Multivariate Bayesian Logistic Regression for Clinical Safety Data

William DuMouchel, PhD
Oracle Health Sciences

4th Seattle Symposium in Biostatistics: Clinical Trials
23 November 2010

ORACLE

Logistic Regression for Subgroup Analyses of Multiple Events

- Start from a Set of *Medically Related* Events to Study
  - Set of ad-hoc events, or all events within a MedDRA SOC
- Fit Logistic Regressions to each AE as a Response
  - Use exactly the same predictor model for each AE
    - Age, gender, concomitant medication, medical history, etc.
  - Include treatment and interactions with treatment as predictors
  - Generate parameter estimates for predictors and interactions
- Empirical Bayes Shrinkage of Estimated Coefficients
  - Coefficients of each predictor borrow strength across AEs
  - Overall treatment and interaction effects shrink toward 0

Multiple Medically Related Events

- Consider Ten MedDRA PTs
  - *Anuria Dry mouth Hyperkalaemia Micturition urgency Nocturia Pollakiuria Polydipsia Polyuria Thirst Urine output increased*
  - All seemed somewhat Treatment related in 2x2 analyses
- Want Ten Separate Estimates of Treatment Effect
  - But some or many of them may have a common cause
    - Common side effects of diuretics
  - Analyze them with a common statistical model
  - Do they have similar responses to various predictor variables?
  - Let the data decide how much they should “borrow strength from each other

Multivariate Bayesian Logistic Regression (MBLR)

- Use the Same Covariates to Predict all 10 Responses (in Addition to Treatment vs. Placebo Estimates)
  - Sex (F, M)
  - Race (Black, White, Other)
  - Age Group ($\leq$ 55, 55 to 65, 65 to 75, > 75)
  - Indication (4 Trials w/ Indication 1, 4 Trials with Indication 2)
    - We could have used Trial itself as a predictor instead of Indication
  - Renal Medical History (Yes, No)
- Five Covariates Need 8 Degrees of Freedom

Rationale for Use of Covariates

- Since all Trials Were Randomized, Why Adjust for Covariates? Won't They all Balance Out Anyway?
  - Depending on sample sizes, will not be perfect balance
  - If covariates have strong effects, adjustment for them will reduce residual variance and therefore Treatment effect uncertainty
  - Less focus on a single pre-specified model for safety analyses than for efficacy analyses
- Main Rationale—Treatment by Covariate Interactions
  - Estimating Treatment x Covariate interactions in a safety analysis is equivalent to searching for vulnerable subgroups
  - MBLR– cross every covariate with the Treatment effect

Rationale for EB Model Across Events

- Coping with Fine Granularity of Adverse Event Data
  - Compare T vs. C on 10 varieties of renal or dehydration issues
  - Approach 1—separate analyses of all 10 events
    - Small counts lead to non significant comparisons
    - Adjustment for multiple comparisons further reduces sensitivity
  - Approach 2—define a single event as union of the 10 events
    - Significant differences may be washed out by the pooling
    - Even if significant, little information about original 10 differences
- Compromise Approach—EB Hierarchical Model
  - 10 individual estimates that “borrow strength” from each other
  - Estimate separate vector of coefficients for each AE
    - But a prior distribution shrinks corresponding coefficients across AEs toward each other
    - The amount of shrinkage is controlled by certain prior variances that are also estimated from the data
    - Treatment-Covariate interaction effects, which are *apriori* less likely, are also shrunk toward the null hypothesis value of 0

Defining Regression Effect Estimates

- Include every Treatment by Covariate Interaction
- Statistical Model for P_{ik} = Prob(Event k in i^{th} Patient)
  - X_{ig} = g^{th} covariate; T_i = Treatment arm indicator

$$P_{ik} = 1/[1 + \exp(-Z_{ik})]$$

$$Z_{ik} = \alpha_{0k} + \Sigma_{1 \le g \le G} X_{ig} \alpha_{gk} + T_i (\beta_{0k} + \Sigma_{1 \le g \le G} X_{ig} \beta_{gk})$$

Bayesian Shrinkage Models

- Statistical Validity of Searching for Extreme Differences
  - Most significant adverse event or patient subgroup
- Classical Approach to Post-Hoc Interval Estimates
  - Maintain centers of CI at observed differences
  - Expand widths of every CI
  - Expansion is greater the more differences you look at
    - If you look at too many, the CI's are too wide to be useful
- Bayesian Approach
  - Requires a prior distribution for differences
    - Can estimate it from the multiple observed differences available
  - Centers of CI's are "shrunk" toward average or null difference
    - High-variance differences shrink the most
  - Widths of CI's usually shrink a little too
    - The more you look at, the better you can model the prior dist.

Prior Distributions for Coefficients

- Two-Stage Hierarchical Model
  - Covariate main effects α_{gk} shrink toward means across issues
  - Treatment main effects β_{0k} shrink toward each other
  - Treatment interactions β_{gk} also shrink toward 0
  - Four prior standard deviations control amount of shrinkage
    - Let $\phi = (\sigma_A, \sigma_0, \sigma_B, \tau)$; prior distributions uniform (0, d = 1.5)
  - $\{\alpha_{0k}, A_g, B_0\}$ have uniform priors $(-\infty, +\infty)$
  - Remaining parameters have prior distributions:

$$\alpha_{gk}|A_g \sim N(A_g, \sigma_A^2), \qquad k = 1, \ldots, K;\ g = 1, \ldots, G$$

$$\beta_{0k}|B_0 \sim N(B_0, \sigma_0^2), \qquad k = 1, \ldots, K$$

$$\beta_{gk}|B_g \sim N(B_g, \sigma_B^2), \qquad k = 1, \ldots, K;\ g = 1, \ldots, G$$

$$B_g \sim N(0, \tau^2), \qquad g = 1, \ldots, G$$

Computational Approach

- Prefer Not To Use MCMC Methodology
  - Commercial software designed for non-statistician users
  - Convergence and non-exact-repeatability are issues
  - Scale-up problem: several hundred regression parameters
- Approximate Posterior Distributions
  - Discrete approximation of posterior of $\phi = (\sigma_A, \sigma_0, \sigma_B, \tau)$
    - $\{\pi_s , \phi_s \;\; s = 1, \ldots, S\}$ defines S-point discrete distribution, $\Sigma_s \pi_s = 1$
  - Normal approximation to $P(\theta \mid \text{data}, \phi)$
    - $\theta = (A_1, \ldots, A_G, B_0, \ldots, B_G, \alpha_{01}, \ldots, \alpha_{GK}, \beta_{01}, \ldots, \beta_{GK})$
    - Can use modified logistic regression likelihood for $P(\theta \mid \text{data}, \phi)$
    - Log $P(\theta \mid \text{data}, \phi_s)$ concave and easy to maximize, ~ $N(\theta_s, V_s)$
- Posterior of θ Approximately $N(\mu, V)$
  - $\mu = \Sigma_s \pi_s \theta_s \qquad V = \Sigma_s \pi_s [V_s + (\theta_s - \mu)(\theta_s - \mu)^t]$

Computing $\{\pi_s, \phi_s \; s = 1, \ldots, S\}$

- Density $P(\phi_s \,|\text{data}) \approx g(\phi_s) \propto P(\text{data} \mid \phi_s, \theta_s)\, \det(V_s)^{1/2}$
  - Maximized likelihood × (approx. factor for integrating out θ)
- Steepest Ascent (Numerical Derivs) to Maximize $g(\phi)$
- Construct Response Surface Design Around Maximum
  - 16 point central composite design at each of two radii
  - Fit 4-D quadratic response surface model (rsm) to log $g(\phi)$
  - Use fitted surface to rescale the 33-point design and refit rsm
- Adjust Final 33 Values of π_s so that Means and Variances of Discrete Dist. Match Continuous Estimates from RSM
  - Minimize K-L = $\Sigma_s\, g(\phi_s) \log(g(\phi_s)/\pi_s)$ subject to constraints
- Complete Estimation ≈10 Seconds if θ Has 200 Elements

Comparing MBLR to "Standard" LR

- Logistic Regression on Rare Events with Several Covariates and Interactions Can Often Fail to Get Reasonable Answers
  - Certain combinations of covariates seem to predict perfectly, leading to coefficient estimates that diverge to + or – infinity
  - Related terms: *Separation, Sparsity, Nonidentifiability*
  - Gelman et al (2008 *Annals of Applied Statistics*)
    - Suggests using a very weak prior distribution on the coefficients to get more reasonable answers and prevent divergence
    - Calls method *Regularized Logistic Regression (RLR)*
- Comparisons of MBLR to RLR
  - RLR: same model as MBLR except that $\sigma_A = \sigma_0 = \sigma_B = \tau = 5$
  - Typically, MBLR estimates of prior standard deviations < 1

Back to the Example

Issue	Treatment Events	Comparator Events	Chi Statistic	Shrunken Odds Ratio
Anuria	8	0	2.61	9.02
Dry mouth	308	65	11.42	4.31
Hyperkalaemia	218	162	1.34	1.15
Micturition urgency	13	3	2.18	3.10
Nocturia	19	7	1.95	2.17
Pollakiuria	193	34	9.55	4.98
Polydipsia	49	4	5.63	8.88
Polyuria	100	17	6.89	4.93
Thirst	543	66	18.38	8.16
Urine output increased	13	1	2.92	6.45
Patient counts:	Treatment = 3,110		Comparator = 2,642	

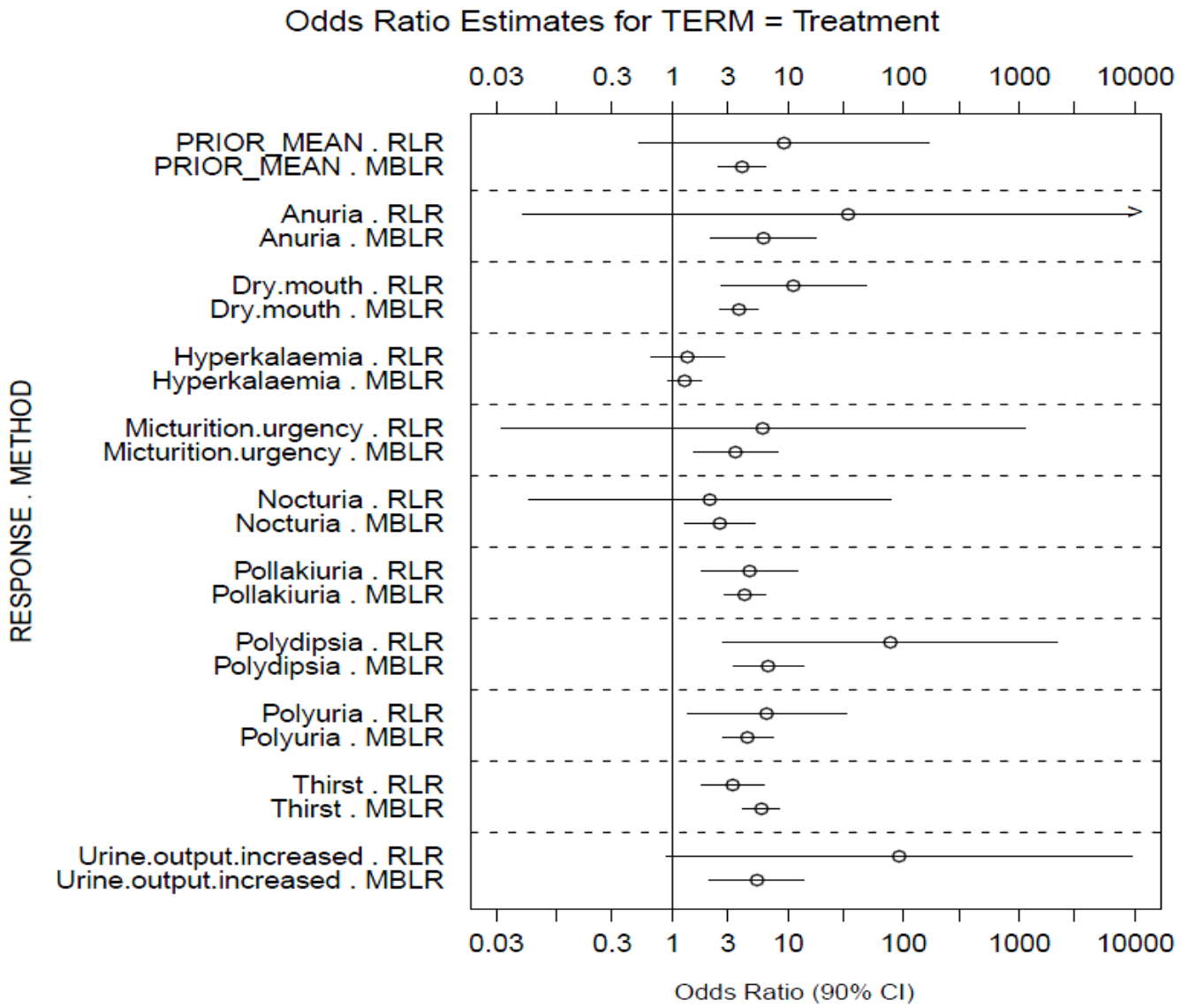
Statistics for 10 Issues Related to Dehydration/Renal Function for a Pool of 8 Trials

Covariate Patient Counts

	Treatment	Comparator
Gender=F	908	685
Gender=M	2202	1957
Race=Black	323	216
Race=White	2516	2228
Race=Other	271	198
Indication=1	796	347
Indication=2	2314	2295
Renal History=Y	190	191
Renal History=N	2920	2451
Age=50 or Under	382	348
Age=51 to 65	1089	902
Age=66 to 75	948	820
Age=Over 75	691	572
All Patients	3110	2642

Distribution of Patients by Covariates and Treatment Arm

Treatment Effects: RLR vs MBLR



ORACLE

Robustness to Post-Hoc Selection

- Simulation Study of Bayesian Estimation
  - Draw “true parameters” from the prior distributions 1000 times
  - Estimate main and interaction effects each time
    - Get both MBLR and RLR estimates
- Focus on Estimating the “Most Significant” Interaction
  - 80 Interactions (8 covariates x 10 response events)
  - For each simulation, select β_{gk} that has *largest* b_{gk}/se_{gk}
  - Compare accuracy of estimates and confidence limits

	SIM.COEF	SD.SIMC	BIAS	RMSE	Z.SCORE	CI.05	CI.95
MBLR	1.7651	0.6094	0.0005	0.2923	-0.0052	0.067	0.056
RLR	1.7445	0.5981	0.2184	0.4330	0.5794	0.008	0.135

Safety Analyses of Clinical Data

- Analysis of Drug Trial Adverse Event Data Is Challenging
  - Small event counts since trials are sized for efficacy
  - Multiple comparisons issues
- Combined Analyses of Multiple Trials Is Important
  - CDISC data standards make pooling data easier
  - This is a form of pooled-data meta-analysis
- Multivariate Bayesian Logistic Regression (MBLR)
  - Multivariate estimation of many possibly medically related AEs
  - Borrowing strength as a solution to the granularity problem
  - Search for vulnerable subgroups involves post-hoc selection
  - Bayesian shrinkage provides multiple-comparisons robustness