

Bayesian Logistic Regression for Medical Claims Data using CPU and GPU

Ivan Zorych, Patrick Ryan, David Madigan
on behalf of the OMOP Research Team

Miami, JSM 2011
August 2, 2011

Outline

- OMOP project (omop.fnih.org)
- Typical issues
- Bayesian logistic regression approach to observational data, numerical experiments
- CPU and GPU implementations; benchmarks
- Conclusions and future work

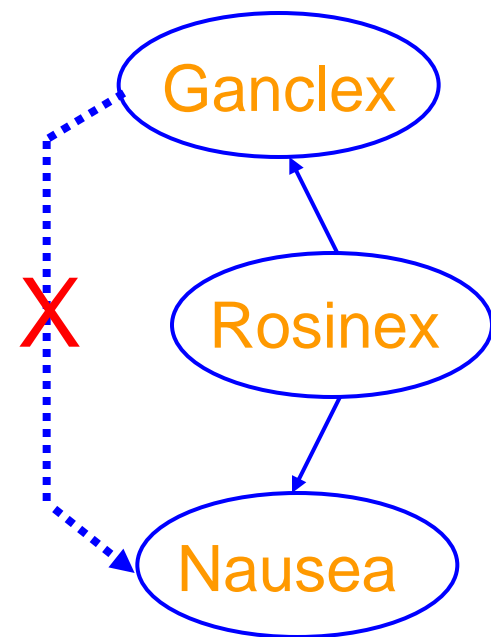
OMOP project

- One of the goals of the Observational Medical Outcomes Partnership is to define methods that can assess the feasibility and utility of using observational data to identify and evaluate associations between drugs and health-related conditions.
- OMOP established an open-source library of 10 Health Outcomes of Interest (HOI) definitions for use in observational studies. These ten HOIs are a subset of all conditions that are of importance due to their historical associations with drug toxicities, their medical significance, and/or public health implications. There is little consensus for best practice in defining HOIs in observational databases, as observational studies for the same outcome often use different definitions.
- OMOP HOI: Angioedema, Aplastic Anemia, Acute Liver Injury, Bleeding, GI Ulcer Hospitalization, Hip Fracture, Hospitalization, Acute Myocardial Infarction, Mortality after Myocardial Infarction, Acute Renal Failure
- The goal of this surveillance analysis is to monitor the relationship between specific drugs and a specific outcome of interest

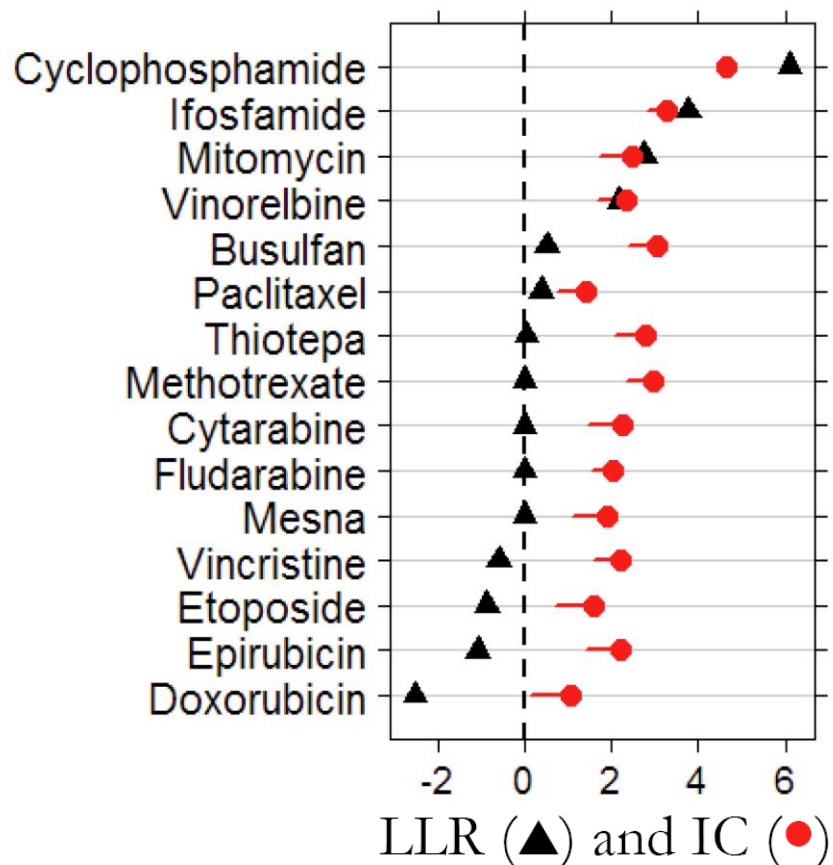
Innocent bystander problem

- Contingency table analysis ignores effects of drug-drug association on drug-AE association

	Rosinex		No Rosinex		Total	
	Nausea	No Nausea	Nausea	No Nausea	Nausea	No Nausea
Ganclex	81	9	1	9	82	18
No Ganclex	9	1	90	810	99	811
OR	1		1		37.3	



Confounding, real AERS data*



- ADR: hemorrhagic cystitis, diffuse inflammation of the bladder leading to dysuria, hematuria, and hemorrhage;

ADR most often seen in female cancer patients as a complication of therapy;

Drugs: anticancer drugs and mesna;

Mesna is an adjuvant used in cancer chemotherapy involving cyclophosphamide and ifosfamide.

*Caster, Noren, Bate, Madigan, 2010

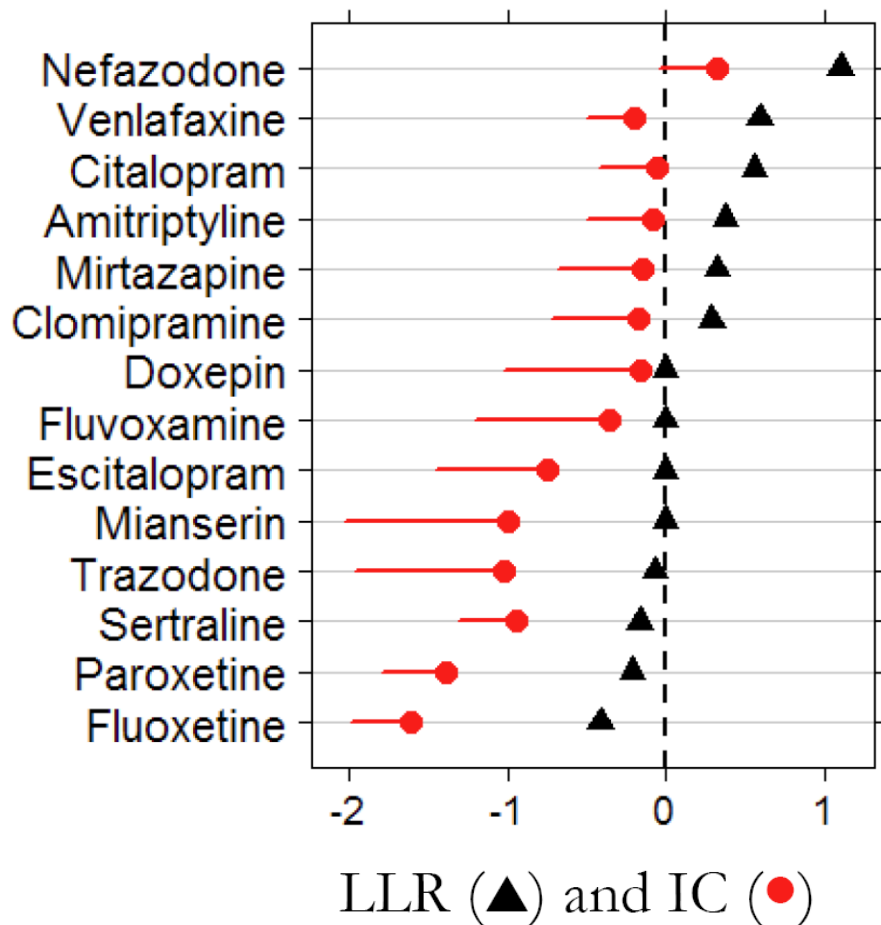
Masking

- Typical DP measures are based on

$$\frac{\Pr(AE|Drug)}{\Pr(AE)}$$

- Masking: effect when the background rate of the ADR, $\Pr(AE)$, is distorted due to massive reporting with other drug(s);
- Example: Rhabdomyolysis and Cerivastatin (Baycol, Lipobay)
Cerivastatin is a synthetic member of the class of statins;
- Cerivastatin was voluntarily withdrawn from the market worldwide in 2001 due to reports of fatal rhabdomyolysis.

Masking, real AERS data*



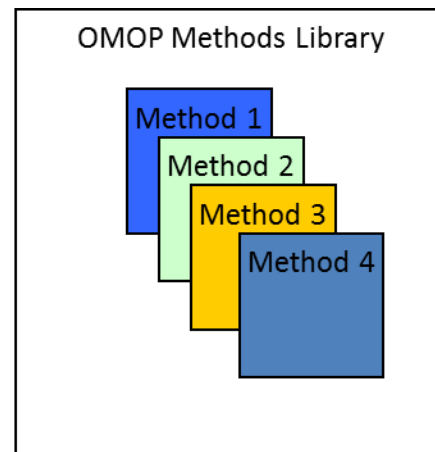
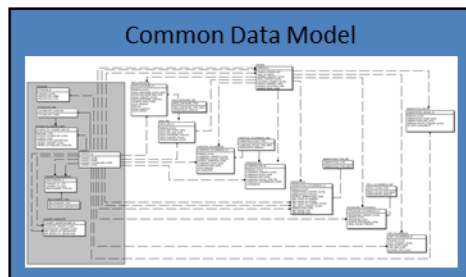
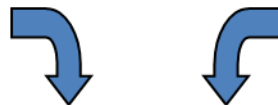
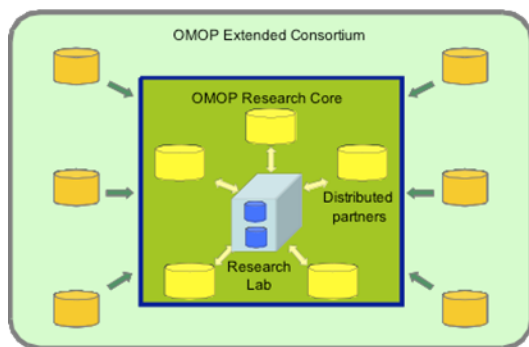
- ADR: Rhabdomyolysis, rapid breakdown of skeletal muscle;
- Drugs: a set of anti-depressant drugs;

*Caster, Noren, Bate, Madigan, 2010

Regression and SRS data

- Drug safety data sets are sparse
- Typical AERS report contains just a few drugs and a few adverse events
- Dependent variable in the regression:
Y=1 if AE is present, 0 otherwise;
- Number of independent variables = number of drugs + sex/age/year info

OMOP Experiment Overview



Health Outcomes of Interest

- Angioedema
- Aplastic Anemia
- Acute Liver Injury
- Bleeding
- GI Ulcer Hospitalization
- Hip Fracture
- Hospitalization
- Myocardial Infarction
- Mortality after MI
- Renal Failure

Drugs

- ACE Inhibitors
- Amphotericin B
- Antibiotics
- Antiepileptics
- Benzodiazepines
- Beta blockers
- Bisphosphonates
- Tricyclic antidepressants
- Typical antipsychotics
- Warfarin

Non-specified conditions

- All outcomes in condition terminology
- ‘Labeled events’ as reference
 - Warning
 - Precautions
 - Adverse Reactions
 - Postmarketing Experience

'Ground truth' assumed for Monitoring Health Outcomes of Interest

Outcome	ACE Inhibitors	Amphotericin B	Antibiotics: erythromycins, sulfonamides, tetracyclines	Antiepileptics: carbamazepine, phenytoin	Benzodiazepines	Beta blockers	Bisphosphonates: alendronate	Tricyclic antidepressants	Typical antipsychotics	Warfarin
Angioedema	True positive' risk	Negative control'		Negative control'	Negative control'	Negative control'				Negative control'
Aplastic Anemia	Negative control'	Negative control'	Negative control'	True positive' risk	Negative control'	Negative control'	Negative control'	Negative control'		Negative control'
Acute Liver Injury		Negative control'	True positive' risk		Negative control'					
Bleeding			Negative control'				Negative control'			True positive' risk
Hip Fracture	Negative control'	Negative control'			True positive' risk	Negative control'				Negative control'
Hospitalization	True positive' benefit									
Myocardial Infarction			Negative control'		Negative control'		Negative control'	True positive' risk	True positive' risk	
Mortality after MI		Negative control'		Negative control'		True positive' benefit				Negative control'
Renal Failure		True positive' risk	Negative control'	Negative control'	Negative control'	Negative control'	Negative control'	Negative control'	Negative control'	Negative control'
GI Ulcer Hospitalization	Negative control'			Negative control'		Negative control'	True positive' risk		Negative control'	

Legend	Total
True positive' benefit	2
True positive' risk	9
Negative control'	44

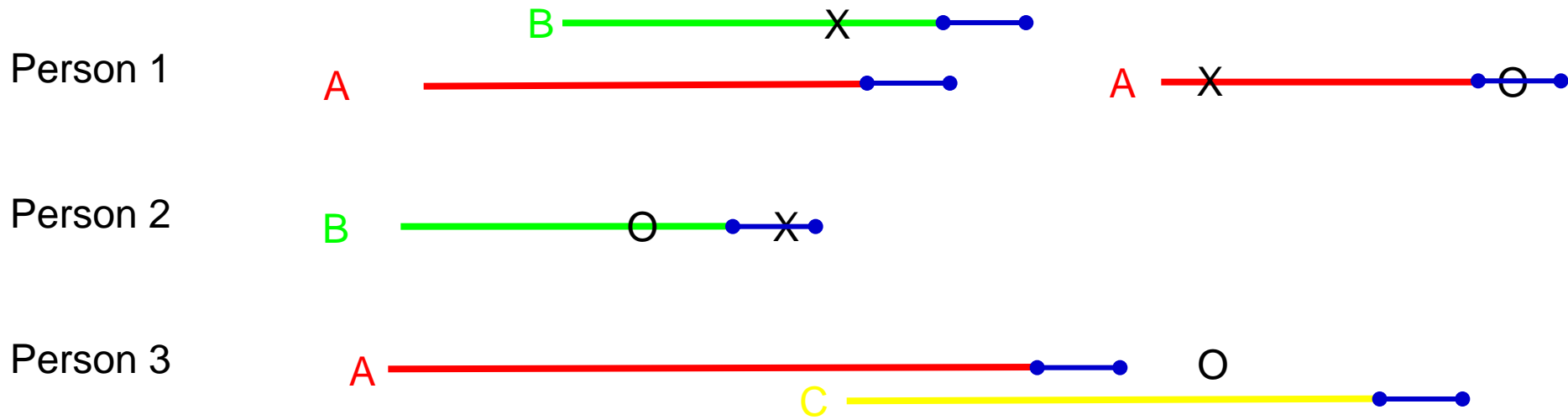
Longitudinal observational data

- Health claims records, electronic medical records
- Information on drug prescriptions, doctor visits, hospitalization

Regression and observational data

- It is relatively easy to apply regression to AERS/SRS data.
- Situation with temporal data is more complicated. We need to create predictors for the regression analysis.

Logistic regression for observational data



prevalent X:

Y	A	B	C
1	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1

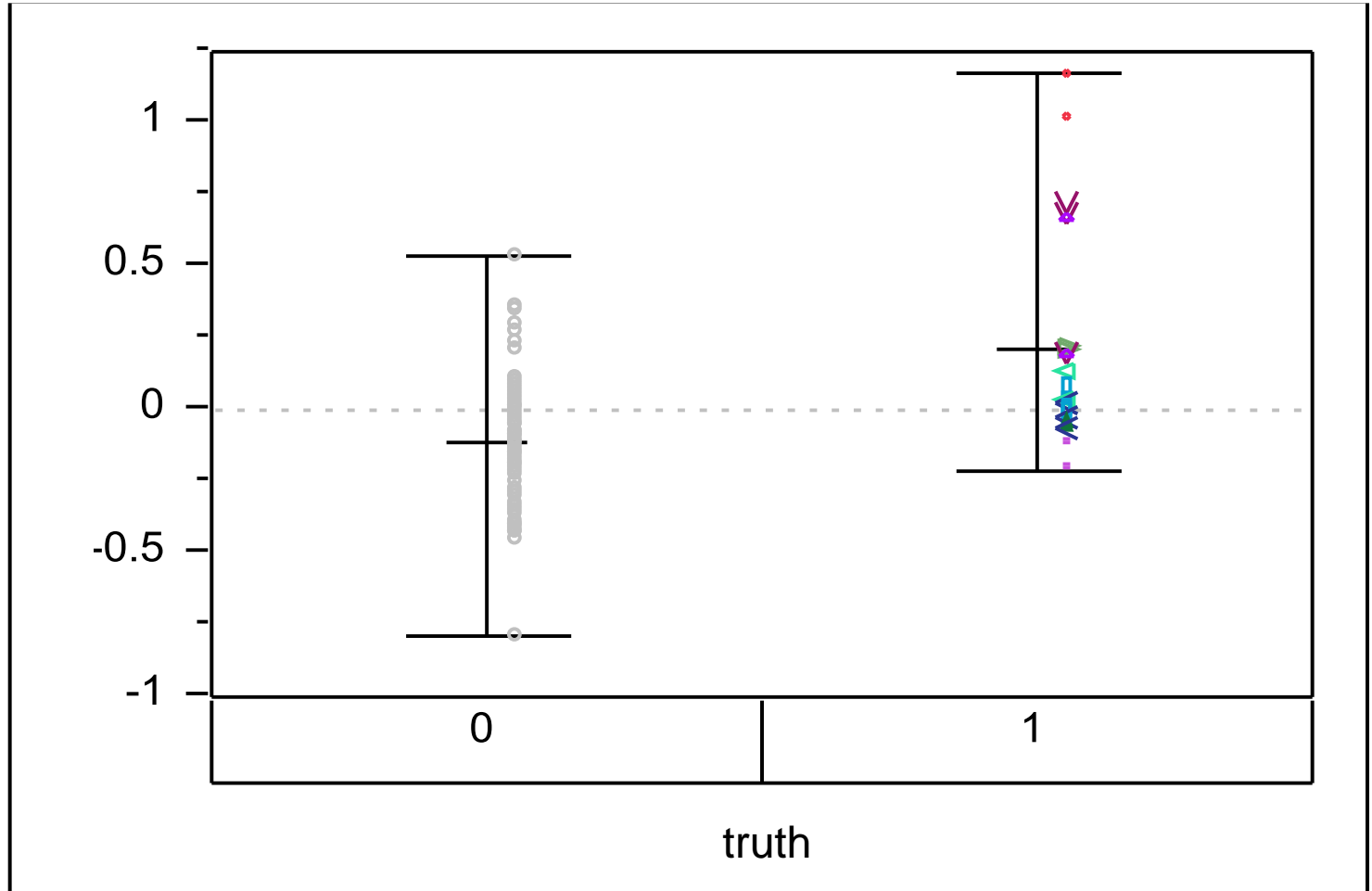
incident X:

Y	A	B	C
1	1	1	0
1	0	1	0
0	0	0	1

Bayesian Logistic Regression

- Two shrinkage methods
 - Ridge regression - Gaussian prior
$$p(\beta_j | \tau) \sim N(0, \tau)$$
 - Lasso regression - Laplace prior
$$p(\beta_j | \lambda) = \lambda / 2 \exp\{-\lambda |\beta_j|\}$$
- Choosing hyperparameter λ
 - Decide how much to shrink
 - Cross-validation: choose prior to fit left-out data

Real Data: MSLR



Measures of Performance:

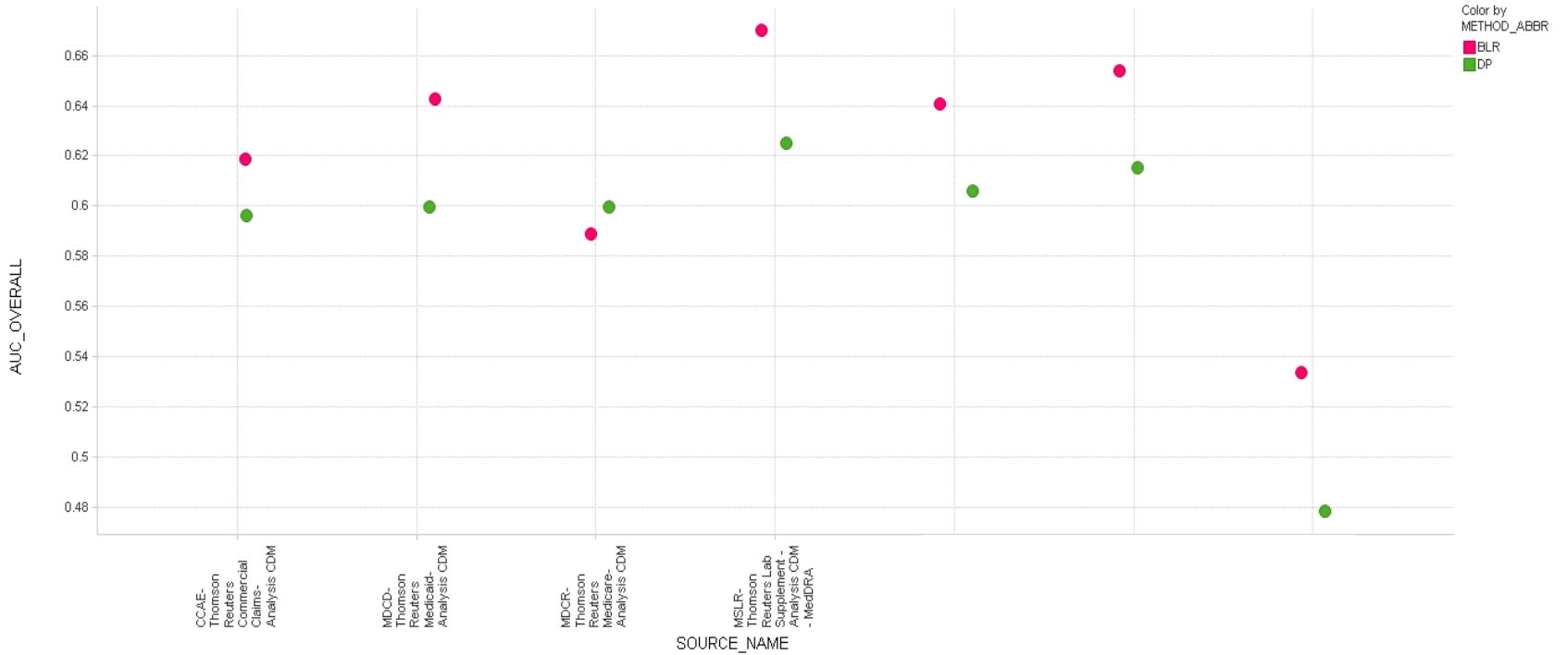
- Mean Average Precision:

Drug	Condition	Original Values		Sorted Values		$P^{(k)}$
		x_i	y_i	$x_{(i)}$	$y_{(i)}$	
D1	C1	5	1	9	1	1/1=1
	C2	0	1	8	1	2/2=1
	C3	9	1	7	0	
D2	C1	8	1	5	1	3/4=0.75
	C2	4	1	4	1	4/5=0.8
	C3	3	0	3	0	
D3	C1	1	0	2	0	
	C2	2	0	1	0	
	C3	7	0	0	1	5/9=0.55
Total Score						(1+1+0.75+0.8+0.55)/5 =0.82

- Area under ROC curve, etc.

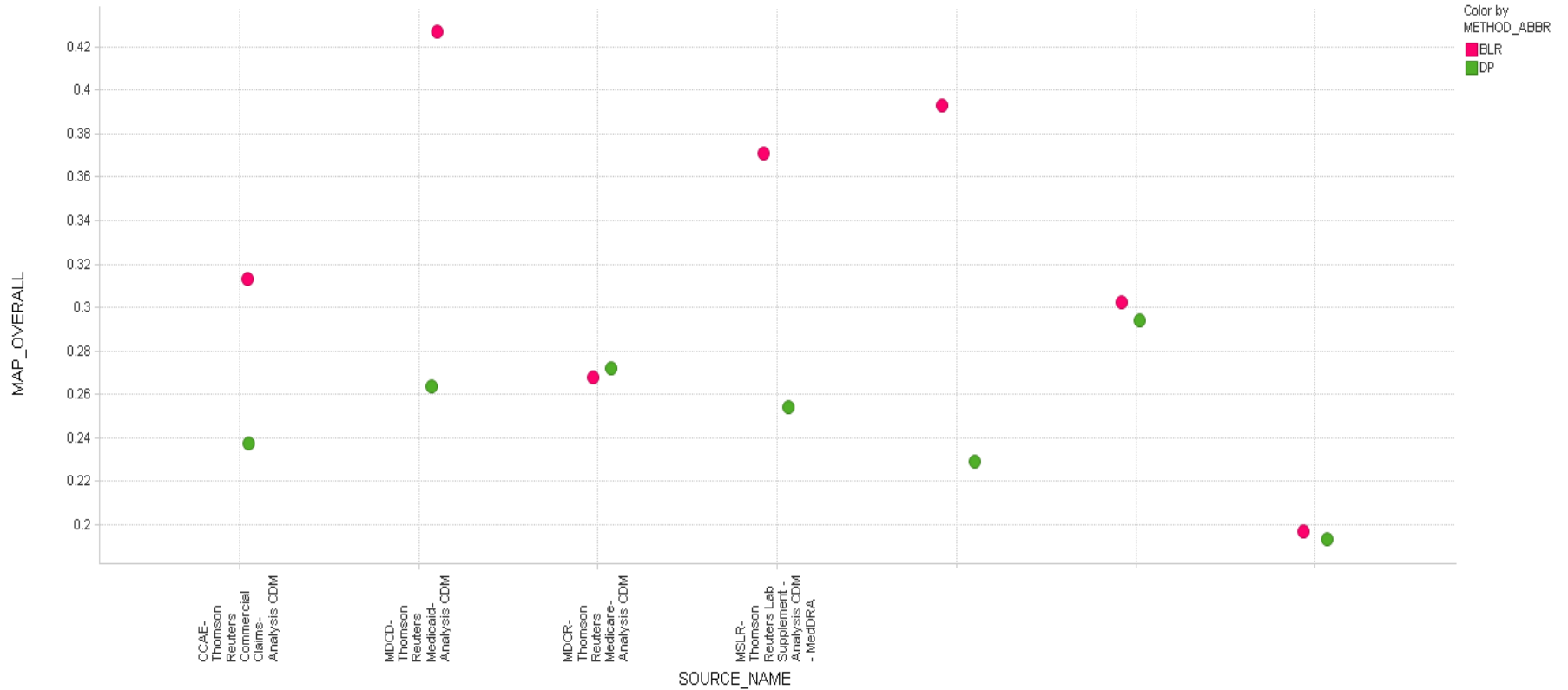
AUC for multiple databases

HOI performance by source



MAP for multiple databases

HOI performance by source



Cyclic Coordinate Descent

$$L(\beta) = p(\beta | D) \propto \prod_{i=1}^n \frac{1}{1 + \exp(-\beta^T x_i y_i)} p(\beta)$$

$$-l(\beta) = \prod_{i=1}^n \ln(1 + \exp(-\beta^T x_i y_i)) + \sum_{j=1}^d (\ln \sqrt{\tau} + 0.5 * \ln(2\pi) + \frac{\beta_j^2}{2\tau_j})$$

- Negated log-posterior is convex – many optimization methods are applicable
- Cyclic coordinate descent begins by setting all variables to initial value
- Then it sets the first variable to a values that makes the objective function smaller (relaxation)
- And so on...
- Multiple passes are made over all variables until some convergence criterion is met

One step from optimization point of view

Find β_j^{new} that makes negated log-likelihood smaller

$$g(z) = \sum_{i=1}^n f(r_i + (z - \beta_j)x_{ij}y_i) + \frac{z^2}{2\tau_j}$$

where $r_i = \beta_j^T x_i y_i$

$$f(w) = \ln(1 + \exp(-w))$$

$$g(z) \approx g(\beta_j) + g'(\beta_j)(z - \beta_j) + \frac{1}{2}G(\beta_j)(z - \beta_j)^2$$

$$\Delta v_j = - \frac{\sum_{i=1}^n \frac{-x_{ij}y_i}{1 + \exp(r_i)} + \frac{\beta_j}{\tau_j}}{\left(\sum_{i=1}^n x_{ij}^2 F(r_i, \Delta_j | x_{ij}) \right) + \frac{1}{\tau_j}}$$

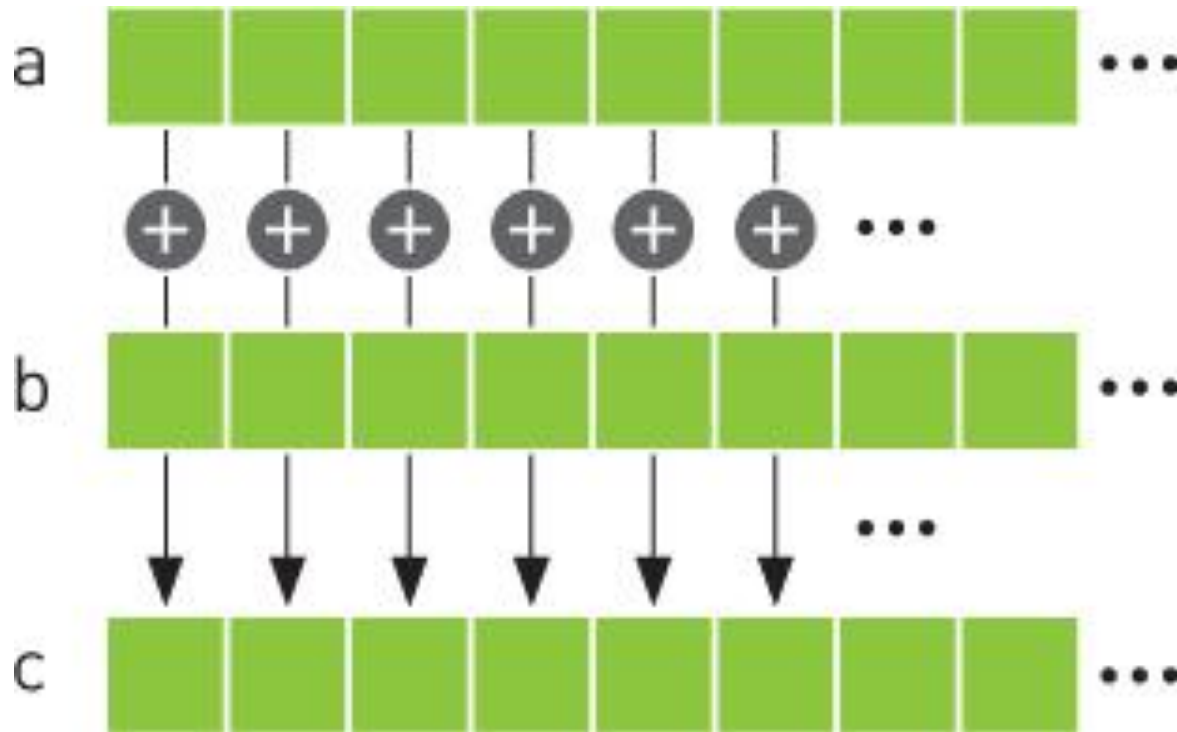
Algorithm

Initialize $\beta_j = 0, \Delta_j = 1$ for $j=1,2,\dots,d$; and $r_i = 0$ for $i=1,2,\dots,n$

while (eps > EPS) do; // i.e. until convergence

```
  for j=1,2,...,d do;
     $\Delta\beta_j = \min(\max(\Delta v_j, -\Delta_j), \Delta_j)$ 
    for i = 1,2,...,n do;
       $\Delta r_i = \Delta\beta_j x_{ij} y_i, r_i = r_i + \Delta r_i$ 
    end;
     $\Delta_j = \max(2|\Delta\beta_j|, \Delta_j / 2)$ 
  end;
end;
```

Example of parallelism: vector additions or multiplication*



* J. Sanders and E. Kandrot, 2010

CUDA, NVIDIA GPU



- CUDA: Compute Unified Device Architecture by NVIDIA
- GPU: Graphical processing unit
- CUDA program: comprises phases that are executed either on CPU (host) or GPU (device).
- Parts of the code with large amount of data parallelism are executed on the device using CUDA kernels.
- Data parallelism: step in each dimension requires similar calculations; observations from different records participate in a 'symmetric' way.
- Kernel functions generate a large number of threads to utilize parallelism in data.

Structure of the GPU/CPU code

Kernel_1: Calculates tentative step Δv_j

$$\Delta v_j = - \frac{\sum_{i=1}^n \frac{-x_{ij}y_i}{1 + \exp(r_i)} + \frac{\beta_j}{\tau_j}}{\left(\sum_{i=1}^n x_{ij}^2 F(r_i, \Delta_j | x_{ij}) \right) + \frac{1}{\tau_j}}$$

Kernel_2: Updates r_i and Δr_i

$$\Delta r_i = \Delta \beta_j x_{ij} y_i, \quad r_i = r_i + \Delta r_i$$

Kernel_3: Check for convergence

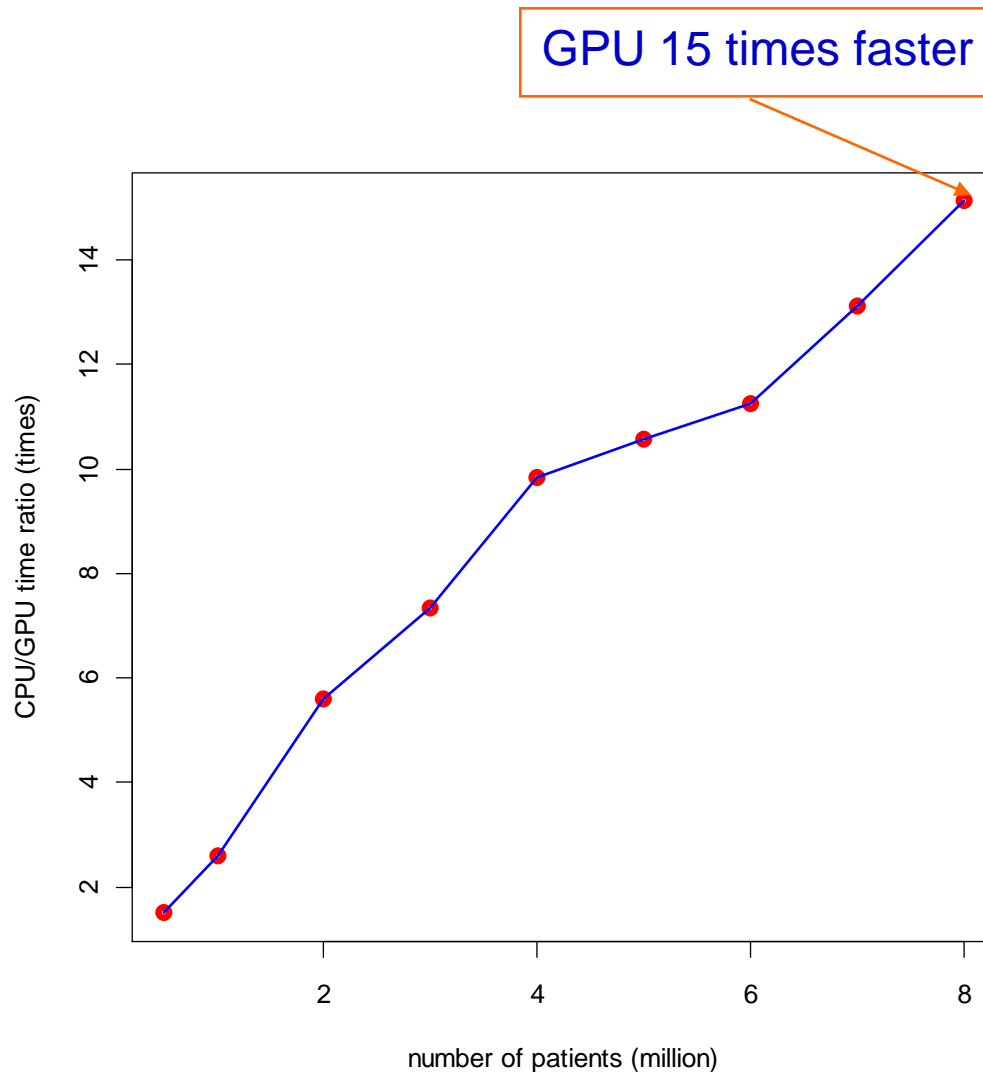
$$\left(\sum_{i=1}^n |\Delta r_i| \right), \left(1 + \sum_{i=1}^n |r_i| \right)$$

Comparisons: GPU vs. CPU

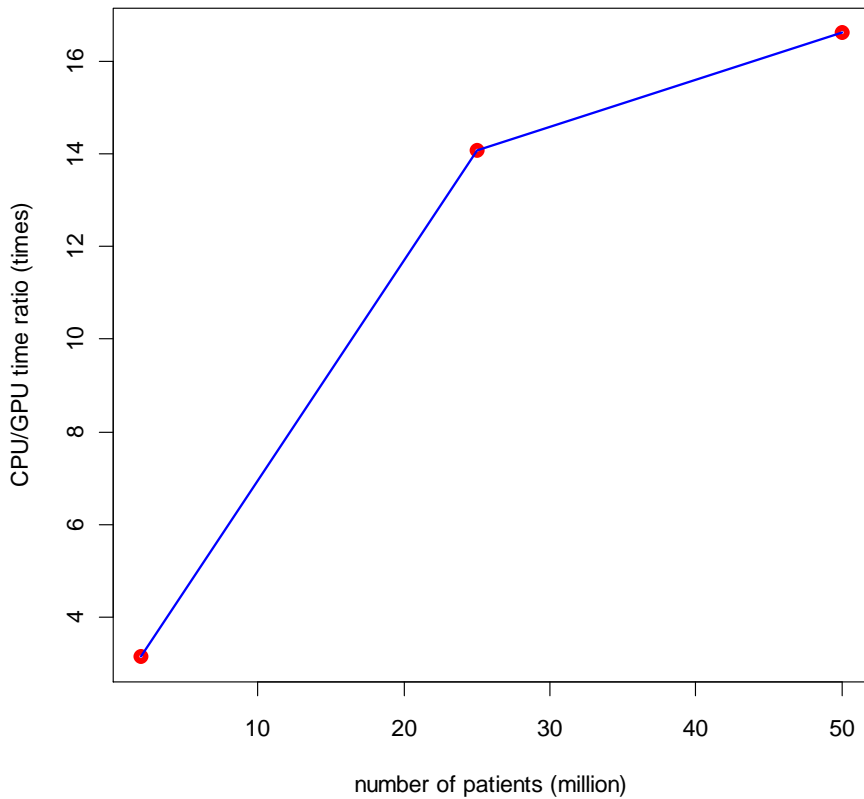
X-axis: number of observations in the regression model (in millions)

Y-axis: ratio of CPU time to GPU time

Number of variables - approximately 1300



Another data set:



Data	GPU	CPU	ratio
2 mln. 1279 var. 7231579 en.	7 sec.	22 sec.	3.14
25 mln. 1405 var. 90404159 en.	25 sec.	352 sec.	14.08
50 mln. 1430 var. 180791616 en.	44 sec.	730 sec.	16.59

Hardware information

NVIDIA GTX 480:

- CUDA Cores 480; Graphics Clock (MHz) 700 MHz
- Processor Clock (MHz) 1401 MHz
- Memory Clock (MHz) 1848 Standard Memory Config 1536 MB GDDR5 Memory Bandwidth (GB/sec) 177.4

CPU info: Core i7 980X @ 3.33 GHz

PassMark - CPU Mark
High End CPUs - Updated 30th of July 2011



Across different CPUs

Data	GPU GTX 480	CPU Core i7 980X @3.33	Amazon Cloud	A.Cloud to GPU ratio	Consumer Corei3 U330 @1.2	Corei3 to GPU ratio
2 mln. 1279 var. 7231579 en.	7 sec.	22 sec.	39 sec.	5.6	52 sec.	7.4
25 mln 1405 var. 90404159 en.	25 sec.	352 sec.	597 sec.	23.8	830 sec (est.)	33 (est.)
50 mln. 1430 var. 180791616 en.	44 sec.	730 sec.	1245 sec.	28.3	1722 sec. (est)	39 (est)

Conclusions and future work

- Advantages over low-dimensional tables
 - Correct confounding and mask effect
 - Analyze multiple drugs/vaccines simultaneously
- Limitations
 - Build separate model for each AE
 - Ignore dependencies between AEs
 - Fail to adjust for unmeasured/unrecorded factors
 - health status, unreported drugs, etc.

Conclusions and future work

- Protopathic bias should be addressed
- Incorporation of drug hierarchy
- Similar/related conditions should be modeled jointly
- Extending to several GPUs (current version uses one GPU)

Acknowledgments

Mark Khayter, Christian Reich, Marc Suchard