Online Appendix

Sample terms and CDC codes by descending order of cosine distance from the term "hypertension"

Term	Number of Unique Cases	Mean Per Patient	Max Per Patient	Cosine Distance
hypertension	342	20.11	98	1
hyperlipidemia	247	12.17	60	0.7693
disease	367	20.07	196	0.6308
benign	276	11.56	100	0.6136
496	29	3.31	27	0.5982
essential	195	11.32	59	0.581
stable	305	8.26	79	0.5775
sympt	5	1.4	3	0.5641
chronic	310	12.88	122	0.5639
hypercholesterolemia	115	10.28	49	0.5535
gerd	91	8.08	44	0.5363
27800	33	3.15	11	0.5328
necnos	24	4.42	26	0.5295
ckd	66	4.58	24	0.5255
disorders	37	3.68	13	0.5116
atherosclerosis	61	4.62	24	0.4919
4011	122	6.39	45	0.4908
Insuff	7	1.14	2	0.4879
controlled	243	5.95	48	0.4835
mellitus	143	9.53	46	0.4806
nephropathy	10	7.2	21	0.4804
5854	7	3.43	11	0.4785
hypothyroidism	80	15.47	55	0.4779
60000	14	4.36	26	0.4735
5852	8	3.12	9	0.4732
pneumo	10	1.1	2	0.4727
obesity	93	7.33	40	0.4699
41400	31	3.48	12	0.4625
sciatica	15	4.33	12	0.46
sment	4	3.5	5	0.46
2722	59	2.98	7	0.4595
assessment	260	9.01	62	0.4581
53081	47	4.53	28	0.4573
icd4359	7	5.86	16	0.4562

Sample terms and CDC codes by descending order of cosine distance from the term "cardiac"

Term	Number of Unique Cases	Mean Per Patient	Max Per Patient	Cosine Distance
cardiac	246	12.36	170	1
om1	10	8.1	48	0.7863
lad	63	12.03	196	0.7358
svg	12	14.08	86	0.7141
multivessel	9	3.56	10	0.681
biventricular	20	15.05	74	0.6807
22513	4	7.5	26	0.6621
acs	16	2.31	20	0.6536
chronically	44	2.2	14	0.6415
rca	30	17.87	206	0.6124

ntg	6	3.33	6	0.6109
instent	5	4.8	8	0.6108
lcx	8	9.62	23	0.6105
32813	4	1.75	4	0.6055
cardiacwise	5	4	8	0.6004
catheterization	86	8.17	84	0.5984
stent	53	11.89	73	0.5982
stenting	19	5.58	20	0.593
stenoses	12	5.58	32	0.5902
leads	49	2.67	20	0.5863
cath	93	8.56	92	0.5778
xience	7	1.71	3	0.5753
circumflex	27	7.96	32	0.5665
coronary	200	15.11	190	0.5562
mildmod	5	2.2	3	0.5497
stents	17	11.29	92	0.5495
biv	6	10	25	0.5463
11868	4	1.5	2	0.5452
ptca	13	11.23	88	0.5446
hypo	14	4.14	32	0.5414
lvedp	4	3.5	10	0.5358
occlusion	31	5.81	44	0.5305