A Bayesian self-controlled method for drug safety surveillance in large-scale longitudinal data

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Shawn E. Simpson David Madigan

Introduction

- Ensuring drug safety begins with extensive pre-approval clinical trials
- This process continues *after* approval when drugs are in widespread use: **post-marketing surveillance**
- Drugs taken by more people, for longer periods of time, and in different ways than in pre-approval trials
- May identify adverse health outcomes associated with drug exposure that were not previously detected

1997

2004

Statistical Objectives

- Identify drug-condition pairs that may be linked
- Find drug interactions linked with conditions
- Estimate the strength of these associations
- Fundamental Difficulties
  - Large size: Millions of people, 10000's of conditions
  - High dimension: 10000's of drugs, millions of interactions

Current System: FDA AERS

- Current approach to surveillance is based on the FDA's **Adverse Event Reporting System (AERS)**
- Anyone can voluntarily submit a report on adverse events (AEs) that may be related to drug exposures
- Difficulties with AERS
  - Messy – spelling errors, etc.
  - Bias – underreporting, duplicate reports, media
  - Unreliable temporal information
- Multiple drugs and AEs may be listed on one report

		AE j: Yes	AE j: No
Drug i	Yes	a	b
	No	c	d
		Total:	N

- 15000 drugs × 16000 AEs = 240 million tables
- Most AEs do not occur with most drugs; small counts in a

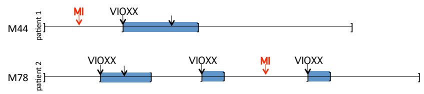
- FDA uses 2 × 2 summaries, applies Bayesian shrinkage methods to deal with variability due to small counts
- Limitations
  - No adjustment for confounding drugs
  - Ignores interactions
  - May not utilize temporal information

Longitudinal Healthcare Databases

- **Sentinel Initiative** - FDA plans to establish an *active* surveillance system using data from healthcare information holders
- **Observational Medical Outcomes Partnership (OMOP)** - Researching methods for analyzing healthcare databases to evaluate safety profiles of drugs on the market
- Advantages
  - Automated
  - Better temporal data
- Disadvantages
  - Little baseline data
  - No OTC information
- Potential analysis techniques: maxSPRT, cohort methods, case control, case-crossover, self-controlled case series …

Self-Controlled Case Series

- Method developed to estimate relative incidence of AEs to assess vaccine safety [Farrington, 1995]
- **One** drug, **one** adverse event (AE)



- Person i observed for τ_i days; (i,d) is their dth day
- y_{id} = # of events observed on (i,d)
- x_{id} = 1 if exposed to drug on (i,d), 0 otherwise

- Events arise according to a non-homogeneous Poisson process, exposure modulates the event rate
- Intensity on (i,d) = $e^{\phi_i + \beta x_{id}}$

$$y_{id} \mid x_{id} \sim \text{Poisson}(\, e^{\phi_i+\beta x_{id}} \,)$$

$$L_i = P(\, y_{i1}, \ldots, y_{i\tau_i} \mid x_{i1}, \ldots, x_{i\tau_i} \,) = P(\, \mathbf{y}_i \mid \mathbf{x}_i \,) = \prod_{d=1}^{\tau_i} P(\, y_{id} \mid x_{id} \,)$$

Condition to remove ϕ_i

- Could use ML to get estimates, but drug effect β is of interest and the ϕ_i's are *nuisance* parameters
- Condition on sufficient statistic $n_i = \Sigma\, y_{id}$

$$n_i \mid \mathbf{x_i} \sim \text{Poisson}\Big(\sum_d e^{\phi_i+\beta x_{id}} \Big)$$

- Conditional likelihood for i

$$L_i^c = P(\, \mathbf{y}_i \mid \mathbf{x_i}, n_i \,) = \frac{P(\, \mathbf{y}_i \mid \mathbf{x_i} \,)}{P(\, n_i \mid \mathbf{x_i} \,)} \propto \prod_{d=1}^{\tau_i} \left(\frac{e^{\beta x_{id}}}{\sum_{d'} e^{\beta x_{id'}}} \right)^{y_{id}}$$

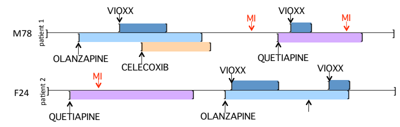
- Maximize $l^c = \sum \log L_i^c$ to get $\hat{\beta}_{CMLE} \longrightarrow$ consistent, asymptotically Normal [Cameron and Trivedi, 1998]

Data Reduction to Cases Only

- If i has no events ($\mathbf{y}_i = \mathbf{0}$) then $L_i^c = 1$, so we **only need cases** (i.e. $n_i \geq 1$) in the analysis
- SCCS does within-person comparison of event rate during exposure to event rate while unexposed ('self-controlled')

Multiple Drugs and Interactions

- We extend the model to one AE and **multiple** drugs



- Intensity on (i,d) = $e^{\phi_i + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_{id}} = e^{\phi_i + \beta_1 x_{id1} + \cdots + \beta_p x_{idp}}$

$$\mathbf{x}_{id} = (\, x_{id1}, \ldots, x_{idp} \,)^{\mathsf{T}} \qquad \boldsymbol{\beta} = (\, \beta_1, \ldots, \beta_p \,)^{\mathsf{T}}$$

- x_{idj} = 1 if exposed to drug j, 0 otherwise
- Intensity with drug interactions, time-varying covariates:

$$e^{\{\phi_i + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_{id} + \sum_{r \neq s} \gamma_{rs}\, x_{idr}\, x_{ids} + \boldsymbol{\alpha}^{\mathsf{T}} \mathbf{z}_{id}\}}$$

Bayesian Extension of SCCS

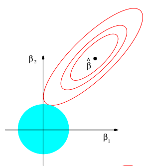
- Longitudinal databases have 10000's of potential drugs
- Intensity model: $e^{\text{(main effects)} + \text{(2-way interactions)}}$

  $\longrightarrow$ high dimensionality with *millions of predictors*
- Standard ML leads to overfitting; need to regularize
- **Our approach** – put a prior on $\boldsymbol{\beta}$ parameters to shrink the estimates toward zero, smooth out estimation, and reduce overfitting

1. Normal prior (*ridge regression*)

$$\beta_j \sim \text{N}(0, \sigma_\beta^2)$$

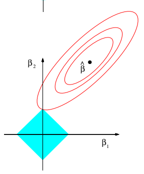
$$\text{max lik subject to} \quad \sum_{j=1}^{p} \beta_j^2 \leq s$$



2. Laplacian prior (*lasso*)

$$\beta_j \sim \text{Laplace}(0, 1/\lambda)$$

$$\text{max lik subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq s$$



- Convex optimization: Posterior modes via cyclic coordinate descent [Genkin et al, 2007]
- Handles millions of predictors in logistic case (BBR)

Results: OMOP Methods Evaluation

- Methods evaluation:
  - Chose 10 drugs, 10 conditions of interest
  - 9 drug-condition pairs with a *true association*
  - Pairs determined to have no link serve as *negative controls*
- Evaluation is based on **mean average precision (mAP) score**: measures the degree to which a method maximizes 'true positives' while minimizing 'false positives'

MSLR database (1.5M people)

Method	mAP score
22 PRR	0.2251486
22 OR	0.2280057
23 BCPNN	0.209197
22 EBGM	0.2173618
23 CHI-SQ	0.2144175
22 PRR05	0.2046662
22 ROR05	0.2046221
12 BCPNN05	0.1832317
12 EB05	0.1860902
SCCS (1 AE, 1 drug)	**0.2216072**
Bayesian SCCS, Normal prior, precision 1 (1 AE, 1 drug)	**0.26065568**
Bayesian Logistic Regression, Normal prior, precision 1 (1 AE, multiple drugs)	0.2665139
Case-Control	0.186743

Further Work

- Hierarchical modeling of drugs into drug classes

$$e^{\overbrace{\beta_1 x_1 + \ldots + \beta_j x_j}^{\text{drug class [1]}} + \ldots + \overbrace{\beta_k x_k + \ldots + \beta_p x_p}^{\text{drug class [D]}}}$$

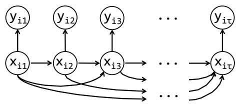
$$\text{N}(\mu_{[1]}, \sigma_{[1]}^2) \quad \cdots \quad \text{N}(\mu_{[D]}, \sigma_{[D]}^2)$$

- Hierarchical modeling of conditions into classes

$$\text{condition class (c)} \left[\begin{array}{l} y_1 \sim e^{\boxed{\beta_1} x_1 + \ldots + \boxed{\beta_p} x_p} \\ \vdots \\ y_m \sim e^{\boxed{\beta_1} x_1 + \ldots + \boxed{\beta_p} x_p} \end{array} \right.$$

$$\text{N}(\mu_1^{(c)}, \sigma_{(1)}^2) \cdots \text{N}(\mu_p^{(c)}, \sigma_{(c)}^2)$$

- Relax independence assumptions to allow dependence between events



- Allow events to influence future exposures

References

- Cameron and Trivedi (1998) Regression Analysis of Count Data. Cambridge University Press.
- Farrington (1995) "Relative incidence estimation from case series for vaccine safety evaluation," *Biometrics*, Vol. 51, No. 1, pg. 228-235.
- Genkin et al. (2007) "Large-scale Bayesian logistic regression for text categorization," *Technometrics*, Vol. 49, No. 3, pg. 291-304.