

Supplemental Information: Hypertension regular expression

Regular expression 1: `.*(?!pulm\w*\W*\w+\W+)hypertension.*` (case insensitive)

OR

Regular expression 2: `.*(?!pulm\w*\W*\w+\W+)HTN.*` (case insensitive)

Supplemental Table 1: Full Set of Features by Category with Descriptions

The table separates all features into key groups used for simple algorithms and to train random forests. Document counts were separately included with the regular expression (RegEx) category, NLP-based concepts, or only once when we used both note-based categories. Lines divide the three subcategories of “ICD9, medications, all BP” features which we also used to train random forests individually. The explanation for each feature is given alongside its column identifier as used in the accompanying KNIME module example input.

Category	Descriptor
ICD9, medications, all BP	date-ICD9 count
	unique ICD9 count
	hypertension ICD9 count
	unique hypertension ICD9 count
	hypertension ICD9 count normalized
	hypertension ICD9 count unique normalized
	unique hypertension ICD9 count normalized
	unique hypertension ICD9 count unique normalized
	meds count
	unique meds count
	HYPERTENSION meds count
	unique hypertension meds count
	hypertension meds count normalized
	hypertension meds count unique normalized
	unique hypertension meds count normalized
	unique hypertension meds count unique normalized
	max age
	vital reading time span (days)
	visits with vitals count (days)
hypertensive BP count (days)	
median systolic (all)	
median diastolic (all)	
hypertensive BP count normalized	
vitals density	
All vitals with separate outpatient and inpatient blood pressures	outpatient visits with vitals count (days)
	outpatient hypertensive BP count (days)
	median systolic (outpatient)
	median diastolic (outpatient)
	outpatient hypertensive BP count normalized
	outpatient vitals density
	inpatient visits with vitals count (days)

inpatient hypertensive BP count (days)
median systolic (inpatient)
median diastolic (inpatient)
inpatient hypertensive BP count normalized
inpatient vitals density
median pulse (all)
median pulse (outpatient)
median pulse (inpatient)

Document counts

PL count
DS count
HPC count
All document count

Regular expression counts and normalized features

PL hypertension regular expression document matches
DS hypertension regular expression document matches
HPC hypertension regular expression document matches
All doc hypertension regular expression document matches
PL hypertension regular expression document matches document normalized
DS hypertension regular expression document matches document normalized
HPC hypertension regular expression document matches document normalized
All doc hypertension regular expression document matches document normalized

NLP-based concepts and normalized features

PL concept count
DS concept count
HPC concept count
All document concept count
PL hypertension concept count
DS hypertension concept count
HPC hypertension concept count
All document hypertension concept count
PL hypertension concept count document normalized
DS hypertension concept count document normalized
HPC hypertension concept count document normalized
All doc hypertension concept count document normalized
PL hypertension concept count concept normalized
DS hypertension concept count concept normalized
HPC hypertension concept count concept normalized
All doc hypertension concept count concept normalized

Supplemental Table 2: Full Set of Hypertension-related ICD9 Billing Codes Used

CODE CODE DESC

405.99 SECOND HYPERTENSION NEC

405 SECONDARY HYPERTENSION

405.9 SECOND HYPERTENSION NOS

401 ESSENTIAL HYPERTENSION

401.1 BENIGN HYPERTENSION

401.0 MALIGNANT HYPERTENSION

405.91 RENOASC HYPERTENSION

405.0 MAL SECOND HYPERTENSION

401.9 HYPERTENSION NOS

Supplemental Table 3: Full Set of Hypertension-related Medication Strings Used (Based on MEDI-HPS)

RXCUI	DRUG_DESC	RXCUI (cont. 1)	DRUG_DESC (cont. 1)	RXCUI (cont. 2)	DRUG_DESC (cont. 2)
217597	hydralazine	5036	guanethidine	865575	metoprolol
71515	prazosin	262242	eprosartan	6860	methyclothiazid e
5033	guanabenz	35296	ramipril	3827	enalapril
151890	propranolol	203344	metoprolol	831221	diltiazem
1009015	amlodipine / olmesartan	136411	sildenafil	202463	hydralazine
754809	prazosin	5487	hydrochlorothia zide	226918	orlistat
225030	fosinopril	202509	nitroglycerin	744629	amlodipine
745969	hydrochlorothia zide	218710	potassium	153668	irbesartan
10602	timolol	9260	reserpine	745974	hydrochlorothia zide
19484	bisoprolol	220669	hydrochlorothia	151155	diltiazem

			zide		
202510	nitroglycerin	216778	clonidine	724892	valsartan
151300	nifedipine	750200	amlodipine	215917	clonidine
203489	verapamil	169050	chlorthalidone	225959	nitroglycerin
224934	methyclothiazid e	219878	hydralazine	218744	nitroglycerin
216252	nadolol	151882	terazosin	31555	nebivolol
856468	propranolol	724876	amlodipine	750204	amlodipine
219881	hydralazine	214354	candesartan	142146	bisoprolol fumarate
225752	verapamil	219878	reserpine	744887	amlodipine
6918	metoprolol	491231	hydrochlorothia zide	724884	amlodipine
225191	reserpine	758540	aliskiren	196460	betaxolol
202369	betaxolol	542847	hydrochlorothia zide	203798	metolazone
215673	betaxolol	8565	polythiazide	151482	captopril

196500	perindopril	823939	hydrochlorothiazide	196472	lisinopril
328141	Angelica sinensis preparation	175128	hydrochlorothiazide	744633	amlodipine
7973	penbutolol	29046	lisinopril	540618	diltiazem
217642	reserpine	6876	methyldopa	216253	timolol
203644	lisinopril	218102	indapamide	151549	nadolol
644	amiloride	215905	hydrochlorothiazide	217597	hydrochlorothiazide
151294	captopril	284747	reserpine	5495	hydroflumethiazide
6916	metolazone	323502	bosentan	219797	reserpine
744621	olmesartan	154991	enalapril	9997	spironolactone
83367	atorvastatin	581642	sildenafil	744891	benazepril
750200	atorvastatin	216909	methyclothiazide	203123	enalapril maleate
705130	nitroglycerin	203477	chlorthalidone	321064	olmesartan

591568	lloprost	17767	amlodipine	216675	trichlormethiazide
49276	doxazosin	214621	hydrochlorothiazide / metoprolol	203275	hydrochlorothiazide
5470	hydralazine	809019	valsartan	215910	carteolol
203211	diltiazem hydrochloride	202941	clonidine	220391	hydralazine
2396	chlorothiazide	7417	nifedipine	750236	amlodipine
20352	carvedilol	214618	hydrochlorothiazide / lisinopril	10828	trimethaphan
196468	doxazosin	6673	mecamylamine	9259	rescinnamine
175128	enalapril	219882	hydralazine	217653	guanadrel
151486	lisinopril	324042	hydrochlorothiazide / spironolactone	750196	atorvastatin
214866	trandolapril	708361	orlistat	8332	pindolol

	/ verapamil				
218936	deserpidine	218353	telmisartan	135481	candesartan cilexetil
155091	fenoldopam	1627	timolol	203273	hydrochlorothiazide
220391	reserpine	151723	carvedilol	218369	hydrochlorothiazide
491095	diltiazem	215919	clonidine	215918	clonidine
758540	hydrochlorothiazide	750228	amlodipine	8629	prazosin
202525	nitroglycerin	33910	isradipine	78678	Rauwolfia preparation
744871	amlodipine	1202	atenolol	809027	valsartan
54635	phentolamine	11170	verapamil	4109	ethacrynic acid
3327	diazoxide	749838	telmisartan	541854	nitroglycerin
202530	nitroglycerin	219882	reserpine	38454	trandolapril
8787	propranolol	2409	chlorthalidone	218747	nitroglycerin

599	alseroxylon	4316	felodipine	8149	phenoxybenzamine
744625	amlodipine	50166	fosinopril	202523	nitroglycerin
190465	sildenafil	744871	benazepril	750232	amlodipine
758548	aliskiren	750224	atorvastatin	40114	guanfacine
491231	timolol	202508	nitroglycerin	266604	metyrosine
541103	bumetanide	215871	hydrochlorothiazide	750224	amlodipine
1808	bumetanide	805852	candesartan	203238	triamterene
6877	methyldopate	72210	quinapril	744891	amlodipine
152434	torseamide	214357	captopril / hydrochlorothiazide	66977	bepidil
750236	atorvastatin	142432	clonidine hydrochloride	1369	bendroflumethiazide
62349	ethacrynate	10600	timolol	724884	valsartan
169050	reserpine	202507	nitroglycerin	744875	amlodipine

151177	nitroglycerin	215870	hydrochlorothiazide	203494	diltiazem
	bisoprolol				
214317	/ hydrochlorothiazide	151172	nitroglycerin	656315	atenolol
216251	nadolol	220004	spironolactone	10763	triamterene
219881	hydrochlorothiazide	2116	carteolol	218748	nitroglycerin
83818	irbesartan	724892	amlodipine	358263	tadalafil
151594	metirosine	750228	atorvastatin	152413	atenolol
542720	nifedipine	218749	nitroglycerin	218750	nitroglycerin
71512	phenoxybenzamine hydrochloride	23742	hydrochlorothiazide	219412	hydrochlorothiazide
217643	hydrochlorothiazide	151490	nicardipine	23742	triamterene
155033	labetalol	6958	amiloride	750208	atorvastatin

235230	nicardipine hydrochloride	218754	nitroglycerin	545347	nitroglycerin
629300	diltiazem	5034	guanabenz	217749	mecamylamine
215869	captopril	220669	reserpine	217643	reserpine
831206	diltiazem	754803	polythiazide	142130	acebutolol hydrochloride
583143	terazosin	218245	trandolapril	202595	diazoxide
152098	indapamide	823960	hydrochlorothiazide	215672	timolol
342280	eplerenone	541600	nifedipine	225702	minoxidil
809851	quinapril	151689	ethacrynate	219798	hydroflumethiazide
202991	furosemide	298869	eplerenone	151558	losartan
202594	diazoxide	7226	nadolol	8588	potassium
52440	hydrochlorothiazide	258337	hydrochlorothiazide / triamterene	202521	nitroglycerin
216221	carvedilol	219798	reserpine	7008	sotalol

					hydrochloride
142424	amiloride hydrochloride	38413	torseamide	754803	prazosin
724876	valsartan	37798	terazosin	258346	isradipine
214619	hydrochlorothia zide / losartan	203160	losartan potassium	202516	nitroglycerin
215458	candesartan	227278	fosinopril sodium	744887	benazepril
203423	nifedipine	220669	hydralazine	216677	chlorothiazide
93113	penbutolol	215869	hydrochlorothia zide	203490	verapamil
4917	nitroglycerin	214620	hydrochlorothia zide / methyldopa	118463	olmesartan medoxomil
202512	nitroglycerin	218416	amiloride	4603	furosemide
202978	pindolol	3829	enalaprilat	37925	orlistat
203138	verapamil	151195	atenolol /	284747	hydralazine

	hydrochloride		chlorthalidone		
7396	nicardipine	284747	hydrochlorothiazide	218936	hydrochlorothiazide
327503	olmesartan	688640	methyldopa	149	acebutolol
745974	bisoprolol	142132	carteolol hydrochloride	219878	hydrochlorothiazide
69749	valsartan	219734	minoxidil	227549	ethacrynate
153165	atorvastatin	214223	amlodipine / benazepril	21406	coenzyme Q10
7435	nisoldipine	203673	guanethidine	54552	perindopril
216251	bendroflumethiazide	214287	benazepril / hydrochlorothiazide	758548	hydrochlorothiazide
82027	hydralazine hydrochloride	203794	metolazone	542509	timolol
823960	losartan	220005	spironolactone	83515	eprosartan
823968	lisinopril	746962	timolol	54980	mibefradil

700402	aliskiren	215868	captopril	325646	aliskiren
217642	hydrochlorothiazide	10601	timolol	805852	hydrochlorothiazide
745969	bisoprolol	215871	captopril	220675	moexipril
224931	polythiazide	9631	acebutolol	724888	valsartan
49737	esmolol	203191	metoprolol tartrate	744621	amlodipine
6984	minoxidil	823939	irbesartan	6185	labetalol
405349	diltiazem	216252	bendroflumethiazide	40138	iloprost
809019	hydrochlorothiazide	218753	nitroprusside	744875	benazepril
219797	hydroflumethiazide	220212	atenolol	75207	bosentan
82084	propranolol hydrochloride	688640	hydrochlorothiazide	219881	reserpine
7476	nitroprusside	220081	nisoldipine	214536	enalapril / hydrochlorothiazide

					zide
220212	chlorthalidone	5764	indapamide	744625	olmesartan
218856	carteolol	151448	esmolol	215870	captopril
744629	olmesartan	353022	eprosartan	152440	labetalol
227749	trichlormethiazide	220778	verapamil	809851	hydrochlorothiazide
203778	phenoxybenzamine	218088	benazepril	1520	betaxolol
216652	valsartan	151131	nifedipine	754809	polythiazide
151318	methyldopa	750204	atorvastatin	751613	nebivolol
656315	chlorthalidone	750232	atorvastatin	221002	bisoprolol
220005	hydrochlorothiazide	17276	hydrochlorothiazide	73494	telmisartan
262418	ramipril	220309	diltiazem	26296	guanadrel
214622	hydrochlorothiazide / moexipril	541019	captopril	71974	torseamide

750208	amlodipine	215868	hydrochlorothiazide	71556	guanfacine
24853	fenoldopam	151435	metoprolol	823968	hydrochlorothiazide
744883	benazepril	214212	amiloride / hydrochlorothiazide	353022	hydrochlorothiazide
52440	triamterene	219868	atenolol	668310	carvedilol phosphate
750196	amlodipine	491311	nifedipine	151133	nifedipine
17276	spironolactone	224921	enalapril	30131	moexipril
402957	metolazone	261438	perindopril	303263	tadalafil
749838	hydrochlorothiazide	151317	spironolactone	151174	nitroglycerin
724888	amlodipine	744883	amlodipine	216909	deserpidine
303838	hydralazine	202693	labetalol hydrochloride	58927	amlodipine

152380	spironolactone	1998	captopril	18867	benazepril
1436	bepiridil	235758	benazepril hydrochloride	8153	phentolamine
35208	quinapril	218416	hydrochlorothiazide	285061	enalaprilat
203016	felodipine	203276	hydrochlorothiazide	7442	nitric oxide
202706	minoxidil	151877	chlorthalidone	991208	bendroflumethiazide
219882	hydrochlorothiazide	83213	mibefradil	52175	losartan
151178	nitroglycerin	72260	perindopril erbumine	203538	bumetanide
744633	olmesartan	2599	clonidine	3443	diltiazem
261415	fosinopril / hydrochlorothiazide	809027	hydrochlorothiazide	801793	coenzyme Q10
404773	amlodipine	219412	lisinopril		

/ atorvastatin

220391

hydrochlorothia

10772

zide

trichlormethiazi

de

Supplemental Table 4: Clinical Note Counts and Extracted Hypertension Regular Expression Matches and Concepts.

Our population had 43,262 notes with 28,264 PL, 14,072 HPC, and 926 DS. Problem lists were both the most numerous and the source resulting in the largest number of regular expression matches and concepts.

	All Document Types	Problem Lists	History & Physicals/Clinic Notes	Discharge Summaries
Note counts	43262	28264	14072	926
Max per individual	356	264	142	29
Min per individual	2	0	0	0
Median per individual (IQR)	47 (25-92)	31 (14-60.5)	15 (8-29)	0 (0-2)
hypertension regular expression match note counts	17392	11557	5489	346
Max per individual	265	203	88	10
Min per individual	0	0	0	0
Median per individual (IQR)	5 (1-38)	0 (0-24)	3 (0-12)	0 (0-0)
Concept counts	11695623	11441802	240452	13369
Max per individual	598893	594178	4585	412
Min per individual	5	0	0	0
Median per individual (IQR)	4356 (1086-16741)	4055 (947.5-16040)	230 (102-431.5)	0 (0-24.5)
Hypertension concept counts	401147	393363	7332	452
Max per individual	43056	42787	256	17
Min per individual	0	0	0	0
Median per individual (IQR)	17 (0-523.5)	3 (0-495.5)	1 (0-15)	0 (0-0)

IQR = Interquartile range

Supplemental Table 5: Full Set of UMLS Hypertension Concepts

CUI	CUI_PN
1303117	Moderate hypertension control (finding)
20540	Hypertension, Malignant [Disease/Finding]
421190	Poor hypertension control (finding)
235222	Diastolic hypertension (disorder)
155583	Benign essential hypertension (disorder)
421189	Good hypertension control (finding)
85580	Hypertension, Essential
20538	Hypertensive disorder, systemic arterial (disorder)
264637	Benign hypertension (disorder)
221155	systolic hypertension
1541808	Hypertension (& [essential])
455527	History of - hypertension (situation)

Supplemental Table 6: Feature Comparison Between Cases and Controls

Feature	Cases		Controls	
	Median	IQR	Median	IQR
date-ICD9 count	136.00	(75.00-271.50)	65.00	(33.75-139.00)
unique ICD9 count	51.00	(29.00-82.00)	32.50	(16.00-54.25)
hypertension ICD9 count	7.00	(1.00-18.00)	0.00	(0.00-0.00)
unique hypertension ICD9 count	2.00	(1.00-2.00)	0.00	(0.00-0.00)
hypertension ICD9 count normalized	0.05	(0.01-0.10)	0.00	(0.00-0.00)
hypertension ICD9 count unique normalized	0.13	(0.04-0.29)	0.00	(0.00-0.00)
unique hypertension ICD9 count normalized	0.01	(0.00-0.02)	0.00	(0.00-0.00)
unique hypertension ICD9 count unique normalized	0.02	(0.01-0.04)	0.00	(0.00-0.00)
meds count	462.00	(206.50-1,017.00)	135.00	(60.75-346.25)
unique meds count	84.00	(43.50-147.00)	41.50	(20.00-76.00)
hypertension meds count	61.00	(21.50-174.00)	0.00	(0.00-5.00)
unique hypertension meds count	7.00	(3.00-13.00)	0.00	(0.00-2.00)
hypertension meds count normalized	0.15	(0.08-0.22)	0.00	(0.00-0.02)
hypertension meds count unique normalized	0.84	(0.35-1.66)	0.00	(0.00-0.06)
unique hypertension meds count normalized	0.01	(0.01-0.03)	0.00	(0.00-0.01)
unique hypertension meds count unique normalized	0.09	(0.05-0.14)	0.00	(0.00-0.03)
max age	65.00	(56.00-75.00)	47.00	(37.00-60.00)
vital reading time span (days)	2412.00	(1,808.50-3,227.00)	2081.00	(1,209.75-2,824.25)
visits with vitals count (days)	30.00	(16.00-52.00)	17.00	(9.00-30.00)
hypertensive BP count (days)	9.00	(4.00-18.00)	1.00	(0.00-3.00)
median systolic (all)	131.00	(124.00-138.00)	120.00	(110.00-125.00)
median diastolic (all)	76.00	(70.00-80.00)	72.00	(68.00-78.00)
hypertensive BP count normalized	0.34	(0.19-0.51)	0.04	(0.00-0.20)
vitals density	0.01	(0.01-0.02)	0.01	(0.01-0.02)
outpatient visits with vitals count (days)	19.00	(9.00-29.50)	10.00	(5.00-18.00)
outpatient hypertensive BP count (days)	6.00	(2.00-10.50)	0.00	(0.00-2.00)
median systolic (outpatient)	131.00	(124.00-140.00)	120.00	(111.75-126.00)
median diastolic (outpatient)	76.00	(70.00-81.50)	72.25	(69.38-79.50)
outpatient hypertensive BP count normalized	0.34	(0.17-0.56)	0.00	(0.00-0.18)
outpatient vitals density	0.01	(0.00-0.01)	0.01	(0.00-0.01)
inpatient visits with vitals count (days)	11.00	(4.00-21.50)	6.00	(2.00-13.25)
inpatient hypertensive BP count (days)	3.00	(1.00-7.00)	0.00	(0.00-1.00)
median systolic (inpatient)	130.00	(121.00-138.00)	118.00	(108.00-125.00)
median diastolic (inpatient)	74.00	(67.00-80.00)	70.00	(65.38-76.00)
inpatient hypertensive BP count normalized	0.28	(0.06-0.50)	0.00	(0.00-0.14)
inpatient vitals density	0.004	(0.002-0.010)	0.003	(0.001-0.007)
median pulse (all)	75.00	(70.00-83.00)	76.50	(72.00-83.00)
median pulse (outpatient)	75.00	(69.00-82.50)	76.00	(71.00-81.13)
median pulse (inpatient)	76.00	(68.38-84.00)	76.00	(68.00-84.00)
PL count	41.00	(20.00-72.50)	22.00	(11.00-41.25)

DS count	1.00	(0.00-2.00)	0.00	(0.00-1.00)
HPC count	21.00	(11.00-34.00)	11.50	(6.00-18.00)
All doc count	62.00	(32.00-108.00)	35.00	(18.00-64.75)
PL hypertension RegEx doc matches	18.00	(1.00-42.50)	0.00	(0.00-0.00)
DS hypertension RegEx doc matches	0.00	(0.00-1.00)	0.00	(0.00-0.00)
HPC hypertension RegEx doc matches	9.00	(3.00-18.00)	0.00	(0.00-2.00)
All doc hypertension RegEx doc matches	29.00	(7.00-63.00)	0.00	(0.00-2.00)
PL hypertension RegEx doc matches doc normalized	0.67	(0.07-0.85)	0.00	(0.00-0.00)
DS hypertension RegEx doc matches doc normalized	0.00	(0.00-0.50)	0.00	(0.00-0.00)
HPC hypertension RegEx doc matches doc normalized	0.55	(0.20-0.78)	0.00	(0.00-0.16)
All doc hypertension RegEx doc matches doc normalized	0.62	(0.20-0.78)	0.00	(0.00-0.06)
PL concept count	7346.00	(2,099.00-23,406.00)	1526.50	(381.75-6,951.75)
DS concept count	4.00	(0.00-35.00)	0.00	(0.00-12.00)
HPC concept count	298.00	(156.00-591.00)	137.50	(61.75-273.75)
All doc concept count	7650.00	(2,459.00-24,003.50)	1709.50	(499.75-7,162.25)
PL hypertension concept count	325.00	(34.00-999.00)	0.00	(0.00-0.00)
DS hypertension concept count	0.00	(0.00-1.00)	0.00	(0.00-0.00)
HPC hypertension concept count	11.00	(2.00-25.50)	0.00	(0.00-0.00)
All doc hypertension concept count	335.00	(39.50-1,019.50)	0.00	(0.00-0.00)
PL hypertension concept count doc normalized	11.14	(1.78-18.44)	0.00	(0.00-0.00)
DS hypertension concept count doc normalized	0.00	(0.00-0.50)	0.00	(0.00-0.00)
HPC hypertension concept count doc normalized	0.60	(0.16-1.00)	0.00	(0.00-0.00)
All doc hypertension concept count doc normalized	7.10	(1.38-12.37)	0.00	(0.00-0.00)
PL hypertension concept count concept normalized	0.05	(0.01-0.09)	0.00	(0.00-0.00)
DS hypertension concept count concept normalized	0.00	(0.00-0.03)	0.00	(0.00-0.00)
HPC hypertension concept count concept normalized	0.04	(0.01-0.06)	0.00	(0.00-0.00)
All doc hypertension concept count concept normalized	0.05	(0.02-0.08)	0.00	(0.00-0.00)

Supplemental Table 7: Full Result Table Including AUC, Sensitivity, and PPV Across Individual Inputs, Simple Algorithms, and Random Forests.

Feature	Inputs	Binary Threshold (Present/Absent)					
		AUC		Sensitivity		PPV	
		Median	(CI)	Median	(CI)	Median	(CI)
date-ICD9 count		0.686	(0.656-0.766)	1.000	(0.991-1)	0.566	(0.535-0.638)
unique ICD9 count		0.643	(0.624-0.71)	0.991	(0.973-1)	0.570	(0.533-0.638)
hypertension ICD9 count		0.908	(0.902-0.911)	0.841	(0.823-0.85)	0.947	(0.918-0.979)
unique hypertension ICD9 count		0.898	(0.894-0.906)	0.841	(0.823-0.85)	0.947	(0.918-0.979)
hypertension ICD9 count normalized		0.907	(0.901-0.911)	0.841	(0.823-0.85)	0.947	(0.918-0.979)
hypertension ICD9 count unique normalized		0.909	(0.903-0.913)	0.841	(0.823-0.85)	0.947	(0.918-0.979)
unique hypertension ICD9 count normalized		0.890	(0.884-0.904)	0.841	(0.823-0.85)	0.947	(0.918-0.979)
unique hypertension ICD9 count unique normalized		0.894	(0.891-0.907)	0.841	(0.823-0.85)	0.947	(0.918-0.979)
meds count		0.745	(0.725-0.799)	0.992	(0.991-1)	0.566	(0.535-0.636)
unique meds count		0.699	(0.673-0.754)	0.992	(0.991-1)	0.566	(0.535-0.633)
hypertension meds count		0.907	(0.881-0.916)	0.944	(0.92-0.961)	0.759	(0.727-0.813)
unique hypertension meds count		0.887	(0.85-0.909)	0.944	(0.92-0.961)	0.759	(0.727-0.813)
hypertension meds count normalized		0.900	(0.876-0.918)	0.944	(0.92-0.961)	0.759	(0.727-0.813)
hypertension meds count unique normalized		0.910	(0.883-0.922)	0.944	(0.92-0.961)	0.759	(0.727-0.813)
unique hypertension meds count normalized		0.838	(0.783-0.853)	0.944	(0.92-0.961)	0.759	(0.727-0.813)
unique hypertension meds count unique normalized		0.890	(0.844-0.902)	0.944	(0.92-0.961)	0.759	(0.727-0.813)
max age		0.785	(0.761-0.811)	1.000	(1-1)	0.568	(0.538-0.638)
vital reading time span (days)		0.600	(0.567-0.648)	0.992	(0.991-1)	0.563	(0.533-0.633)
visits with vitals count (days)		0.666	(0.661-0.733)	1.000	(0.984-1)	0.568	(0.538-0.638)
hypertensive BP count (days)		0.854	(0.847-0.874)	0.953	(0.944-0.964)	0.701	(0.662-0.742)
median systolic (all)		0.802	(0.774-0.814)	1.000	(1-1)	0.571	(0.538-0.638)
median diastolic (all)		0.591	(0.544-0.62)	0.992	(0.981-1)	0.568	(0.53-0.633)
hypertensive BP count normalized		0.825	(0.809-0.85)	0.953	(0.944-0.964)	0.701	(0.662-0.742)
vitals density		0.621	(0.566-0.642)	1.000	(0.991-1)	0.563	(0.538-0.638)
outpatient visits with vitals count (days)		0.685	(0.665-	0.981	(0.965-	0.565	(0.536-

		0.744)		0.991)		0.629)
outpatient hypertensive BP count (days)	0.841	(0.823-0.859)	0.897	(0.89-0.909)	0.713	(0.691-0.748)
median systolic (outpatient)	0.775	(0.748-0.787)	0.981	(0.965-0.991)	0.565	(0.536-0.629)
median diastolic (outpatient)	0.578	(0.507-0.615)	0.981	(0.965-0.991)	0.565	(0.536-0.629)
outpatient hypertensive BP count normalized	0.798	(0.758-0.829)	0.897	(0.89-0.909)	0.713	(0.691-0.748)
outpatient vitals density	0.629	(0.6-0.66)	0.981	(0.965-0.991)	0.565	(0.536-0.629)
inpatient visits with vitals count (days)	0.624	(0.608-0.691)	0.953	(0.935-0.964)	0.563	(0.548-0.637)
inpatient hypertensive BP count (days)	0.763	(0.749-0.819)	0.752	(0.738-0.832)	0.759	(0.712-0.798)
median systolic (inpatient)	0.736	(0.721-0.764)	0.953	(0.935-0.964)	0.563	(0.548-0.637)
median diastolic (inpatient)	0.591	(0.561-0.597)	0.953	(0.935-0.964)	0.563	(0.548-0.637)
inpatient hypertensive BP count normalized	0.741	(0.734-0.799)	0.752	(0.738-0.832)	0.759	(0.712-0.798)
inpatient vitals density	0.593	(0.582-0.649)	0.953	(0.935-0.964)	0.563	(0.548-0.637)
median pulse (all)	0.435	(0.393-0.473)	0.992	(0.991-1)	0.563	(0.533-0.633)
median pulse (outpatient)	0.450	(0.439-0.498)	0.981	(0.965-0.991)	0.568	(0.538-0.633)
median pulse (inpatient)	0.486	(0.439-0.498)	0.944	(0.921-0.956)	0.571	(0.54-0.626)
PL count	0.670	(0.628-0.732)	0.992	(0.991-1)	0.566	(0.533-0.633)
DS count	0.611	(0.567-0.677)	0.558	(0.449-0.591)	0.697	(0.615-0.743)
HPC count	0.670	(0.616-0.728)	0.981	(0.965-0.984)	0.556	(0.533-0.635)
All doc count	0.674	(0.63-0.739)	0.992	(0.984-1)	0.571	(0.538-0.631)
PL hypertension RegEx doc matches	0.855	(0.842-0.868)	0.757	(0.72-0.787)	0.943	(0.92-0.952)
DS hypertension RegEx doc matches	0.676	(0.627-0.696)	0.389	(0.299-0.394)	0.952	(0.889-1)
HPC hypertension RegEx doc matches	0.844	(0.812-0.879)	0.900	(0.879-0.916)	0.731	(0.676-0.793)
All doc hypertension RegEx doc matches	0.896	(0.875-0.921)	0.950	(0.935-0.963)	0.725	(0.685-0.796)
PL hypertension RegEx doc matches doc normalized	0.852	(0.837-0.87)	0.757	(0.72-0.787)	0.943	(0.92-0.952)
DS hypertension RegEx doc matches doc normalized	0.677	(0.63-0.696)	0.389	(0.299-0.394)	0.952	(0.889-1)
HPC hypertension RegEx doc matches doc normalized	0.840	(0.8-0.861)	0.900	(0.879-0.916)	0.731	(0.676-0.793)
All doc hypertension RegEx doc matches doc normalized	0.903	(0.877-0.92)	0.950	(0.935-0.963)	0.725	(0.685-0.796)
PL concept count	0.710	(0.687-0.751)	1.000	(0.984-1)	0.568	(0.538-0.631)
DS concept count	0.611	(0.577-0.681)	0.527	(0.439-0.567)	0.697	(0.618-0.742)
HPC concept count	0.690	(0.665-	0.976	(0.965-1)	0.556	(0.531-

		0.763)				0.633)
All doc concept count	0.709	(0.692-0.755)	1.000	(0.992-1)	0.568	(0.538-0.633)
PL hypertension concept count	0.856	(0.851-0.868)	0.782	(0.752-0.795)	0.931	(0.919-0.944)
DS hypertension concept count	0.661	(0.614-0.682)	0.346	(0.271-0.364)	0.948	(0.879-1)
HPC hypertension concept count	0.883	(0.877-0.911)	0.841	(0.823-0.879)	0.891	(0.849-0.894)
All doc hypertension concept count	0.908	(0.89-0.933)	0.906	(0.897-0.944)	0.863	(0.828-0.885)
PL hypertension concept count doc normalized	0.860	(0.855-0.871)	0.782	(0.752-0.795)	0.931	(0.919-0.944)
DS hypertension concept count doc normalized	0.663	(0.617-0.682)	0.346	(0.271-0.364)	0.948	(0.879-1)
HPC hypertension concept count doc normalized	0.886	(0.881-0.914)	0.841	(0.823-0.879)	0.891	(0.849-0.894)
All doc hypertension concept count doc normalized	0.914	(0.901-0.938)	0.906	(0.897-0.944)	0.863	(0.828-0.885)
PL hypertension concept count concept normalized	0.869	(0.851-0.877)	0.782	(0.752-0.795)	0.931	(0.919-0.944)
DS hypertension concept count concept normalized	0.661	(0.617-0.682)	0.346	(0.271-0.364)	0.948	(0.879-1)
HPC hypertension concept count concept normalized	0.887	(0.879-0.914)	0.841	(0.823-0.879)	0.891	(0.849-0.894)
All doc hypertension concept count concept normalized	0.914	(0.909-0.942)	0.906	(0.897-0.944)	0.863	(0.828-0.885)
Sum of hypertension ICD9 and meds	0.92358	(0.894-0.95)	0.96748	(0.945-1)	0.77397	(0.718-0.83)
Sum of normalized hypertension ICD9 and meds	0.92961	(0.901-0.954)	0.96748	(0.945-1)	0.77397	(0.718-0.83)
Sum of unique normalized hypertension ICD9 and meds	0.92808	(0.901-0.951)	0.96748	(0.945-1)	0.77397	(0.718-0.83)
Sum of hypertension ICD9 and BP	0.9202	(0.885-0.947)	0.98214	(0.959-1)	0.70988	(0.644-0.767)
Sum of normalized hypertension ICD9 and BP	0.86057	(0.827-0.903)	0.98214	(0.959-1)	0.70988	(0.644-0.767)
Sum of unique normalized hypertension ICD9 and BP	0.8961	(0.867-0.93)	0.98214	(0.959-1)	0.70988	(0.644-0.767)
Sum of hypertension meds and BP	0.92954	(0.895-0.956)	1	(1-1)	0.67251	(0.608-0.731)
Sum of normalized hypertension meds and BP	0.89672	(0.866-0.929)	1	(1-1)	0.67251	(0.608-0.731)
Sum of unique normalized hypertension meds and BP	0.93979	(0.913-0.967)	1	(1-1)	0.67251	(0.608-0.731)
Sum of hypertension ICD9, meds, and BP	0.93547	(0.903-0.961)	1	(1-1)	0.67251	(0.608-0.731)
Sum of normalized hypertension ICD9, meds, and BP	0.91513	(0.888-0.942)	1	(1-1)	0.67251	(0.608-0.731)
Sum of unique normalized hypertension ICD9, meds, and BP	0.94757	(0.923-0.971)	1	(1-1)	0.67251	(0.608-0.731)
Sum of hypertension ICD9, meds, BP, and RegEx	0.940	(0.906-0.961)	1.000	(1-1)	0.631	(0.574-0.699)
Sum of hypertension ICD9, meds, BP, and RegEx normalized	0.940	(0.912-0.96)	1.000	(1-1)	0.631	(0.574-0.699)
Sum of normalized hypertension ICD9, meds, BP, and RegEx doc normalized	0.953	(0.932-0.969)	1.000	(1-1)	0.631	(0.574-0.699)
Sum of unique normalized hypertension ICD9, meds, BP, and RegEx	0.959	(0.935-	1.000	(1-1)	0.631	(0.574-

doc normalized		0.977)				0.699)
		(0.91-				(0.609-
Sum of hypertension ICD9, meds, BP, and concepts	0.936	0.961)	1.000	(1-1)	0.663	0.734)
		(0.912-				(0.609-
Sum of hypertension ICD9, meds, BP, and concepts normalized	0.941	0.964)	1.000	(1-1)	0.663	0.734)
		(0.911-				(0.609-
Sum of hypertension ICD9, meds, BP, and concepts doc normalized	0.940	0.965)	1.000	(1-1)	0.663	0.734)
		(0.897-				(0.609-
Sum of normalized hypertension ICD9, meds, BP, and concepts concept normalized	0.929	0.955)	1.000	(1-1)	0.663	0.734)
		(0.92-				(0.609-
Sum of unique normalized hypertension ICD9, meds, BP, and concepts doc normalized	0.949	0.971)	1.000	(1-1)	0.663	0.734)
		(0.925-				(0.609-
Sum of unique normalized hypertension ICD9, meds, BP, and concepts concept normalized	0.953	0.977)	1.000	(1-1)	0.663	0.734)
		(0.893-				(0.718-
Count of hypertensive ICD9 and med	0.92323	0.951)	0.96748	(0.945-1)	0.77397	0.83)
		(0.742-				(0.718-
Count of hypertensive ICD9 and med 1 of 2	0.78711	0.821)	0.96748	(0.945-1)	0.77397	0.83)
		(0.85-				(0.923-
Count of hypertensive ICD9 and med 2 of 2	0.88169	0.921)	0.82143	0.882)	0.95699	0.989)
		(0.881-				(0.644-
Count of hypertensive ICD9 and BP	0.90879	0.939)	0.98214	(0.959-1)	0.70988	0.767)
		(0.659-				(0.644-
Count of hypertensive ICD9 and BP 1 of 2	0.70373	0.74)	0.98214	(0.959-1)	0.70988	0.767)
		(0.851-				(0.915-
Count of hypertensive ICD9 and BP 2 of 2	0.87786	0.912)	0.81452	0.871)	0.95745	0.981)
		(0.805-				(0.608-
Count of hypertensive med and BP	0.83929	0.878)	1	(1-1)	0.67251	0.731)
		(0.62-				(0.608-
Count of hypertensive med and BP 1 of 2	0.65789	0.69)	1	(1-1)	0.67251	0.731)
		(0.785-				(0.779-
Count of hypertensive med and BP 2 of 2	0.82473	0.87)	0.91228	0.948)	0.83206	0.879)
		(0.913-				(0.608-
Count of hypertensive ICD9, med, and BP	0.93462	0.962)	1	(1-1)	0.67251	0.731)
		(0.62-				(0.608-
Count of hypertensive ICD9, med, and BP 1 of 3	0.65789	0.69)	1	(1-1)	0.67251	0.731)
		(0.788-				(0.77-
Count of hypertensive ICD9, med, and BP 2 of 3	0.83275	0.868)	0.95161	0.983)	0.82222	0.872)
		(0.849-				(0.937-
Count of hypertensive ICD9, med, and BP 3 of 3	0.87691	0.914)	0.79832	0.86)	0.96739	0.99)
		(0.905-				(0.574-
Count of hypertensive ICD9, med, BP, and RegEx	0.944	0.965)	1.000	(1-1)	0.631	0.699)
		(0.557-				(0.574-
Count of hypertensive ICD9, med, BP, and RegEx 1 of 4	0.588	0.613)	1.000	(1-1)	0.631	0.699)
		(0.723-				(0.703-
Count of hypertensive ICD9, med, BP, and RegEx 2 of 4	0.764	0.81)	0.983	(0.965-1)	0.750	0.819)
		(0.824-				(0.828-
Count of hypertensive ICD9, med, BP, and RegEx 3 of 4	0.874	0.908)	0.927	0.958)	0.878	0.925)
		(0.836-				(0.938-1)
Count of hypertensive ICD9, med, BP, and RegEx 4 of 4	0.879	0.914)	0.795	0.856)	0.969	
		(0.911-				(0.609-
Count of hypertensive ICD9, med, BP, and concept	0.951	0.971)	1.000	(1-1)	0.663	0.734)
		(0.602-				(0.609-
Count of hypertensive ICD9, med, BP, and concept 1 of 4	0.640	0.686)	1.000	(1-1)	0.663	0.734)
		(0.751-				(0.732-
Count of hypertensive ICD9, med, BP, and concept 2 of 4	0.802	0.854)	0.974	0.992)	0.784	0.862)
		(0.868-				(0.882-
Count of hypertensive ICD9, med, BP, and concept 3 of 4	0.910	0.936)	0.925	0.958)	0.924	0.966)
		(0.832-				(0.937-1)
Count of hypertensive ICD9, med, BP, and concept 4 of 4	0.875		0.785	(0.71-	0.969	

		0.906)		0.846)		
ICD9 category (random forest)	0.899	(0.864-0.927)	0.984	(0.953-1)	0.615	(0.553-0.683)
meds category (random forest)	0.888	(0.848-0.924)	0.982	(0.951-1)	0.627	(0.553-0.693)
vitals category (random forest)	0.865	(0.823-0.917)	1.000	(1-1)	0.588	(0.538-0.638)
RegEx category (random forest)	0.900	(0.861-0.933)	1.000	(0.982-1)	0.622	(0.558-0.689)
concept category (random forest)	0.928	(0.895-0.953)	1.000	(0.983-1)	0.594	(0.536-0.653)
ICD9, meds, all BP (random forest)	0.955	(0.934-0.975)	1.000	(0.992-1)	0.624	(0.55-0.692)
ICD9, meds, vitals (all, outpatient, inpatient) (random forest)	0.961	(0.938-0.98)	1.000	(1-1)	0.603	(0.54-0.667)
ICD9, meds, vitals (all, outpatient, inpatient), RegEx (random forest)	0.967	(0.948-0.985)	1.000	(1-1)	0.601	(0.543-0.655)
ICD9, meds, vitals (all, outpatient, inpatient), concept (random forest)	0.976	(0.95-0.984)	1.000	(1-1)	0.579	(0.523-0.649)
ICD9, meds, vitals (all, outpatient, inpatient), RegEx, concepts (random forest)	0.968	(0.951-0.985)	1.000	(1-1)	0.596	(0.544-0.67)

Supplemental Table 8: Full Set of Comparisons Between Mean and Median of Correctly and Incorrectly Classified

Individuals.

This table contains the absolute average difference (Avg. Diff) and median difference (Med. Diff.) between true negatives and false negatives (TN-FN), false positives and true positives (FP-TP), false negatives and true positives (FN-TP), and true negatives and false positives (TN-FP). The feature identifiers are along the left side and large difference values highlight features that are very different between the sets compared. All cells where at least one of the items was zero are listed as 'N/A'.

Category Measures	TN-FN		FP-TP		FN-TP		TN-FP	
	Avg. Diff	Med. Diff.	Avg. Diff	Med. Diff.	Avg. Diff	Med. Diff.	Avg. Diff	Med. Diff.
date-ICD9 count	1.6	1.8	415.5	1.0	29.6	1.0	22.6	3.7
unique ICD9 count	1.6	1.3	149.9	1.0	8.6	1.0	11.0	4.6
hypertension ICD9 count	N/A	N/A	2.9	1.0	N/A	N/A	N/A	N/A
unique hypertension ICD9 count	N/A	N/A	15.9	1.0	N/A	N/A	N/A	N/A
hypertension ICD9 count normalized	N/A	N/A	695.9	1.0	N/A	N/A	N/A	N/A
hypertension ICD9 count unique normalized	N/A	N/A	230.6	1.0	N/A	N/A	N/A	N/A
unique hypertension ICD9 count normalized	N/A	N/A	7694.8	1.0	N/A	N/A	N/A	N/A
unique hypertension ICD9 count unique normalized	N/A	N/A	2307.3	1.0	N/A	N/A	N/A	N/A
meds count	1.3	11.9	1460.3	1.0	33.0	1.0	57.8	1.8
unique meds count	1.9	2.5	182.1	1.0	8.4	1.0	11.6	2.2
hypertension meds count	4.7	N/A	44.7	1.0	2.0	1.0	102.2	N/A
unique hypertension meds count	4.0	N/A	5.5	1.0	1.0	1.0	22.0	N/A
hypertension meds count normalized	3.6	N/A	45.0	1.0	22.2	1.0	1.8	N/A
hypertension meds count unique normalized	8.7	N/A	4.1	1.0	4.2	1.0	8.8	N/A
unique hypertension meds count normalized	3.1	N/A	744.5	1.0	92.5	1.0	2.6	N/A
unique hypertension meds count unique normalized	7.4	N/A	65.0	1.0	16.6	1.0	1.9	N/A
max age (by last vital)	1.3	1.0	43.5	1.0	59.4	1.0	1.0	1.3
vital reading time span (days)	1.4	1.4	4015.4	1.0	2861.7	1.0	2.0	1.1
visits with vitals count (days)	1.8	1.3	69.0	1.0	8.4	1.0	14.4	4.8
hypertensive BP count (days)	3.0	5.0	5.1	1.0	5.1	1.0	8.7	1.5
median systolic (all)	1.0	1.1	113.3	1.0	120.2	1.0	1.1	1.1
median diastolic (all)	1.2	1.0	70.1	1.0	65.3	1.0	1.1	1.0
hypertensive BP count normalized	5.3	4.0	14.6	1.0	46.2	1.0	1.7	3.2
vitals density	1.2	1.1	68.5	1.0	401.2	1.0	7.2	5.5
outpatient visits with vitals count (days)	2.0	1.2	50.1	1.0	7.7	1.0	13.0	3.8
outpatient hypertensive BP count (days)	2.0	3.6	4.0	1.0	4.5	1.0	9.0	1.3

median systolic (outpatient)	1.0	1.1	115.2	1.0	121.8	1.0	1.1	1.1
median diastolic (outpatient)	1.1	1.0	71.5	1.0	67.6	1.0	1.1	1.0
outpatient hypertensive BP count normalized	4.0	3.0	12.3	1.0	34.0	1.0	1.4	3.0
outpatient vitals density	1.4	1.2	55.2	1.0	255.9	1.0	6.5	4.3
inpatient visits with vitals count (days)	1.0	1.4	20.5	1.0	1.1	1.0	18.5	14.0
inpatient hypertensive BP count (days)	N/A	12.0	1.3	N/A	N/A	1.0	8.0	N/A
median systolic (inpatient)	1.3	1.1	106.4	1.0	107.7	1.0	1.3	1.1
median diastolic (inpatient)	1.5	1.0	66.4	1.0	58.4	1.0	1.3	1.0
inpatient hypertensive BP count normalized	N/A	8.8	16.2	N/A	N/A	1.0	2.3	N/A
inpatient vitals density	1.4	1.0	314.1	1.0	4141.0	1.0	9.3	16.0
median pulse (all)	1.0	1.1	69.1	1.0	72.1	1.0	1.1	1.1
median pulse (outpatient)	1.0	1.1	67.7	1.0	72.7	1.0	1.1	1.1
median pulse (inpatient)	1.0	1.0	67.8	N/A	73.7	1.0	1.1	N/A
PL count	1.8	2.0	139.5	1.0	15.7	1.0	15.5	2.8
DS count	N/A	2.0	3.7	N/A	N/A	1.0	N/A	N/A
HPC count	4.7	1.8	42.3	1.0	9.4	1.0	21.0	1.9
PL hypertension RegEx document matches	N/A	N/A	5.4	N/A	N/A	1.0	25.1	N/A
DS hypertension RegEx document matches	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HPC hypertension RegEx document matches	N/A	N/A	8.9	1.0	N/A	1.0	18.0	N/A
PL hypertension regular expression document matches document normalized	N/A	N/A	10.9	N/A	N/A	1.0	1.6	N/A
DS hypertension regular expression document matches document normalized	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HPC hypertension regular expression document matches document normalized	N/A	N/A	3.5	1.0	N/A	1.0	1.2	N/A
All document count	2.3	1.9	183.7	1.0	25.1	1.0	17.1	2.6
All doc hypertension regular expression doc matches document normalized	N/A	N/A	4.1	1.0	N/A	1.0	1.4	N/A
All doc hypertension regular expression document matches	N/A	N/A	23.0	1.0	N/A	1.0	23.6	N/A
PL concept count	4.8	1.4	26185.6	1.0	274.4	1.0	456.1	9.6
PL hypertension concept count	N/A	N/A	97.7	1.0	N/A	N/A	105.6	N/A
DS concept count	N/A	N/A	16.3	N/A	N/A	N/A	N/A	N/A
DS hypertension concept count	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HP concept count	5.1	1.4	580.7	1.0	112.1	1.0	26.4	3.5
HP hypertension concept count	1.0	N/A	1.3	1.0	14.4	1.0	19.0	N/A
All document concept count	4.8	1.4	26937.0	1.0	331.7	1.0	392.0	9.2
All document hypertension concept count	24.0	N/A	99.0	1.0	24.7	1.0	102.0	N/A
PL hypertension concept count document normalized	N/A	N/A	1.7	1.0	N/A	N/A	6.8	N/A
DS hypertension concept count document normalized	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HPC hypertension concept count document normalized	4.7	N/A	10.3	1.0	43.3	1.0	1.1	N/A
All doc hypertension concept count document normalized	56.0	N/A	1.3	1.0	254.5	1.0	6.0	N/A
PL hypertension concept count concept normalized	N/A	N/A	84.2	1.0	N/A	N/A	4.3	N/A

DS hypertension concept count concept normalized	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HPC hypertension concept count concept normalized	5.1	N/A	100.2	1.0	367.4	1.0	1.4	N/A	N/A
All doc hypertension concept count concept normalized	115.9	N/A	82.1	1.0	2474.0	1.0	3.8	N/A	N/A