

Indication Crosswalk

Ben Berger

June 5, 2017

Goal:

Match descriptive condition categories to dataset of drug approvals.

Issues:

Cortellis indication categories are too fine, as are 3-digit ICD-9 groupings, and it is not obvious how we should proceed in aggregating them into coarser groupings.

What we have:

1. Dataset of drug approvals with Cortellis Indications.

brandname	condition
Nexterone	ventricular tachycardia
Cotellic	stage iv melanoma
Opana IR	pain
Myozyme	pompes disease
Vectibix	colorectal tumor
Elaprase	hunter syndrome

2. Crosswalk of Cortellis Indications -> 3-digit ICD-9 codes.

icd9	condition
46	measles virus infection
706	acne
42	acquired immune deficiency syndrome
253	acromegaly
255	addisons disease
79	adenovirus infection

Sorted by ICD-9 code....

icd9	condition
1	vibrio cholerae infection
2	salmonella typhi infection
3	salmonella infection
4	shigella infection
4	shigella flexneri infection
4	shigella sonnei infection

3. ICD-9 to HCC or CCS crosswalks.

- ICD-9 -> HCC crosswalk from NBER

The first three digits of each ICD-9 correspond to the three-digit groupings used in the Indication -> ICD-9 crosswalk. Note these 3-digit ICD-9 groupings **do not** in general map to a single HCC. Also note that only HCCs that are included in CMS's risk-adjustment model in a particular year are present.

icd	icddesc	hcc
0031	Salmonella septicemia	2
00322	Salmonella pneumonia	115
00323	Salmonella arthritis	39
00324	Salmonella osteomyelitis	39
0064	Amebic lung abscess	115
0074	Cryptosporidiosis	6

- ICD-9 -> CCS crosswalk from NBER

Unlike the HCC crosswalk, the CCS crosswalk is year-invariant (ICD-9 codes map to the same HCC regardless of year) and includes a much larger set of ICD-9 codes. Three-digit ICD-9 groupings **do not** in general map to a single CCS. CCS codes, category labels, and associated ICD-9s are listed *here*. CCS categories tend to be granular enough to distinguish between related conditions while still aggregating into useful categories (e.g. all leukemias are coded CCS 39).

icd	ccs	ccsdesc	icddesc
0010	135	Intest infct CHOLERA D/T VIB	CHOLERA
0011	135	Intest infct CHOLERA D/T VIB	EL TOR
0019	135	Intest infct	CHOLERA NOS
0020	135	Intest infct	TYPHOID FEVER
0021	135	Intest infct	PARATYPHOID FEVER A
0022	135	Intest infct	PARATYPHOID FEVER B
0023	135	Intest infct	PARATYPHOID FEVER C
0029	135	Intest infct	PARATYPHOID FEVER NOS
0030	135	Intest infct	SALMONELLA ENTERITIS
0031	2	Septicemia	SALMONELLA SEPTICEMIA

Summary Statistics:

There are 1941 unique Cortellis indications in the Indication -> 3-digit ICD-9 Crosswalk.

Of these, 307 indications are represented in the drug approval data.

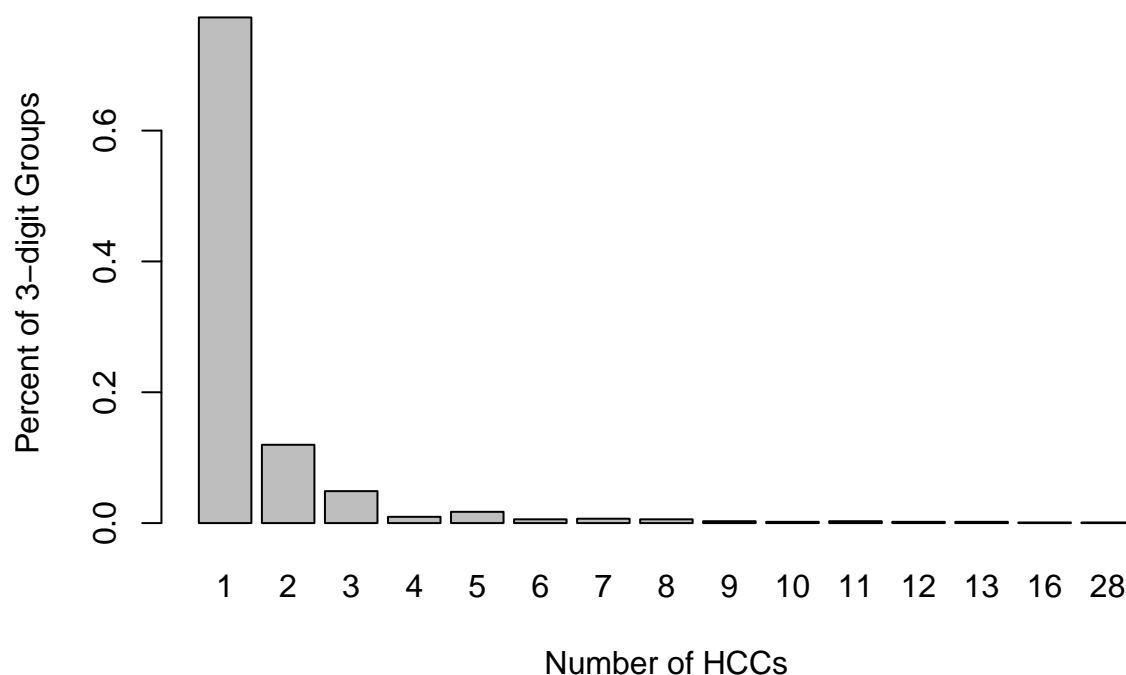
76 drug approvals map to a unique 3-digit ICD-9 code.

10 breakthrough drug approvals map to a unique 3-digit ICD-9 code.

On average, a 3-digit ICD-9 group contains 3.9 drugs, with a median of 2 .

On average, a 3-digit ICD-9 group corresponds to 1.58 unique CCS codes.

Number of Associated HCCs for 3-digit ICD-9 Groupings



Recommendation:

I recommend we aggregate indications into CCS codes. Although there are over 1,900 Cortellis indications in the Indication -> 3-digit ICD-9 crosswalk, only around 300 unique indications are present in the drug approval data.

We can either do this by

1. Looking up the entire 4-or-5-digit ICD-9 code for the indication and mapping that to a CCS code using a crosswalk, or
2. Looking up the CCS code alone.

I am happy to work on this and don't expect it to take ridiculously long.