LEAP: The latent exchangeability prior for borrowing information from historical data

Ethan M. Alt*

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

*with X. Chang, X. Jiang, Q. Liu, M. Mo, H. A. Xia, and J. G. Ibrahim

Slides available online

- These slides are available on my Web site, accessible from scanning the QR code below.

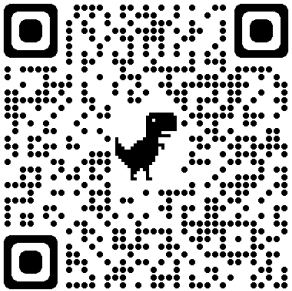


Figure: Scan this QR code for a link.

Motivation

Case Study: The ESTEEM Trials

- The ESTEEM I [5] and ESTEEM II [5] trials were two phase 3 clinical trials for patients with moderate-to-severe plaque psoriasis.
- Patients were randomized 2:1 to Apremilast, an oral phosphodiesterase 4 (PDE4) inhibitor, or placebo.
- The ESTEEM I trial randomized 562 subjects to Apremilast and 282 subjects to placebo.
- In the ESTEEM II trial, 274 subjects were randomized to Apremilast arm and 137 were assigned to placebo.
- We wish to borrow the most relevant (i.e., exchangeable) patients from the ESTEEM I trial.

Current approaches

Propensity Score Approaches

- In the Bayesian information borrowing literature, propensity score (PS) integrated priors have been proposed.

- The PS is defined as the conditional probability a subject enrolls in the current study given covariates, i.e.,

$$e_i = \Pr(S_i = 1|\boldsymbol{x}_i) \text{ and } e_{0i} = \Pr(S_{0i} = 1|\boldsymbol{x}_{0i}),$$

  where $S_i, S_{0i} \in \{0, 1\}$ denotes the study ID (1 = current; 0 = historical) for the current and historical data participants, respectively.

- The goal behind PS methods is to use matching, weighting, or stratification techniques to assess exchangeability.

Propensity Score Approaches

- PS approaches are typically two-stage approaches:
  1. Design stage
     - The PS is estimated.
     - A PS method (e.g., matching, stratification) is applied.
     - Individuals who are matched or within a stratum are considered exchangeable.
  2. Analysis stage
     - A treatment-only model is estimated.
     - Bayesian approaches integrate the matched or stratified data with popular priors like power priors [2], commensurate priors [1], and meta-analytic predictive (MAP) priors [7].

Propensity Score Approaches

- In the causal inference (CI) literature, it is shown that PS approaches result in balance in the baseline covariates, so that exchangeability can be assumed.
- The underlying DAG when using PS approaches in clinical trials lacks real-world intuition.
- In many cases, it is more desirable to condition on covariates.
- Moreover, these approaches treat the PS as known and fixed, which is at odds with Bayesian thinking.
- In fact, in the causal inference literature, Bayesian approaches that incorporate propensity scores have been criticized due to the lack of a Bayesian justification for the PS model (i.e., a joint model for the outcome and PS).

LEAP: The general methodology

The Latent Exchangeability Prior

- The LEAP is predicated under the assumption that the historical data follow a finite mixture model (FMM), where one of the components of the mixture is the same density as the current data.

- Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ denote the parameters for each of the K components in a FMM and let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$ denote the mixture probabilities.

- The LEAP is given by

$$\pi_{\text{LEAP}}(\boldsymbol{\theta}, \boldsymbol{\gamma}) \propto \prod_{i=1}^{n_0} \left\{ \sum_{k=1}^{K} \gamma_k f(y_{0i} | \boldsymbol{\theta}_k) \right\} \pi_0(\boldsymbol{\theta}), \tag{1}$$

  where π_0 is an initial prior for $\boldsymbol{\theta}$.

The Latent Exchangeability Prior (LEAP)

- We assume WLOG that the parameter vector for the current data is $\boldsymbol{\theta}_1$.
- Thus, the joint density of the current data is given by $f(\boldsymbol{y}|\boldsymbol{\theta}_1) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta}_1)$.
- The parameter $\gamma_1 = \Pr(c_{0i} = 1)$ quantifies the marginal probability of being exchangeable with the current data.
- **Theorem:** The initial prior π_0 in (1) can sometimes be improper for the first component, but must otherwise be proper.

The Latent Exchangeability Prior

- It is more convenient to work with the LEAP in (1) using the latent class representation.
- Let $c_{0i} \in \{1, \ldots, K\}$ denote to which component subject i of the historical data belongs.
- The latent class representation of the LEAP is

$$\pi_{\text{LEAP}}(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{c}_0) = \prod_{i=1}^{n_0} \prod_{k=1}^{K} \{\gamma_k f(y_{0i}|\boldsymbol{\theta}_k)\}^{c_{0ik}}, \tag{2}$$

  where $\boldsymbol{c}_{0i} = (c_{0i1}, \ldots, c_{0ik})$ is a multinomial vector and $c_{0ik} = 1\{c_{0i} = k\}$.

- Thus, (2) indicates that individuals with $c_{0i} = 1$ are exchangeable with the current data.

Connection with Bayesian Model Averaging

- Bayesian model averaging (BMA) is considered a gold standard in handling model uncertainty.
- For the ESTEEM trials, our uncertainty arises in not knowing who among the ESTEEM I trial is exchangeable with the ESTEEM II trial.
- Ideally, we would place prior probabilities on all $L = K^{n_0}$ partitions of the historical data, say, π_{0l}, $l = 1, \ldots, L$.
- Given the l^{th} partition, say, $\boldsymbol{y}_{0l}$, we may compute the posterior as

$$p(\boldsymbol{\theta}_1|\boldsymbol{y}, \boldsymbol{y}_{0l}) \propto \mathcal{L}(\boldsymbol{\theta}_1|\boldsymbol{y})\pi(\boldsymbol{\theta}_1|\boldsymbol{y}_{0l}).$$

- Finally, we average over these partitions so that our posterior is given by

$$p(\boldsymbol{\theta}_1|\boldsymbol{y}, \boldsymbol{y}_0) = \frac{\sum_{l=1}^{L} \pi_{0l} p(\boldsymbol{\theta}_1|\boldsymbol{y}, \boldsymbol{y}_{0l})}{\sum_{m=1}^{M} \pi_{0m} \int p(\boldsymbol{\theta}_1|\boldsymbol{y}, \boldsymbol{y}_{0m}) d\boldsymbol{\theta}_1}.$$

Connection with Bayesian Model Averaging

- Unfortunately, the classical BMA approach is not practical in general.
- Let $\pi(\mathcal{D}_{0l})$ denote the probability from the PMF $f(\boldsymbol{c}_0)$ corresponding with the l^{th} partition, $\mathcal{D}_{0l}$. Let $\mathcal{D}_{0lk} = \{i \in \mathcal{D}_{0l} : c_{0i} = k\}$.
- Let $\pi(\boldsymbol{\theta}, \boldsymbol{\gamma}|\mathcal{D}_{0l}) \propto \pi_0(\boldsymbol{\theta}, \boldsymbol{\gamma}) \prod_{k=1}^{K} \gamma_k^{n_{0k}} L(\boldsymbol{\theta}_k|\mathcal{D}_{0lk})$ be the conditional LEAP corresponding to partition $\mathcal{D}_{0l}$.
- We refer to the $\pi(\mathcal{D}_{0l})$'s as "prior partition probabilities." The prior for $\boldsymbol{\theta}$ may then be expressed as

$$\pi(\boldsymbol{\theta}, \boldsymbol{\gamma}|D_0) = \sum_{l=1}^{L} \pi(\mathcal{D}_{0l}) \pi(\boldsymbol{\theta}, \boldsymbol{\gamma}|\mathcal{D}_{0l}). \tag{3}$$

- Thus, the LEAP is conceptually equivalent to BMA with prior partition probabilities induced by the mixture model and initial prior of the historical data.

BMA: Poisson Example

- Suppose the historical data are $(y_{01}, y_{02}, y_{03}) = (1, 2, 6)$.
- Let $n_{0k} = \sum_{i=1}^{n_0} 1\{c_{0i} = k\}$ and let $\bar{y}_{0k} = n_{0k}^{-1} \sum_{i=1}^{n_0} (y_{0i} 1\{c_{0i} = k\})$.
- We elicit $\gamma \sim \text{Dir}(0.9, 0.9)$ and $\theta_k \sim \text{Gamma}(0.1, 0.1)$, yielding

$$\pi(\boldsymbol{\theta}|\boldsymbol{c}_0, D_0) \propto f_{\text{Dir}}(\gamma|\boldsymbol{n}_0 + 0.9 \times \mathbf{1}) \prod_{k=1}^{K} f_\Gamma\left(\theta_k | n_{0k}\bar{y}_{0k} + 0.1, n_{0k} + 0.1\right)$$

- It can be shown that

$$\pi(\boldsymbol{c}_0|D_0) \propto B(n_{01} + 0.9, n_{02} + 0.9) \times \prod_{k=1}^{2} \frac{\Gamma\left(n_{0k}\bar{y}_{0k} + 0.1\right)}{(n_{0k} + 0.1)^{n_{0k}\bar{y}_{0k}+0.1}}$$

BMA: Poisson Example

(c_{01}, c_{02}, c_{03})	$E[\theta_1 \mid D_0, \boldsymbol{c}_0]$	$\pi(\boldsymbol{c}_0 \mid D_0)$	$E[\theta_1 \mid D, D_0, \boldsymbol{c}_0]$	$p(\boldsymbol{c}_0 \mid D, D_0)$
$(1,1,1)$	2.94	0.319	1.84	0.412
$(2,2,2)$	1.00	0.319	1.50	0.108
$(1,1,2)$	1.48	0.092	1.50	0.259
$(2,2,1)$	5.55	0.092	1.90	0.017
$(1,2,1)$	3.38	0.020	1.83	0.019
$(2,1,2)$	1.91	0.020	1.54	0.045
$(1,2,2)$	1.00	0.068	1.45	0.105
$(2,1,1)$	3.86	0.068	1.91	0.035

Table: Prior and posterior means and partition probabilities for a Poisson model with $n = 10$, $\bar{y} = 1.5$, $\boldsymbol{y}_0 = (1,2,6)$, $\gamma_1 \sim \text{Beta}(0.9, 0.9)$, and $\theta_k \sim \text{Gamma}(0.1, 0.1)$. The overall posterior mean is $E[\theta_1 \mid D, D_0] = 1.66$.

BMA: Poisson Example

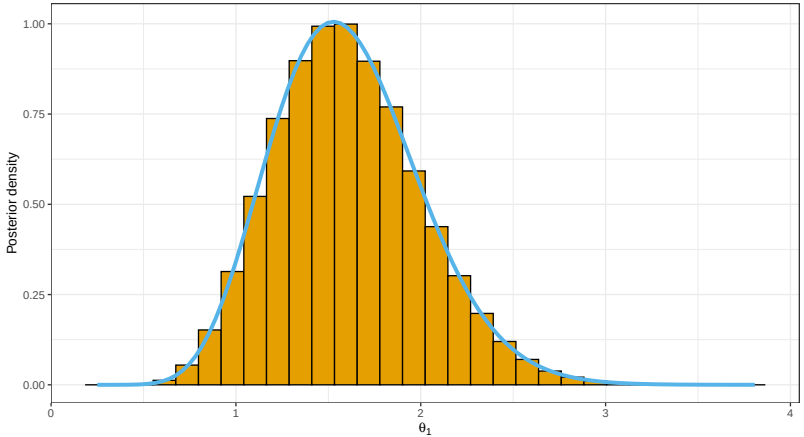


Figure: Comparison of MCMC and BMA approaches.

LEAP: Asymptotic Properties

- Asymptotically, the LEAP is capable of pooling the current and historical data sets together under full exchangeability.
- To see this, Rousseau and Mengersen (2011) [6] showed that if $\gamma_k = 0$ for some k and $\gamma \sim \text{Dir}(\eta_0)$ with $\eta_{0j} < 1$ for every j, then $E[\gamma_{0k}|D_0] \to 0$ as $n_0 \to \infty$.
- This illustrates both the potential power gains robustness property of the LEAP.
  1. If no one is exchangeable with the current data, we have $E[\gamma_1|D, D_0] \to 0$ as $n_0 \to \infty$.
  2. If everyone is exchangeable with the current data, we have $E[\gamma_1|D, D_0] \to 1$.
  3. If a fraction of individuals are exchangeable with the current data, it is feasible to obtain power gains.

LEAP: Asymptotic Properties

- Consider $y_i \sim \text{Pois}(\theta_1)$, $y_{0i}|c_{0i} \sim \text{Pois}(\theta_{c_{0i}})$, $c_{0i} \sim \text{Cat}(\gamma_1)$, $\gamma_1 \in \{0.0, 0.5, 1.0\}$, $n = 1,000$.
- We report results for a single simulated data set with $K = 2$.
- We specify $\theta_k \sim \text{Gamma}(0.1, 0.1)$, $\gamma_1 \sim \text{Beta}(0.9, 0.9)$.

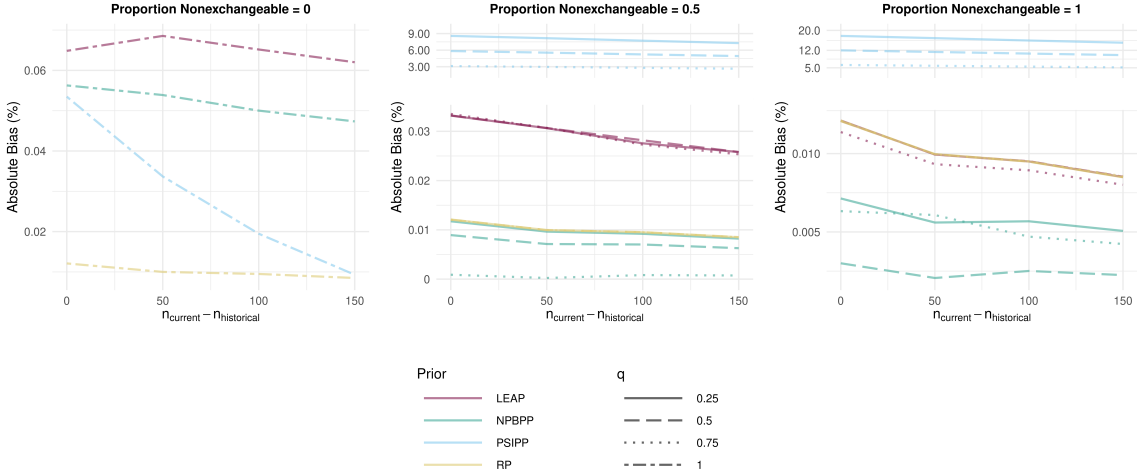
γ_1	n_0	$E[\gamma_1 \mid D, D_0]$	$\text{SD}(\gamma_1 \mid D, D_0)$
0.0	100	0.058	0.045
	1000	0.014	0.011
0.5	100	0.486	0.074
	1000	0.473	0.034
1.0	100	0.926	0.081
	1000	0.981	0.026

Simulation Studies

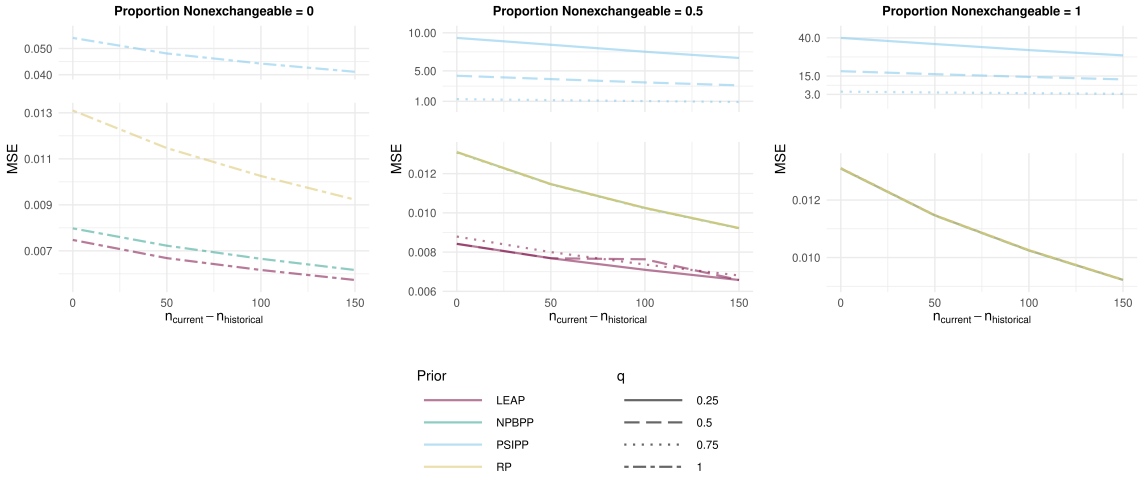
Simulation Studies

- We assume $K = 2$ for the historical data and
  - $y_i \sim N(\boldsymbol{x}_i'\beta_1, \sigma_1^2), \quad i = 1, \ldots, n_0 + h, \quad h \in \{0, 50, 100, 150\}$
  - $y_{0i}|c_{0i} \sim N(\boldsymbol{x}_{0i}'\beta_{c_{0i}}, \sigma^2_{c_{0i}}), \quad i = 1, \ldots, n_0 = 278$
  - $\gamma_1 = \Pr(c_{0i} = 1) = 1 - \Pr(c_{0i} = 2) \in \{0.0, 0.5, 1.0\}$
  - $\boldsymbol{x}_i \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$
  - $\boldsymbol{x}_{0i}|c_{0i} \sim N(\boldsymbol{\mu}_{c_{0i}}, \boldsymbol{\Sigma}_1)$
  - $\boldsymbol{\mu}_2 = q \times \boldsymbol{\mu}_1$ and $\beta_2 = q \times \beta_1, \;\; q \in \{0.25, 0.50, 0.75, 1.0\}$
- Values for β_1, σ_1^2, $\boldsymbol{\mu}_1$, and $\boldsymbol{\Sigma}_1$ were taken as the MLEs of the ESTEEM II trial.
- We elicited $\gamma_1 \sim \text{Beta}(0.9, 0.9)$, $\beta_k \sim N(0, 10^2\boldsymbol{I}_p)$, $\sigma_k^2 \sim \text{Half-Normal}(0, 10^2)$.

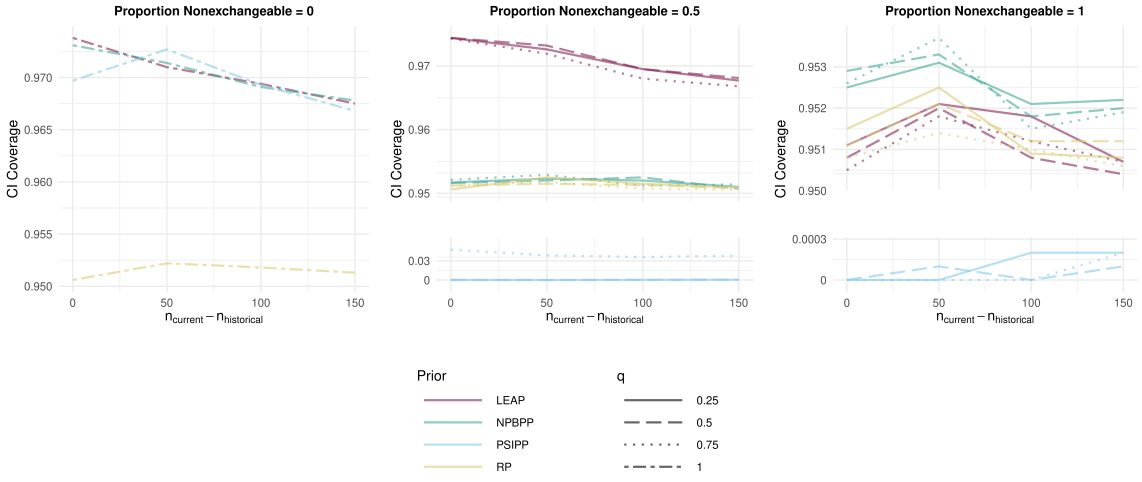
Simulation Results: Percent Absolute Bias



Simulation Results: Mean Squared Error



Simulation Results: 95% Credible Interval Coverage



Application to the ESTEEM Trials

ESTEEM Studies: Data Analysis

- Patient characteristics for patients in the ESTEEM studies are presented in the Table below

	ESTEEM I		ESTEEM II		
Characteristic	**N**	**Placebo**	**N**	**Placebo**	**Apremilast**
Age	282	46.5 $\pm$ 12.7	411	45.7 $\pm$ 13.4	45.3 $\pm$ 13.1
Smoker Category	282		411		
Current user		92 (33%)		61 (45%)	101 (37%)
Not a current user		190 (67%)		76 (55%)	173 (63%)
Prior use of Systemic Therapies	282		411		
N		132 (47%)		64 (47%)	117 (43%)
Y		150 (53%)		73 (53%)	157 (57%)
Baseline PASI Score	282	19.4 (7.4)	411	20.0 (8.0)	18.9 (7.1)
% change in total PASI Score	278	-16.7 (31.5)	405	-15.8 (41.3)	-50.9 (34.0)

: Patient characteristics and of ESTEEM I and ESTEEM II. Continuous covariates show Mean $\pm$ SD; binary covariates show *N* (%).

ESTEEM Studies: Data Analysis

- The number of historical controls is $n_{00} = 278$, and the number of current controls and current treated are given respectively by $n_{10} = 137$ and $n_{11} = 274$.
- We wish to augment the control arm without exceeding the current data treatment arm.
- Thus, we wish to borrow no more than $\tilde{n}_{00} = n_{11} - n_{10} = 137$ controls from the historical data.
- This implies a constraint: $\gamma_1 \leq \frac{137}{278} \approx 0.49$.
- To accommodate this constraint, we derive a partially truncated Dirichlet (PTD) density, which constrains the first element of the vector.

ESTEEM Studies: Data Analysis

- We elicit:
  - $\beta_k \sim N_p(0, 10^2 \boldsymbol{I}_p)$.
  - $\sigma_k \sim$ Half-Normal$(0, 10^2)$.
  - $\gamma \sim$ PTD $(\boldsymbol{\alpha} = 0.95 \times \mathbf{1}, a = 0, b = 0.49)$.
- We compare our approach with
  1. A normalized version of the partial borrowing power prior (Ibrahim et al., 2015) [3], $a_0 \sim U(0, 0.49)$.
  2. A propensity score integrated power prior (PSIPP) of Lu et al. (2022) [4].
  3. A noninformative reference prior.

ESTEEM Studies: Data Analysis

Prior	DIC	Post. Mean	Post. SD	95% CI
LEAP ($K = 2$)	4063.27	-31.4	3.35	$(-38.0, -24.9)$
LEAP ($K = 3$)	4063.26	-31.5	3.27	$(-37.9, -25.0)$
PBNPP	4062.09	-31.9	3.09	$(-37.8, -25.8)$
PSIPP	4091.21	-28.1	2.69	$(-33.3, -22.8)$
Reference	4064.04	-32.1	3.45	$(-38.9, -25.4)$

Table: Summary of the posterior density of the treatment effect (mean difference in % change in PASI score) using the ESTEEM I historical controls and ESTEEM II data sets. DIC = deviance information criterion; Post. Mean = posterior mean; Post. SD = posterior standard deviation, CI = credible interval.

ESTEEM Studies: Data Analysis

	Prior							
	LEAP ($K = 2$)		LEAP ($K = 3$)		PBNPP		Reference	
Parameter	Post. Mean	Post. SD	Post. Mean	Post. SD	Post. Mean	Post. SD	Post. Mean	Post. SD
β_0	−21.23	3.915	−21.37	3.826	−21.63	3.755	−17.97	3.946
β_1	−31.39	3.353	−31.49	3.273	−31.86	3.085	−32.12	3.448
β_2	− 1.06	1.742	− 1.13	1.720	− 0.73	1.611	− 0.48	1.752
β_3	− 1.98	1.493	− 2.08	1.660	− 1.65	1.390	− 2.29	1.539
β_4	− 2.23	3.369	− 2.04	3.372	− 2.23	3.334	− 4.32	3.404
β_5	11.54	3.245	11.93	3.173	12.50	3.216	10.11	3.324
β_6	1.48	1.628	1.48	1.585	1.38	1.563	1.29	1.724
σ_1	35.44	1.145	35.41	1.112	35.27	1.150	35.71	1.237
γ_1	0.40	0.088	0.41	0.083				
a_0					0.34	0.095		

Table: Posterior means and standard deviations from the posterior densities of the ESTEEM-2 trial. The parameter $\sigma_1 = \tau_1^{-1/2}$ is the standard deviation of the outcome. Post. Mean = posterior mean; Post. SD = posterior standard deviation.

ESTEEM Studies: Data Analysis

- The number of historical controls is $n_{00} = 278$, and the number of current controls and current treated are given respectively by $n_{10} = 137$ and $n_{11} = 274$.
- We wish to augment the control arm without exceeding the current data treatment arm.
- Thus, we wish to borrow no more than $\tilde{n}_{00} = n_{11} - n_{10} = 137$ controls from the historical data.
- This implies a constraint: $\gamma_1 \leq \frac{137}{278} \approx 0.49$.
- To accommodate this constraint, we derive a partially truncated Dirichlet (PTD) density, which constrains the first element of the vector.

Conclusion

Conclusion

- We have developed a new class of priors called latent exchangeability priors.
- The LEAP is applicable when only a fraction of individuals in an external data set are exchangeable with the current data set.
- Efficiency gains can be made under partial exchangeability, which was not observed for the normalized power prior.
- Further avenues of research:
  - Treating K as random.
  - A LEAP for time-to-event data.
  - Using PS to elicit a prior for $\gamma_{1i} = \Pr(c_{0i} = 1 | \boldsymbol{x}_{0i})$.

Resources

- Software for implementing the proposed method is available at https://tinyurl.com/leapcode.
- A pre-print of the paper is available at https://tinyurl.com/leapprior.

Figure: Scan this QR code for the pre-print.

References I

[1] B. P. Hobbs, D. J. Sargent, and B. P. Carlin. Commensurate Priors for Incorporating Historical Information in Clinical Trials Using General and Generalized Linear Models. *Bayesian Analysis*, 7(3):639–674, Aug. 2012.

[2] J. G. Ibrahim and M.-H. Chen. Power prior distributions for regression models. *Statistical Science*, 15(1):46–60, Feb. 2000.

[3] J. G. Ibrahim, M.-H. Chen, M. Lakshminarayanan, G. F. Liu, and J. F. Heyse. Bayesian probability of success for clinical trials using historical data. *Statistics in Medicine*, 34(2):249–264, Jan. 2015.

[4] N. Lu, C. Wang, W.-C. Chen, H. Li, C. Song, R. Tiwari, Y. Xu, and L. Q. Yue. Propensity score-integrated power prior approach for augmenting the control arm of a randomized controlled trial by incorporating multiple external data sources. *Journal of Biopharmaceutical Statistics*, 32(1):158–169, 2022.

References II

[5] K. Papp, K. Reich, C. L. Leonardi, L. Kircik, S. Chimenti, R. G. B. Langley, C. Hu, R. M. Stevens, R. M. Day, K. B. Gordon, N. J. Korman, and C. E. M. Griffiths. Apremilast, an oral phosphodiesterase 4 (PDE4) inhibitor, in patients with moderate to severe plaque psoriasis: Results of a phase III, randomized, controlled trial (Efficacy and Safety Trial Evaluating the Effects of Apremilast in Psoriasis [ESTEEM] 1). *Journal of the American Academy of Dermatology*, 73(1):37–49, July 2015.

[6] J. Rousseau and K. Mengersen. Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):689–710, 2011.

[7] H. Schmidli, S. Gsteiger, S. Roychoudhury, A. O'Hagan, D. Spiegelhalter, and B. Neuenschwander. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4):1023–1032, 2014.