Discussion of three talks on (covariate) adaptive designs

Bhramar Mukherjee

University of Michigan

bhramar@umich.edu

Go Get Them!

Mueller

Hired!

Mukherjee

Ji

Kim

Thall

Papers 1 and 2

- A Bayesian Covariate-Adjusted Response-Adaptive Design with Biomarkers for Targeted Therapies in Cancer: ***Kim and Eickhoff***

- From bench to bedside: The application of differential protein networks on Bayesian adaptive designs for trials with targeted therapies: ***et al, Ji.*** *(changed to ARRM)*

The main theme

- Response-adaptive: Optimize patient allocation by randomizing patients to the superior treatment based on biomarker profile.

- Covariate Adjusted: Identify sub-groups of patients who would respond better to targeted therapy based on biomarker profile and response.

The new feature

- Biomarker profile groups G are initially unknown.
- Define a partition on the subspace generated by *K* potentially correlated biomarkers in a data-adaptive way.
- The partition changes as data is coming in, as opposed to sub-groups being defined at the onset.
- *J* therapies including targets. Target therapy often does not do best for target sub-groups. Does it work for non-targets?
- Allowing uncertainty on sub-group definition is a good thing to do, extracting information from both outcome and biomarkers.

The Advisor or the Student: Who adjusted for covariates better?

- Talked about the past (49 slides)
- Always worked on early phase clinical trials.
- Used Bayesian methods in this paper, usually frequentist
- Adjusted for biomarker profile by partial least squares logistic regression
- Acronym: BCARA-PLSLR

- Talked about the future (20 slides)
- Became interested in early phase trials after thesis work.
- Remained Bayesian
- Used a truly flexible (vague?) Bayesian model to cluster biomarker profiles to define sub-groups
- Acronym: ARRM (America's recovery and reinvestment method)

KyungMann Kim

Bayesian logistic regression model:

$$y_{tj} = \begin{cases} 1 & \text{patient } t \text{ on treatment } j \text{ has a response} \\ 0 & \text{no response} \end{cases}$$

$X_t = (X_{1t},\ldots,X_{Kt})^{\mathrm{T}}$ denotes the biomarker profile of patient t

$$\Pr(y_{tj} = 1 \mid X_t) = \psi\left(\theta_j^{\mathrm{T}} X_t\right) = \frac{\exp\left(\theta_j^{\mathrm{T}} X_t\right)}{1+\exp\left(\theta_j^{\mathrm{T}} X_t\right)}$$

$$y_{tj} = \begin{cases} 1 & \text{if } z_{tj} > 0 \\ 0 & \text{if } z_{tj} \le 0 \end{cases} \text{ where } \begin{cases} z_{tj} = \theta_j^{\mathrm{T}} X_t + \varepsilon_t \\ \varepsilon_t \sim N(0,\lambda_t) \\ \lambda_t = (2\omega_t)^2 \\ \omega_t \sim KS \end{cases}$$

Trial conduct

- Start with equal randomization to each arm.

- Response adaptive randomization with allocation proportion for patient i+1 to treatment j based on posterior probability of treatment j being **superior** to any other treatment j', given biomarker profile X for patient i+1.

Defining profile group

- Dimension reduction of the biomarker space using the leading ``PLS components'' extracted from both biomarker and outcome information.
- Say leading PLS captures most variation.

$$\delta_t^{(i)} = \begin{cases} 1 & \text{if } b_{1t}^{(i)} > 0, \text{ i.e., patient } t \text{ has a "positive" biomarker profile} \\ 0 & \text{if } b_{1t}^{(i)} \leq 0, \text{ i.e., patient } t \text{ has a "negative" biomarker pofile} \end{cases}$$

Stopping Rule

$$p_{jg}^{(i)} = \Pr\Big(\bigcap_{j' \neq j} \mathrm{E}\Big(\sum_{t=1}^{i} y_{tj} \mid X^{(i)}; \theta_j\Big) > \mathrm{E}\Big(\sum_{t=1}^{i} y_{tj'} \mid X^{(i)}; \theta_{j'}\Big) \mid \delta_1^{(i)} = \ldots = \delta_i^{(i)} = g \Big)$$

If $\max_j p_{jg}^{(i)} > \delta$, the trial is stopped after patient i, and treatment $j^* = \arg\max_j p_{jg}^{(i)} > \delta$ is selected for biomarker group g.

If a pre-specified maximum sample size $N_{\max}$ is reached with $p_{jg}^{(N_{\max})} \leq \delta$ for all $g = 1,\ldots,G$ and $j = 1,\ldots,J$, the trial is declared as inconclusive.

Questions/comments

- Why is the value of "g" same for all patients

$$p_{jg}^{(i)} = \Pr\Big(\Big| \underset{j' \neq j}{\quad} \mathrm{E}\Big(\sum_{t=1}^{\cdot} y_{tj} \mid X^{(i)};\theta_j\Big) > \mathrm{E}\Big(\sum_{t=1}^{\cdot} y_{tj'} \mid X^{(i)};\theta_{j'}\Big) \mid \delta_1^{(i)} = \ldots = \delta_i^{(i)} = g\Big)$$

Drug *j* is expected to work if you have target *g*, does it not work if you do not have target *g*?

- How to define partition with more than one PLS components? Penalized PLS to select feature?
- Uncertainties in defining sub-groups, are they measured?
- PLS is sensitive to outliers, robustify?
- The PLS models have interactions, sample sizes?

Yuan Ji

- Change biomarker subgroups during course of the trial according to patients' differential response to treatment.

- Random partition of the covariates subspace based on the extracted principal components

- ARRMS: Adaptive Reassessment and RandoMization Scheme

Yuan Ji

2. Fit a Bayesian model $\prod_i p(y_i \mid x_i = x, t_i = t, \theta) \cdot \pi(\theta)$ to the data of n patients from step 1, denoted as $(\mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n})$.
3. Compute for patient $n+1$,

$$q(t) = Pr(y_{i+1} = 1 \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}, x_{i+1}, t_{i+1} = t) =$$

$$\int Pr(y_{i+1} = 1 \mid x_{i+1}, t_{i+1} = t, \theta) p(\theta \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}) d\theta$$

.

4. Randomize patient $n+1$ to treatment t with probability $\propto f(q(t))$, where $f()$ is a nondecreasing function.

Ji's attempt towards dimension reduction and biomarker summarization

Choice 1 : Fit a regression (logit, probit) with variable selection of $x_i = (x_{i1}, \ldots, x_{ik})$, k being the number of biomarkers.

Choice 2 : Dimension reduction (e.g., PCA) of x_i as $c(x_i)$, where dim$c(x_i) <<$ dimx_i. Fit a regression $p(y_i \mid c(x_i), t_i, \theta)$. – No variable selection.

We took choice 2 (more details on this later).

Magic five PCs are chosen in simulation.

CART-type model for binary outcomes

Let $\Pi = (B_1, B_1, \ldots, B_M)$ be a random partition of $\mathbb{X} = R^k$;

$[\theta \mid \Pi]$:

$$\theta_{j,m} \mid \Pi \overset{iid}{\sim} \text{Beta}(a, b) \qquad j = 1, 2, 3, \;\; m = 1, \ldots, M$$

$[Y \mid X, \textbf{treatment}, \Pi, \theta]$:

$$Y_i \mid X_i, Treatment = j, \Pi, \theta \sim Bernoulli(\theta_{j,m_{X_i}}), \quad m_{X_i} = (m : X_i \in B_m)$$

With a simple random partition:
[Pictures here] and explain it is constructed by randomly selecting one dimension and iteratively splitting one subset with respect to the conditional mean. Also explain that the probability model is constrained to

Constrain: EXPECTED SAMPLE SIZE$(B_m) > 30$

Interesting idea, feasible?

- M=50,000 subsets in partition to start with?
- Random partition is defined after extracting PC from biomarkers.
- Dependence structure of PCs through CART?
- P[y|x,treatment, parameters] or
  P[y|x, treatment, partition, parameter], what is the likelihood?
- End deliverable: Partition with uncertainty, parameter estimate?

Question/Comments

- PC extraction and Construction of Random Partition, can it be accomplished jointly?
- Priors on the random partition: Are the results sensitive?
- Is there danger of overfitting and identifying too many sub-groups? What N and K are typical? (Phase II or III?)
- Do the treatment work beyond targets, do you learn? Are you selecting biomarkers that are predictive of response?
- No comparison with alternatives.

Pathway based clustering: Too fuzzy !

Pathway-based Adaptive Reassessment and RAndomization (P-ARRA)
The rationale:

- Since the treatments target certain pathways, we want to define biomarker groups (disease subpopulations) based on the activations of these pathways.
- Such a definition requires a graphical model $G_i(X_i)$, where G_i denotes a graph on the biomarkers for patient i, $i = 1, ..., n$.
- Applying a probability model with (random) partition, we obtain $T(< n)$ clusters $\{S_1, \ldots, S_T\}$.
- Treat the T clusters as biomarker groups and apply BATTLE or I-SPY2.

Wisely taken out of the talk !

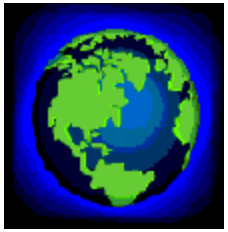
Paper 3: Thall et al, 2011

- ***Optimizing the Bolus and Concentration of a Drug Administered by Continuous Infusion***
  (to appear in *Biometrics*)
- Thoughtful joint modeling of efficacy and toxicity
- Highly motivated by the AIS Phase I trial (too specific?)
- Utility elicited as a function of patient outcome
- Prior elicitation and calibration
- Covariate adjusted? Nay!

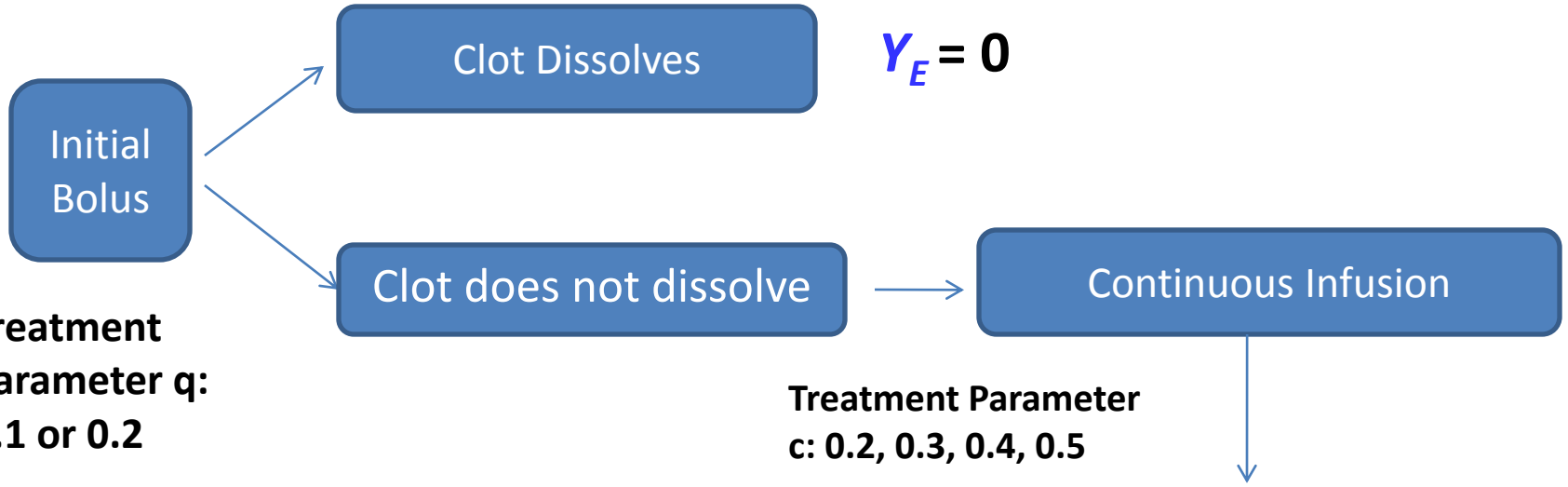
In the speaker's own words:

- ***``My plan is to explain the motivating trial and method, and then elucidate its meaning in the context of our position as a species on a tiny planet orbiting a star located at the edge of one of billions of galaxies in the cosmos".***

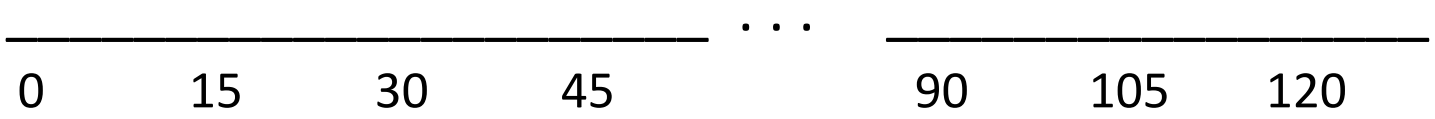
- (Personal communications, July 19, 2011)



The Design



Toxicity event Y_T is a binary variable that depends on c, q and Y_E



Times of infusion and evaluation (minutes)

Y_E = Time to dissolve the clot is *interval censored up to 120 minutes and administratively censored thereafter*

A highly parametric joint model

- $p_0(c,q)$ = Pr(bolus dissolves the clot instantly)

  $Pr(Y_E = 0) = 1 - exp\,(- \alpha_0 c^{\alpha 1} q^{\alpha 2})$

- The pdf of Y_E is the discrete-continuous mixture

$$f_E(y, c, q, \boldsymbol{\alpha}) = p_0(c, q, \boldsymbol{\alpha}) \mathbf{1}(y = 0) +$$

$$\{1 - p_0(c, q, \boldsymbol{\alpha})\} \lambda(y, c, q, \boldsymbol{\alpha})\ e^{-\Lambda(y,c,q,\boldsymbol{\alpha})} \mathbf{1}(y > 0)$$

- The model for the toxic event SICH

$$\pi_T(Y_E, c, q, \boldsymbol{\beta}) = \Pr(Y_T = 1 \mid Y_E, c, q, \boldsymbol{\beta})$$

$$= 1 - \exp[-\{\beta_0 + \beta_2 c^{\beta_1} q + \beta_3 c^{\beta_1} (1 - q)(Y_E \wedge 1) + \beta_4 \mathbf{1}(Y_E > 1)\}]$$

A complex looking likelihood function

For $Y_T = 0$ or 1,

Dissolve the clot with the bolus at $Y_E = 0$

$$\begin{aligned}
\mathcal{L}(\boldsymbol{Y}|c,q,\boldsymbol{\theta}) &= \left[p_0(c,q,\boldsymbol{\alpha})\,\pi_T(0,c,q,\boldsymbol{\beta})^{Y_T}\left\{1-\pi_T(0,c,q,\boldsymbol{\beta})\right\}^{1-Y_T}\right]^{\mathbf{1}(Y_E=0)} \\
&\times \prod_{m=1}^{M}\left[\pi_{E,T}(I_{E,m},1 \mid c,q,\boldsymbol{\theta})^{Y_T}\,\pi_{E,T}(I_{E,m},0 \mid c,q,\boldsymbol{\theta})^{1-Y_T}\right]^{\mathbf{1}(Y_E\in I_{E,m})} \\
&\times \left[\left\{1-F_E(1|c,q,\boldsymbol{\alpha})\right\}\pi_T(1,c,q,\boldsymbol{\beta})^{Y_T}\left\{1-\pi_T(1,c,q,\boldsymbol{\beta})\right\}^{1-Y_T}\right]^{\mathbf{1}(Y_E>1)}
\end{aligned}$$

Dissolve the clot during some 15-minute interval

Fail to Dissolve the clot within 120 minutes

Optimal design: utility function

- Elicit a utility $U(Y_E, Y_T) = U(\boldsymbol{Y})$ from the physicians
- Average utility for a given set of treatment parameters c and q

$$u(c, q, \boldsymbol{\theta}) = \mathrm{E}_{\mathbf{Y}}\{U(\boldsymbol{Y}) \mid c, q, \boldsymbol{\theta}\} = \sum_{y_T=0}^{1} \int_{y_E=0}^{\infty} U(\boldsymbol{y})\, f_{E,T}(\boldsymbol{y} \mid c, q, \boldsymbol{\theta}) dy_E$$

- Optimal Design: over the 4 x 2 eight treatments

$$u(c, q)^{opt}(\mathcal{D}_n) = \underset{c,q}{\operatorname{argmax}}\, \mathrm{E}_{\boldsymbol{\theta}}\{u(c, q, \boldsymbol{\theta}) \mid \mathcal{D}_n\}$$

Trial Conduct (Up to a pre-specified N_{max} patients)

1) Treat 1st patient at lowest pair (*c, q*) = (.20, .10)

2) Treat each patient at the ***optimal*** (*c, q*) pair, that maximizes the posterior expected utility

3) Do not skip untried (*c, q*) pairs when escalating

4) If no (*c, q*) pair is acceptable (low efficacy or high toxicity)

→ Stop the trial

5) Select the ***optimal*** (*c, q*) pair at the end of the trial

Eliciting utility

0	1-15	16-30	31-45	46-60	61-75	76-90	91-105	106-120	>120
100	95	90	85	80	75	70	60	50	30
7	6.5	6	5	4.5	4	2	1	0	0

Time

No SICH

SICH Occurs

Beliefs: Are we allowed to question them? Same for each patient?

Other utility functions: Fixed quantities of the information matrix, are they relevant here?

Eliciting priors: Imaginative or Imaginary?

a. Elicited Prior Mean Probabilities

		$c = 0.20$	$c = 0.30$	$c = 0.40$	$c = 0.50$
$q = .10$	$\mathrm{E}\{p_0(c, q, \boldsymbol{\theta})\}$	0.10	0.15	0.15	0.25
	$\mathrm{E}\{F_E(\frac{1}{2} \mid c, q, \boldsymbol{\theta})\}$	0.25	0.30	0.45	0.50
	$\mathrm{E}\{F_E(1 \mid c, q, \boldsymbol{\theta})\}$	0.35	0.45	0.60	0.70
	$\mathrm{E}\{\pi_T(0, c, q, \boldsymbol{\theta})\}$	0.02	0.03	0.03	0.03
	$\mathrm{E}\{\pi_T(1, c, q, \boldsymbol{\theta})\}$	0.04	0.06	0.08	0.12
$q = .20$	$\mathrm{E}\{p_0(c, q, \boldsymbol{\theta})\}$	0.15	0.20	0.25	0.30
	$\mathrm{E}\{F_E(\frac{1}{2} \mid c, q, \boldsymbol{\theta})\}$	0.40	0.45	0.50	0.60
	$\mathrm{E}\{F_E(1 \mid c, q, \boldsymbol{\theta})\}$	0.50	0.60	0.70	0.80
	$\mathrm{E}\{\pi_T(0, c, q, \boldsymbol{\theta})\}$	0.02	0.03	0.03	0.03
	$\mathrm{E}\{\pi_T(1, c, q, \boldsymbol{\theta})\}$	0.04	0.06	0.08	0.12
$q = .10$ or $.20$	$\mathrm{E}\{\pi_T(.50, q, 1(Y_E > 1), \boldsymbol{\theta})\} = .15$				

Interesting calibration of prior variance: what is non-informative?

- Generate pseudo data with large samples, 50 for each of the 8 treatment combinations given those probabilities.
- Use non-informative prior with the pseudo likelihood to give a pseudo posterior
- Calibrate actual prior variance used for analysis to yield a given effective sample size
- Set the actual prior mean used for analysis to posterior mean of pseudo sample

Questions/comments

- Dependence on Patient accrual, rate of events?
- Very thorough sensitivity analysis under different configurations of design and model parameters.
- Can things be formulated in simpler way to make computing Leśś Miserable! Inclusion of covariates?
- Ad-hocery: How do we sell or justify it?
- 11 parameters, 36 patients, non-informative prior, are we asking for too much for too less?

Session Summary

- All three talks very much in line with what is being pursued in clinical setting.
- Thanks to two of the speakers for sending their slides and thanks to the third for not sending much.
- Combine Phases II and III for covariate adjusted designs?
- Bayesian D-optimality, c-optimality are they relevant?
- Final Analysis: Missing data, Covariate adjustment, Sub-group?
- A discussant should never be assigned 45 minutes.
- Thank you for coming to the session!

Open the floor for fist fights!!