

Supplementary Materials:

Table S1 Ontologies: We use 19 ontologies from UMLS, BioPortal and elsewhere for building a lexicon for annotation, and for normalizing drugs and diseases.

UMLS	
CPT	Current Procedural Terminology, 2011
CST	COSTART, 1995
HCPCS	Healthcare Common Procedure Coding System, 2011
ICD10CM	International Classification of Diseases, 10th Edition, Clinical Modification, 2011_01
ICD9CM	International Classification of Diseases, Ninth Revision, Clinical Modification, 2012
LNC	LOINC, 236
MDR	Medical Dictionary for Regulatory Activities Terminology (MedDRA), 14.0
MEDLINEPLUS	MedlinePlus Health Topics, 20110108
MSH	Medical Subject Headings, 2012_2011_09_09
NCI	NCI Thesaurus, 2011_02D
NDFRT	National Drug File, 2011_09_06
RXNORM	RxNorm Vocabulary, 11AA_110906F
SNOMED CT	SNOMED Clinical Terms, 2011_07_31
UMD	UMDNS: product category thesaurus, 2011
VANDF	Veterans Health Administration National Drug File, 2011_07_29
WHO	WHO Adverse Reaction Terminology, 1997
OMIM	Online Mendelian Inheritance in Man, 2011_06_08
OTHER	
ATC	Anatomical Therapeutic Classification
DO	Human Disease Ontology

Table S2 Terms used to identify drugs and adverse events from EHRs (ddi_concepts.csv): Drugs are defined as active ingredients from the RxNorm ontology that are also present in DrugBank and the ATC ontology. To prepare the term list for these active ingredients, we start with the RxNorm concept and include synonyms, trade names and other forms of the drug.

We study 14 adverse events based on the availability of known drug interactions causing them, their population event rate in STRIDE and our ability to successfully detect their presence in text from EHRs. To prepare the term list for adverse events, we start at the UMLS concept that best describes it and—using the *is-a* relationship hierarchy in the SNOMED CT ontology to reach relevant parents and siblings—we gather all terms and their synonyms that define the adverse event.

We then process these sets to remove terms that have multiple meanings that can cause ambiguity, followed by manual curation. The file (ddi_concepts.csv) lists the final set of terms used to define our set of 1,165 drugs and 14 adverse events.

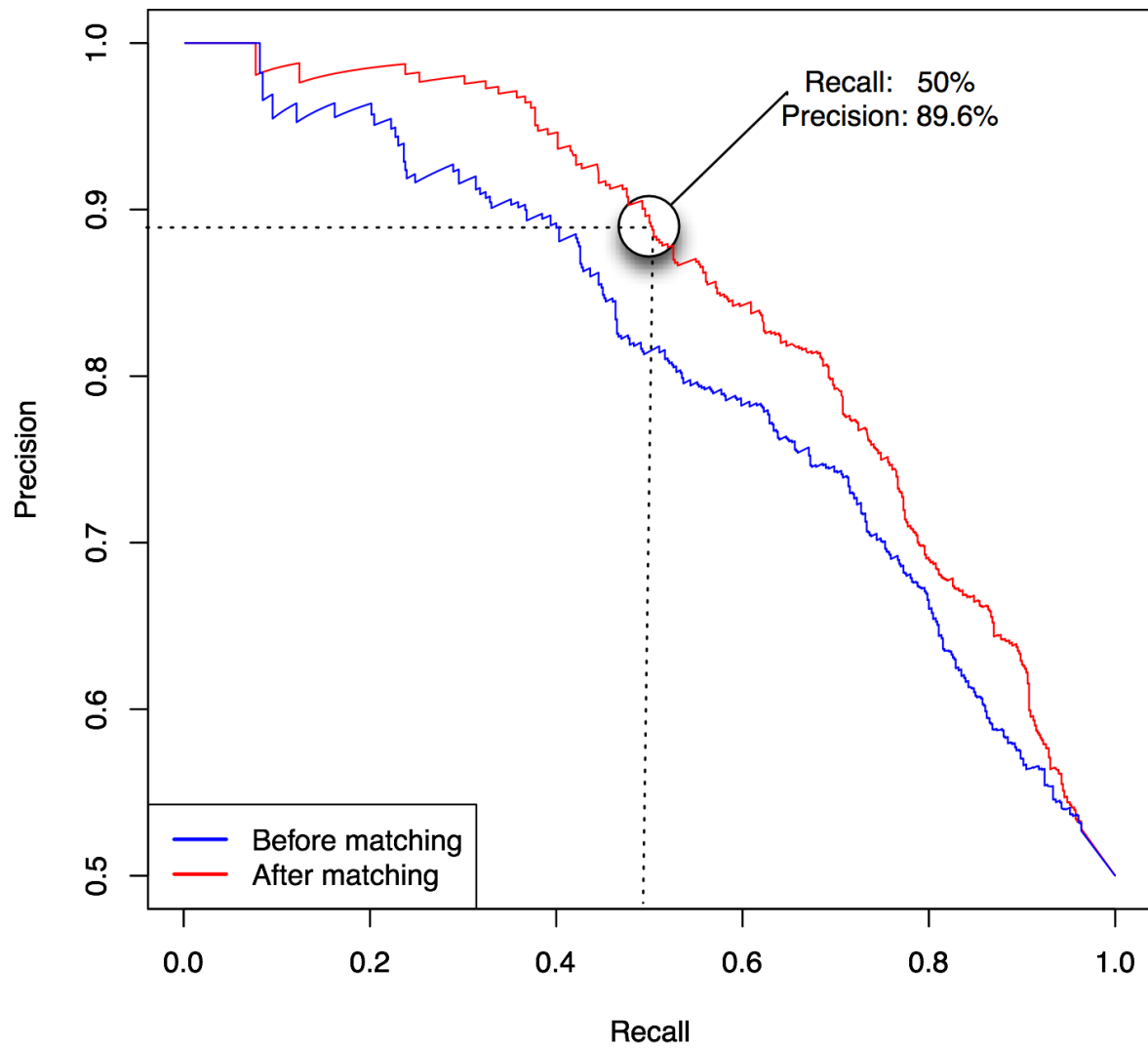
S3 Validation results: Accuracy measures for validating the automated annotation mechanism (compared with curated annotations) as determined using 16 different conditions curated for the i2b2 Obesity NLP reference set and an additional 9 conditions that were manually curated. [Atherosclerotic cardiovascular disease; Gastroesophageal reflux disease; Progressive multifocal leukoencephalopathy; TP – True Positives; FP – False Positives; FN – False Negatives; TN – True Negatives; SENS – Sensitivity; SPEC – Specificity (Recall); PPV – Positive Predictive Value (Precision); F1 – F Measure]

CONDITION	TP	FP	FN	TN	SENS	SPEC	PPV	F1
Asthma	161	13	0	1058	100%	99%	93%	96%
Atherosclerotic cardiov. dis.	523	78	153	464	77%	86%	87%	82%
Congestive heart failure	373	32	142	669	72%	95%	92%	81%
Depression	171	45	5	1013	97%	96%	79%	87%
Diabetes mellitus	740	16	83	390	90%	96%	98%	94%
Gallstones/cholecystectomy	31	1	165	1039	16%	100%	97%	27%
Gastroesophageal reflux dis.	178	13	9	1027	95%	99%	93%	94%
Gout	134	9	8	1082	94%	99%	94%	94%
Hypercholesterolemia	275	4	242	707	53%	99%	99%	69%
Hypertension	709	23	202	296	78%	93%	97%	86%
Hypertriglyceridemia	13	0	15	1208	46%	100%	100%	63%
Obstructive sleep apnea	65	2	109	1055	37%	100%	97%	54%
Osteoarthritis	139	75	62	954	69%	93%	65%	67%
Peripheral vascular disease	106	21	60	1049	64%	98%	83%	72%
Venous insufficiency	23	18	8	1186	74%	99%	56%	64%
Obesity	301	0	195	727	61%	100%	100%	76%
Myocardial infarction	200	53	115	869	63%	94%	79%	70%
Venous thrombosis	94	106	5	1032	95%	91%	47%	63%
Acute renal failure	88	111	11	1027	89%	90%	44%	59%
Cardiac valve fibrosis	14	223	4	996	78%	82%	6%	11%
QT prolongation	13	27	4	1193	76%	98%	33%	46%
Pancytopenia	12	2	0	1223	100%	100%	86%	92%
Bladder cancer	6	2	3	1226	67%	100%	75%	71%
Rhabdomyolysis	3	0	1	1233	75%	100%	100%	86%
Prog. multif. leukoenc..pathy	0	0	0	1237	NA	NA	NA	NA
GRAND TOTAL:	4372	874	1601	23960	0.74	0.96	0.79	0.71

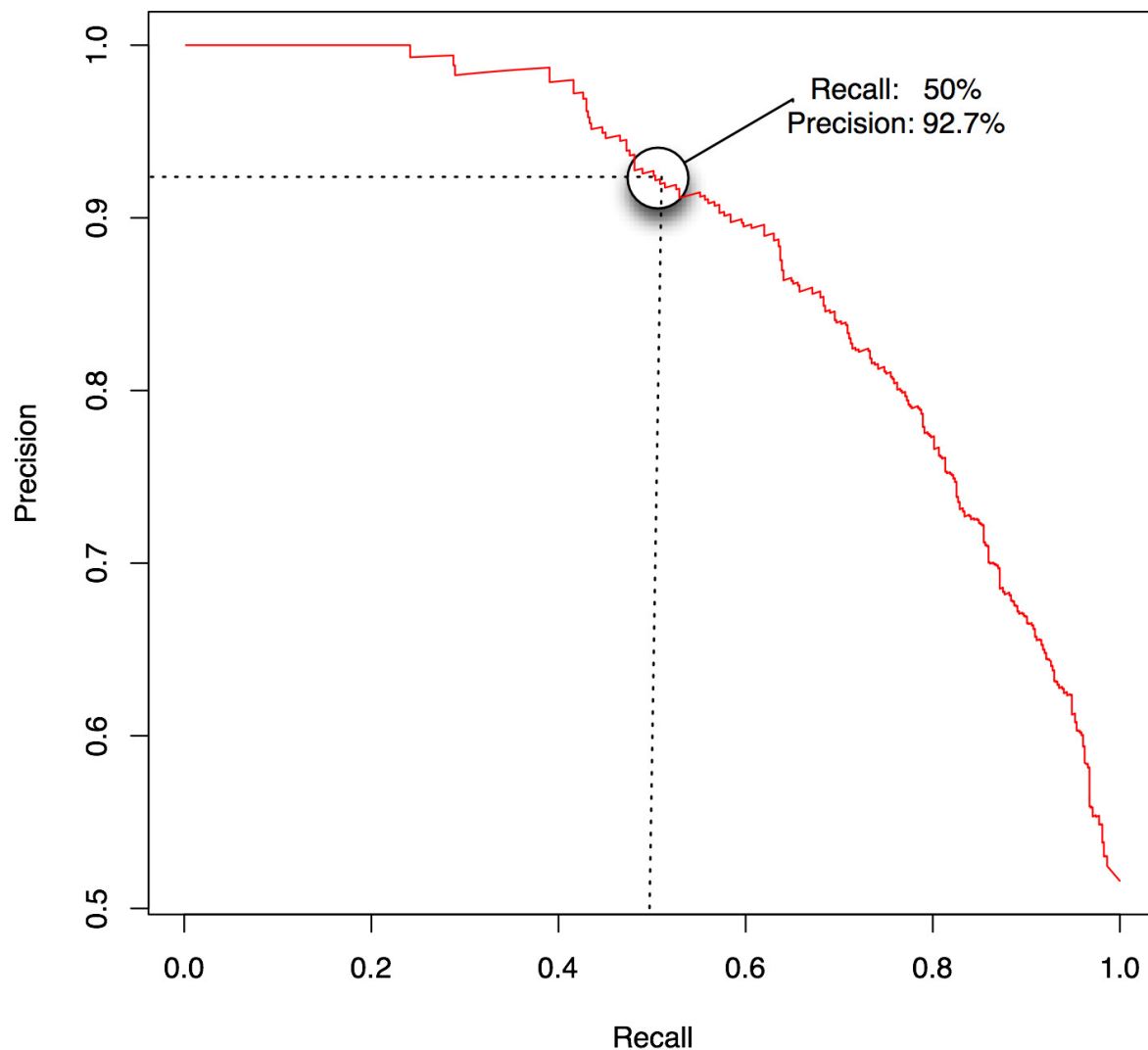
S4 Mapping from adverse event concepts in EHRs to MedDRA codes: In order to compare our method on EHRs with established methods on FAERS, we use the following mapping from EHR event concepts to MedDRA PTs that are used to code FAERS.

Event	MedDRA PT
Cardiac Arrhythmias	Arrhythmia
Neutropenia	Neutropenia
Bradycardia	Bradycardia
Hypoglycemia	Hypoglycaemia
Acute Renal Failure	Renal Failure acute
Hyperkalemia	Hyperkalaemia
Hyperglycemia	Hyperglycaemia
Nephrotoxicity	-
Pancytopenia	Pancytopenia
Hypokalemia	Hypokalaemia
Serotonin Syndrome	Serotonin syndrome
QT Prolongation	Long QT syndrome
Parkinsonian Symptoms	Parkinsonism
Rhabdomyolysis	Rhabdomyolysis

S5 Precision-Recall curves showing the values of precision and recall that can be obtained by varying the threshold for $UOR_{0.25}$ and $AOR_{0.25}$, for 10 events [$UOR_{0.25}$ – Lower Bound of the Unadjusted Odds Ratio 95% Confidence Interval; $AOR_{0.25}$ – Lower Bound of the Adjusted Odds Ratio 95% Confidence Interval]



Precision Recall curves for STRIDE using $UOR_{0.25}$ and $AOR_{0.25}$. Adjusting improves performance at all points.



Precision Recall curves for PAMF using AOR_{025} .

S6 Population event rate of all drug-drug-event tuples in our study (event_rate.csv.tar.gz)

We provide the 2-by-2 table counts for all drug-drug-event tuples in our study that have an exposure of at least 100 patients, along-with UOR_{025} and the rate of the event for the drug combination—defined as the fraction of patients on the two drugs, who are subsequently afflicted with the adverse event. We also calculate the AOR_{025} for those tuples that are signaled positive by EB05 in FAERS (threshold=1.5) and by UOR_{025} (threshold=4.7) in STRIDE. The 'label' field indicates whether the tuple was used as a positive test case, a negative test case or was unused, in our gold standard.

S7 New additions to Drugs.com List of interactions signaled in FAERS (EB05>1.5) and EHRs (UOR₀₂₅> 4.7 and AOR₀₂₅> 1.1) using STRIDE data up to May 2011, that were added into Drugs.com between May 2011 and Nov 2012. [UOR₀₂₅ – Lower Bound of the Unadjusted Odds Ratio 95% Confidence Interval; AOR₀₂₅ – Lower Bound of the Adjusted Odds Ratio 95% Confidence Interval]

Adverse Event	Drug1	Drug2	AOR ₀₂₅
hypoglycemia	levofloxacin	rosiglitazone	3.807794
hypoglycemia	insulin glargine	sitagliptin	4.210419
hypoglycemia	gemfibrozil	rosiglitazone	11.33265
hypoglycemia	metformin	Orlistat	4.662141
pancytopenia	fludarabine	alemtuzumab	15.75846
serotonin syndrome	tramadol	sertraline	4.482547
serotonin syndrome	tramadol	sumatriptan	2.933818
serotonin syndrome	tramadol	venlafaxine	10.57121
serotonin syndrome	tramadol	duloxetine	37.77019
serotonin syndrome	tramadol	Linezolid	1.778889
serotonin syndrome	sertraline	sumatriptan	15.51812
serotonin syndrome	sertraline	venlafaxine	8.091355
serotonin syndrome	sertraline	duloxetine	16.6861
serotonin syndrome	sertraline	Linezolid	2.539668
serotonin syndrome	sumatriptan	venlafaxine	9.134166
serotonin syndrome	sumatriptan	duloxetine	17.03185
serotonin syndrome	venlafaxine	duloxetine	18.60719
serotonin syndrome	venlafaxine	Linezolid	8.424942
serotonin syndrome	duloxetine	Linezolid	7.657865
serotonin syndrome	duloxetine	eletriptan	14.99389
serotonin syndrome	clomipramine	sertraline	53.09415
serotonin syndrome	fluoxetine	tramadol	6.34216
serotonin syndrome	fluoxetine	sumatriptan	10.39615
serotonin syndrome	fluoxetine	venlafaxine	10.38797
serotonin syndrome	fluoxetine	duloxetine	20.14702
serotonin syndrome	fluoxetine	Linezolid	3.250727
serotonin syndrome	fluoxetine	metoclopramide	3.759478
serotonin syndrome	metoclopramide	sertraline	1.39371
serotonin syndrome	metoclopramide	venlafaxine	3.954983
serotonin syndrome	amitriptyline	sertraline	9.856795
serotonin syndrome	amitriptyline	venlafaxine	6.721252
serotonin syndrome	amitriptyline	duloxetine	20.32482
serotonin syndrome	amitriptyline	Linezolid	2.302513
serotonin syndrome	amitriptyline	fluoxetine	5.344813
parkinsonian symptoms	clonazepam	levodopa	188.8748
parkinsonian symptoms	amantadine	metoclopramide	5.418087
parkinsonian symptoms	levodopa	pramipexole	391.2834
parkinsonian symptoms	levodopa	metoclopramide	103.986
cardiac arrhythmias	digoxin	dobutamine	8.220163
hyperkalemia	potassium chloride	Lisinopril	1.632412
hyperkalemia	potassium chloride	fosinopril	1.308203
hyperkalemia	benazepril	Losartan	1.329921
hyperkalemia	lisinopril	eplerenone	1.202691
hyperkalemia	ramipril	Losartan	2.303345
hyperkalemia	enalapril	potassium chloride	1.279014
hyperkalemia	enalapril	triamterene	1.339857
hyperkalemia	enalapril	Losartan	1.424779
hyperkalemia	enalapril	Heparin	1.220514
hyperkalemia	heparin	Lisinopril	1.8721