

# Indication Crosswalk

*Ben Berger*

*June 16, 2017*

## Goal:

Match descriptive condition categories to dataset of drug approvals with associated Cortellis indications.

## Changes:

Andrew has provided us with a crosswalk of Cortellis indications to several different condition coding standards including ICD-9, ICD-10, MEDDRA, and others.

## Issues:

### Crosswalk Preview:

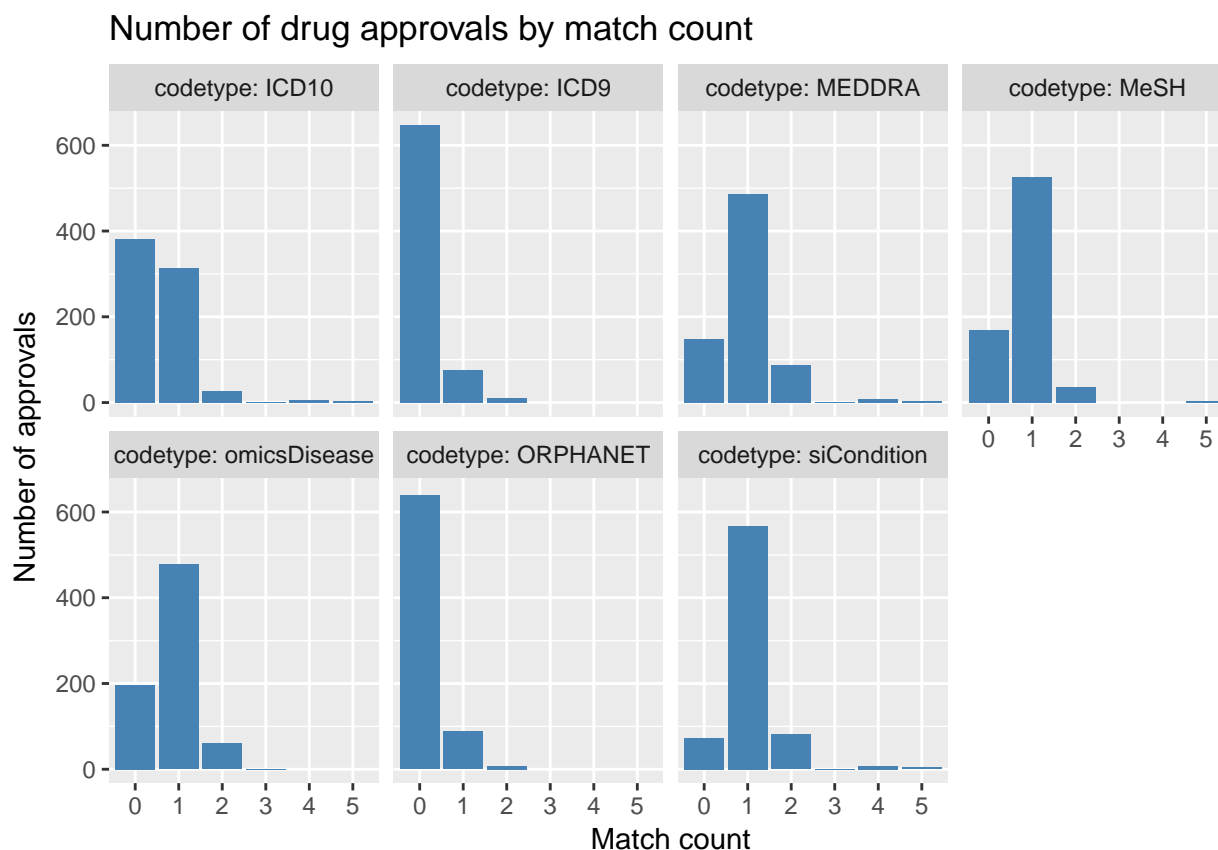
indicationID	ICD10	ICD9	MEDDRA	MeSH	ORPHANET	omicsDisease	siCondition
1	B05	55	10027011	D008457	NA	-145534850	1716
3	L70	706.1	10000496	NA	NA	NA	443
4	NA	NA	10000565	D000163	NA	-956122641	1124
5	NA	NA	10000599	D000172	963	-945923203	525
6	NA	NA	10001130	D000224	85138	-113788509	1264
7	NA	NA	10060931	D000258	NA	-251245854	1704
8	NA	NA	10001052	NA	NA	NA	1146
8	NA	NA	10038687	NA	NA	NA	1146
9	R45.5	NA	10020400	NA	NA	-569686292	62
9	R45.5	NA	10001488	D000374	NA	-569686292	62

1. Many ICD-9 and ICD-10 codes are missing.
  - Although not all of the indication IDs are used in the drug approval dataset, few approvals match to one or more ICD-9 codes. ICD-10 coverage is better (about 300 approvals are matched on).
2. Coverage on MEDDRA, MeSH and siCondition codes is good, albeit not perfect, but it is unclear if these are useful from a practical standpoint.
3. Andrew's crosswalk links some Cortellis indications to multiple codings of the same variable. See indication IDs 8 and 9 above.

## Drug Approval Matching:

I first match each drug approval to the coding standards in the crosswalk. All drug approvals match to at least one indication ID in the crosswalk, but, as shown above, cells in the row may be blank. Then, because some approvals match to multiple entries in the crosswalk, I count the number of unique codes for each coding standard and each drug. For example, Alvimopan matches to zero ICD-9s, zero ICD-10s, two MEDDRA codes, etc. For clarity, I'll refer to this as match count.

Below, for each coding standard, I plot the number of drug approvals by match count. Ideally, all approvals would match to exactly one code (match count = 1), but many match to zero or 2+ (several codings for one indication). Additionally, the plot is replicated at the end of the document for only the set of breakthrough drug approvals.



ICD-9 matching is particularly bad. ICD-10 matching is quite a bit more successful, but still fails to match one-to-one on even half the drug approvals. MEDDRA and siCondition are the most successful, but unlike ICD-9 and ICD-10 it is unclear how we could group these variables into meaningful categories. Deciding how to best employ these coding standards will be critical moving forward.

Number of breakthrough drug approvals by match count

