Machine Learning for Adverse Drug Event Detection

David Page
Dept. of Biostatistics & Medical Informatics
University of Wisconsin-Madison

Bringing Variety of ML Approaches to Bear on Adverse Drug Events

- Regularized Regression
- Random Forests
- Support Vector Machines
- Graphical Model Learning (Bayes nets, Markov nets, dynamic Bayes nets, continuous-time models)
- Deep Learning (deep neural nets, restricted Boltzman machines)
- Relational Learning

Data: EHR or Claims Data in a Relational Data Warehouse

Demographics

Patient ID	Gender	Birthdate
P1	M	3/22/1963

Diagnoses

Patient ID	Date	Physician	Symptoms	Diagnosis
P1	1/1/2001	Smith	palpitations	hypoglycemic
P1	2/1/2001	Jones	fever, aches	influenza

Lab Results

Patient ID	Date	Lab Test	Result
P1	1/1/2001	blood glucose	42
P1	1/9/2001	blood glucose	45

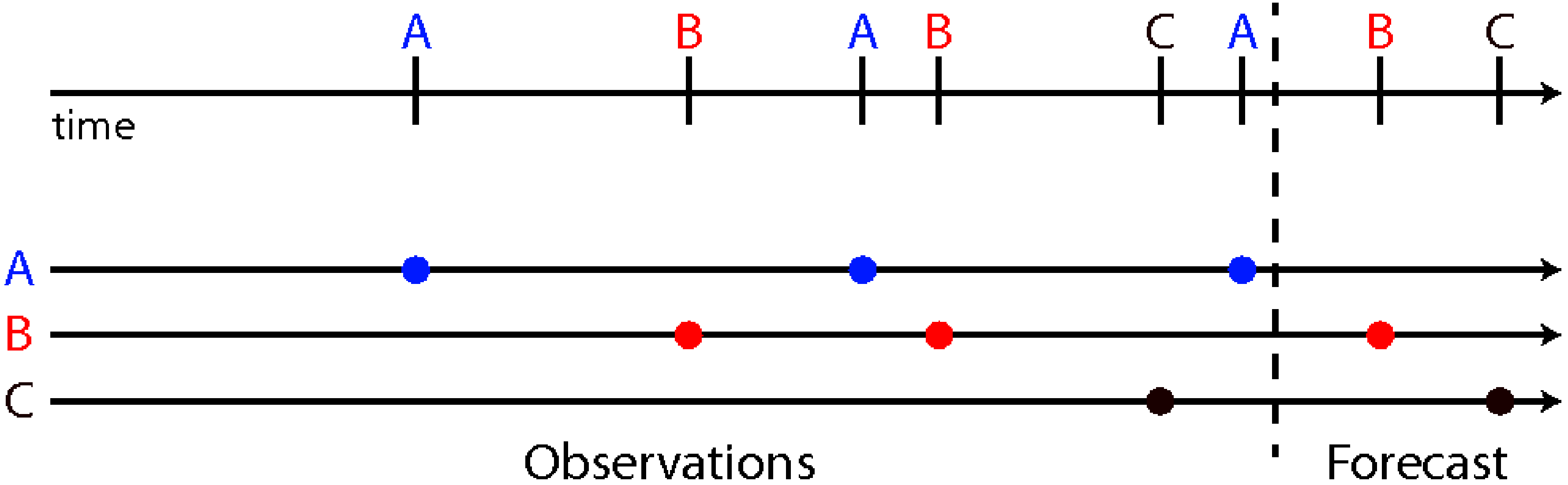
Vitals

Patient ID	Date	Observation	Result
P1	1/1/2001	Height	5'11
P2	1/9/2001	BMI	34.5

Medications

Patient ID	Date Prescribed	Date Filled	Physician	Medication	Dose	Duration
P1	5/17/1998	5/18/1998	Jones	Prilosec	10mg	3 months

Alternative View of Patient Data: Irregularly-Sampled Time Series



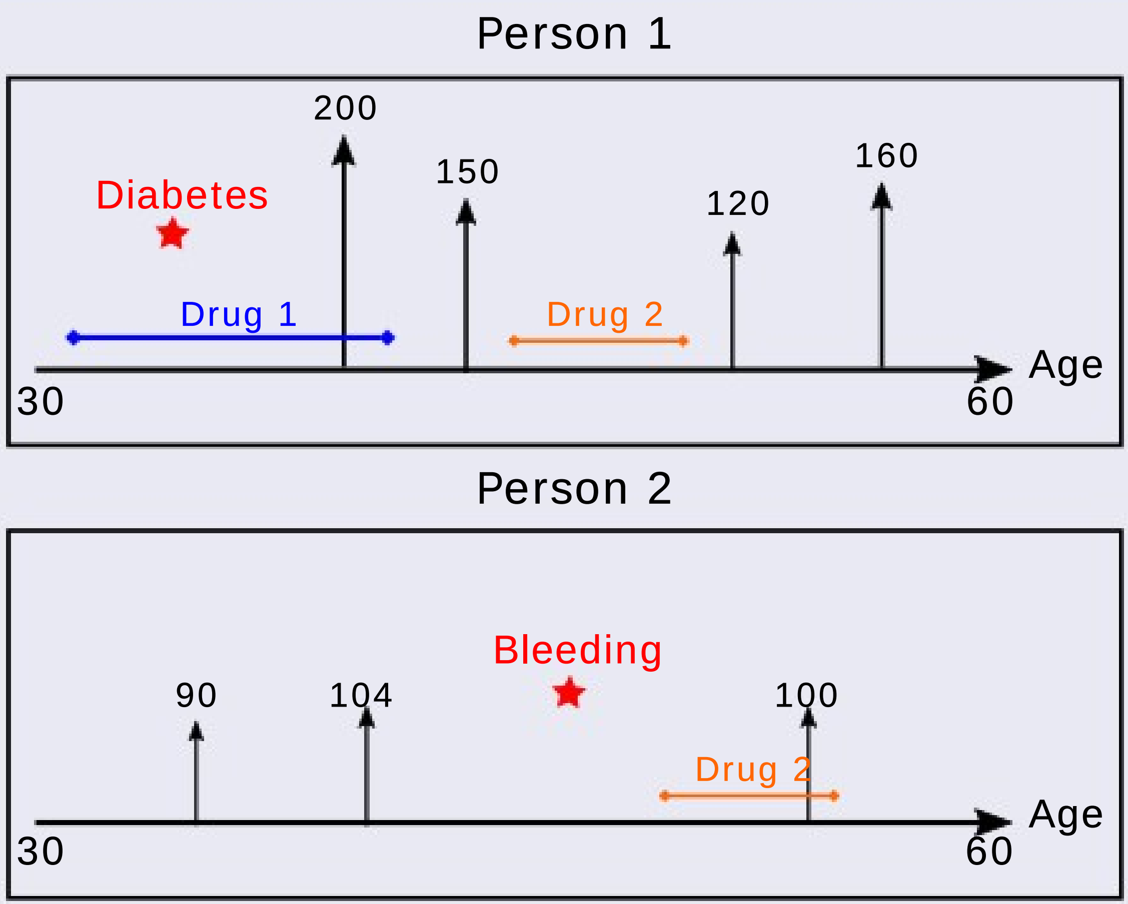
But Most ML Algorithms Expect:

- Single Table (Spreadsheet), or
- Regularly-Sampled Time Series

- Another Challenge: ML Algorithms aim for accurate prediction, not causal discovery

Extending SCCS to Numerical Response

Electronic Health Records (EHRs)



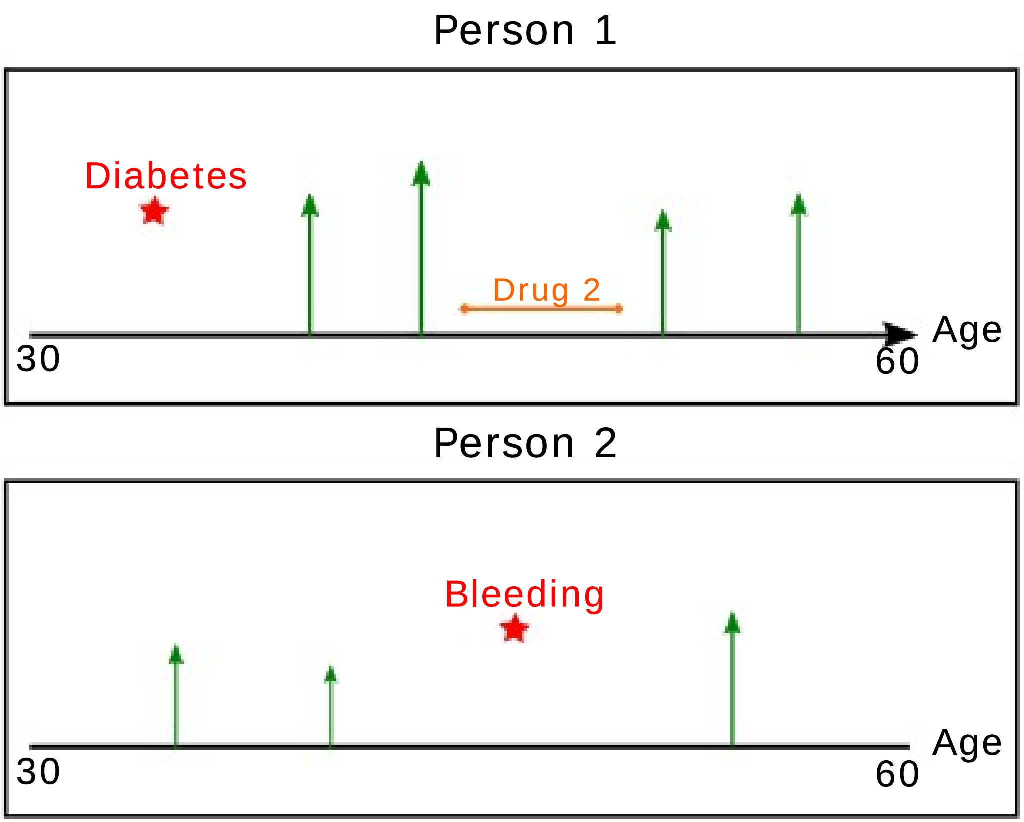
Properties

- Longitudinal
- Observational

Applications

- Adverse Drug Reaction (ADR) discovery
- Computational Drug Repositioning (CDR)

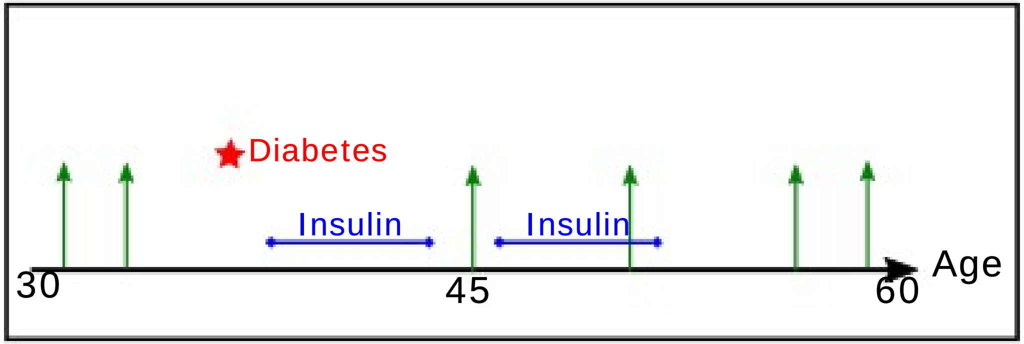
A Critical Intuition: Underlying Baseline



Baseline: Blood sugar level under no influence of any drugs.

Fixed Effect Model

Person 1

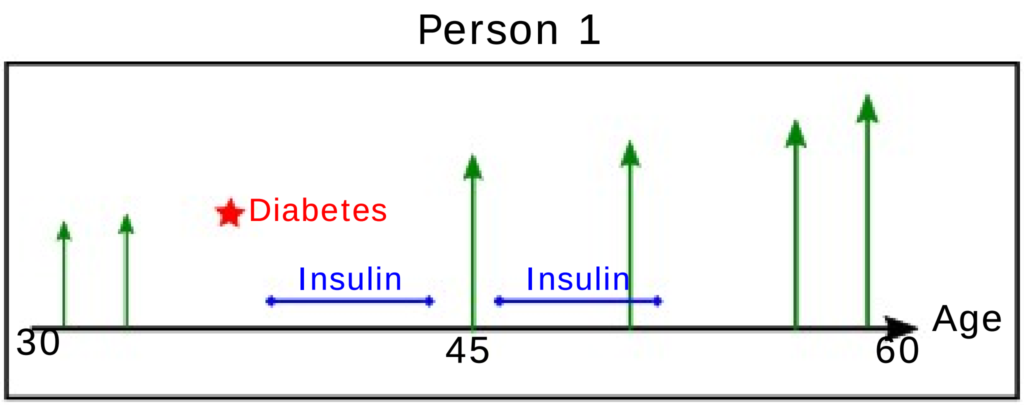


- Fixed Effect Model (Frees, 2004):

$$y_{ij} \mid x_{ij} = \alpha_i + \beta^T x_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma^2).$$

- $\dim \beta = \#$ drugs

Time-Varying Baseline



- Time-Varying Baseline, add regularization to minimize change In consecutive t_{ij} values:

$$y_{ij} \mid x_{ij} = t_{ij} + \beta^T x_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma^2).$$

More Ground Truth Available for Glucose *Lowering*

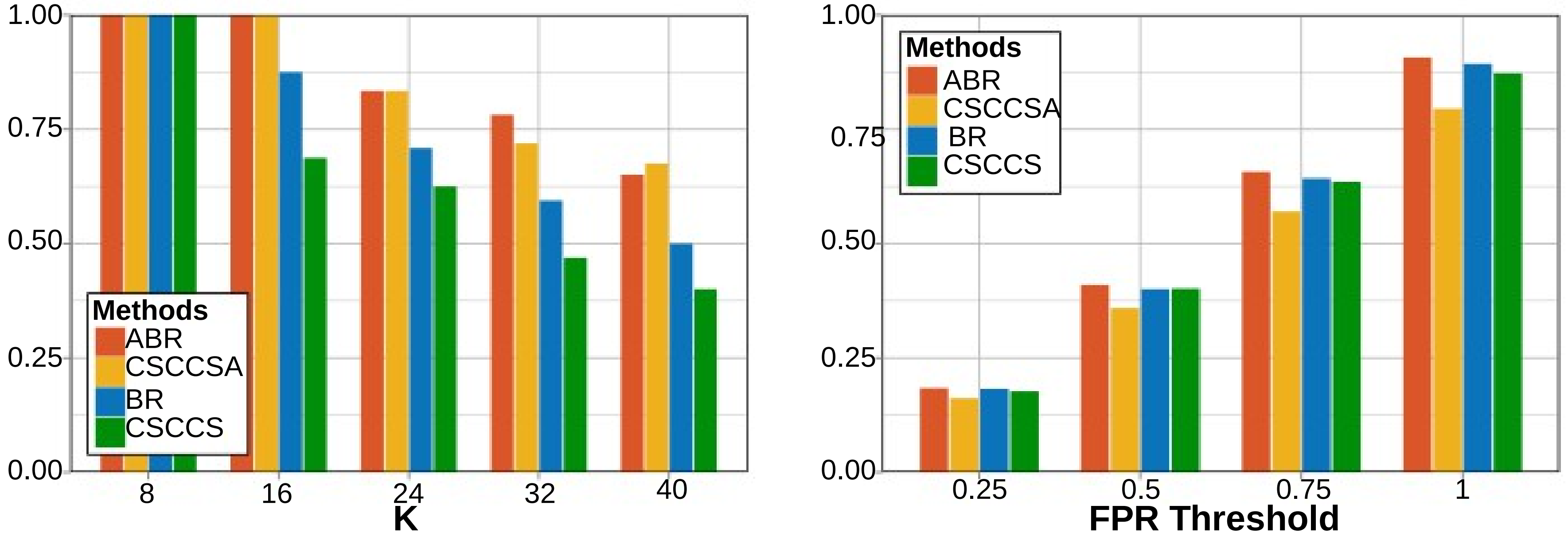


Figure: Left: Precision at K among the top-forty drugs generated by the four models; Right: Partial AUCs on the top-forty drugs generated by the four models.

- Sample size: 219306.
- Number of drug candidates: 2980.

Recovery of Known Glucose Lowering Agents

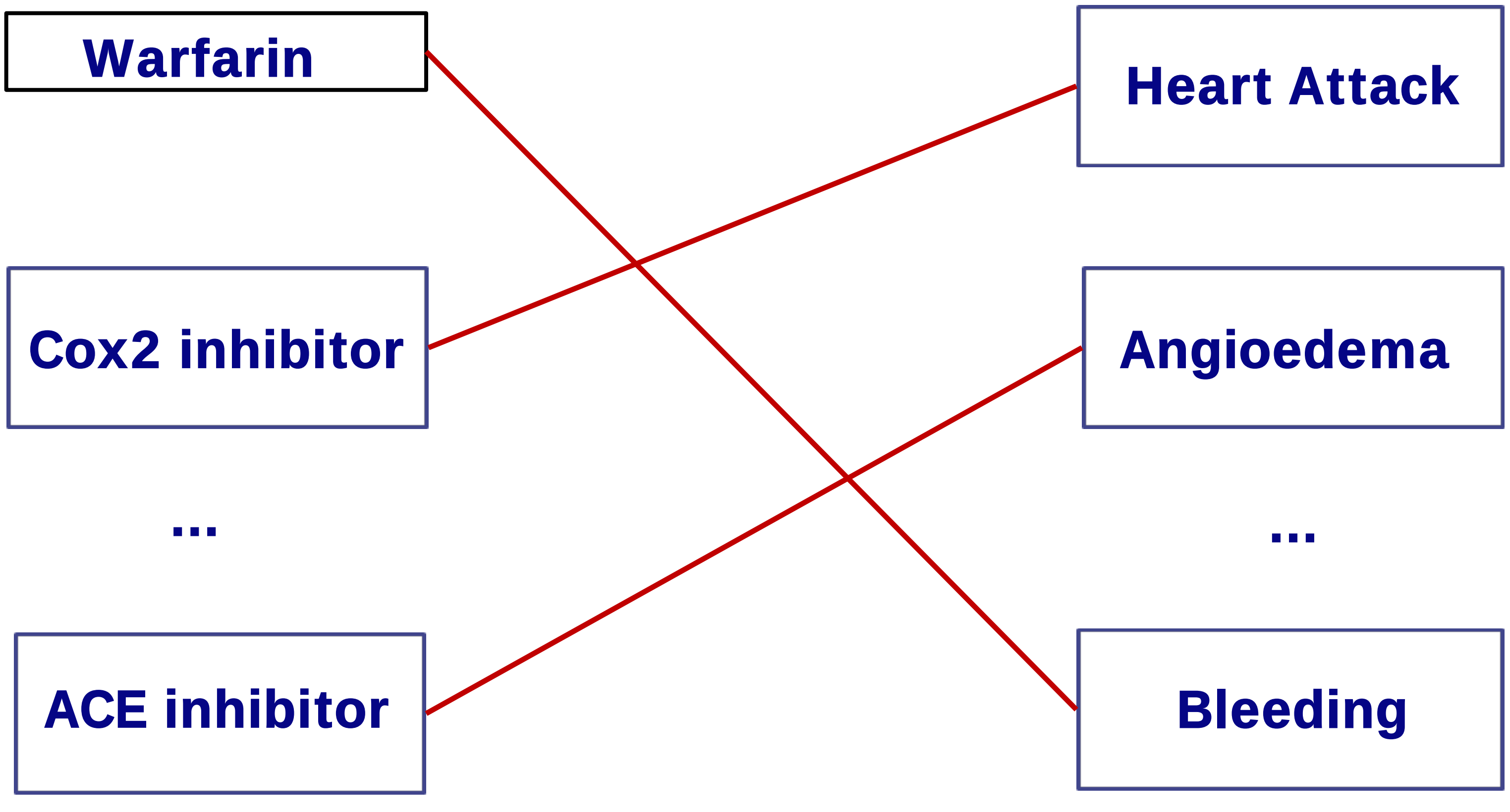
INDX	CODE	DRUG NAME	SCORE	COUNT
1	4485	HUMALOG	-11.786	124
2	7470	PIOGLITAZONE HCL	-10.220	3075
3	8437	ROSIGLITAZONE MALEATE	-9.731	1019
4	4837	INSULN ASP PRT/INSULIN ASPART	-9.658	258
5	6382	NEEDLES INSULIN DISPOSABLE	-9.464	2827
6	4171	GLUCOTROL XL	-8.117	2853
7	4106	GLIMEPIRIDE	-7.940	3384
8	160	ACTOS	-7.721	1125
9	824	AVANDIA	-6.802	1239
10	9152	SYRING W-NDL DISP INSUL 0.5ML	-6.623	4186
11	4132	GLUCOPHAGE	-6.322	6736
12	4184	GLYBURIDE	-6.021	8879
13	4170	GLUCOTROL	-5.721	1259
14	4208	GLYNASE	-5.670	591
15	416	AMARYL	-5.599	2240
16	4107	GLIPIZIDE	-5.563	9993
17	844	AXID	-4.682	189
18	2830	DILTIAZEM	-4.297	1021
19	4806	INSULIN GLARGINE HUM.REC.ANLOG	-4.175	4213
20	5787	METFORMIN HCL	-4.147	19584
21	2824	DILAUDID	-4.076	39
22	5786	METFORMIN	-3.890	3838
23	7731	PRAVACHOL	-3.532	1700
24	1760	CELEXA	-3.517	1473
25	4497	HUM INSULIN NPH/REG INSULIN HM	-3.501	1829
26	9889	URSODIOL	-3.132	376
27	4813	INSULIN NPL/INSULIN LISPRO	-2.972	623
28	4133	GLUCOPHAGE XR	-2.845	765
29	6445	NEURONTIN	-2.615	1418
30	6656	NPH HUMAN INSULIN ISOPHANE	-2.500	2874
31	9379	THIAMINE HCL	-2.383	341
32	1636	CARDURA	-2.198	1079
33	1218	BLOOD SUGAR DIAGNOSTIC DRUM	-2.073	2593
34	8025	PROZAC	-2.037	1525
35	8316	REZULIN	-1.895	444
36	9136	SYRINGE & NEEDLE INSULIN 1 ML	-1.885	3542
37	4802	INSULIN	-1.812	1526
38	7674	POTASSIUM CHLORIDE	-1.779	9842
39	4804	INSULIN ASPART	-1.752	2476
40	1200	BLOOD-GLUCOSE METER	-1.719	5289

INDX	CODE	DRUG NAME	SCORE	COUNT
1	4802	INSULIN	47	635
2	8316	REZULIN	50	120
3	824	AVANDIA	59	449
4	416	AMARYL	65	503
5	5226	LANTUS	66	33
6	5789	METFORMIN HYDROCHLORIDE	75	10
7	4485	HUMALOG	81	63
8	4132	GLUCOPHAGE	86	1813
9	4811	INSULIN NPH	88	19
10	144	ACTIGALL	90	34
11	1389	CAL	90	45
12	4171	GLUCOTROL XL	90	701
13	9155	SYRNG W-NDL DISP INSUL 0.333ML	95	29
14	4116	GLUCAGON	97	121
15	6652	NOVOLOG	98	51
16	160	ACTOS	106	480
17	6646	NOVOFINE 31	106	31
18	4813	INSULIN NPL/INSULIN LISPRO	108	118
19	8437	ROSIGLITAZONE MALEATE	109	332
20	4170	GLUCOTROL	113	641
21	9889	URSODIOL	113	123
22	5052	KAY CIEL	114	23
23	4118	GLUCAGON HUMAN RECOMBINANT	115	227
24	2521	DARVOCET-N	116	11
25	7470	PIOGLITAZONE HCL	121	705
26	5786	METFORMIN	125	2149
27	10366	ZINC SULFATE	130	34
28	4500	HUMULIN	135	33
29	4172	GLUCOVANCE	136	115
30	7471	PIOGLITAZONE HCL/METFORMIN HCL	136	16
31	6382	NEEDLES INSULIN DISPOSABLE	137	649
32	4184	GLYBURIDE	144	1354
33	4208	GLYNASE	145	115
34	4210	GLYSET	148	7
35	4163	GLUCOSE	159	1778
36	5977	MINIPRESS	163	19
37	7946	PROPANTHELINE	163	6
38	1602	CARBOCAINE	182	63
39	1305	BUDEPRION SR	185	9
40	6657	NPH INSULIN	185	43

Existing Methods' Limitations

- Response or candidate *conditions* must be pre-specified (though might be many)
- No consideration of *context* – ADE might only arise when patient
  - is taking another drug (drug interaction)
  - has specific properties, such as low weight or specific genetic variation

Most Current Approaches



What We Would Like:

Warfarin

Cox2 inhibitor

...

ACE inhibitor

Cox2 inhibitor(P,D) → hypertension(P)

older(P,55) , vioxx(D)

PatientID	Gender	Birthdate
P1	M	3/22/63

PatientID	Date	Physician	Symptoms	Diagnosis
P1	1/1/01	Smith	palpitations	hypoglycemic
P1	2/1/03	Jones	fever, aches	influenza

PatientID	Date	Lab Test	Result
P1	1/1/01	blood glucose	42
P1	1/9/01	blood glucose	45

PatientID	SNP1	SNP2	…	SNP 1M
P1	AA	AB		BB
P2	AB	BB		AA

PatientID	Date Prescribed	Date Filled	Physician	Medication	Dose	Duration
P1	5/17/98	5/18/98	Jones	prilosec	10mg	3 months

EMR

Reverse Machine Learning

- We already know who is on drug, and we want to find the condition it causes
- But we don’t know which condition
  - Might not even have predicate for condition in our vocabulary
  - Assume only that we can build condition definition from vocabulary as a clause body
- Treat drug use as *target concept*, and learn to predict that based on events *after* drug initiation

Use Rule Learning (ILP)

- If *antibiotics(P)* and *bleeding(P)* then *warfarin(P)*

- If *age_at_least(P,55)* and *hypertension(P)* then *vioxx(P)*

Using ML to Find Subgroups of Patients on Drug Based on Common Events Afterward

- Rule consequent specifies drug and rule antecedent specifies ADE
- Reverse of what we normally expect
- Richer condition definitions
- Can identify events that don’t correspond neatly to single condition
- Can identify drug interactions

SCCS-Like Scoring of Models

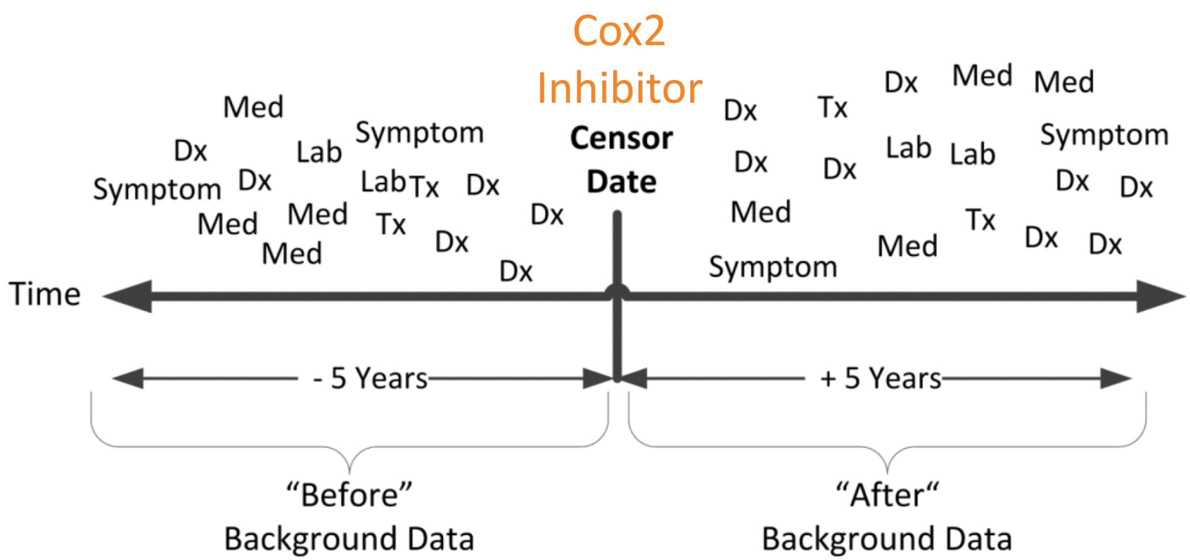
- Search for events that occur more frequently after drug initiation than before
- Example scoring function:

$$P(t_c > t_d \mid c,d)$$

- Could normalize, dividing by:

$$P(t_C > t_d \mid C,d)\ P(t_c > t_D \mid c,D)$$

Temporal filtering and Scoring Functions



$$\text{CASE}_{\text{After}} - \text{CASE}_{\text{Before}}$$

where now a CASE is person on a drug rather than person experiencing event

Results

Rules for Cox2(A) :-	Pos	Neg	Total	P-value
diagnoses(A,_,'790.29','Abnormal Glucose Test, Other Abn Glucose',_).	333	137	470	6.80E-20
diagnoses(A,_,'V54.89','Other Orthopedic Aftercare ',_).	403	189	592	8.59E-19
diagnoses(A,_,'V58.76','Aftcare Foll Surg Of The Genitourinary Sys',_).	287	129	416	6.58E-15
diagnoses(A,_,'V06.1','Diphtheria-Tetanus-Pertussis,Comb(Dtp)(Dtap)',_).	211	82	293	2.88E-14
diagnoses(A,_,'959.19','Other Injury Of Other Sites Of Trunk ',_).	212	89	301	9.86E-13
diagnoses(A,_,'959.11','Other Injury Of Chest Wall',_).	195	81	276	5.17E-12
diagnoses(A,_,'V58.75','Aftcar Foll Surg Of Teeth, Oral Cav, Dig Sys',_).	236	115	351	9.88E-11
diagnoses(A,_,'V58.72','Aftercare Following Surgery Nervous Syst, Nec',_)	222	106	328	1.40E-10
diagnoses(A,_,'410','Myocardial Infarction',_).	212	100	312	2.13E-10
diagnoses(A,_,'790.21','Impaired Fasting Glucose ',_).	182	80	262	2.62E-10

Test Summary Statistics

Rule	+	-	
+	838	333	1171
-	987	1492	2479
	1825	1825	3650

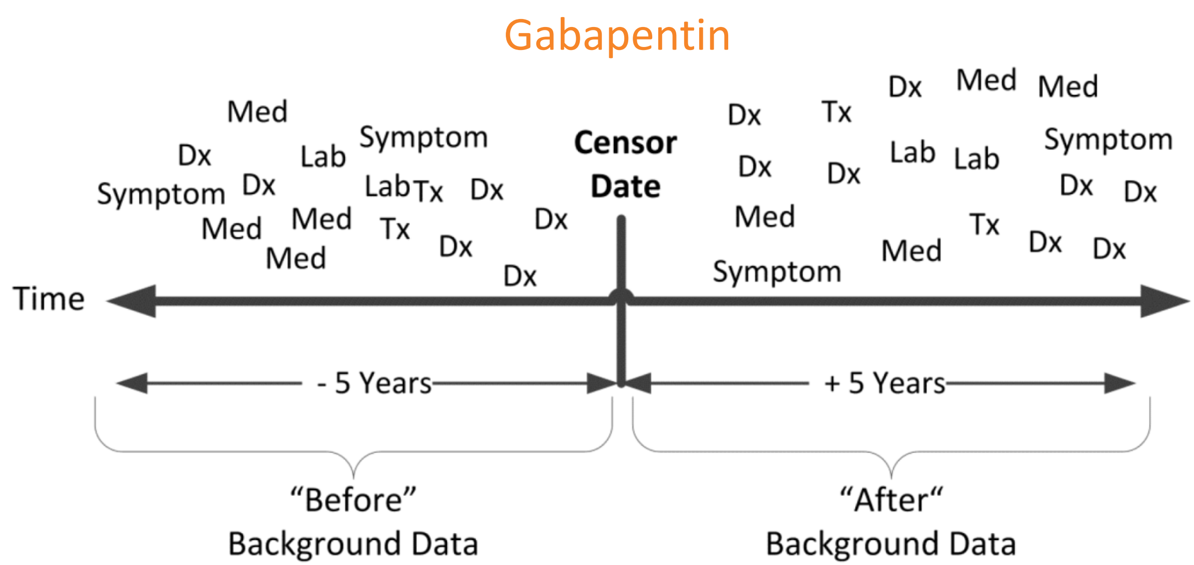
Accuracy = 0.638

Testset Recall/Precision/F1/Dsq2best = 0.459 0.716 0.559 0.373

Testset ROC_x/ROC_y/Dsq2best = 0.182 0.459 0.326

- Using only diagnoses → Accuracy = 0.63
- Using diagnoses, medications, labs → Accuracy = 0.78

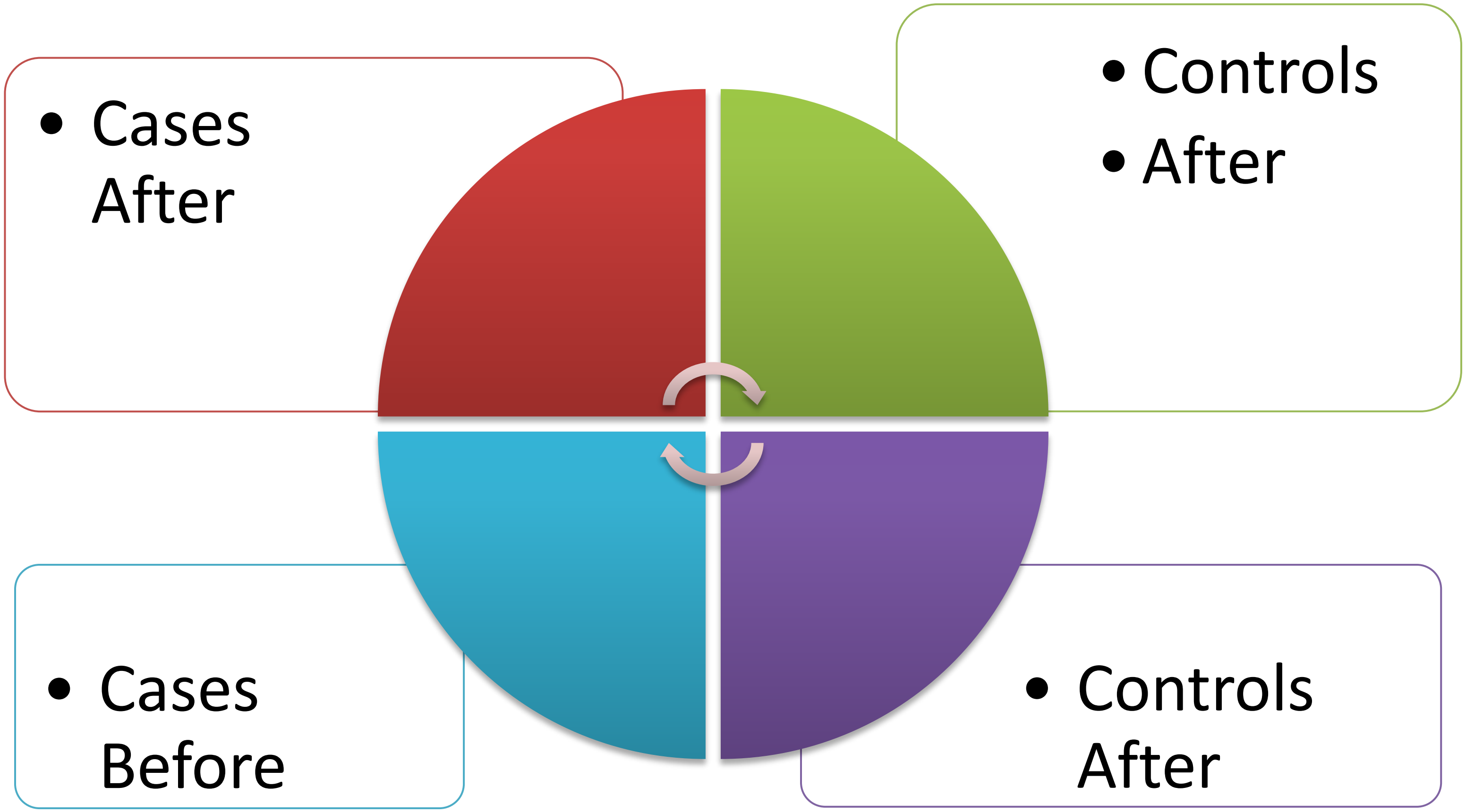
Recent Work on Generic vs. Brand Comparison



$$(\text{CASE}_{\text{After}} - \text{CTRL}_{\text{After}}) - (\text{CASE}_{\text{Before}} - \text{CTRL}_{\text{Before}})$$

where Censor Date is 2005 (time CASEs were switched from brand to generic)

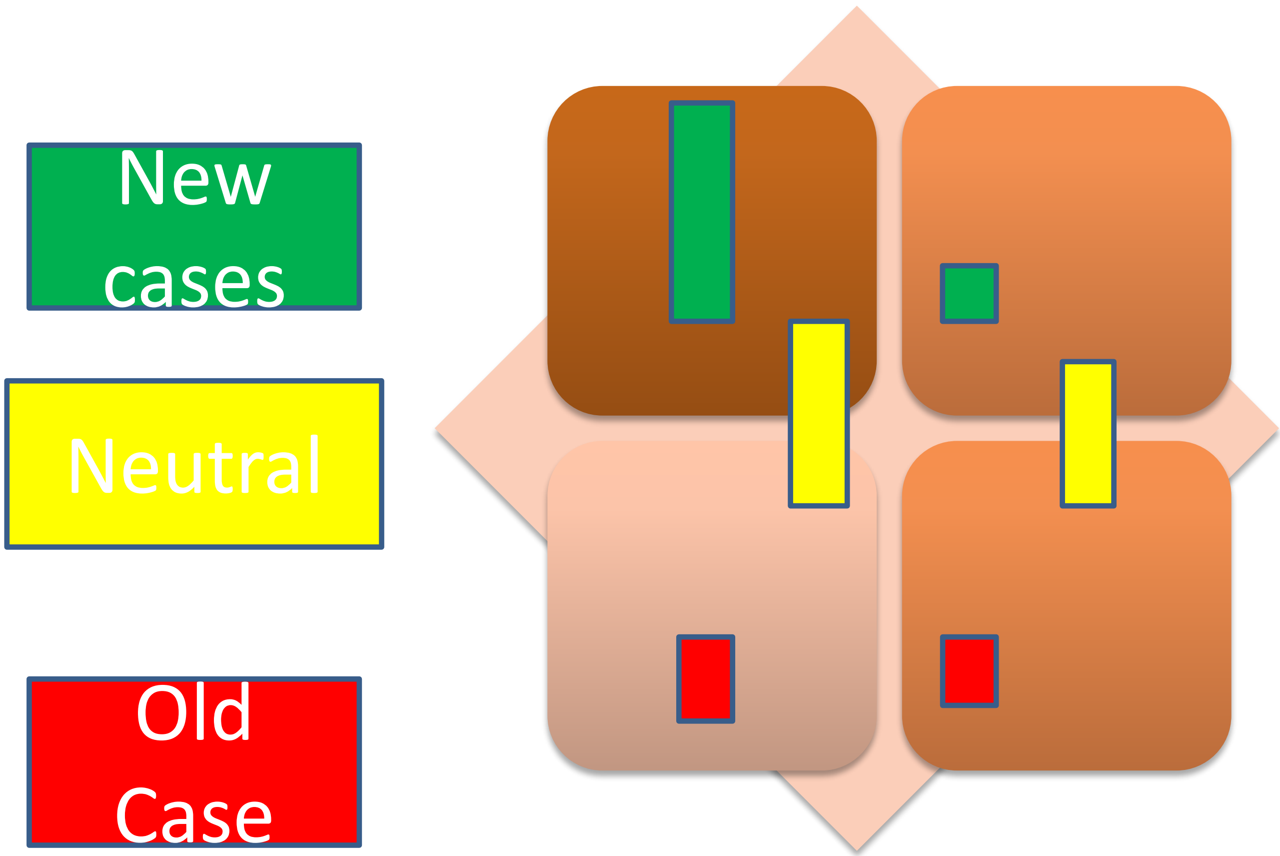
Cases and controls



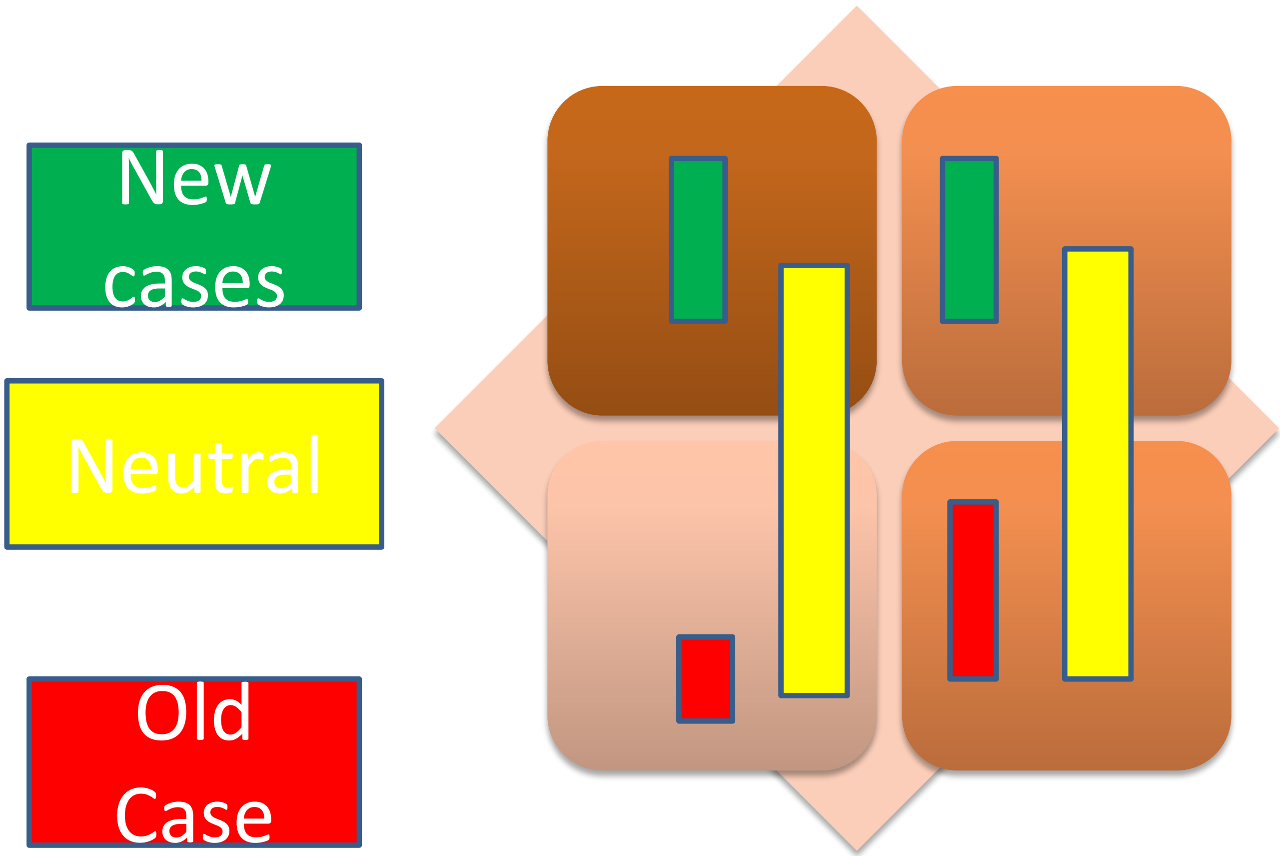
Reverse Learning

- - Can we detect who on Generic Gabapentin?
- - Each Patient is two examples
- - Confounders:
  - Marshfield policy change 2005
  - Most patients were switched to generic
  - Made unrelated changes to reporting system:
    - spurious, but highly predictive

Scoring: Informative Rule



Scoring: Less Informative

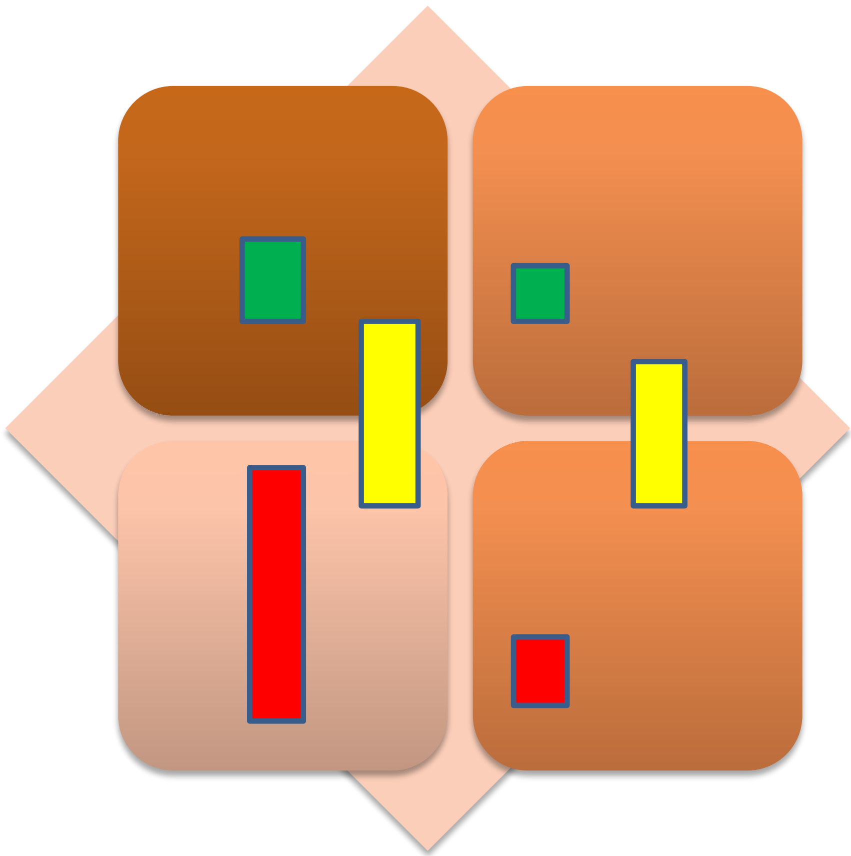


Scoring: Informative?

New cases

Neutral

Old Case



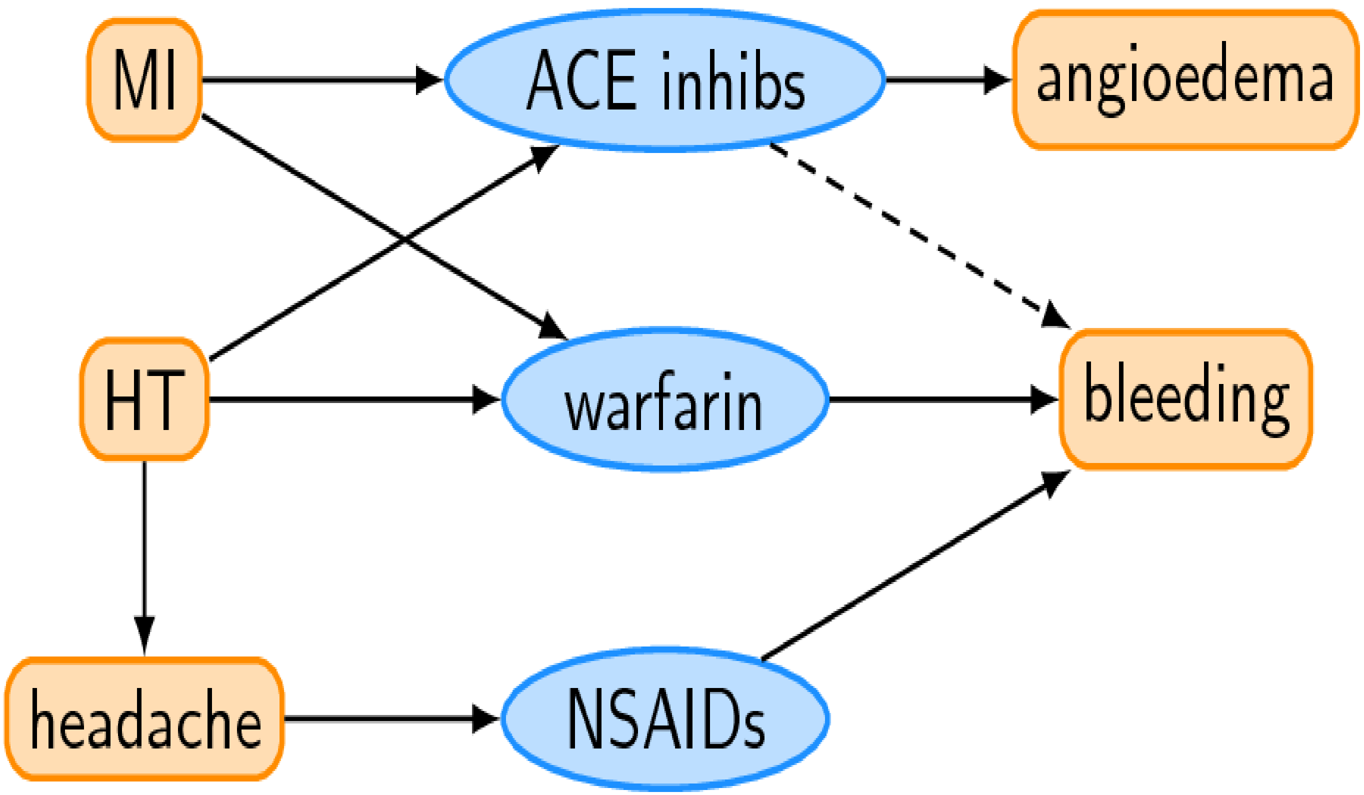
Biggest Challenges Now

- Evaluation: Few known cases of generic vs. brand differences for rediscovery evaluations
- Temporal confounding: adding controls (people not on drug) removed obvious ones
  - Prescription transmitted electronically
  - ICD code “other non-operative exam”
- But what about newer results such as hyperlipidemia, lidoderm, or levoquin?

Future Work

- Further addressing confounding, temporal and otherwise
- One approach: Incorporating learned rules as nodes in a graphical model taking time into account
- Finding new ways to evaluate, such as text mining to associate with recent findings in literature

Motivation



MI
ACE inhibs
angioedema
HT
warfarin
bleeding
headache
NSAIDs

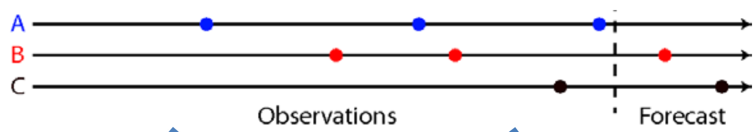
Continuous-time Graphical Models

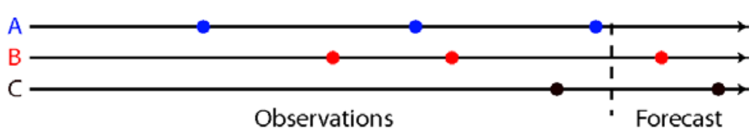
Continuous-time, discrete-state, with piecewise-constant transition rates

Point process: piecewise-continuous conditional intensity model (PCIM)

(Gunawardana et al., NIPS 2011)

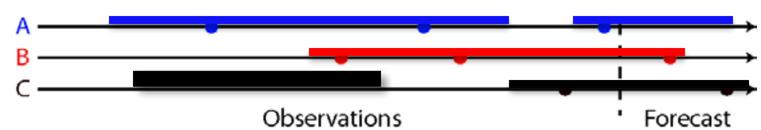
Continuous-time Bayesian networks (CTBNs) (Nodelman et al, UAI 2002)





Model of *Events*

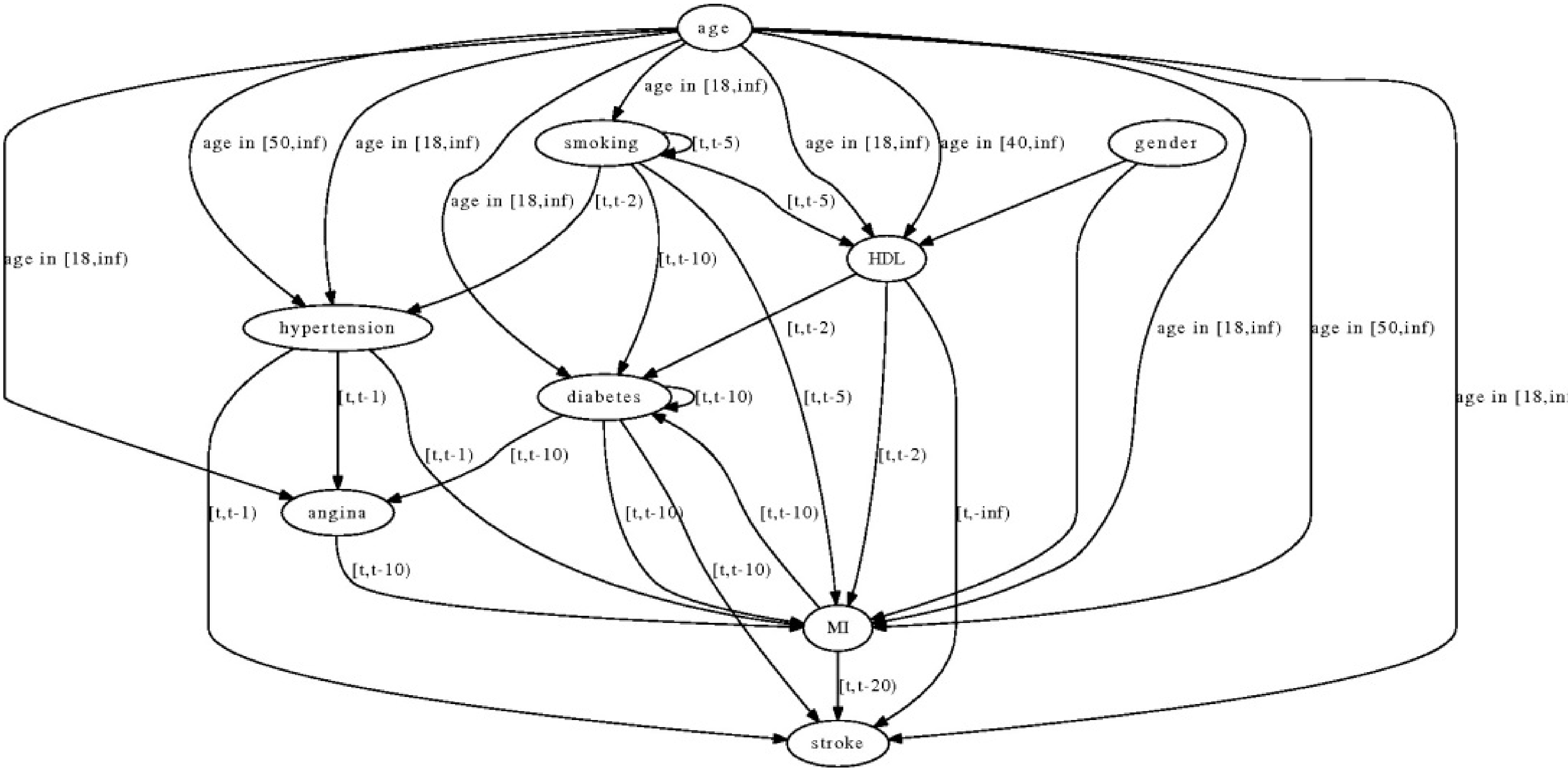
Point Processes



Model of *Persistent State*

CTBNs

Example CTBN or Point Process Structure



Goal: recover network-dependent event rates – measured by **test set log likelihood**

Conclusion

- ML has at least the potential to bring new approaches to ADE Detection task
- ML needs lessons from Epidemiology
- ADE Detection task provides exciting variant to the hot topic within ML of causal discovery
  - Pearl, Robbins, Cooper, GSS: under what conditions or assumptions can we guaranteed-correctly infer causal relationships from observational data?
  - Here: as in ML, we don’t expect to be 100% accurate. What methods let us most accurately rank causal relationships from observational data?

Thanks

- Peggy Peissig
- Rick Hansen
- Michael Caldwell
- Vitor Santos Costa
- Charles Kuang
- Aubrey Barnard
- Jeremy Weiss
- FDA
- NIGMS