{ij} : region location

Distribution:

- $\text{vec}(\mathbf{Y}_i^z) \sim N(\mathbf{1}_{n_i^z} \otimes \boldsymbol{\mu}_z, \boldsymbol{\Psi}_i^z \otimes \boldsymbol{\Sigma}_z)$
- Intra-region:
 - Mean: $\boldsymbol{\mu}_z = \mathbb{E}[y_{ij}|Z = z]$
 - Covariance: $\boldsymbol{\Sigma}_z = \text{cov}(y_{ij}|Z = z)$
- Inter-region:
 - Independence across patients
 - Independence across tissue classes
 - Within the same patient and the same tissue class

$$\text{cov}(y_{ij}, y_{ij'}) = \psi(s_{ij}, s_{ij'}) \boldsymbol{\Sigma}_z$$

Note: separable correlation between biomarker and location

Model

Inter-region correlation:

- Independence across patients
- Independence across tissue classes
- Within the same patient and the same tissue class

$$\text{cov}(y_{ij}, y_{ij'}) = \psi(s_{ij}, s_{ij'}) \Sigma_z$$

- For features from identical class we assume:
- Inter-region correlation for identical feature = $\psi(s_{ij}, s_{ij'})$
- Inter-region cross-correlations = intra-region cross-correlation scaled by $\psi(s_{ij}, s_{ij'})$
- separability reduces the degrees of freedom in the covariance from $\frac{1}{2} n_i^z m(n_i^z m + 1)$ to $\frac{1}{2} n_i^z (n_i^z + 1) + \frac{1}{2} m(m + 1)$

Inter-Region Correlation Models

- Compound symmetry

$$\psi(s_{ij}, s_{ij'}; \phi) \equiv \phi$$

- Spatial dependence

- Exponential:

$$\psi(s_{ij}, s_{ij'}; \phi) = \exp\left\{-\frac{d}{\phi}\right\}, \quad d = \text{dist}(s_{ij}, s_{ij'})$$

- Spherical:

$$\psi(s_{ij}, s_{ij'}; \phi) = \begin{cases} 1 - \frac{2}{\pi} \left(\frac{d}{\phi} \sqrt{1 - \left(\frac{d}{\phi}\right)^2} + \sin^{-1} \frac{d}{\phi} \right) & d < \phi \\ 0 & d \geq \phi \end{cases}$$

- General structure

- Anisotropic models, etc.

Simultaneous Bayesian Classification: ROI inference

- Data: train \mathcal{Y} , test \mathcal{Y}_0
- The simultaneous Bayesian classification rule is

$$\hat{\mathbf{d}}_0 = \arg \min_{\mathbf{d} \in \mathcal{D}} \sum_{\mathbf{d}_k \in \mathcal{D}} L_\alpha(\mathbf{d}_k, \mathbf{d}) p(\mathbf{d}_k | \mathcal{Y}_0, \mathcal{Y})$$

- $p(\mathbf{d}_k | \mathcal{Y}_0, \mathcal{Y})$: the joint posterior classification probability
- $L_\alpha(\mathbf{d}_k, \mathbf{d})$: the joint weighted 0-1 loss:

$$L_\alpha(\mathbf{d}_k, \mathbf{d}) = \sum_k \{\alpha \{\text{false negative}\} + (1 - \alpha) \{\text{false positive}\}\}$$

- \mathcal{D} : the set of all the possible class configurations
- requires a prior probability for each possible configurations

$$\Pr(z_{N+1} = \mathbf{d}_k) = p^l (1 - p)^{n_{N+1} - l},$$

where $l = \text{number of tumor ROIs given by } \mathbf{d}_k$.

- Hyperparameter p is fixed at the estimated rate of tumor incidence in the presence of the training data.

Simultaneous Bayesian Classification: ROI inference

- Data: train \mathcal{Y} , test \mathcal{Y}_0
- The simultaneous Bayesian classification rule is

$$\widehat{\mathbf{d}}_0 = \arg \min_{\mathbf{d} \in \mathcal{D}} \sum_{\mathbf{d}_k \in \mathcal{D}} L_\alpha(\mathbf{d}_k, \mathbf{d}) p(\mathbf{d}_k | \mathcal{Y}_0, \mathcal{Y})$$

- $p(\mathbf{d}_k | \mathcal{Y}_0, \mathcal{Y})$: the joint posterior classification probability
- $L_\alpha(\mathbf{d}_k, \mathbf{d})$: the joint weighted 0-1 loss:

$$L_\alpha(\mathbf{d}_k, \mathbf{d}) = \sum_k \{\alpha \{\text{false negative}\} + (1 - \alpha) \{\text{false positive}\}\}$$

- \mathcal{D} : the set of all the possible class configurations
- requires a prior probability for each possible configurations

$$\Pr(z_{N+1} = \mathbf{d}_k) = p^l (1 - p)^{n_{N+1} - l},$$

where $l = \text{number of tumor ROIs given by } \mathbf{d}_k$.

- Hyperparameter p is fixed at the estimated rate of tumor incidence in the presence of the training data.

Maximum *a posteriori* classifier \neq Minimum risk classifier *Bayesian*

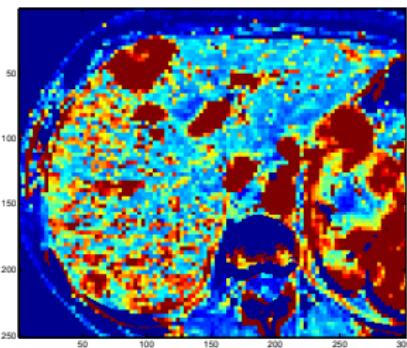
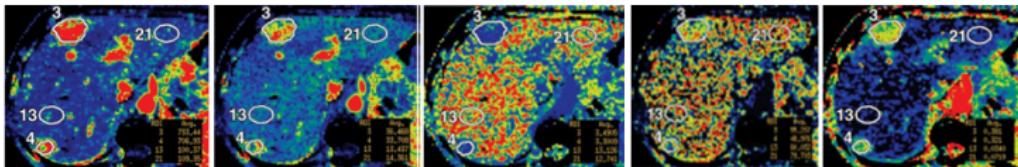
Detection of ROIs with Metastases using Perfusion Biomarkers

- True positive rate (TPR): proportion of correctly identified tumor
- False positive rate (FPR): proportion of falsely identified normal

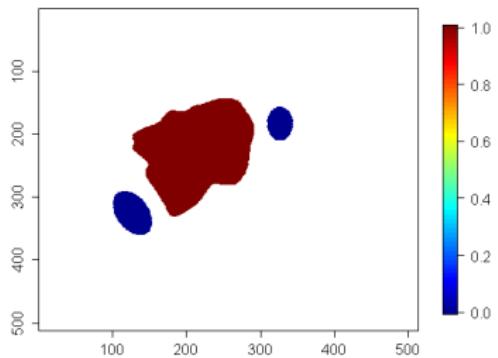
	method	TPR	FPR
Conventional	Bayesian quadratic	0.76	0.18
	Support vector machine	0.88	0.19
Simultaneous	(1) $\alpha = 0.5$ (equal cost)	0.96	0.07
	(1) $\alpha = 0.8$ (prefer FP)	0.96	0.11
	(2) $\alpha = 0.5$ (equal cost)	0.96	0.04
	(2) $\alpha = 0.8$ (prefer FP)	1.00	0.19

(1) compound symmetry; (2) exponential decay

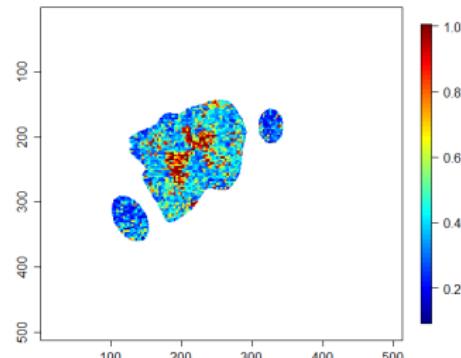
Voxel-level Posterior Probability maps from integration of perfusion features



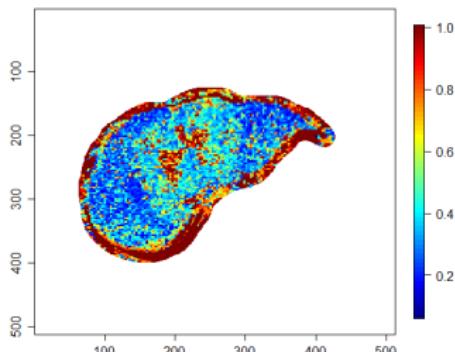
Subject 8, slice 2



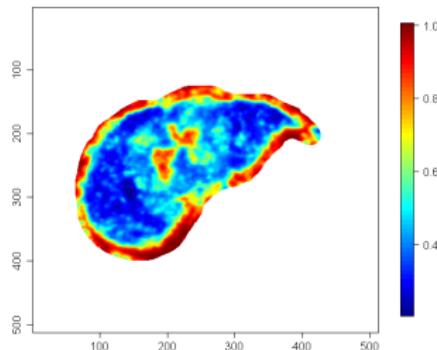
Bayesian discriminant analysis, subject 8, slice 2

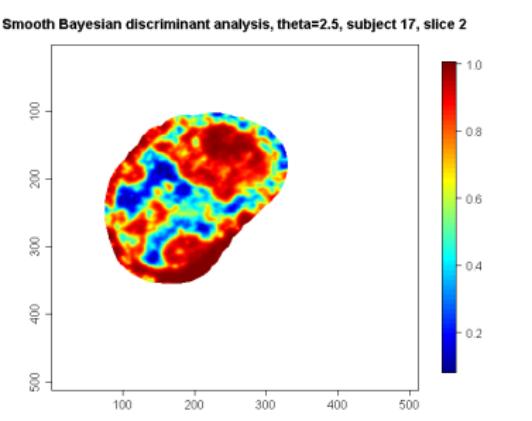
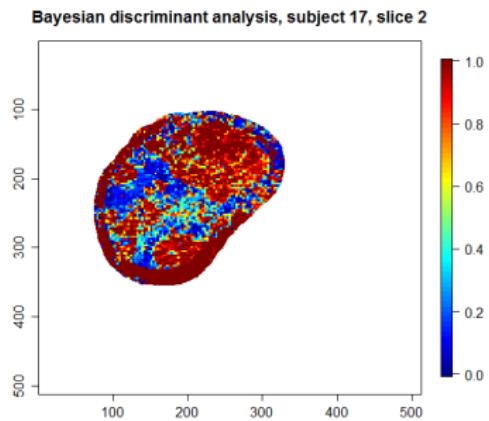
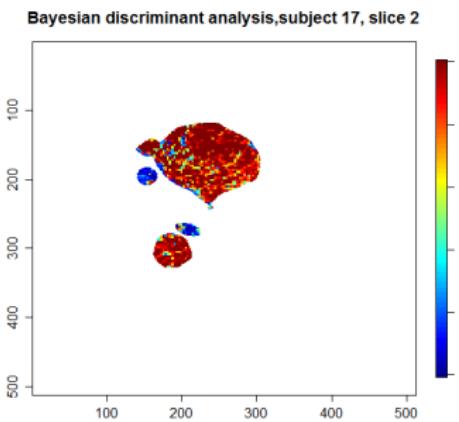
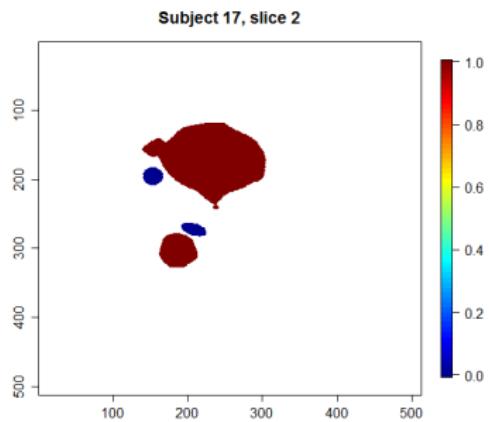


Bayesian discriminant analysis, subject 8, slice 2

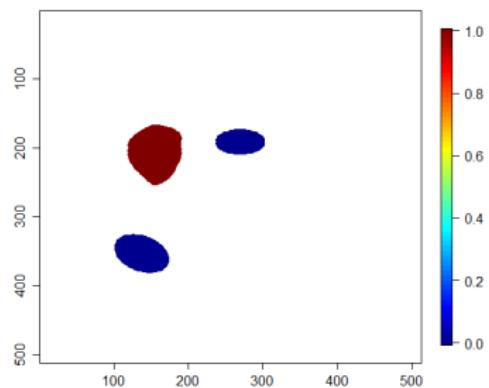


Smooth Bayesian discriminant analysis, theta=2.5, subject 8, slice 2

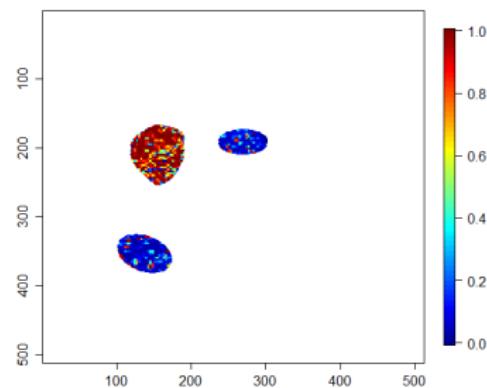




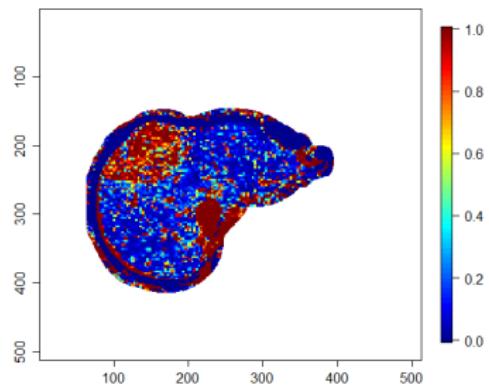
Subject 19, slice 2



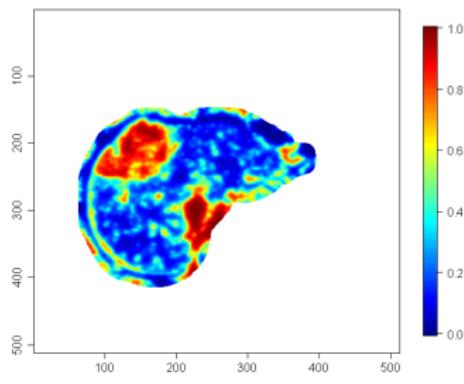
Bayesian discriminant analysis, subject 19, slice 2



Bayesian discriminant analysis, subject 19, slice 2

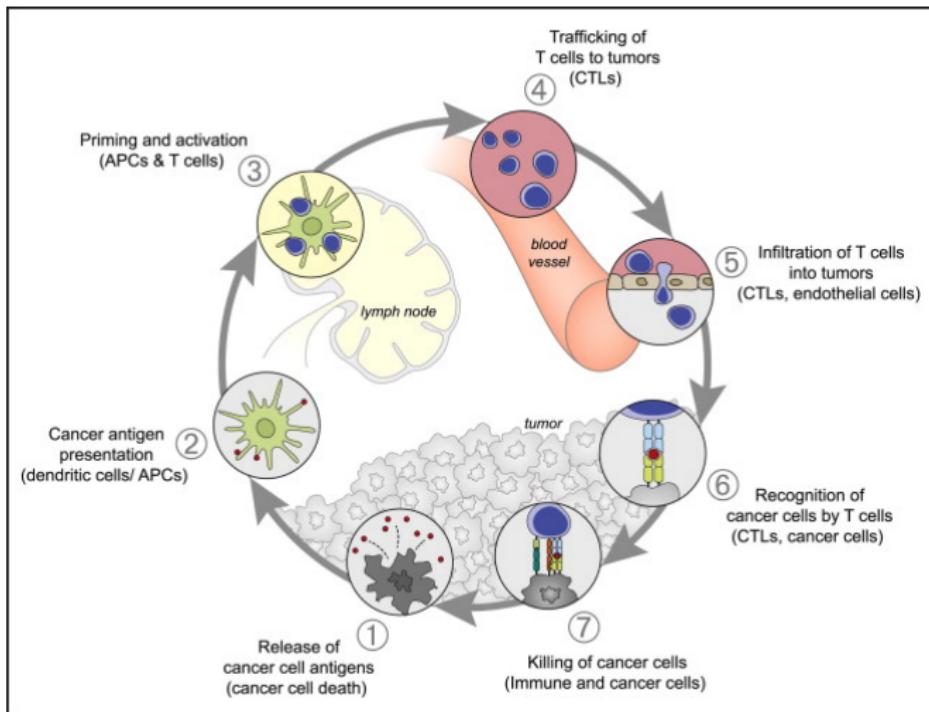


Smooth Bayesian discriminant analysis, theta=2.5, subject 19, slice 2

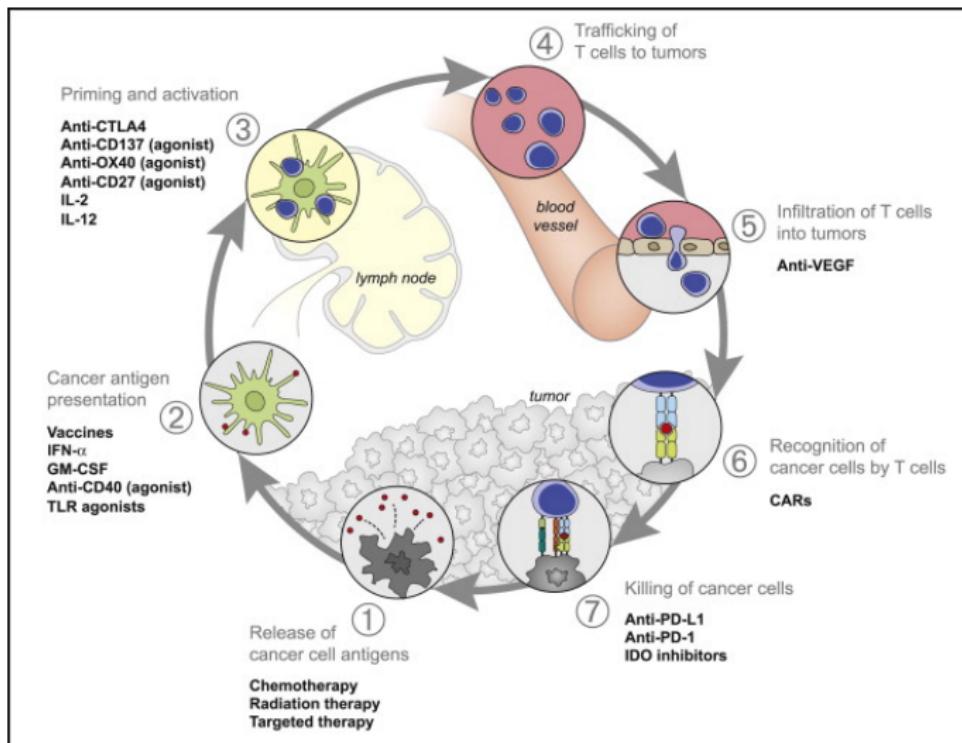


Current models of clinical oncology

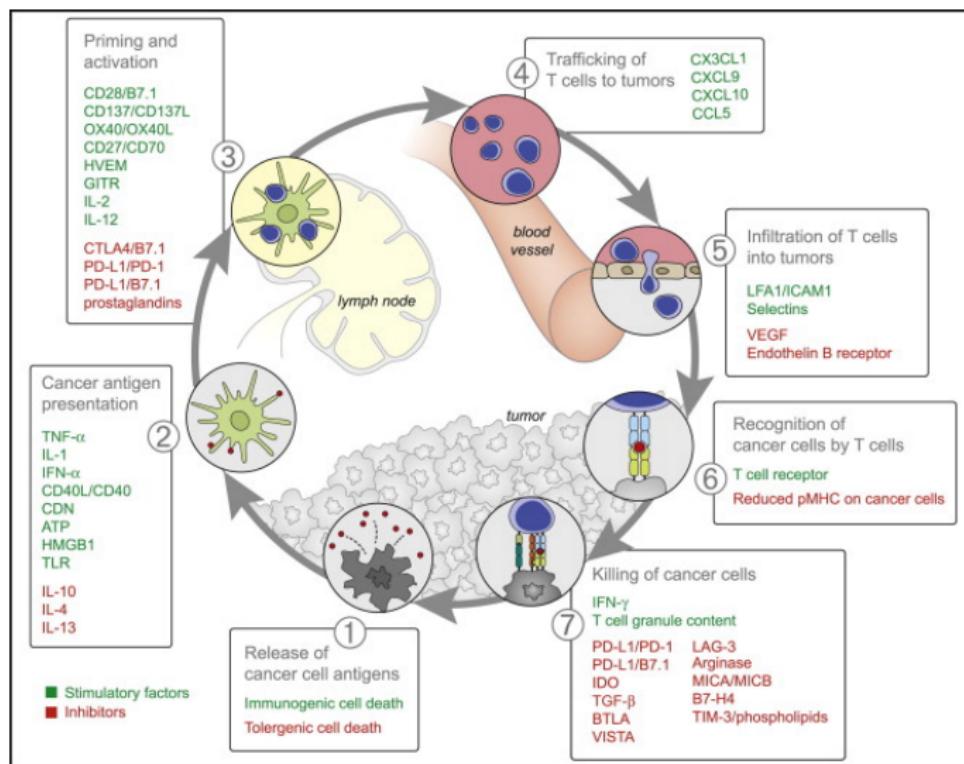
Seminal Model of Immuno-oncology



- Chen and Mellman (2013). “Oncology meets immunology: the cancer-immunity cycle”



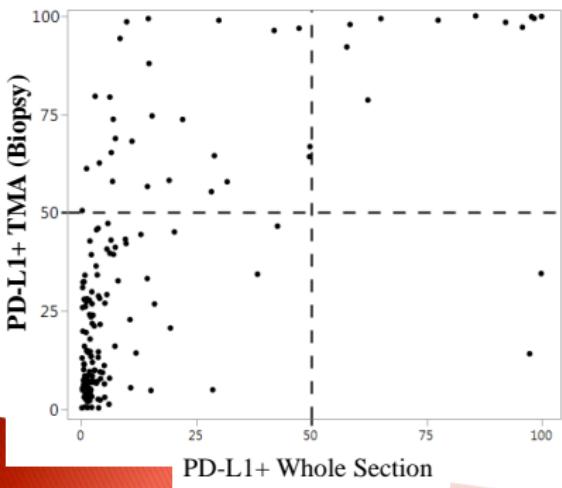
- Chen and Mellman (2013). “Oncology meets immunology: the cancer-immunity cycle”



- Chen and Mellman (2013). “Oncology meets immunology: the cancer-immunity cycle”

Limitations of IHC immune pathology: PDL1 positivity

176 Lung cancer patients treated with resection. Samples were scored for PDL1+ positivity



Median (IQR) of % Tumor PDL1+

	% Tumor PDL1+ TMA (biopsy)	% Tumor PDL1+ Whole Section
T1 (n=81)	18.8 (6.7-34.6)	2.3 (1.1-6.3)
T2 (n=71)	28.1 (9.1-58.1)	3.8 (1.7-14.4)
T3/4 (n=21)	20.7 (6.3-78.6)	6.0 (2.1-22.1)

SCIENTIFIC REPORTS



OPEN

Development of an Immune-Pathology Informed Radiomics Model for Non-Small Cell Lung Cancer

Received: 25 September 2017

Accepted: 9 January 2018

Published online: 31 January 2018

Chad Tang¹, Brian Hobbs², Ahmed Amer^{3,4}, Xiao Li², Carmen Behrens³, Jaime Rodriguez Canales⁴, Edwin Parra Cuentas⁴, Pamela Villalobos⁴, David Fried⁷, Joe Y. Chang¹, David S. Hong⁵, James W. Welsh¹, Boris Sepesi⁶, Laurence Court⁷, Ignacio I. Wistuba⁶ & Eugene J. Koay¹

With increasing use of immunotherapy agents, pretreatment strategies for identifying responders and non-responders is useful for appropriate treatment assignment. We hypothesize that the local immune micro-environment of NSCLC is associated with patient outcomes and that these local immune features exhibit distinct radiologic characteristics discernible by quantitative imaging metrics. We assembled two cohorts of NSCLC patients treated with definitive surgical resection and extracted quantitative parameters from pretreatment CT imaging. The excised primary tumors were then quantified for percent tumor PDL1 expression and density of tumor-infiltrating lymphocyte (via CD3 count) utilizing immunohistochemistry and automated cell counting. Associating these pretreatment radiomics parameters with tumor immune parameters, we developed an immune pathology-informed model (IPIM) that separated patients into 4 clusters (designated A-D) utilizing 4 radiomics features. The IPIM designation was significantly associated with overall survival in both training (5 year OS: 61%, 41%,

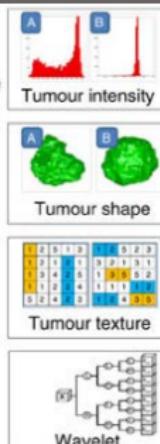
* co-first authors

Current Applications of Cancer “Radiomics”

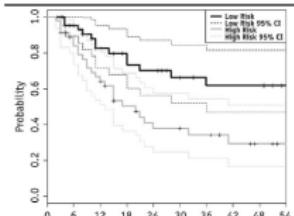
Standard of care imaging



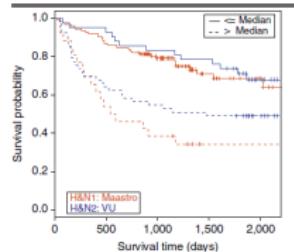
Feature extraction



Survival association

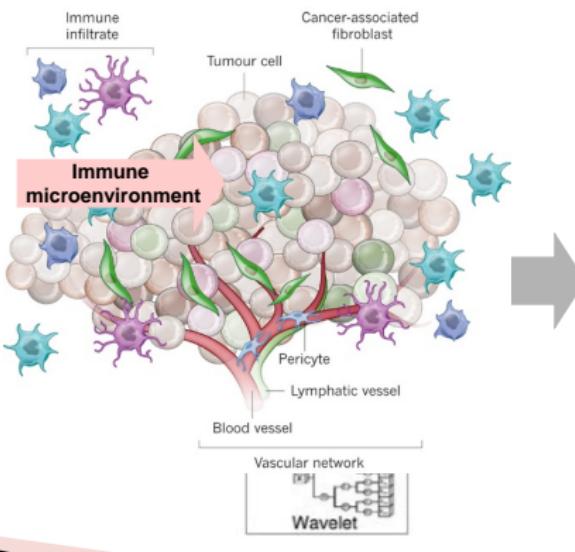


Survival association

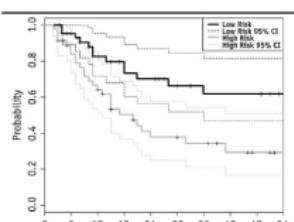


Radiomics Signatures of Immune Environment

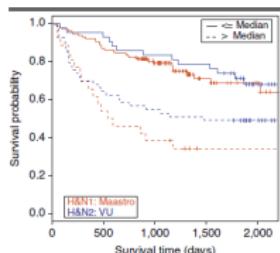
Standard of care imaging



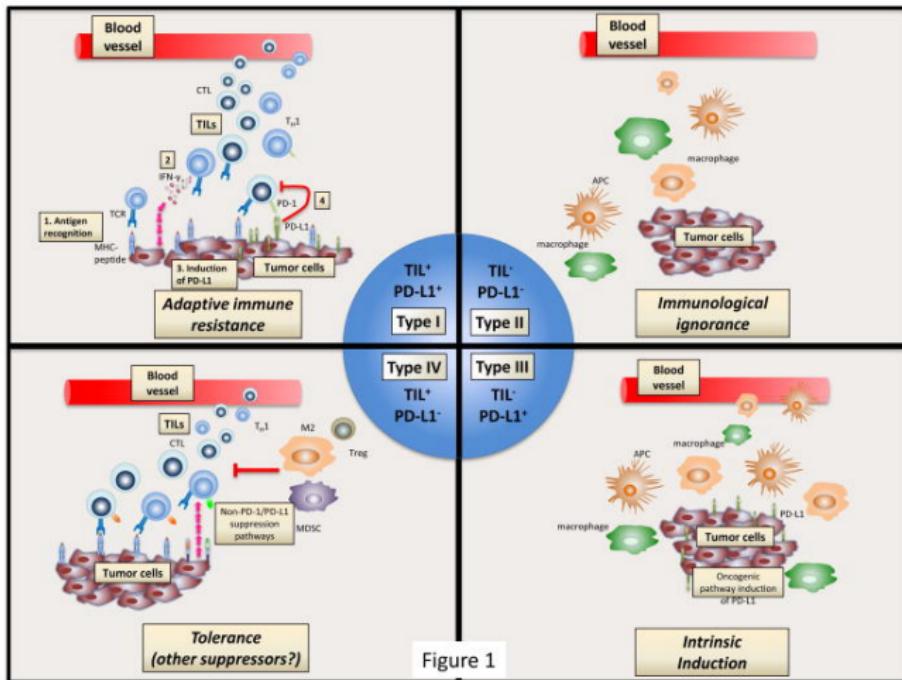
Survival association



Survival association

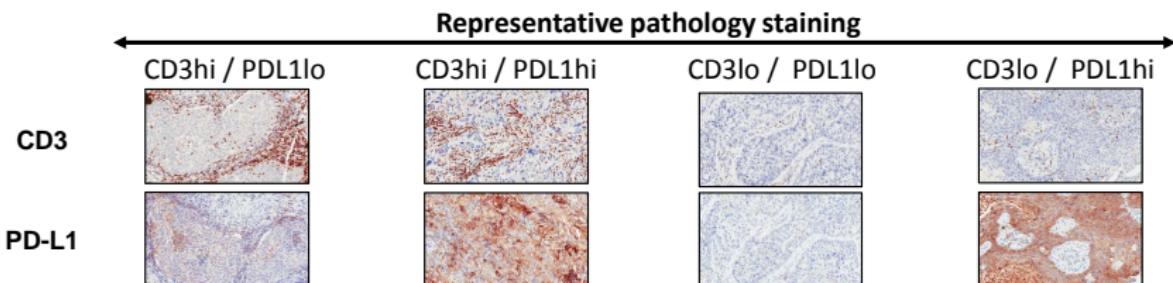
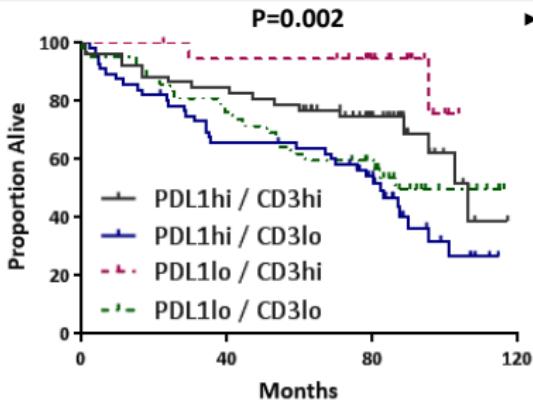
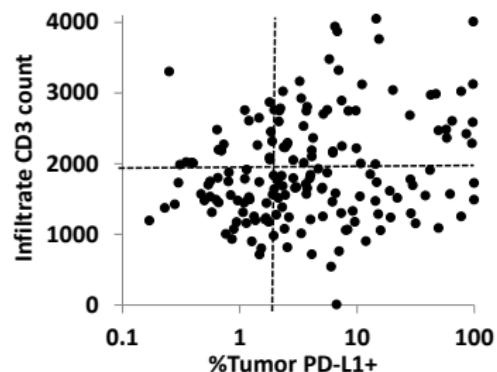


Immune Phenotypes of NSCLC



- Teng MW, Ngiow SF, et al. *Cancer Res* (2015) "Classifying Cancers Based on T-cell Infiltration and PD-L1"

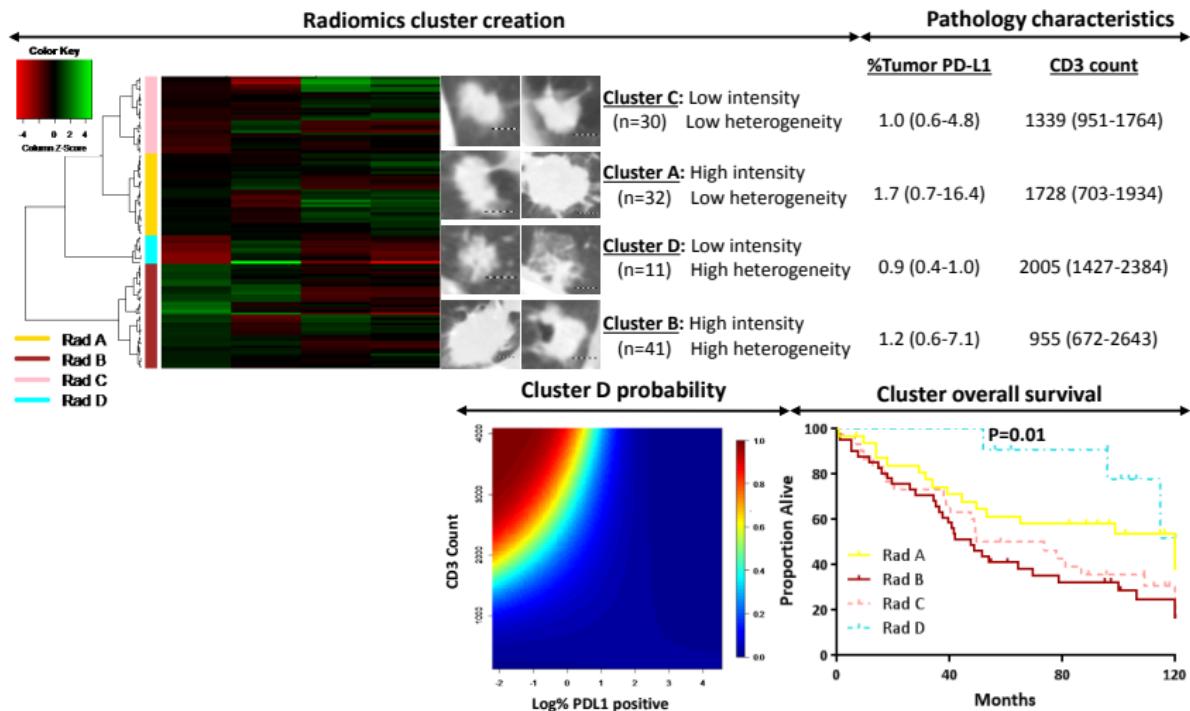
Immune Phenotypes of NSCLC



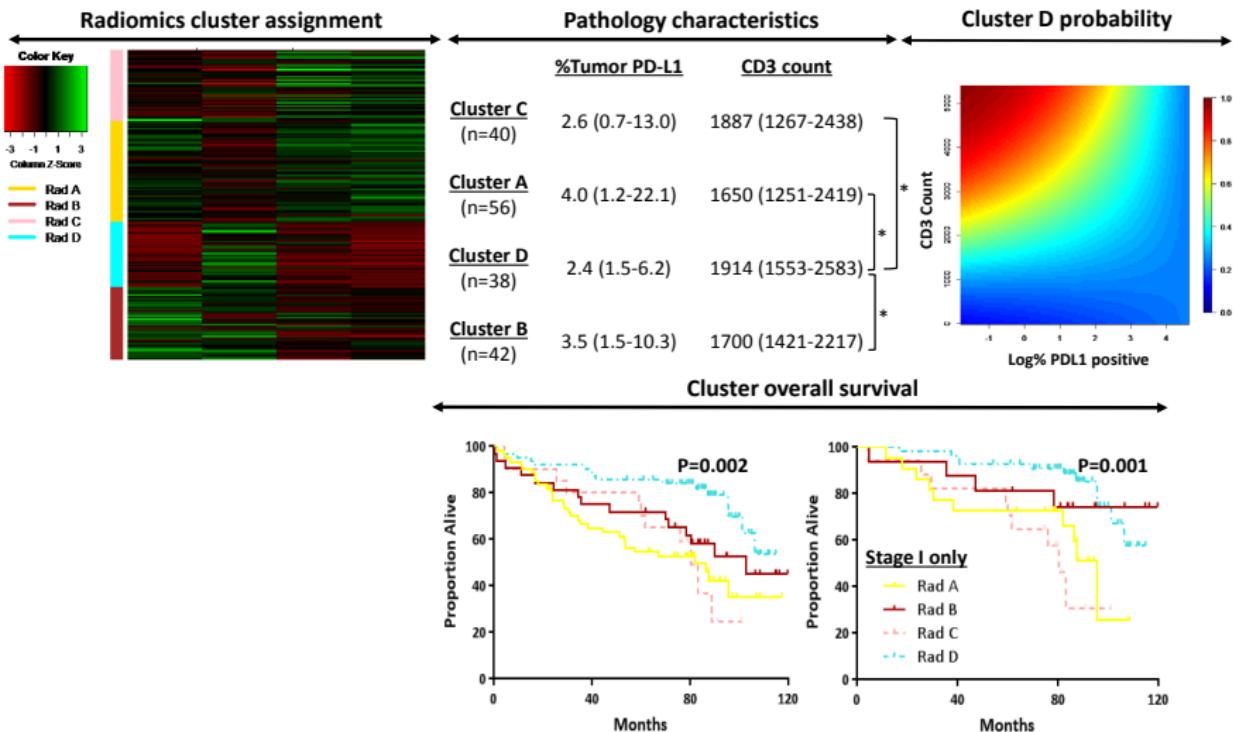
- Teng MW, Ngiow SF, et al. *Cancer Res* (2015) "Classifying Cancers Based on T-cell Infiltration and PD-L1"

Imaging Models of Immune Phenotypes

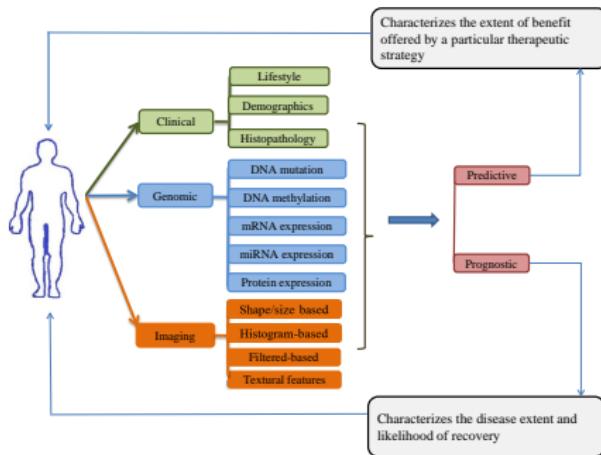
Radiomics signatures of Immune Phenotypes



Radiomics signatures of Immune Phenotypes



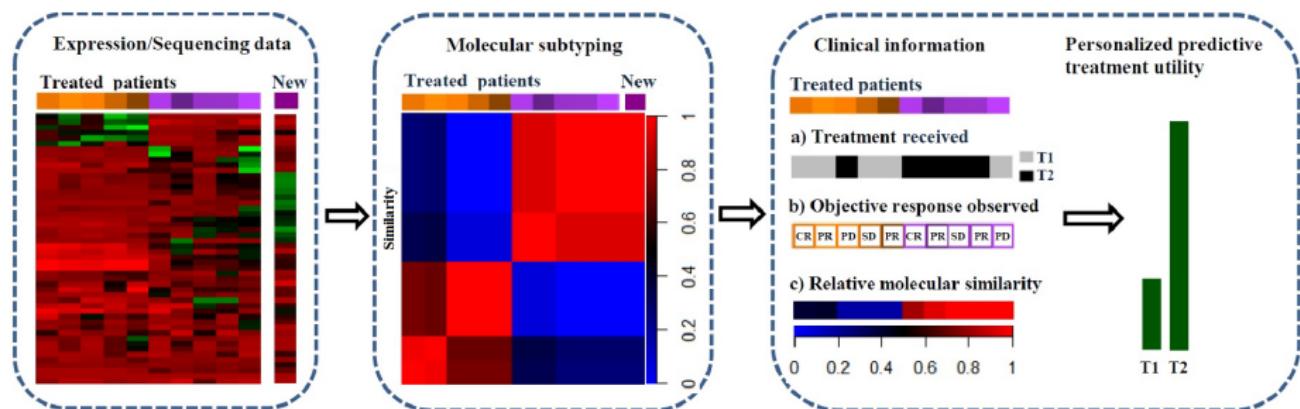
Understanding tumor/patient heterogeneity



- Ma, Stingo, Hobbs. *Biometrics*, (2016). Treatment Selection based on Personalized Predictive Treatment Utilities
- Ma, Hobbs, Stingo. *Stat. Methods in Med. Res.*, (2017). Treatment Selection based on Personalized Predictive Failure-Time
- Ma, Stingo, Hobbs. *submitted*, (2017). Bayesian personalized treatment selection strategies that integrate predictive with prognostic determinants.
- Huang & Hobbs *submitted* (2017). Estimating mean local posterior predictive benefit for biomarker-guided treatment strategies

Bayesian partial exchangeability frameworks for prec med

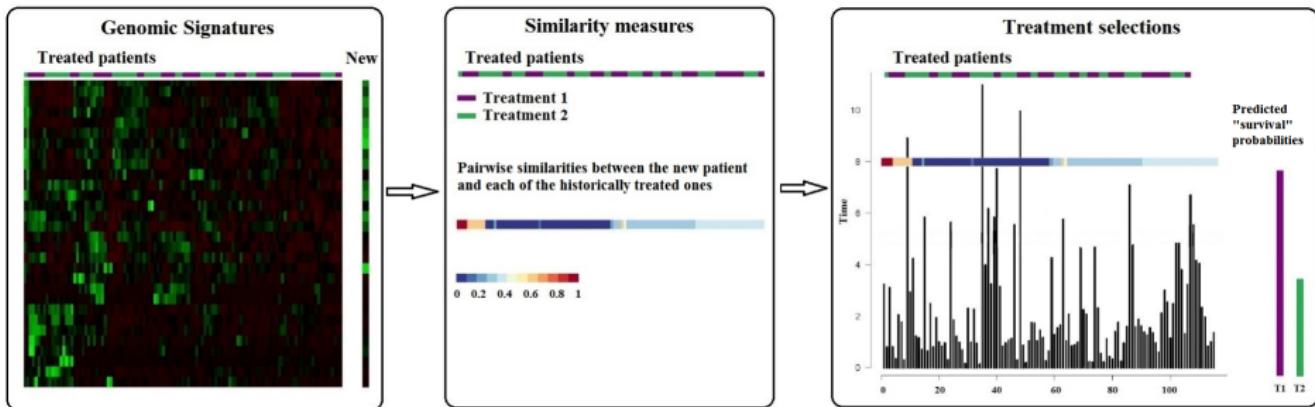
Ma, Stingo, Hobbs. *Biometrics*, (2016). Treatment Selection based on Personalized Predictive Treatment Utilities



- Quantifying similarities from clinical/molecular derived candidate features
- Characterizing pairwise partial statistical exchangeability
- Bayesian prediction models for treatment selection

Bayesian partial exchangeability frameworks for prec med

Ma, Hobbs, Stingo. Stat. Methods in Med. Res, (2017).
Treatment Selection based on Personalized Predictive Failure-Time

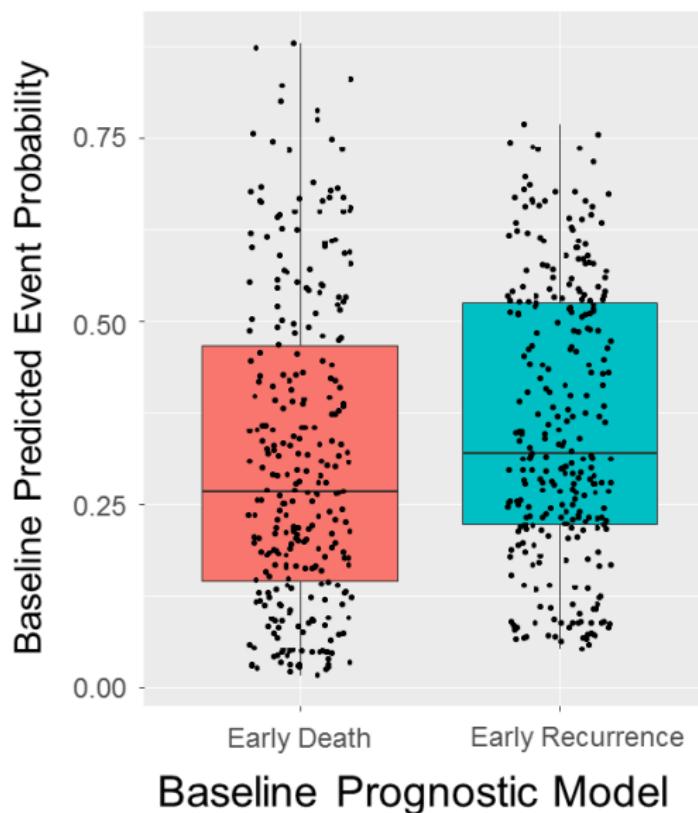


- Optimal treatment selection based on Bayesian predictive failure time
- Partial exchangeability based on tumor/patient characteristics, pairwise similarity
- Predict the probability of prolonging treatment failure

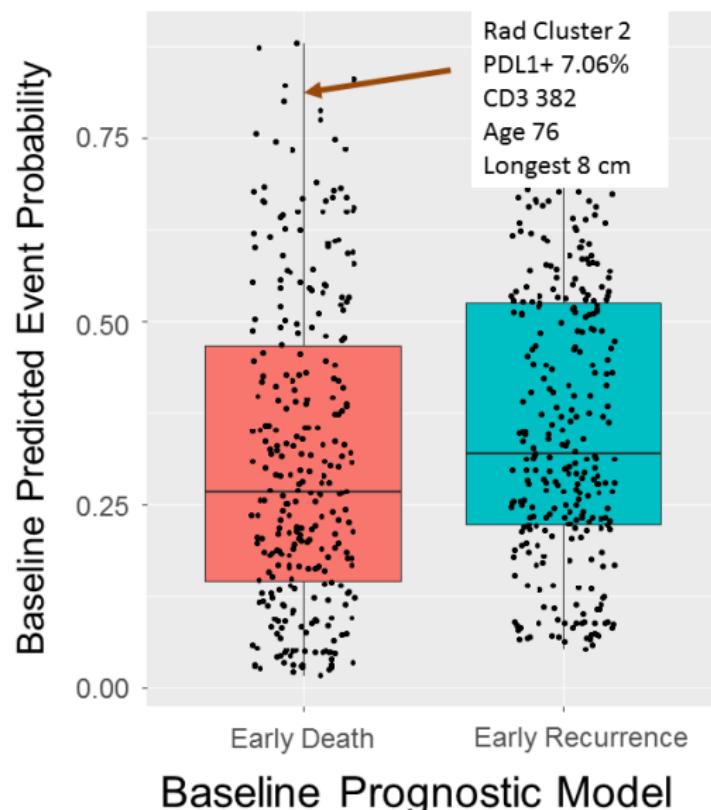
Clinical, Immune Pathology, Radiomics

Integrative Prognostic Model for NSCLC (n=411)

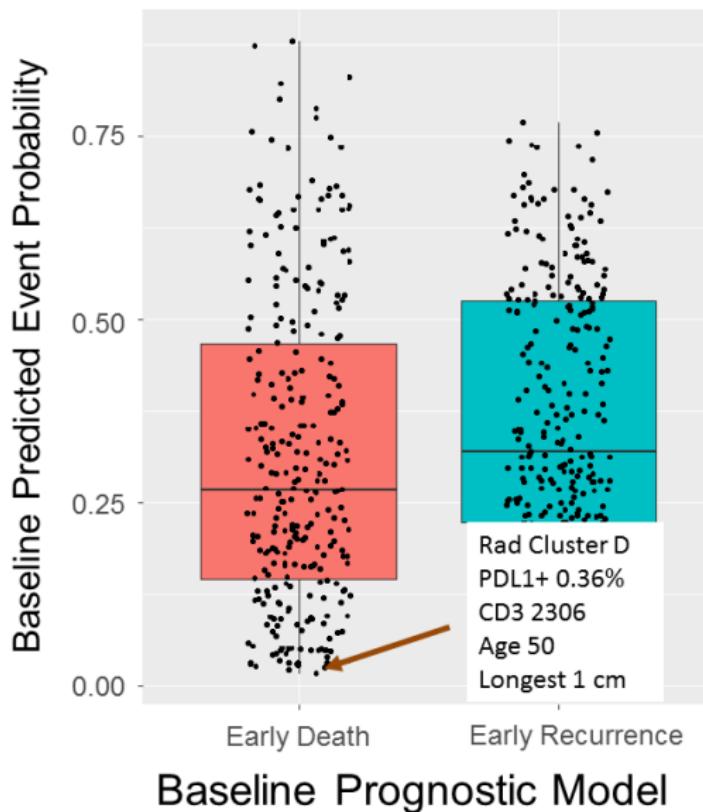
Integrative Predicted Event Probabilities



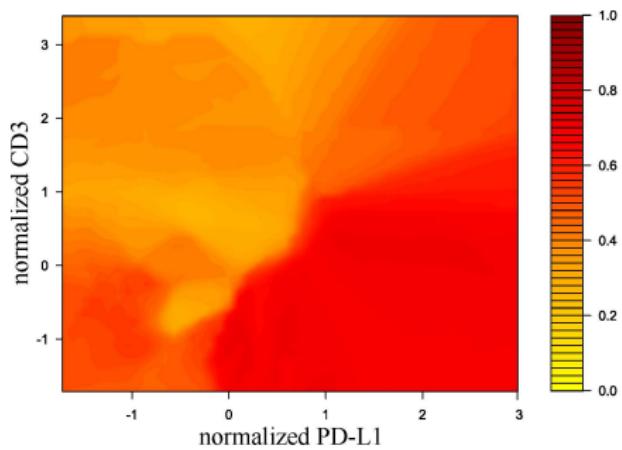
Integrative Predicted Event Probabilities



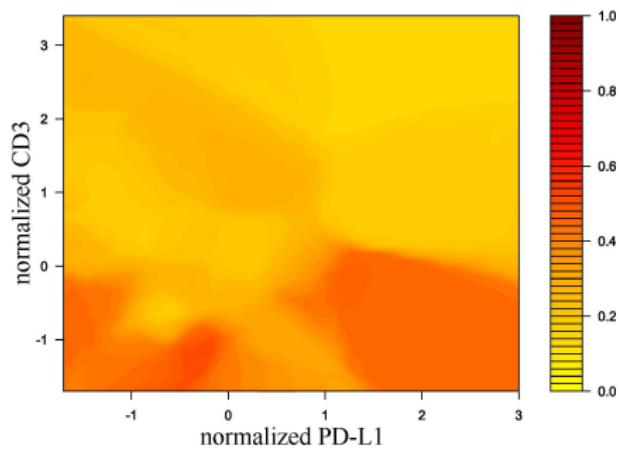
Integrative Predicted Event Probabilities



Baseline Prognostic Model of Local Immune Environment

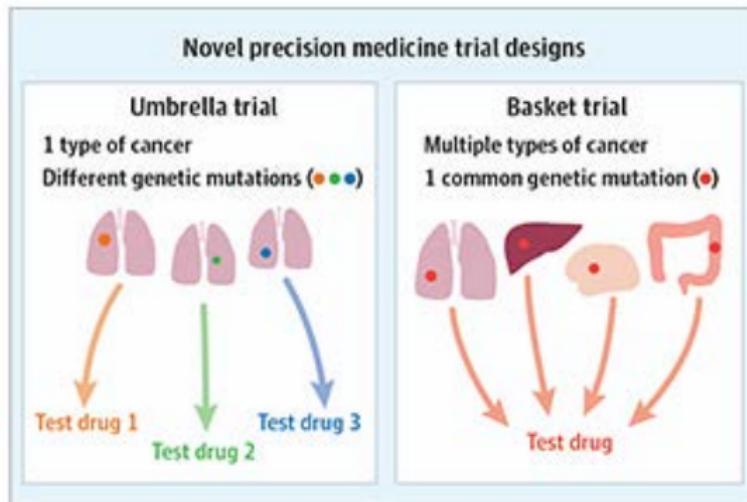


(a) Train data set, N=109



(B) Validation data set, N=167

Designs for “Precision” Medicine



Remarks on single-arm oncology trials

Guidelines for Revised Analytical Framework for pre-confirmatory trials

- Statistical models should focus characterizing “treatment effects” as predicted versus observed benefit
- Require formulation of prognostic models integrating biology with clinical knowledge
- Must consider heterogeneity intrinsic to the tumor microenvironment
- Requires model for phase III endpoint Mediation!!

Intermediate phase trials in Oncology

Case Study: Vemurafenib non-melanoma basket trial

- Histology-independent phase 2 basket trial of vemurafenib in BRAF V600 mutation-positive non-melanoma cancers
- Six pre-specified cancer cohorts
- Adaptive Simon two-stage design with primary endpoint of response rate at week 8
- Basket-wise testing: $H_0 : \pi = 0.15$ vs. $H_1 : \pi = 0.35$

Case Studies

Case Study: Vemurafenib non-melanoma basket trial

Baskets	Enrolled	Evaluable	Responders	Posterior probability $Pr(\pi > 0.15)$ based on response only
NSCLC	20	19	8	0.998
CRC (vemu)	10	10	0	0.068
CRC (vemu + cetu)	27	26	1	0.039
Bile Duct	8	8	1	0.472
ECD or LCH	18	14	6	0.995
ATC	7	7	2	0.847

- Bayesian Posterior Probability $Pr(\pi > 0.15 | Data) > \theta$, with θ fixed to control type I error at 0.10
- data reported in article: "Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations," *NEJM* (2015)

Basket Design Dilemma

- Implicit to the concept of a basket trial is exchangeable treatment effects across baskets
- early basket trials have been criticized [JCO Cunanan 2017] for implementing basketwise analysis strategies which
 - failed to convey to the extent of statistical evidence for exchangeability across subtypes/baskets
 - ignore additional sources of inter-patient heterogeneity, either observed or unobserved in the study
 - in the presence of imbalanced enrollment, basketwise analyses fail to elucidate evidential measures of effect in small baskets
- conversely, pooling patients across baskets under the assumption of inter-patient exchangeability induces bias and limits the designs power for identifying favorable subtypes in the presence of heterogeneity of effect across basket labels.

Bayesian Modeling to assess exchangeable effects across baskets/subtypes, is it useful?

Borrowing Information across Subgroups in Phase II Trials: Is It Useful?

Boris Freidlin and Edward L. Korn

Table 3. Empirical probabilities of rejecting the null hypothesis: 10 subgroups (no interim monitoring, 25 patients per subgroup, 10,000 replications)

Design	True response rate in each subgroup									
	0.1	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Case 1	0.1	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Subgroup-specific analyses	0.096	0.905	0.908	0.907	0.909	0.908	0.910	0.908	0.909	0.912
HB model 1 moderate borrowing	0.096	0.905	0.908	0.907	0.909	0.908	0.910	0.908	0.909	0.912
HB model 1 strong borrowing	0.099	0.892	0.892	0.894	0.897	0.896	0.903	0.895	0.895	0.898
HB model 2 (Berry et al.; ref. 13)	0.099	0.905	0.908	0.907	0.908	0.912	0.910	0.908	0.909	0.911
Case 2	0.1	0.1	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Subgroup-specific analyses	0.096	0.098	0.908	0.907	0.909	0.908	0.910	0.908	0.909	0.912
HB model 1 moderate borrowing	0.096	0.098	0.908	0.907	0.909	0.908	0.910	0.908	0.909	0.912
HB model 1 strong borrowing	0.085	0.087	0.857	0.857	0.861	0.858	0.867	0.859	0.857	0.861
HB model 2 (13)	0.097	0.098	0.904	0.903	0.906	0.905	0.906	0.904	0.905	0.907
Case 3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.3
Subgroup-specific analyses	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.909	0.912
HB model 1 moderate borrowing	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.909	0.912
HB model 1 strong borrowing	0.019	0.022	0.020	0.021	0.023	0.019	0.022	0.022	0.677	0.681
HB model 2 (13)	0.032	0.033	0.031	0.031	0.036	0.032	0.037	0.031	0.783	0.790
Case 4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3
Subgroup-specific analyses	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.098	0.914
HB model 1 moderate borrowing	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.098	0.914
HB model 1 strong borrowing	0.012	0.014	0.013	0.013	0.016	0.012	0.015	0.015	0.014	0.656
HB model 2 (13)	0.029	0.029	0.028	0.028	0.033	0.030	0.034	0.027	0.031	0.747
Case 5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Subgroup-specific analyses	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.098	0.094
HB model 1 moderate borrowing	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.098	0.094
HB model 1 strong borrowing	0.009	0.010	0.010	0.010	0.012	0.007	0.011	0.010	0.010	0.010
HB model 2 (13)	0.022	0.023	0.022	0.023	0.026	0.024	0.026	0.021	0.024	0.023
Case 6	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Subgroup-specific analyses	0.913	0.905	0.908	0.907	0.909	0.908	0.910	0.908	0.909	0.912
HB model 1 moderate borrowing	0.913	0.905	0.908	0.907	0.909	0.908	0.910	0.908	0.909	0.912
HB model 1 strong borrowing	0.908	0.910	0.910	0.910	0.912	0.910	0.917	0.910	0.912	0.911
HB model 2 (13)	0.915	0.906	0.909	0.908	0.910	0.908	0.911	0.909	0.909	0.913

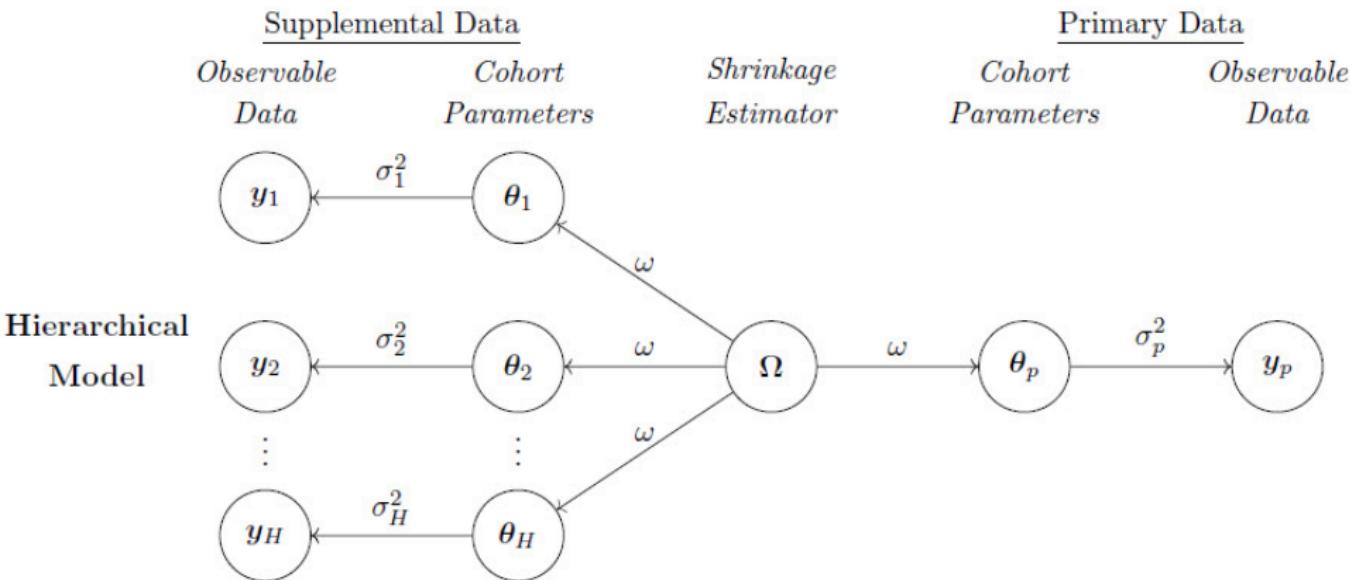
Abbreviation: HB, hierarchical Bayesian.

Table 3. Empirical probabilities of rejecting the null hypothesis: 10 subgroups (no interim monitoring, 25 patients per subgroup, 10,000 replications)

Design	True response rate in each subgroup									
Case 1	0.1	0.3								
Subgroup-specific analyses	0.096	0.095	0.098	0.097	0.099	0.098	0.0910	0.098	0.099	0.0912
HB model 1 moderate borrowing	0.096	0.095	0.098	0.097	0.099	0.098	0.0910	0.098	0.099	0.0912
HB model 1 strong borrowing	0.099	0.092	0.092	0.094	0.097	0.096	0.0903	0.095	0.095	0.0898
HB model 2 (Berry et al.; ref. 13)	0.099	0.095	0.098	0.097	0.098	0.0912	0.0910	0.098	0.099	0.0911
Case 2	0.1	0.1	0.3							
Subgroup-specific analyses	0.096	0.098	0.098	0.097	0.099	0.098	0.0910	0.098	0.099	0.0912
HB model 1 moderate borrowing	0.096	0.098	0.098	0.097	0.099	0.098	0.0910	0.098	0.099	0.0912
HB model 1 strong borrowing	0.085	0.087	0.087	0.087	0.0861	0.0858	0.0867	0.0859	0.0857	0.0861
HB model 2 (13)	0.097	0.098	0.094	0.093	0.096	0.095	0.0906	0.094	0.095	0.0907
Case 3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.3
Subgroup-specific analyses	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.099	0.0912
HB model 1 moderate borrowing	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.099	0.0912
HB model 1 strong borrowing	0.019	0.022	0.020	0.021	0.023	0.019	0.022	0.022	0.077	0.0681
HB model 2 (13)	0.032	0.033	0.031	0.031	0.036	0.032	0.037	0.031	0.783	0.790
Case 4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3
Subgroup-specific analyses	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.098	0.0914
HB model 1 moderate borrowing	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.098	0.0914
HB model 1 strong borrowing	0.012	0.014	0.013	0.013	0.016	0.012	0.015	0.015	0.014	0.0656
HB model 2 (13)	0.029	0.029	0.028	0.028	0.033	0.030	0.034	0.027	0.031	0.0747
Case 5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Subgroup-specific analyses	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.098	0.094
HB model 1 moderate borrowing	0.096	0.098	0.095	0.095	0.100	0.099	0.108	0.094	0.098	0.094
HB model 1 strong borrowing	0.009	0.010	0.010	0.010	0.012	0.007	0.011	0.010	0.010	0.010
HB model 2 (13)	0.022	0.023	0.022	0.023	0.026	0.024	0.026	0.021	0.024	0.023
Case 6	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Subgroup-specific analyses	0.913	0.905	0.908	0.907	0.909	0.908	0.910	0.908	0.909	0.912
HB model 1 moderate borrowing	0.913	0.905	0.908	0.907	0.909	0.908	0.910	0.908	0.909	0.912
HB model 1 strong borrowing	0.908	0.910	0.910	0.910	0.912	0.910	0.917	0.910	0.912	0.911
HB model 2 (13)	0.915	0.906	0.909	0.908	0.910	0.908	0.911	0.909	0.909	0.913

Abbreviation: HB, hierarchical Bayesian.

Conventional Hierarchical Models are limited!



Bayesian hierarchical modeling based on multisource exchangeability

ALEXANDER M. KAIZER, JOSEPH S. KOOPMEINERS*

*Division of Biostatistics, University of Minnesota, A460 Mayo Building, MMC 303 420 Delaware St. SE,
Minneapolis, MN 55455, USA*

koopm007@umn.edu

BRIAN P. HOBBS

The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd. Houston, TX 77030, USA

SUMMARY

Bayesian hierarchical models produce shrinkage estimators that can be used as the basis for integrating supplementary data into the analysis of a primary data source. Established approaches should be considered limited, however, because posterior estimation either requires prespecification of a shrinkage weight for each source or relies on the data to inform a single parameter, which determines the extent of influence or shrinkage from all sources, risking considerable bias or minimal borrowing. We introduce multisource exchangeability models (MEMs), a general Bayesian approach for integrating multiple, potentially non-exchangeable, supplemental data sources into the analysis of a primary data source. Our proposed modeling framework yields source-specific smoothing parameters that can be estimated in the presence of the data to facilitate a dynamic multi-resolution smoothed estimator that is asymptotically consistent while reducing the dimensionality of the prior space. When compared with competing Bayesian hierarchical modeling strategies, we demonstrate that MEMs achieve approximately 2.2 times larger median effective supplemental sample size when the supplemental data sources are exchangeable as well as a 56% reduction in bias when there is heterogeneity among the supplemental sources. We illustrate the application of MEMs using a recently completed randomized trial of very low nicotine content cigarettes, which resulted in a 30% improvement in efficiency compared with the standard analysis.

Keywords: Bayesian hierarchical modeling; Heterogeneous sources of data; Multisource smoothing; Supplementary data.

Kaizer, Koopmeiners, Hobbs (2017) Bayesian hierarchical modeling based on multi-source exchangeability. Biostatistics

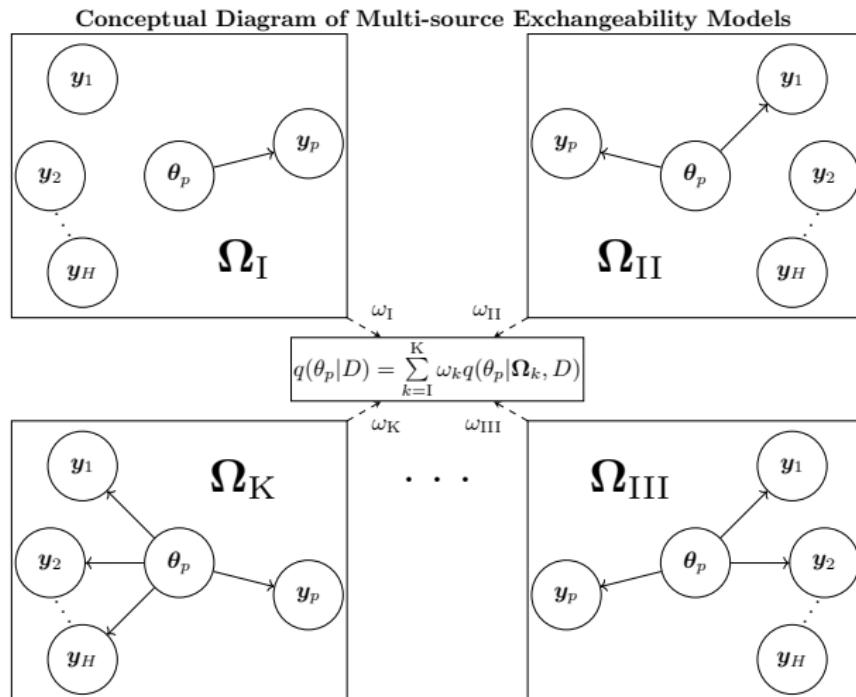


Fig. 1: Each MEM is a combination of supplemental sources assumed exchangeable with the primary cohort in order to estimate the parameters of interest, θ_p , and is contained within each box for Ω_k . Within a box the solid arrows θ_p and the observables, y_h , represent which supplemental sources are assumed exchangeable with the primary cohort within the given MEM.

Bayesian hierarchical modeling based on multisource exchangeability

Alexander M Kaizer, Joseph S Koopmeiners , Brian P Hobbs

Biostatistics, Volume 19, Issue 2, 1 April 2018, Pages 169–184, <https://doi.org/10.1093/biostatistics/kxx051>

Let L represent the likelihood, given the data for Ω_k , $\Theta = (\theta_p, \theta_1, \dots, \theta_H)$, and $\pi(\Theta|\Omega_k)$ denote the prior density of Θ under Ω_k . In the context of MEMs, the integrated marginal likelihood for a particular MEM, given the data are obtained by averaging the likelihood over the posterior distribution for the vector of all model parameters of interest,

$$p(D|\Omega_k) = \int L(\Theta|D, \Omega_k) \pi(\Theta|\Omega_k) d\Theta. \quad (2.1)$$

The posterior model weights for each MEM are given by

$$\omega_k = p(\Omega_k|D) = \frac{p(D|\Omega_k)\pi(\Omega_k)}{\sum_{j=1}^K p(D|\Omega_j)\pi(\Omega_j)}, \quad (2.2)$$

where $\pi(\Omega_k)$ is the prior probability that Ω_k is the true model. The marginal posterior distribution, given the observable data D to be used for inference on θ_p is the weighted average using the posterior model

Bayesian hierarchical modeling based on multisource exchangeability

Alexander M Kaizer, Joseph S Koopmeiners , Brian P Hobbs

Biostatistics, Volume 19, Issue 2, 1 April 2018, Pages 169–184, <https://doi.org/10.1093/biostatistics/kxx031>

Bayesian hierarchical modeling based on MEMs

173

weights of the K MEM posteriors, $p(\theta_p | \Omega_k, D)$:

$$p(\theta_p | D) = \sum_{k=1}^K \omega_k p(\theta_p | \Omega_k, D). \quad (2.3)$$

Unfortunately, BMA quickly becomes highly parameterized, as the number of models grows exponentially with the number of supplementary sources ($K = 2^H$), and prior specification over a model space of large size is problematic. Fernández and others (2001) noted that posterior model weights can be very sensitive to the specification of priors in the model, especially in the absence of strong prior knowledge. In the analysis of limited data obtained from a clinical study, these issues with the conventional BMA approach become critical and motivate our proposed approach. With MEMs, the supplemental sources are assumed to be distinct and independent, therefore we can specify priors with respect to sources instead of models, $\pi(\Omega_k) = \pi(S_1 = s_{1,k}, \dots, S_H = s_{H,k}) = \pi(S_1 = s_{1,k}) \times \dots \times \pi(S_H = s_{H,k})$. This results in drastic dimension reduction in that it necessitates the specification of only H source-specific prior inclusion probabilities in place of 2^H prior model probabilities comprising the entire model space. In Section 3.2, we propose prior weights for the source-inclusion probabilities that result in consistent posterior model weights and yield desirable small sample properties, as evaluated by simulation in Section 4. In contrast, similarly constructed prior weights on the models did not result in consistent posterior model weights.

3. ESTIMATION AND THEORETICAL RESULTS

We now describe posterior inference using MEMs in the Gaussian case and investigate the asymptotic properties of our proposed approach for two classes of source-specific prior model weights. For our primary cohort (P), let $x_{1,p}, x_{2,p}, \dots, x_{n,p}$, denote a sample of i.i.d. and normally distributed observables with mean μ and variance σ^2 . Similarly, for supplemental cohorts $h = 1, \dots, H$, let $x_{1,h}, \dots, x_{n_h,h}$, denote a sample of i.i.d. normally distributed samples with source-specific mean μ_h and variance σ_h^2 . Throughout this section, we assume σ^2 and σ_h^2 are known and define $v = \frac{\sigma^2}{n}$ and $v_h = \frac{\sigma_h^2}{n_h}$ for notational simplicity.

Using the MEM approach, the likelihood is a weighted average of MEMs representing all possible assumptions regarding exchangeability:

$$\sum_{k=1}^K \omega_k L(\boldsymbol{\mu}, \sigma^2 | (s_{1,k}, \dots, s_{H,k})) = \sum_{k=1}^K \omega_k \mathcal{N}(\boldsymbol{\mu}, \sigma^2) \prod_{h=1}^H \{\mathcal{N}(\mu s_{h,k} + \mu_h (1 - s_{h,k}), \sigma_h^2)\}. \quad (3.1)$$

Then, assuming a flat prior on the Gaussian mean, μ_k , as described by (Gelman, 2006), $\pi(\mu_k | \Omega_k) \propto 1$ in (2.1), the Gaussian conditional marginal likelihood can be generally written for any MEM as:

$$\begin{aligned} p(D | \Omega_k) &= \frac{(\sqrt{2\pi})^{(H+1) - \sum_{h=1}^H \{s_{h,k}\}}}{\sqrt{\left(\frac{1}{v} + \sum_{i=1}^H \left\{ \frac{s_{i,k}}{v_i} \right\}\right) \left(\prod_{j=1}^H \left\{ \left[\frac{1}{v_j}\right]^{1-s_{j,k}} \right\}\right)}} \\ &\times \exp \left(-\frac{1}{2} \left[\sum_{l=1}^H \left\{ \frac{s_{l,k}(\bar{x} - \bar{x}_l)^2}{v + v_l + v v_l (\sum_{m \neq l} \{s_{m,k} v_m^{-1}\})} \right. \right. \right. \\ &+ \left. \left. \left. \sum_{l < r}^H \left\{ \frac{s_{l,k} s_{r,k} (\bar{x}_l - \bar{x}_r)^2}{v_l + v_r + v_l v_r (v^{-1} + \sum_{p \neq l,r} \{s_{p,k} v_p^{-1}\})} \right\} \right\} \right] \right). \end{aligned} \quad (3.2)$$

The exponential portion of (3.2) is comprised of the squared deviations between the sources included in Ω_k , such that if all included sources are exchangeable, then $\exp(0) = 1$ and the posterior weights of (2.2) are influenced by the non-exponential terms of (3.2), which do not include sample means, and the priors placed on model weights, $\pi(\Omega_k)$.

The posterior distribution of μ , derived from (3.2) and used for inference with multi-resolution shrinkage of the supplemental cohorts, is a mixture of normal distributions computed using (2.3) and (2.2):

$$\begin{aligned}
 p(\mu|D) &= \sum_{k=1}^K \omega_k p(\mu|D, \sigma^2, (s_{1,k}, \dots, s_{H,k})) \\
 &= \sum_{k=1}^K \omega_k \mathcal{N} \left(\frac{\bar{x} \prod_{h=1}^H \left\{ v_h^{s_{h,k}} \right\} + \sum_{l=1}^H \left\{ \frac{v\bar{x}_l}{v_l} \right\} \prod_{j=1}^H \left\{ v_j^{s_{j,k}} \right\}}{v \left[\sum_{l=1}^H \left\{ \frac{s_{l,k}}{v_l} \right\} \prod_{m=1}^H \left\{ v_m^{s_{m,k}} \right\} \right] + \prod_{r=1}^H \left\{ \frac{s_{r,k}}{v_r} \right\}}, \left(\frac{1}{v} + \sum_{p=1}^H \frac{s_{p,k}}{v_p} \right)^{-1} \right). \tag{3.3}
 \end{aligned}$$

The posterior mean is obtained as the weighted average of the model-specific posterior means. The posterior variance of a mixture of normal distributions is also available in closed form.

3.2. Asymptotic properties

Methods that incorporate supplemental information should endeavor to integrate data from potentially very different sources and arrive at a posterior estimate that minimizes the bias introduced by incorporating the supplemental data. In the case of MEMs, bias arising from using the supplemental data can be minimized if sources that are exchangeable attain a weight of 1 while all other sources attain weight 0.

Using the two prior specifications presented in Section 3.1, we demonstrate the frequentist, asymptotic properties of MEMs. Specifically, we describe the conditions whereby the MEM specification yields asymptotically consistent model-specific weights, resulting in consistent estimation of μ by the posterior mean. The asymptotic properties assume a finite mixture of MEMs represented by Gaussian distributions with known variances and posterior model weights calculated in the MEM framework.

THEOREM 3.1 As $n, n_1, \dots, n_H \rightarrow \infty$, $\omega_{k^*} \rightarrow 1$ for model k^* defined by $(S_1 = s_{1,k^*}, \dots, S_H = s_{H,k^*})$, where $s_{h,k^*} = \mathbb{1}_{\{\mu = \mu_h\}}$ for all $h = 1, \dots, H$ and $\omega_k \rightarrow 0$ for $k \neq k^*$ with priors π_e and π_n .

obtained from a Bayesian model to characterize an additional number of “effective primary” samples effectuated for joint inference, was considered by [Hobbs and others \(2013\)](#) as an extension of prior effective sample size ([Morita and others, 2008](#)). Formally, for any model in which posterior precision is approximately linear in sample size, ESSS is defined as $ESSS = n \left\{ \frac{\mathcal{P}^*(\mathbf{x}_p, \mathbf{x}_1, \dots, \mathbf{x}_H)}{\mathcal{P}^*(\mathbf{x}_p)} - 1 \right\}$, where $\mathcal{P}^*(\mathbf{x}_p)$ is the posterior precision of the reference model with no borrowing from supplemental sources and $\mathcal{P}(\mathbf{x}_p, \mathbf{x}_1, \dots, \mathbf{x}_H)$ is the posterior precision under the joint model which incorporates supplemental information. In the Gaussian case for MEMs, the posterior precision for Ω_k is $\frac{1}{v} + \sum_{h=1}^H \frac{s_{h,k}}{v_h}$, which results in an ESSS that can be calculated exactly as

$$ESSS_{MEM} = n \sum_{k=1}^K \left\{ \omega_k \left[\frac{\frac{1}{v} + \sum_{h=1}^H \frac{s_{h,k}}{v_h}}{\frac{1}{v}} - 1 \right] \right\}. \quad (4.1)$$

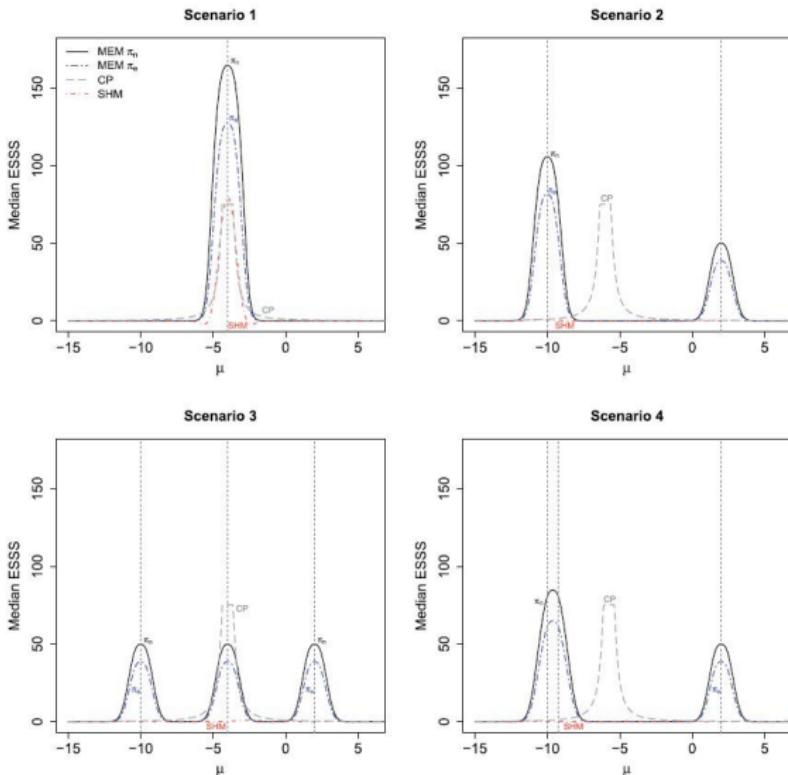


Fig. 2. Median effective supplemental sample size using MEM with π_e , π_B priors, CP, and SHM under each scenario. Dashed vertical lines are used to represent assumed observed values of the supplemental group means for each scenario.

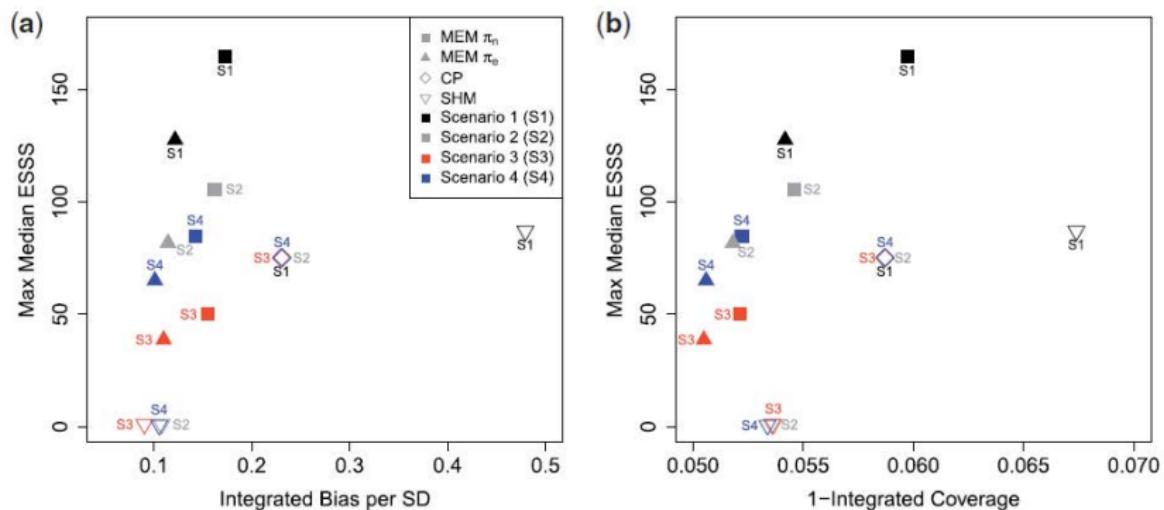


Fig. 3. Plots demonstrating bias versus shrinkage trade-offs using the methods of CP, SHM, and MEM with π_e, π_n source-inclusion priors. Note that CP overlaps for all four scenarios.

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Bayesian Basket Trial Design with Exchangeability Monitoring

Brian P. Hobbs¹ and Rick Landin²

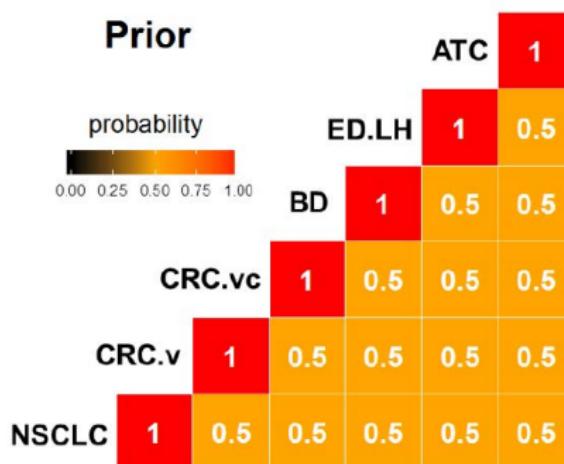
Precision medicine endeavors to conform therapeutic interventions to the individuals being treated. Implicit to the concept of precision medicine is heterogeneity of treatment benefit among patients and patient subpopulations. Thus, precision medicine challenges conventional paradigms of clinical translational which have relied on estimates of population-averaged effects to guide clinical practice. Basket trials comprise a class of experimental designs used to study solid malignancies that are devised to evaluate the effectiveness of a therapeutic strategy among patients defined by the presence of a particular drug target (often a genetic mutation) rather than a particular tumor histology. Acknowledging the potential for differential effectiveness on the basis of traditional criteria for cancer subtyping, evaluations of treatment effectiveness are conducted with respect to the “baskets” which collectively represent a partition of the targeted patient population consisting of discrete subtypes. Yet, designs of early basket trials have been criticized for their reliance on basketwise analysis strategies which suffered from limited power in the presence of imbalanced enrollment as well as failed to convey to the clinical community evidentiary measures for consistent effectiveness among the studied clinical subtypes. This article presents novel methodology for sequential basket trial design formulated with Bayesian monitoring rules. Interim analyses are based a novel hierarchical modeling strategy for sharing information among a collection of discrete, potentially non-exchangeable subtypes. The methodology is demonstrated by analysis as well as permutation and simulation studies based on a recent basket trial designed to estimate the effectiveness of vemurafenib in $BRAF^{V600}$ mutant non-melanoma among six primary disease sites and histologies.

Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: Adaptive clinical trial design; Basket trial design; Bayesian hierarchical model; oncology; precision medicine; sequential design

Bayesian Basket Trial Design with Exchangeability Monitoring

Brian P. Hobbs¹ and Rick Landin²



Bayesian Basket Trial Design with Exchangeability Monitoring

Brian P. Hobbs¹ and Rick Landin²

hyperparameters a and b , upon having observed successes $S = \{S_1, \dots, S_J\}$, Bayes' Theorem yields the following conjugate conditional posterior distribution for the response probability of basket j representing the Bayesian update of $p(\pi | S_{(-j)})$ with likelihood $Bin(S_j | \pi_j, n_j)$,

$$q(\pi_j | S, \Omega_j) \propto \text{Beta}\left(a + \sum_{h=1}^J \Omega_{j,h} S_h, b + \sum_{k=1}^J \Omega_{j,k} (n_k - S_k)\right), \quad (1)$$

where Ω_j represents the j^{th} row of multisource exchangeability matrix Ω . Marginal posterior inference with respect to $\pi_j | S$ averages (1) with respect to the marginal posterior probability of $G = 2^{J-1}$ possible exchangeability configurations of Ω_j . Let $\omega = \{\omega_1, \dots, \omega_G\}$ denote the collection of vectors each of length J and with j^{th} element = 1 that collectively span the sample space of Ω_j . The marginal posterior distribution can be represented by a finite mixture density

$$q(\pi_j | S) \propto \sum_{g=1}^G q(\pi_j | S, \Omega_j = \omega_g) Pr(\Omega_j = \omega_g | S), \quad (2)$$

where the posterior probability of exchangeability configuration ω_g given the observed data follows from Bayes' Theorem in proportion to the marginal density of the data given ω_g and its unconditional prior probability

$$Pr(\Omega_j = \omega_g | S) \propto \frac{m(S_j | \Omega_j = \omega_g, S_{(-j)}) Pr(\Omega_j = \omega_g)}{\sum_{u=1}^G m(S_j | \Omega_j = \omega_u, S_{(-j)}) Pr(\Omega_j = \omega_u)}. \quad (3)$$

Bayesian Basket Trial Design with Exchangeability Monitoring

Brian P. Hobbs¹ and Rick Landin²

Let $B()$ denote the beta function. Given an exchangeability configuration Ω_j , the marginal density of S_j may be obtained by integrating the likelihood of $\pi_j | S_j$ with respect to $p(\pi | S_{(-j)})$,

$$m(S_j | \Omega_j, S_{(-j)}) \propto \frac{B\left(a + \sum_{h=1}^J \Omega_{j,h} S_h, b + \sum_{k=1}^J \Omega_{j,k} (n_k - S_k)\right)}{B(a, b)} \times \prod_{i=1}^J \left(\frac{B(a + S_i, b + n_i - S_i)}{B(a, b)} \right)^{1 - \Omega_{j,i}}. \quad (4)$$

The Bayesian model is complete given specification of a vector comprising the unconditional prior probabilities of all possible pairwise exchangeability configurations, $Pr(\Omega)$, which is challenging given the high dimensionality of the MEM sample domain. Defining $Pr(\Omega)$ as the product of prior exchangeability probabilities for each unique basket pair, however, reduces the dimension from $\prod_{k=1}^{J-1} 2^k$ to $J(J - 1)/2$ yielding feasibility and thereby offering an advantage with respect to conventional Bayesian model averaging

$$Pr(\Omega) = Pr(\Omega_{1,2} = 1) \times Pr(\Omega_{1,3} = 1) \times \cdots \times Pr(\Omega_{J-1,J} = 1). \quad (5)$$

By the Kolmogorov definition of conditional probability, the prior exchangeability probabilities for all 2^{J-1} configurations of Ω_j in (3) follow from (5) as $Pr(\Omega_j = \omega) = \prod_{i=1}^J Pr(\Omega_{j,i} = 1)^{I(\omega_i=1)} \times \{1 - Pr(\Omega_{j,i} = 1)\}^{(1 - I(\omega_i=1))}$, where $I()$ is the indicator function and ω represents one vector of length J within the sample domain of Ω_j .

2.3. Posterior Probability and Effective Sample Size

The MEM Bayesian model specification facilitates posterior inference with respect to all possible pairwise exchangeability relationships among J subtypes. The framework facilitates estimation of disjointed subpopulations comprised of meta-subtypes or singleton subtypes and thereby offers additional flexibility when compared to SEM specifications. The uncontrolled basket study considered herein is devised with the intention of testing the hypothesis that the response probability for a targeted intervention exceeds a null value, which we denote π_0 , while acknowledging the potential for heterogeneity in effectiveness in accordance with the pre-specified basket partitions. Within the MEM framework, this testing procedure follows from the cumulative density function (cdf) of the marginal posterior distribution (2). Specifically, the posterior probability that π_j exceeds π_0 may be computed as the weighted average of cdfs for all possible exchangeability configurations,

$$Pr(\pi_j > \pi_0 | S) = \sum_{g=1}^G Pr(\Omega_j = \omega_g | S) \left\{ 1 - \frac{\int_0^{\pi_0} u^{a+\sum_{h=1}^J \omega_{g,h} S_h - 1} (1-u)^{b+\sum_{h=1}^J \omega_{g,h} (n_h - S_h) - 1} du}{B(a + \sum_{h=1}^J \omega_{g,h} S_h, b + \sum_{h=1}^J \omega_{g,h} (n_h - S_h) - 1)} \right\}. \quad (6)$$

Measurement of the extent to which information has been shared across subtypes in the context of a Bayesian analysis is best characterized by the effective sample size (ESS) of the resultant posterior distribution [see e.g. 17, 18, 19, 10]. The ESS quantifies information content in relation to the number of observables that would be required to obtain the level of posterior precision achieved by the candidate posterior distribution when analyzed using a vague “reference” or maximum entropy prior. Given a specific MEM, the ESS of the conditional posterior in (1) can be derived as

$$ESS(\Omega_j) = a + b + \sum_{h=1}^J \Omega_{j,h} n_h. \quad (7)$$

The marginal posterior ESS may be approximated by the weighted average of the individual MEM-specific values of ESS as $ESS = \sum_{g=1}^G Pr(\Omega_j = \omega_g | S) ESS(\Omega_j)$. Because the initial beta prior distribution carries the effective information of $a + b$ patients, the MEM model facilitates a minimum ESS of $a + b + n_j$ for posterior inference of $\pi_j | S$, which is

The posterior probability of the i, j^{th} pairwise exchangeability relationship, referred to hereafter as the posterior exchangeability probability (PEP) matrix, can be obtained from $Pr(\Omega | S)$ by evaluating the union of MEMs for which $\Omega_{i,j} = 1$ over the sample domain of \mathcal{O}

$$Pr(\Omega_{i,j} = 1 | S) = \sum_{\Omega \in \mathcal{O}} I(\Omega_{i,j} = 1) Pr(\Omega | S). \quad (8)$$

There are $\prod_{k=1}^{J-1} 2^k / 2$ MEM configurations within the space of \mathcal{O} for which $\Omega_{i,j} = 1$.

Similarly, using $Pr(\Omega | S)$ in place of $Pr(\Omega_j | S)$, the mixture posterior distribution in (2) and posterior cdf in (6) may be obtained by evaluating the union of MEMs that characterize a particular row configuration among all possible row vectors of Ω_j ,

$$Pr(\Omega_j = \omega | S) = \sum_{\Omega \in \mathcal{O}} I(\Omega_j = \omega) Pr(\Omega | S). \quad (9)$$

There are $\prod_{k=1}^{J-1} 2^k / 2^{J-1}$ occurrences of each possible row configuration within the space of \mathcal{O} .

In the basket trial setting investigators may desire to maximize data learning by adopting an empirical Bayesian (EB) approach as an alternative to full Bayesian inference. Attenuating prior influence, the EB paradigm when applied to the MEM framework conducts inference using the MEM posterior formulation conditionally with respect to the maximizer of the marginal density, which we denote $\hat{\Omega}$,

$$\hat{\Omega} = \arg \max_{\Omega \in \mathcal{O}} \left\{ \prod_{j=1}^J m(S_j | \Omega_j, S_{(-j)}) \right\}. \quad (10)$$

Case Study Analysis: Vemurafenib non-melanoma basket trial

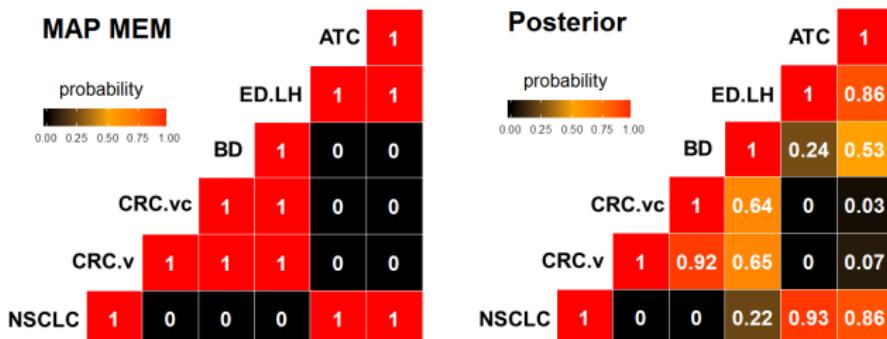
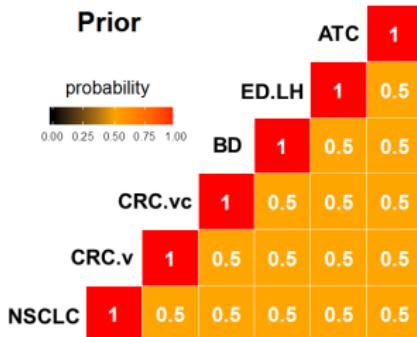


Figure 2. Prior, MAP, and PEP that result from Bayesian inference using the observed vemurafenib basket trial data

Table 1. Case Study Analysis of the Vemurafenib Basket Trial: The baskets and corresponding observed number of response and evaluable patients (*S/N*) reported in Hyman et al. [12] as well as posterior summaries that results from Bayesian analysis of the baskets individually (Basket-specific) and three MEM approaches. Posterior probabilities are reported under PP; effective sample size under ESS; 95% highest posterior density intervals under HPD.

Basket (<i>S/N</i>)	Basket-specific			MEM Full Bayes			MEM Emp Bayes			MEM const Emp Bayes		
	PP	ESS	HPD	PP	ESS	HPD	PP	ESS	HPD	PP	ESS	HPD
ATC (2/7)	0.847	8	0.04, 0.61	0.970	42.2	0.17, 0.55	1.00	41	0.26, 0.56	0.960	24.5	0.13, 0.60
ED.LH (6/14)	0.995	15	0.20, 0.68	0.999	40.6	0.24, 0.55	1.00	41	0.25, 0.54	0.998	28.0	0.22, 0.61
BD (1/8)	0.472	9	0.00, 0.40	0.352	43.3	0.00, 0.40	0.015	45	0.00, 0.12	0.167	27.0	0.00, 0.26
CRC.vc (1/26)	0.039	27	0.00, 0.14	0.018	41.5	0.00, 0.12	0.015	45	0.00, 0.12	0.030	36.0	0.00, 0.13
CRC.v (0/10)	0.068	11	0.00, 0.17	0.027	40.6	0.00, 0.13	0.015	45	0.005, 0.12	0.054	28.0	0.00, 0.15
NSCLC (8/19)	0.998	20	0.23, 0.64	1.00	40.8	0.24, 0.55	1.00	41	0.25, 0.55	0.999	30.5	0.23, 0.59

Bayesian hierarchical modeling of patient subpopulations: Efficient designs of Phase II oncology clinical trials

Scott M Berry,^a Kristine R Broglio,^a Susan Groshen,^b and Donald A Berry^{a,c}

[Author information](#) ► [Copyright and License information](#) ► [Disclaimer](#)

The publisher's final edited version of this article is available at [Clin Trials](#)

See other articles in PMC that [cite](#) the published article.

Bayesian hierarchical adaptive design

In this design, the four patient groups are considered together in a single, integrated trial, and a Bayesian hierarchical model borrows information across the groups. To accomplish this, we model the θ_i with a normal distribution with unknown mean μ and variance σ^2

$$\theta_i \sim N(\mu, \sigma^2), i = 1, 2, 3, 4$$

A second-level of the distribution (hierarchy) is used to model the unknown mean and variance. The data across the groups will shape the posterior distribution for the mean and variance across groups, thus creating a dynamic amount of borrowing, depending on the similarity across groups. The following prior distributions of μ and σ^2 are

$$\mu \sim N(-1.34, 10^2), \sigma^2 \sim \text{Inverse-Gamma}(0.0005, 0.000005)$$

The parameter σ^2 represents the degree of heterogeneity between the patient groups. At one extreme, when σ^2 is 0, there is complete pooling, with adjustment for the targeted p_1 rates in each group, of the results across the patient groups. At the other extreme, when σ^2 is near infinity, then there is no borrowing across the groups. For values between these two extremes, there is an amount of borrowing consistent with the variability across groups. Because the model is powerful enough to capture such extremes, the model results are sensitive to the prior selection for σ^2 . Our selection reflects a small amount of heterogeneity across the four groups. The prior for σ^2 is equivalent to assuming a prior estimate of $\sigma = 0.1$, but with very little weight, 0.1% of one observation. Given the four patient groups to be observed in the trial, the posterior distribution contributes very little information to the posterior. The prior distribution of μ is essentially noninformative, with a weak prior mean close to the null hypothesis.

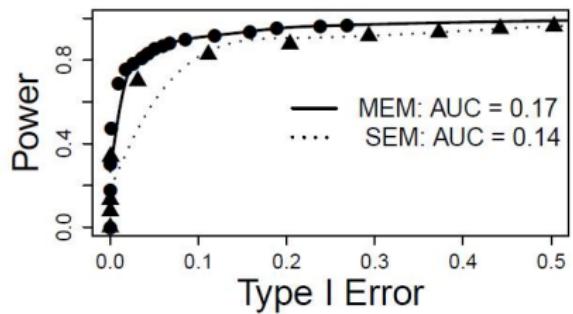
Table 2. Simulation Design: True response probabilities used in our simulation study comparing SEM and MEM methods for basket trial analysis based on the vemurafenib trial.

Simulation Scenario	Basket					
	ATC N=7	ED.LH N=14	BD N=8	CRC.vc N=26	CRC.v N=10	NSCLC N=19
Global Null	0.15	0.15	0.15	0.15	0.15	0.15
Mixed Alternative 1	0.15	0.15	0.15	0.15	0.15	0.45
Mixed Alternative 2	0.15	0.45	0.15	0.15	0.15	0.45
Mixed Alternative 3	0.45	0.45	0.15	0.15	0.15	0.45
Global Alternative	0.45	0.45	0.45	0.45	0.45	0.45

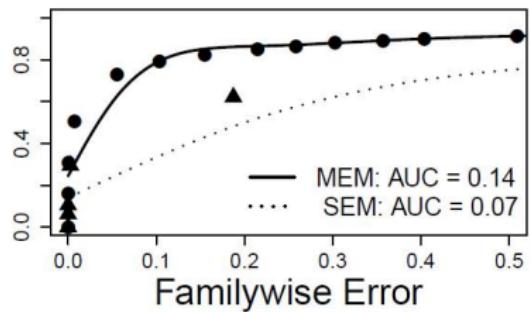
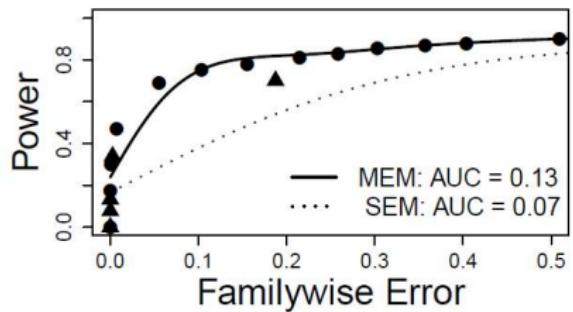
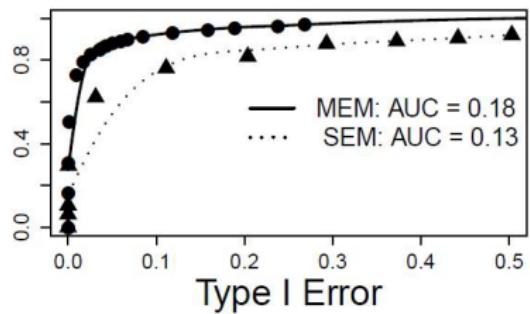
Table 3. Comparing Frequentist Power for basket trial analysis between SEM and MEM methods: Power resulting from our simulation study of the vemurafenib trial for multisource and single-source exchangeability models with type I error controlled at 0.10 under the global null scenario.

Method	Type I error control at 0.1	Posterior Threshold	Power Mixed Alt. 1	Power Mixed Alt. 2	Power Mixed Alt. 3	Power Global Alt.
Multisource Exchangeability Model	basketwise	0.8815	0.909	0.922	0.922	0.995
	familywise	0.9809	0.746	0.787	0.791	0.973
Single-source Exchangeability Model	basketwise	0.9814	0.816	0.746	0.635	0.642
	familywise	0.9952	0.379	0.334	0.268	0.287

Mixed Alternative 1



Mixed Alternative 2



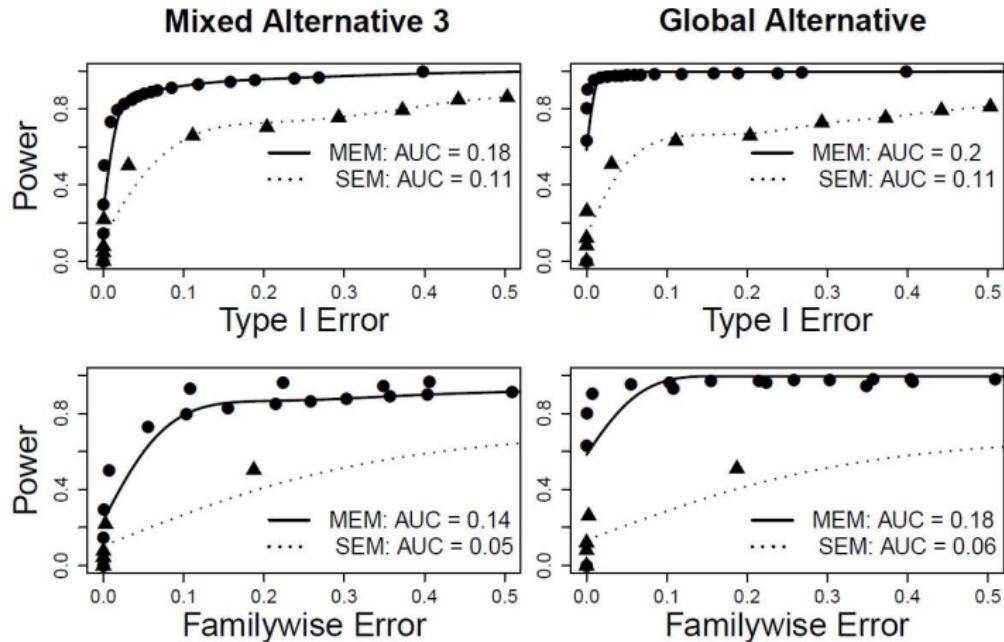


Figure 2. Simulation Study Comparing Full Bayesian MEM to SEM: ROC curves depicting power for each alternative scenario as function of both basketwise (top) and familywise (bottom) type I error under the global null scenario is plotted for both MEM and SEM methods. Points (dots for MEM; triangles for SEM) depict power as a function of type I error for individual posterior decision thresholds. Lines (solid for MEM; dotted for SEM) represent spline interpolations. The area under each ROC curve (AUC) over the type I error domain of $[0, 0.2]$ is given for each subplot.

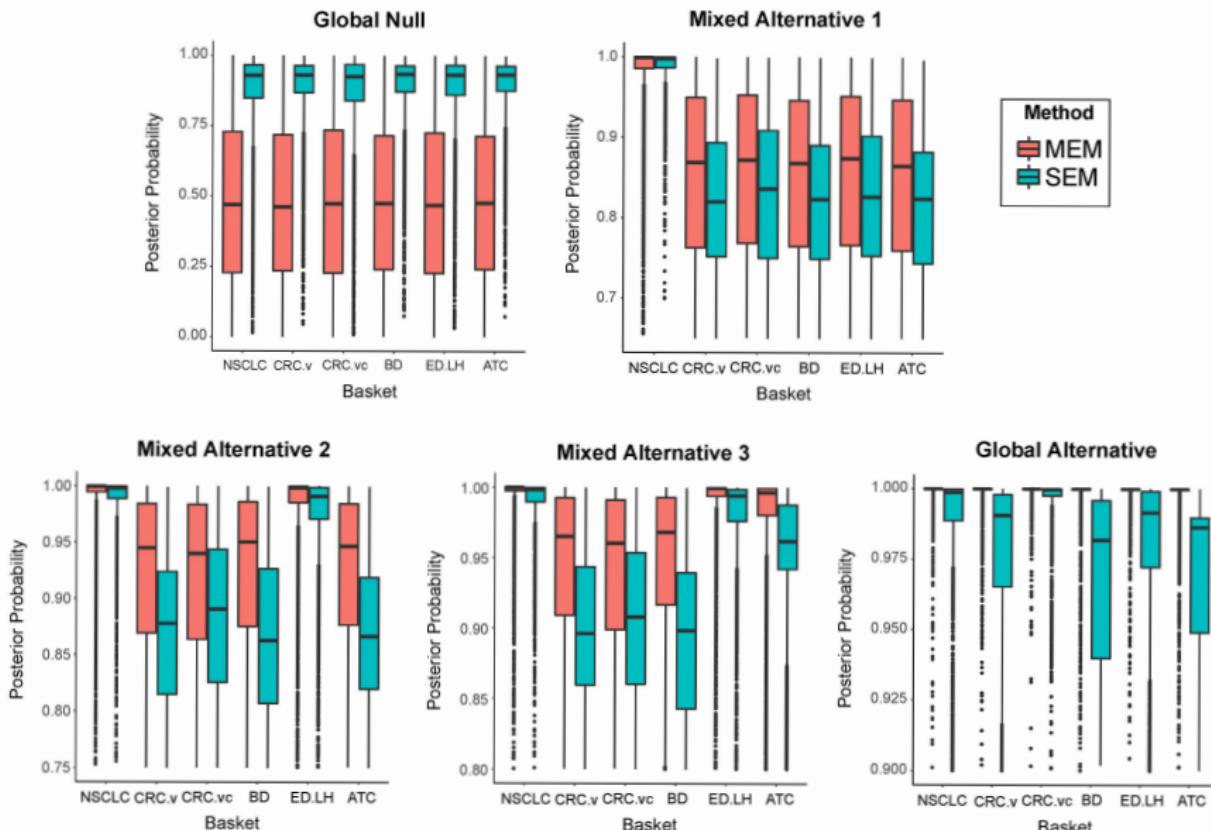
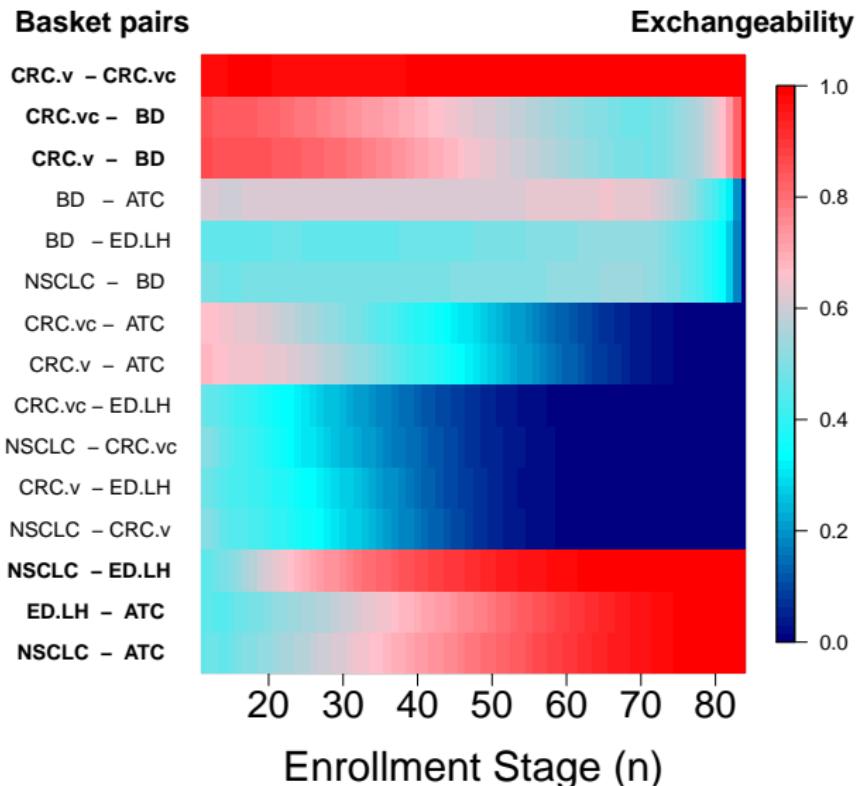


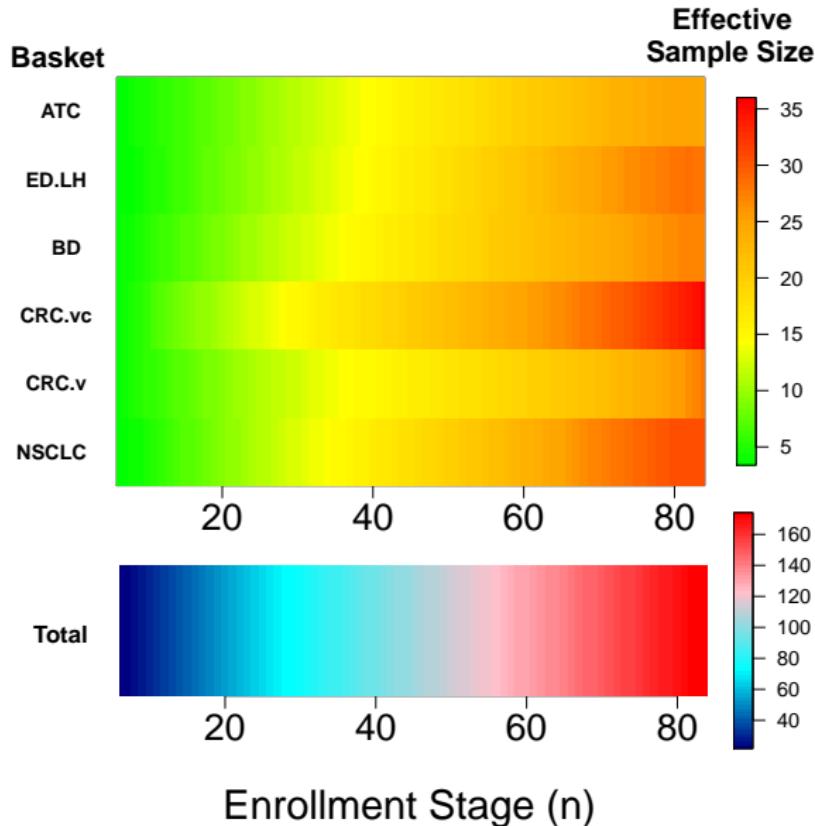
Figure 3. Simulation Study Comparing Full Bayesian MEM to SEM: Boxplots depicting the distribution of posterior probabilities obtained for each simulation scenario for both SEM (blue) and MEM (red) methods.

Sequential Design based on Exchangeability Monitoring with MEM

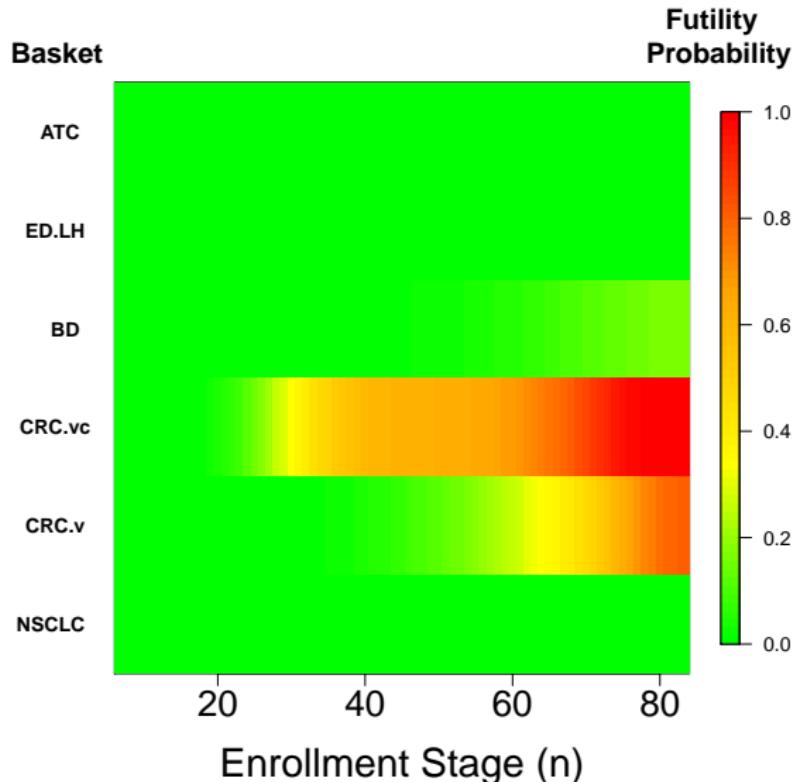
Permutation Study: Vemurafenib non-melanoma basket trial



Permutation Study: Vemurafenib non-melanoma basket trial



Permutation Study: Vemurafenib non-melanoma basket trial



Freidlin and Korn example re-visited

Comparing MEM to subgroup-specific analyses

Scenarios in Freidlin and Korn CCR 2012										
Scenarios	Arm 1	Arm 2	Arm 3	Arm 4	Arm 5	Arm 6	Arm 7	Arm 8	Arm 9	Arm 10
1	0.1	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
2	0.1	0.1	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.3
4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3
5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
6	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3

Frequentist Power for MEM and (subgroup-specific) analyses

Frequentist Size	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 6
10% basketwise Global Null (Sc 5)	0.973 (0.9)	0.967 (0.9)	0.915 (0.9)	0.904 (0.9)	0.980 (0.9)
10% basketwise Single Null (Sc 1)	0.912 (0.9)	0.916 (0.9)	0.796 (0.9)	0.767 (0.9)	0.912 (0.9)
10% familywise Global Null (Sc 5)	0.820 (0.659)	0.807 (0.659)	0.662 (0.659)	0.658 (0.659)	0.852 (0.659)

Comparing MEM to subgroup-specific analyses

Fully Bayesian Evaluation

Priors	Null Scenario					Alternative Scenario				
	1	2	3	4	5	1	2	3	4	6
	0.1	0.1	0.1	0.1	0.6	0.1	0.1	0.1	0.1	0.6
Prior 1	0.1	0.1	0.2	0.2	0.4	0.2	0.2	0.1	0.1	0.4
Prior 2										

10% Average Basketwise Type I error MEM model

Priors	Threshold	Average Type I Error	Average Power
Prior 1	0.890	0.098	0.944
Prior 2	0.895	0.097	0.941

10% Average Familywise Type I error

Prior 1	0.975	0.099	0.802
Prior 2	0.975	0.099	0.794

Recall Case Study A

Case Study: Vemurafenib non-melanoma basket trial

Baskets	Enrolled	Evaluable	Responders	Posterior probability $Pr(\pi > 0.15)$ based on response only	Number (%) of Prior Systemic Therapies		
					≤ 1	2	≥ 3
NSCLC	20	19	8	0.998	11 (55)	4	5
CRC (vemu)	10	10	0	0.068	1	2	7 (70)
CRC (vemu + cetu)	27	26	1	0.039	5 (18)	11	11
Bile Duct	8	8	1	0.472	2	1	5 (63)
ECD or LCH	18	14	6	0.995	9 (50)	7	2
ATC	7	7	2	0.847	5 (71)	1	1

- Are patients with differing treatment histories “statistically exchangeable” as required to infer π ?

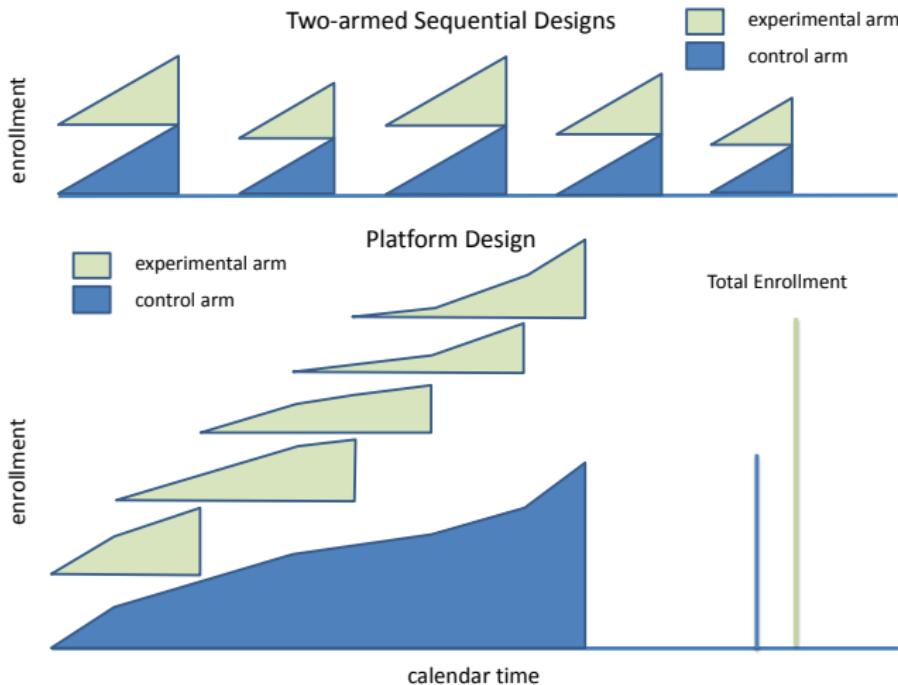
$$Pr(\pi > 0.15 | Data) > \theta$$

- No association between prior therapy reported in Table 1

Platform Designs

- Eliminates inter-trial heterogeneities
- Promotes efficiency through trial consolidation
- Facilitates:
 - Bias Control
 - Randomized intervention comparison *prior to* Phase III
 - Stratification for prognostic balance
 - Distinguish “treatment predictive” from “pre-therapy prognostic features”
 - Identify sub-populations of exceptional responders

Hobbs, Chen, Lee (2016) established screening platforms based on predictive probability

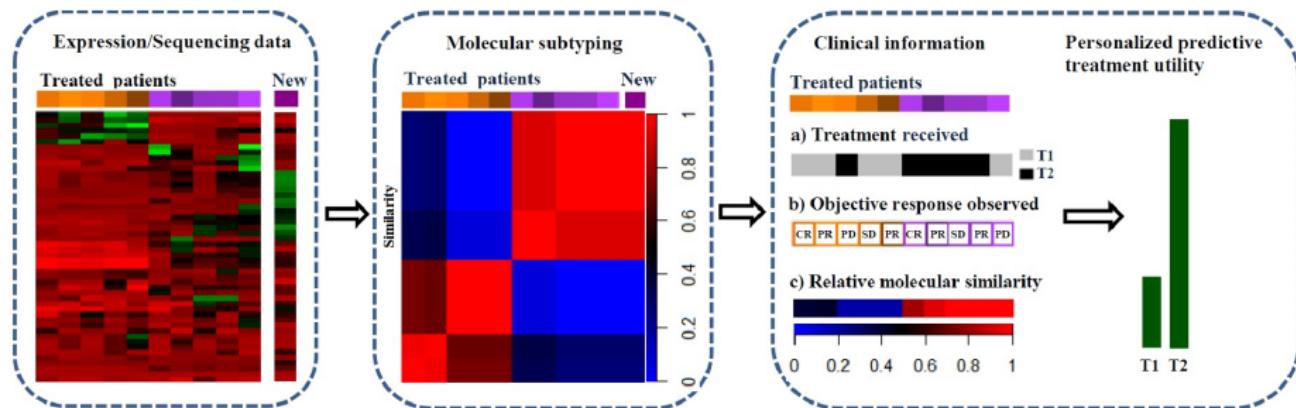


Illustrations of the sequential two-arm and proposed platform-based approaches to randomized

Precision Medicine

Bayesian partial exchangeability frameworks for prec med

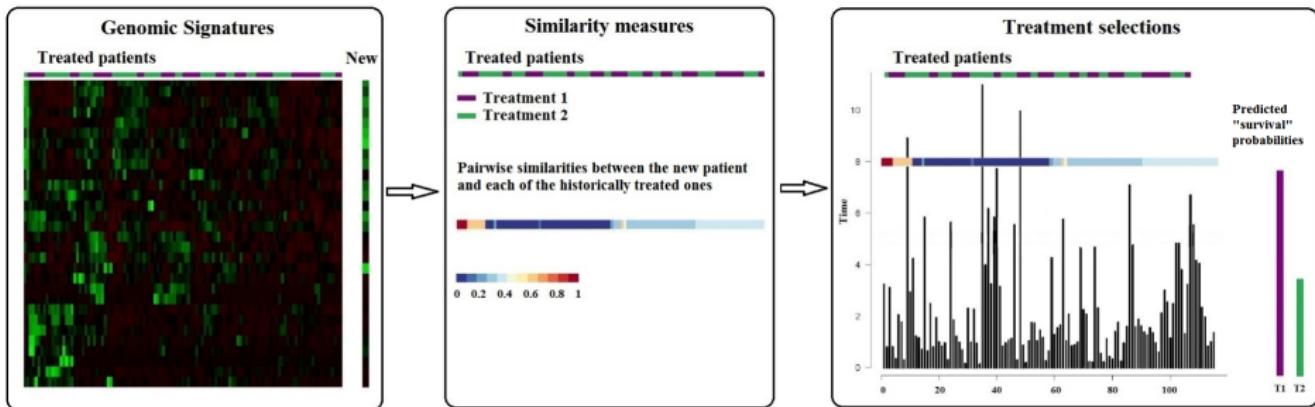
Ma, Stingo, Hobbs. *Biometrics*, (2016). Treatment Selection based on Personalized Predictive Treatment Utilities



- Quantifying similarities from clinical/molecular derived candidate features
- Characterizing pairwise partial statistical exchangeability
- Bayesian prediction models for treatment selection

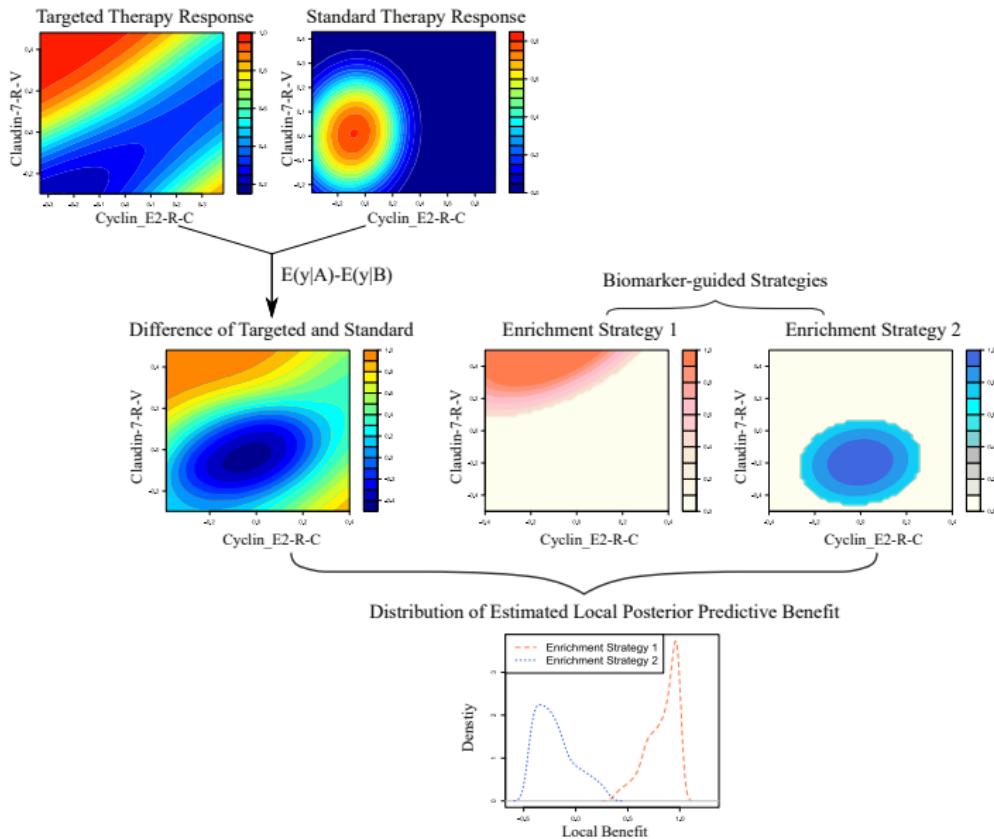
Bayesian partial exchangeability frameworks for prec med

Ma, Hobbs, Stingo. Stat. Methods in Med. Res, (2017).
Treatment Selection based on Personalized Predictive Failure-Time



- Optimal treatment selection based on Bayesian predictive failure time
- Partial exchangeability based on tumor/patient characteristics, pairwise similarity
- Predict the probability of prolonging treatment failure

Huang & Hobbs submitted (2017). Estimating mean local posterior predictive benefit for biomarker-guided treatment strategies



Acknowledgements

Collaborators

Trainees

Caimiao Wei (Pfizer),
Shabnam Azadeh (FDA),
Meilin Huang (Regeneron),
Xiao Li (Gilead),
Kate Shoemaker (Rice),
Yuan Wang (Assist Prof Washington St.)
Junsheng Ma (MD Anderson)

Francesco Stingo, (Univ Florence)
Michele Guindani (UC Irvine)
Chaan Ng (MD Anderson)
Chad Tang (MD Anderson)
Nan Chen (MD Anderson)
Joe Koopmeiners(Minnesota)
Alex Kaizer (Colorado)
Michael Kane (Yale)
Rick Landin (LJPC)