Journal: Drug Safety

Title: A method to combine signals from spontaneous reporting systems and observational healthcare data to detect adverse drug reactions

Authors:

Ying Li[1], Patrick B Ryan[1,2,3], Ying Wei[4*] and Carol Friedman[1*]

1. Department of Biomedical Informatics, Columbia University, New York, New York, 10032
2. Janssen Research and Development, 1125 Trenton Harbourton Rd, Titusville NJ 08560
3. Observational Health Data Sciences and Informatics (OHDSI), New York, NY 10032
4. Department of Biostatistics, Columbia University, New York, New York, 10032

* The last two authors contributed equally

Compliance with ethical standards

Author 1 Ying Li has no conflicts of interest that are directly relevant to the content of this study.

Author 2 Patrick B Ryan is an employee of Janssen Research and Development, and a shareholder of Johnson & Johnson

Author 3 Ying Wei has no conflicts of interest that are directly relevant to the content of this study.

Author 4 Carol Friedman is a consultant for Health Fidelity.

Correspondence
Mailing address:
Ying Li
Department of Biomedical Informatics
Columbia University Medical Center
622 W. 168th Street, Presbyterian Building 20th Floor
New York, NY 10032

Email:
yl2565@columbia.edu

Telephone:
347-277-5572

Fax:
212-305-3302

Electronic Supplementary Material

Box 1 Formulas for two-step LASSO regression

In the first step, standard LASSO is applied to select a set of potential confounders associated with the ADR, denoted by S_1; In the second step, LASSO type regression is used again to select medical conditions that are highly associated with the drug use, and denote them as S_2. In both steps, we used 5-fold cross-validation to select LASSO penalties. Finally, we estimate the conditional association between the ADR and drug adjusting for all the confounders in ($S_1 \cup S_2$). We then use one-sided p-values of the adjusted log odds ratios (log ORs) in the last step as the signal scores.

$$\text{logit}(\text{prob}(\text{ADR} = 1)) = \alpha^{(1)} + \beta^{(1)}\text{Rx} + \sum_{i \in M} \gamma_i^{(1)} C_i \quad (1)$$

$$\text{E}[\omega \text{Rx}] = \alpha^{(2)} + \sum_{i \in M} \omega \gamma_i^{(2)} C_i \quad \text{where}$$

$$w = \sqrt{\text{prob}(\text{ADR} = 1|\text{Rx}, C_i) * (1 - \text{prob}(\text{ADR} = 1|\text{Rx}, C_i))} \quad \text{and } i \in S_1 (2)$$

$$\text{logit}\big(\text{prob}(\text{ADR} = 1)\big) = \alpha^{(3)} + \beta^{(2)}\text{Rx} + \sum_{i \in S_1 \cup S_2} \gamma_i^{(3)} C_i \, (3)$$

Table S1. False positive signals in the EHR

Medication	ADR	a	b	c	d	Pvalue1	Pvalue2
hyoscyamine	GI Bleed	24	976	27614	2094848	0.00	0.00
rosiglitazone	GI Bleed	213	30123	27425	2065701	0.00	0.01
hyoscyamine	ALI	19	981	39029	2083433	0.00	0.01
metaxalone	AMI	81	2214	18371	2102796	0.09	0.04

a: number of patients were exposed to the medication and developed the ADR
b: number of patients were exposed to the medication and not developed the ADR
c: number of patients were not exposed to the medication and developed the ADR
d: number of patients were not exposed to the medication and not developed the ADR
pvalue1: unadjusted one-sided p-value
pvalue2: adjusted one-sided p-value

Table S2. False negative signals in the NYP/CUMC EHR

Medication	ADR	a	b	c	d	Pvalue1	Pvalue2
amlodipine	AMI	963	17010	4854	191352	0.00	1.00
darbepoetin alfa	AMI	134	2160	5683	206202	0.00	1.00
dipyridamole	AMI	102	1510	5715	206852	0.00	0.99
nifedipine	AMI	209	3490	5608	204872	0.00	0.17
Acyclovir	ARF	266	2631	14624	197082	0.00	1.00
allopurinol	ARF	725	2079	14165	197634	0.00	0.19
Captopril	ARF	400	1739	14490	197974	0.00	1.00
cyclosporine	ARF	352	907	14538	198806	0.00	1.00
enalaprilat	ARF	228	1191	14662	198522	0.00	0.85
Ibuprofen	ARF	756	32402	14134	167311	1.00	1.00
Ketorolac	ARF	164	5386	14726	194327	1.00	1.00
Lisinopril	ARF	2815	16984	12075	182729	0.00	0.98
meloxicam	ARF	103	1977	14787	197736	1.00	0.90
Naproxen	ARF	256	6767	14634	192946	1.00	1.00
allopurinol	ALI	164	2926	5935	203323	0.00	1.00
ciprofloxacin	ALI	222	4892	5877	201357	0.00	1.00
cyclosporine	ALI	178	1117	5921	205132	0.00	0.94
Diltiazem	ALI	224	5814	5875	200435	0.00	1.00
fluconazole	ALI	330	4845	5769	201404	0.00	1.00
Ibuprofen	ALI	545	30766	5554	175483	1.00	1.00
Ketorolac	ALI	125	5120	5974	201129	0.98	1.00
lamivudine	ALI	126	1204	5973	205045	0.00	1.00
levofloxacin	ALI	591	11486	5508	194763	0.00	1.00
Lisinopril	ALI	738	19117	5361	187132	0.00	1.00
Naproxen	ALI	134	5921	5965	200328	1.00	1.00
nifedipine	ALI	150	3742	5949	202507	0.00	0.98
Ramipril	ALI	139	3562	5960	202687	0.00	0.99
citalopram	GI BLEED	246	4250	6437	202220	0.00	1.00
clopidogrel	GI BLEED	542	12940	6141	193530	0.00	1.00
escitalopram	GI BLEED	188	3616	6495	202854	0.00	0.72
Ibuprofen	GI BLEED	492	27177	6191	179293	1.00	1.00
Ketorolac	GI BLEED	105	4813	6578	201657	1.00	1.00
Naproxen	GI BLEED	168	5052	6515	201418	0.36	1.00
potassium chloride	GI BLEED	154	2778	6529	203692	0.00	1.00

Sertraline	GI BLEED	256	4850	6427	201620	0.00	0.85

a: number of patients were exposed to the medication and developed the ADR
b: number of patients were exposed to the medication and not developed the ADR
c: number of patients were not exposed to the medication and developed the ADR
d: number of patients were not exposed to the medication and not developed the ADR
pvalue1: unadjusted one-sided p-value
pvalue2: adjusted one-sided p-value

Table S3 Reference set and the AUC performance for the confounding adjustment method and Gamma Poisson Shrinkage (GPS) method on the basis of FAERS from 2004 to 2010

	Reference set		AUC performance	
	Positive	Negative	Confounding adjustment method	GPS (EB05)
Acute renal failure	23	52	0.90	0.76
Acute liver injury	77	33	0.72	0.87
Acute myocardial infarction	34	59	0.72	0.70
Upper GI bleeding	24	63	0.81	0.79

Figure S1 Comparison between single LASSO and double LASSO on the basis of the EHR for acute renal failure

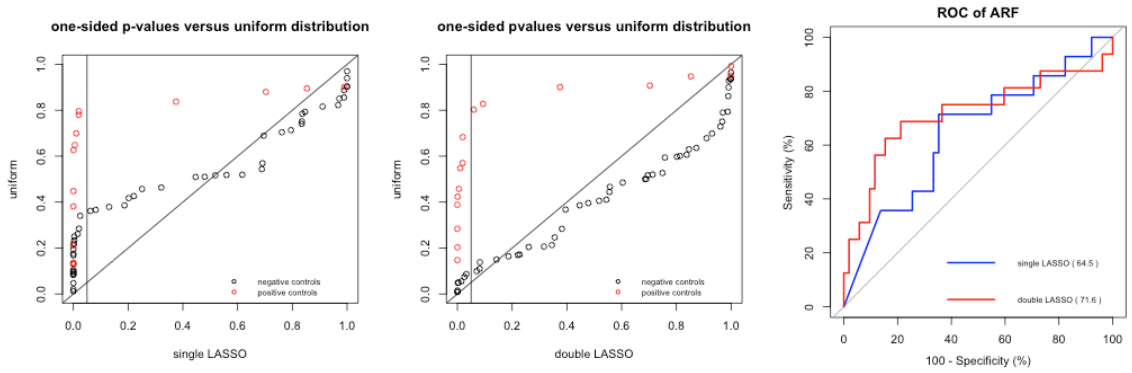

Figure S2 Comparison between single LASSO and double LASSO on the basis of the FAERS for acute renal failure