Hierarchical Bayesian Methods for Combining Efficacy and Safety in Mixed Treatment Comparisons

Hwanhee Hong[1], Haitao Chu[2], and Bradley P. Carlin[3]

Division of Biostatistics, University of Minnesota

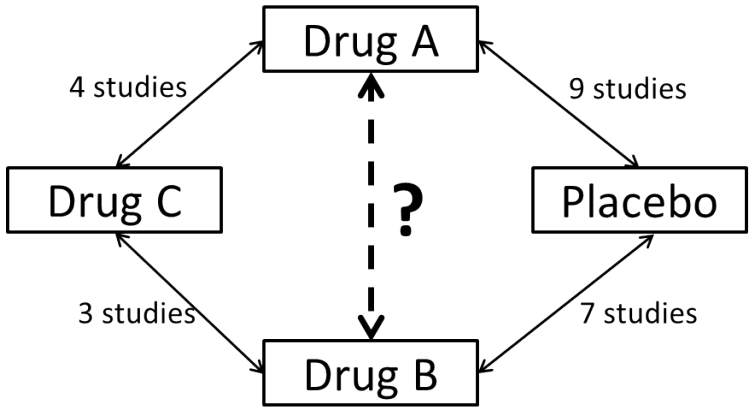
[1] *hong0362@umn.edu*
[2] *chux0051@umn.edu*
[3] *brad@biostat.umn.edu*

FDA-Industry Statistics Workshop, Washington DC
September 13, 2012

Outline

Mixed treatment comparisons (MTCs)

- Mixed treatment comparisons (MTCs) is a meta-analytic statistical technique that incorporates the findings from multiple studies, typically each of treatment against a control (or common comparator) to address the comparative effectiveness and safety of interventions by combining all sources of data.
- MTCs may or may not include head-to-head RCTs of the treatments of primary interest. In such cases, we must rely on *indirect* comparisons that use statistical techniques to incorporate the findings from multiple studies.

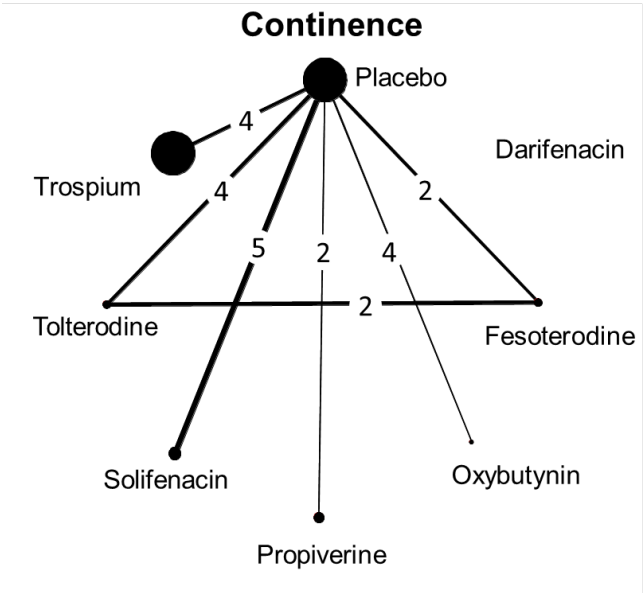


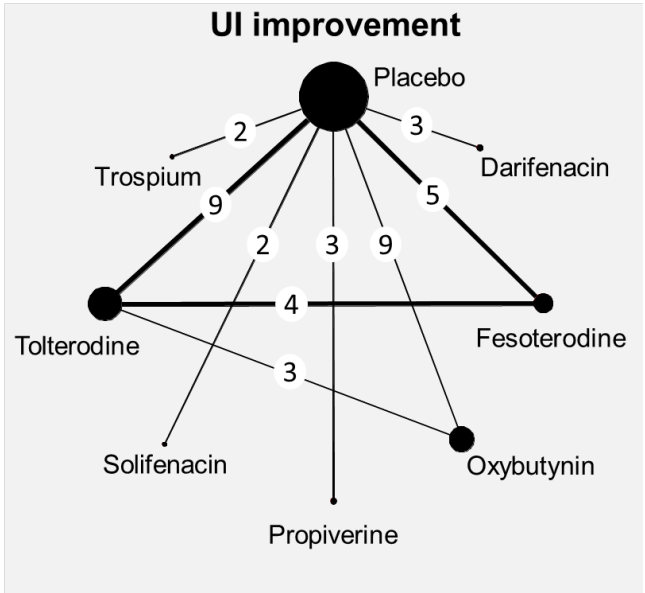
Two major issues in MTCs

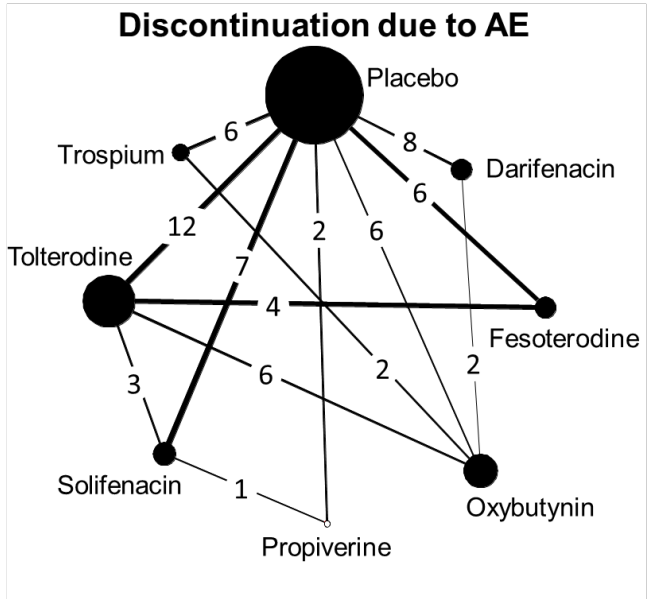
- *Statistical heterogeneity*: effect size variability between studies
- *Evidence inconsistency*: incompatibility that arises between direct and indirect comparisons
  i.e., $LOR_{AB} \neq LOR_{BP} - LOR_{AP}$

In this talk we do not address evidence inconsistency
– to be covered in Sofia's talk!

Urinary incontinence (UI) data

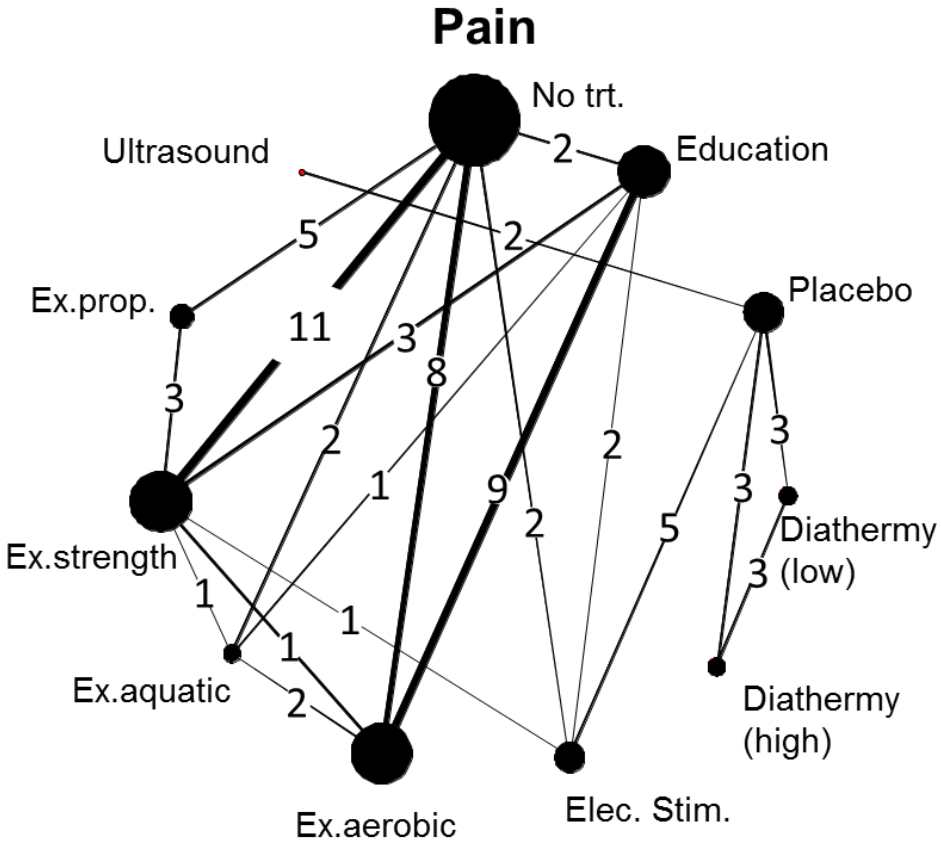


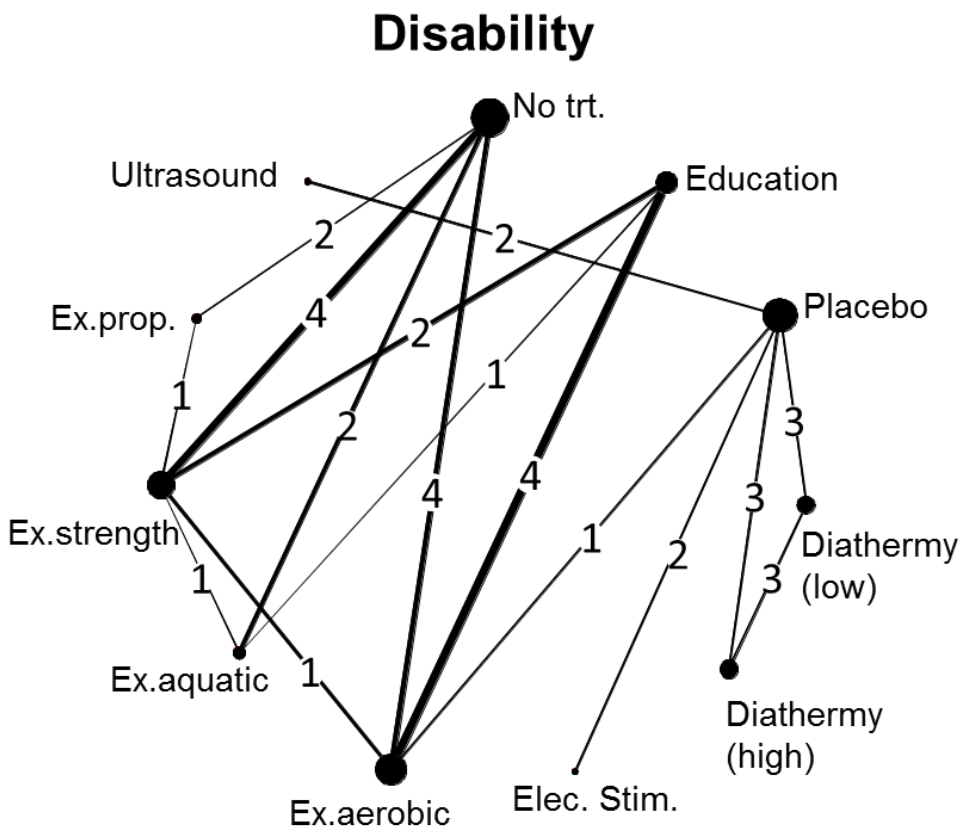




- Our UI data comprise 66 RCTs to compare 8 drugs with respect to
  - 2 efficacy endpoints: continence and UI improvement
  - 1 safety endpoint: discontinuation due to adverse events (AE)
- Each study contains only the number of events and sample sizes in each arm for each outcome; i.e., no individual patient-level data (IPD)

Knee pain secondary to osteoarthritis (OA) data





- 54 RCTs to compare 11 non-drug treatments for 2 continuous outcomes: pain (51) and disability (26); 23 RCTs include both
- Again, only aggregate data (sample means and standard deviations) are reported (no IPD)

Common features of both datasets

- Baseline treatment (control arm) is not the same across studies.
- Many studies report multiple outcomes measured on the same subjects, and correlations across arms and outcomes are likely.
- Missingness rate in each study is high (60-75% for UI, 70-80% for OA), and the missingness may well depend on observed or unobserved factors (say, the results of previous studies).

Binary data likelihood

- Indices
  - $i = 1, \ldots, N$: study
  - $k = 1, \ldots, K$: treatment
  - $\ell = 1, \ldots, L$: outcome
- The data have likelihood

$$y_{ik\ell} \sim Bin(n_{ik\ell}, p_{ik\ell})$$

  - $y_{ik\ell}$: the number of observed events of type ℓ for treatment k in study i
  - $n_{ik\ell}$: sample size (known)
  - $p_{ik\ell}$: probability of having an event (unknown)
- Two types of correlation: between-arm ("Trt") and between-outcome ("Out")

Lu and Ades random effects model (LARE)

- Random effects model allows variability (heterogeneity) between studies

$$\begin{aligned} \text{logit}(p_{ik\ell}) &= \alpha_{iB\ell} && \text{if } k = B \\ \text{logit}(p_{ik\ell}) &= \alpha_{iB\ell} + \delta_{iBk\ell} && \text{if } k \neq B \end{aligned}$$

where

$$\delta_{iBk\ell} \overset{ind}{\sim} N(d_{k\ell} - d_{B\ell}, \tau_\ell^2).$$

  - B indicates the baseline treatment in each study i
  - $\alpha_{iB\ell}$ is the baseline effect
  - $d_{k\ell}$ is the log odds ratio between drug k and placebo
  - $\delta_{iBk\ell} \equiv 0$ when $k = B$

- Priors: $\alpha_{iB\ell}, d_{k\ell} \sim N(0, 100^2)$; $\tau_\ell \sim Unif(0.01, 2)$ $\Leftarrow$ both vague

Limitations of Lu and Ades model

- The reference treatment effect $\alpha_{iB\ell}$ is uninterpretable, since each study may have a different baseline treatment
- Missing arms are not accounted for by the model
- Each study contributes to the likelihood for a different set of treatments, making it hard to construct covariance matrices of contrasts with appropriate priors

Allowing missing data and correlation between outcomes

- We assume that all studies can in principle contain every arm but some of this information is missing for various reasons.
  $\rightarrow$ we can borrow strength from those missing data by imputing them in Bayesian hierarchical models using MCMC algorithms.
  $\rightarrow$ don't need to worry about incompatible baseline treatments

	A	B	C		δ_{iAB}	δ_{iAC}
Study 1	o	o	o	$\rightarrow$	o	o
Study 2	o	o		$\rightarrow$	o	*
Study 3	o		o	$\rightarrow$	*	o

- We incorporate correlation between outcomes
  $\rightarrow$ fit multiple outcomes in one model not in separate models

Contrast-based random effects model (CBRE)

- *Contrast-based* (CB) model contains parameters for relative effects (e.g., log odds ratio).

$$\text{logit}(p_{ik\ell}) = \alpha_{iB\ell} + \delta_{i1k\ell}$$

- **CBRE1**: assume independence between outcomes

$$(\delta_{i12\ell}, \ldots, \delta_{i1K\ell})^T \overset{ind}{\sim} MVN(\mathbf{d}_\ell, \mathbf{\Sigma}_\ell^{Trt}), \quad \ell = 1, \ldots, L$$

  - $\mathbf{\Sigma}_\ell^{Trt}$ is a $(K-1) \times (K-1)$ covariance matrix including between-arm correlations

- **CBRE2**: allow correlation between outcomes but independence between arms

$$(\delta_{i1k1}, \ldots, \delta_{i1kL})^T \overset{ind}{\sim} MVN(\mathbf{d}_k, \mathbf{\Sigma}_k^{Out}), \quad k = 2, \ldots, K$$

  - $\mathbf{\Sigma}_k^{Out}$ is a $L \times L$ covariance matrix including between-outcome correlations

Arm-based random effects model (ABRE)

- *Arm-based* (AB) model contains parameters for absolute effects.

$$\text{logit}(p_{ik\ell}) = \mu_{k\ell} + v_{ik\ell}$$

  - $\mu_{k\ell}$ is the log odds of response under treatment k for outcome ℓ

- **ABRE1**: assume independence between outcomes

$$(v_{i1\ell}, \ldots, v_{iK\ell})^T \overset{ind}{\sim} MVN(\mathbf{0}, \mathbf{\Lambda}_\ell^{Trt}), \quad \ell = 1, \ldots, L$$

  - $\mathbf{\Lambda}_\ell^{Trt}$ is a $K \times K$ covariance matrix including between-arm correlations

- **ABRE2**: allow correlation between outcomes but independence between arms

$$(v_{ik1}, \ldots, v_{ikL})^T \overset{ind}{\sim} MVN(\mathbf{0}, \mathbf{\Lambda}_k^{Out}), \quad k = 1, \ldots, K$$

  - $\mathbf{\Lambda}_k^{Out}$ is a $L \times L$ covariance matrix including between-outcome correlations

Continuous data likelihood

- Indices (same as binary case)
  - $i = 1, \ldots, N$: study
  - $k = 1, \ldots, K$: treatment
  - $\ell = 1, \ldots, L$: outcome
- The data now have likelihood

$$\bar{y}_{ik\ell} \sim N\left(\Delta_{ik\ell}, \frac{\sigma^2_{ik\ell}}{n_{ik\ell}}\right)$$

  - $\bar{y}_{ik\ell}$: observed sample mean measurement in group $ik\ell$ at the end of the study
  - $\Delta_{ik\ell}$: underlying true mean in group $ik\ell$
  - $\sigma_{ik\ell}$: sample standard deviation (assumed known)
  - $n_{ik\ell}$: sample size (known)
- Again, two types of correlation: between-treatment ("Trt") and between-outcome ("Out")

CBRE, ABRE for continuous data

⋆ Contrast-based (CB) model now contains parameters for relative effects (e.g., mean effect differences),

$$\Delta_{ik\ell} = \alpha_{iB\ell} + \delta_{iBk\ell}$$

- **CBRE1**: assume independence between outcomes
- **CBRE2**: allow correlation between outcomes but independence between arms

⋆ Arm-based (AB) model contains parameters for absolute effects,

$$\Delta_{ik\ell} = \mu_{k\ell} + v_{ik\ell}$$

  - $\mu_{k\ell}$ is now the fixed mean effect of treatment k for outcome ℓ

- **ABRE1**: assume independence between outcomes
- **ABRE2**: allow correlation between outcomes but independence between arms

Priors for binary and continuous data models

- $\alpha_{iB\ell} \sim N(a_\ell, \xi_\ell^2)$ (or just $N(0, 100^2)$)
- $\mu_{k\ell}, d_{k\ell} \sim N(0, 100^2)$
- Covariance matrices ($\mathbf{\Sigma}$ and $\mathbf{\Lambda}$) follow vague *InverseWishart*$(\mathbf{\Omega}, \gamma)$ priors of appropriate dimensions

Ranking drugs

- Suppose $P_{k\ell}$ is the marginal probability of experiencing event ℓ under treatment k, obtained by plugging in posterior means of $d_{k\ell}$ and $\alpha_{i1\ell}$ for CB models ($d_{k\ell}$ and $\mu_{k\ell}$ for AB models)
- To integrate their magnitudes and permit different weights on efficacy and safety, we introduce an overall score for drug k, defined as

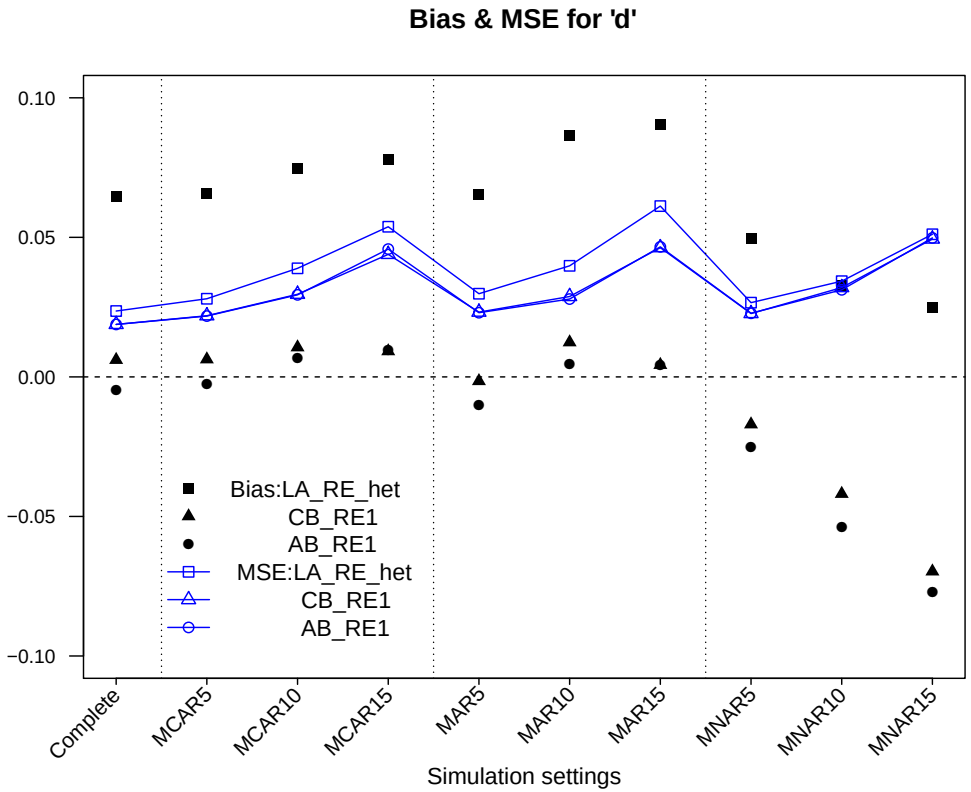
$$S_k = \sum_\ell w_\ell T_{k\ell},$$

  where $T_{k\ell} = P_{k\ell}$ for positive outcomes and $T_{k\ell} = 1 - P_{k\ell}$ for negative outcomes, and w_ℓ is the weight for outcome ℓ such that $\sum_\ell w_\ell = 1$.
- Define:
  "Best1" probability $= Pr\{\text{rank}(S_k) = 1 | \text{data}\}$
  "Best12" probability $= Pr\{\text{rank}(S_k) = 1 \text{ or } 2 | \text{data}\}$

Simulation settings

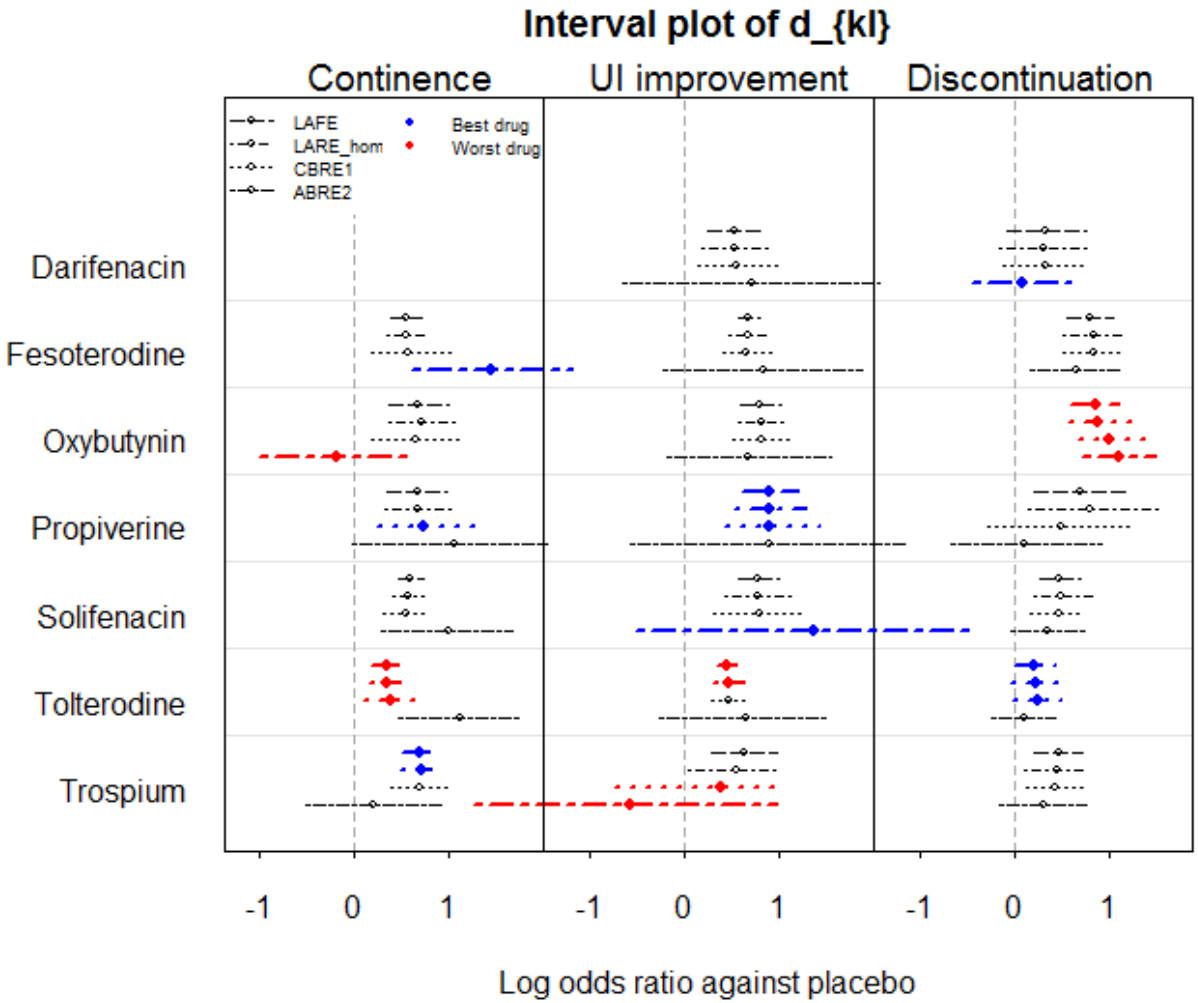
- Compare the performance of LARE, CBRE, and ABRE in terms of bias and MSE under various missingness mechanisms (MCAR, MAR, and MNAR).
- Generate 25 studies comparing two treatments (1, 2) for a single outcome and compare the LARE, CBRE, and ABRE models.
- Drop 5, 10, and 15 arms for treatment 2 under MCAR, MAR, and MNAR missingness mechanisms. Appendix1
- True values in the ABRE model:
  - $(\mu_1, \mu_2) = (-2.5, -1.5)$ which gives $d = \mu_2 - \mu_1 = 1$
  - $(\sqrt{var(v_{i1})}, \sqrt{var(v_{i2})}) = (0.45, 0.65)$
  - $\rho \equiv Corr(v_{i1}, v_{i2}) = 0.692$ Appendix2

Simulation results



- LARE yields a lot larger bias and MSE than our models.
- The results get worse as the amount of missing data gets larger and the degree of missingness increases.

UI data analysis: parameter estimates



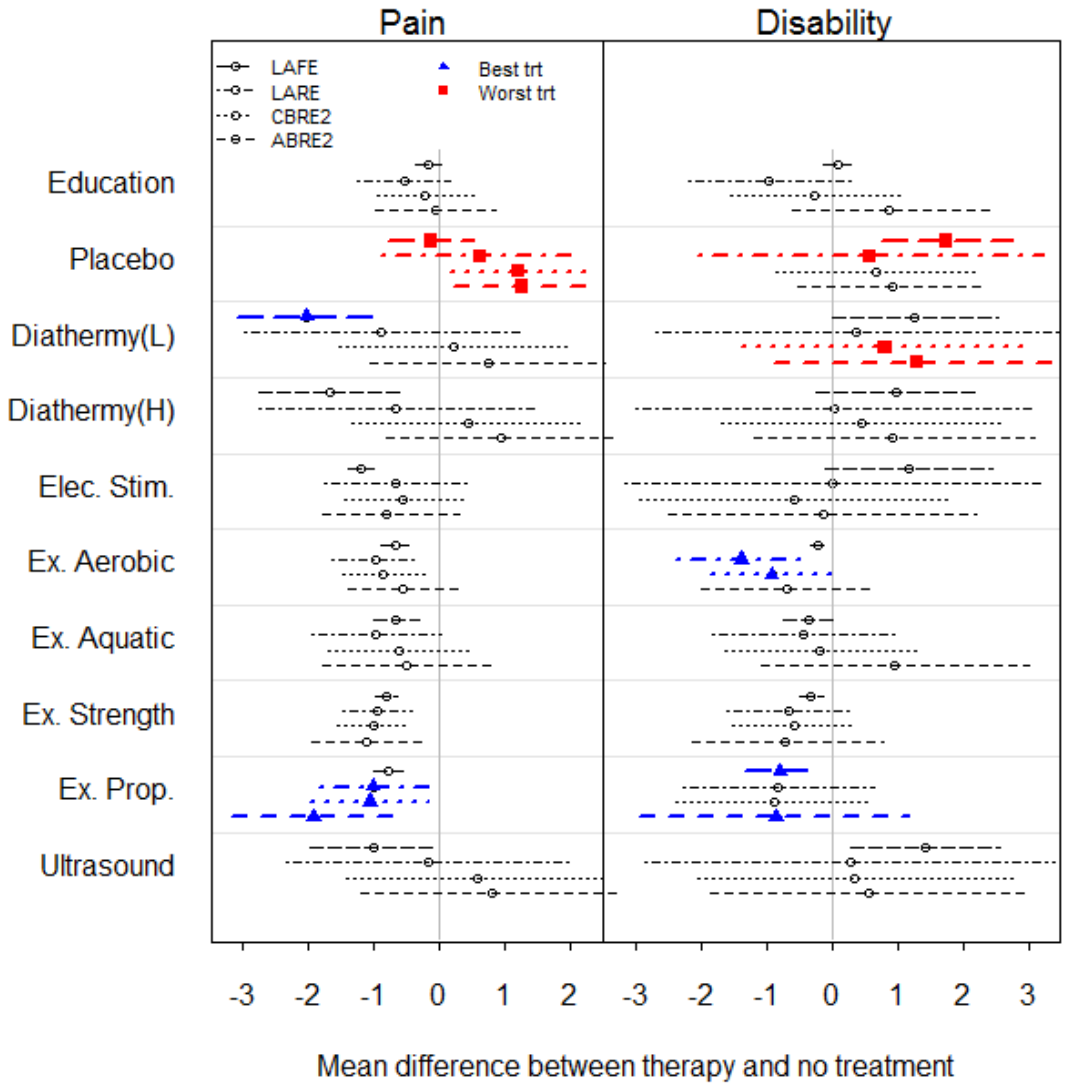
- The best and worst drugs in each model are marked in terms of $d_{k\ell}$.

UI data analysis: decision making

- $S_k = w(w_e P_{k1} + (1 - w_e)P_{k2}) + (1 - w)(1 - P_{k3})$

LARE	Best12.eq (w=0.5)			Best12.eff (w=0.8)			Best12.saf (w=0.2)		
	w_e=0.5	w_e=0.8	w_e=0.2	w_e=0.5	w_e=0.8	w_e=0.2	w_e=0.5	w_e=0.8	w_e=0.2
Placebo	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.000
Darifenacin	0.002	0.000	0.061	0.000	0.000	0.024	0.140	0.030	0.298
Fesoterodine	0.050	0.057	0.059	0.068	0.091	0.072	0.014	0.014	0.015
Oxybutynin	0.444	0.435	0.420	0.611	**0.585**	0.566	0.054	0.070	0.053
Propiverine	**0.603**	0.480	**0.653**	**0.701**	0.570	**0.729**	0.328	0.282	0.362
Solifenacin	0.510	0.363	0.559	0.362	0.235	0.440	**0.545**	0.468	**0.569**
Tolterodine	0.004	0.008	0.006	0.000	0.001	0.000	0.393	0.413	0.348
Trospium	0.387	**0.657**	0.241	0.258	0.520	0.169	0.526	**0.719**	0.354
CBRE1									
Placebo	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.005	0.001
Darifenacin	0.010	0.000	0.095	0.003	0.000	0.052	0.112	0.021	0.264
Fesoterodine	0.128	0.171	0.102	0.179	0.242	0.129	0.021	0.034	0.014
Oxybutynin	0.273	0.259	0.281	0.485	0.435	0.490	0.014	0.021	0.014
Propiverine	**0.795**	**0.708**	**0.774**	**0.781**	**0.699**	**0.751**	**0.629**	**0.590**	**0.644**
Solifenacin	0.544	0.346	0.599	0.398	0.258	0.480	0.542	0.398	0.585
Tolterodine	0.030	0.060	0.019	0.005	0.021	0.002	0.377	0.400	0.298
Trospium	0.220	0.457	0.131	0.150	0.346	0.097	0.303	0.531	0.180

OA data analysis: decision making



The median posteriors of correlations between outcomes are 0.494 (95% BCI 0.18 – 0.71) and 0.377 (0.06 – 0.61) for the CBRE2 and ABRE2 models, respectively, revealing the two outcomes to be positively but weakly correlated.

Discussion and limitations

- Our approaches (incorporating missingness, correlation structure, and arm-based modeling) with binary and continuous outcomes broaden the range of the MTC modeling
- Simulation studies support use of our methods over more standard models under various common missingness mechanisms
- We can apply our models to multiple binary or continuous outcomes
- All our models are fitted under the assumption of consistency.
- In our CB and AB random effect models, we have so far assumed that either between-arm or between-outcome correlations are zero.

Future work

- Two sources of correlation can be incorporated simultaneously in the model by assuming

$$\Delta_{ik\ell} = \mu_{k\ell} + v_{ik} + w_{i\ell}$$

  - $(v_{i1}, \ldots, v_{ik})^T \sim MVN(\mathbf{0}, \mathbf{D}^{Trt})$
  - $(w_{i1}, \ldots, w_{ik})^T \sim MVN(\mathbf{0}, \mathbf{D}^{Out})$
  - v_{ik} and w_{ik} are mutually independent
- We are developing ways of measuring inconsistency, and connecting them to missingness mechanism assumed
- We are currently extending our methods to mixed type of outcomes (a binary safety outcome paired with a continuous efficacy outcome).
- We will ultimately incorporate IPD to borrow more strength, attempt to avoid an ecological fallacy, and investigate how patient-level characteristics impact treatment effects.
  - of keen interest to a firm with access to IPD on its own products that wishes to understand their standing in the marketplace!

References

- Lu G and Ades AE (2006). Assessing evidence inconsistency in mixed treatment comparisons. *JASA*, **101,** 447–459.
- Hong H, Carlin BP, Shamliyan T, Wyman JF, Ramakrishnan R, Sainfort F, and Kane RL (2012). Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. Submitted to *Medical Decision Making*.
- Hong H, Chu H, Shamliyan TA, Wang S, Kane RL, and Carlin BP (2012). A Bayesian missing data framework for multiple continuous outcome mixed treatment comparisons. Research report, Division of Biostatistics, University of Minnesota.
- Spiegelhalter DJ, Abrams KR, and Myles JP (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: Wiley.
- Little, RJ and Rubin, DB (2002). *Statistical Analysis with Missing Data, 2nd Edition*. Chichester: Wiley.

How we generate our partially missing data back

- MAR: missingness depends only on the observed data (X_{obs})

  MNAR: missingness depends on both observed and unobserved data (X_{mis})
- We calculate the 'probability of missing ($p_{i,mis}$)' as follows

$$\text{MAR} : \text{logit}(p_{i,mis}) = \alpha_0 + \alpha_1 \hat{p}_{i1}$$
$$\text{MNAR} : \text{logit}(p_{i,mis}) = \alpha_0 + \alpha_1 \hat{p}_{i1} + \alpha_2 \hat{p}_{i2},$$

  where
  - $\hat{p}_{i1}$ and $\hat{p}_{i2}$ are the observed proportions of events in the two treatment groups
  - α_0, α_1, and α_2 are fixed constants chosen to make the average $p_{i,mis}$ over i to be approximately 0.4; here, for MAR we take $\alpha_0 = 1, \alpha_1 = -3$, while for MNAR we take $\alpha_0 = 1, \alpha_1 = -2, \alpha_2 = -2$.

How to get $\rho = \sigma_1/\sigma_2$? back

In AB_RE1, we set

$$Var[logit(P_{ik})] = \sigma_k^2,$$
$$Cov[logit(P_{i1}), logit(P_{i2})] = \rho\sigma_1\sigma_2.$$

Then, in LA_RE_het and CB_RE1

$$Var[logit(P_{i1})] = Var[\mu_i] = \sigma_1^2,$$
$$Var[logit(P_{i2})] = Var[\mu_i] + Var[\delta_{i2}] = \sigma_1^2 + \sigma^2,$$

$\therefore \sigma^2 = \sigma_2^2 - \sigma_1^2$. And

$$\begin{aligned} Cov[logit(P_{i1}), logit(P_{i2})] &= Cov[\mu_i, \mu_i + \delta_{i2}] \\ &= Var[\mu_i]. \end{aligned}$$

This implies $\rho\sigma_1\sigma_2 = \sigma_1^2$,
$\therefore \rho = \sigma_1/\sigma_2$