

FORMAT DESCRIPTION

=====

drug_names.tsv and drug_atc.tsv contain the automatically generated drug names, and their ATC code. You might be able to retrieve better names via the ATC codes.

meddra_all_se.tsv.gz

1 & 2: STITCH compound ids (flat/stereo, see above)

3: UMLS concept id as it was found on the label

4: MedDRA concept type (LLT = lowest level term, PT = preferred term; in a few cases the term is neither LLT nor PT)

5: UMLS concept id for MedDRA term

6: side effect name

All side effects found on the labels are given as LLT. Additionally, the PT is shown. There is at least one

PT for every LLT, but sometimes the PT is the same as the LLT. LLTs are sometimes too detailed, and therefore

you might want to filter for PT. E.g. for this term:

PT	C0235431	Blood creatinine increased
----	----------	----------------------------

there are several LLT (leftmost number = count in SIDER 4.1)

149	C0151578	LLT	C0151578	Creatinine increased
100	C0235431	LLT	C0235431	Blood creatinine increased
93	C0700225	LLT	C0700225	Serum creatinine increased
2	C0858118	LLT	C0858118	Plasma creatinine increased

All of these LLTs are equivalent for most purposes and to the same PT.

344	PT	C0235431	Blood creatinine increased
-----	----	----------	----------------------------

The mapping was performed by extracting the LLT-->PT relations from UMLS.

meddra_freq.tsv.gz

This file contains the frequencies of side effects as extracted from the labels. Format:

1 & 2: STITCH compound ids (flat/stereo, see above)

3: UMLS concept id as it was found on the label

4: "placebo" if the info comes from placebo administration, "" otherwise

5: a description of the frequency: for example "postmarketing", "rare", "infrequent", "frequent", "common", or an exact percentage

6: a lower bound on the frequency

7: an upper bound on the frequency

8-10: MedDRA information as for meddra_all_se.tsv.gz

The bounds are ranges like 0.01 to 1 for "frequent". If the exact frequency is known, then the lower bound

matches the upper bound. Due to the nature of the data, there can be more than one frequency for the same label,

e.g. from different clinical trials or for different levels of severeness.

meddra_all_indications.tsv.gz

- 1: STITCH compound id (flat, see above)
- 2: UMLS concept id as it was found on the label
- 3: method of detection: NLP_indication / NLP_precondition / text_mention
- 4: concept name
- 5: MedDRA concept type (LLT = lowest level term, PT = preferred term; in a few cases the term is neither LLT nor PT)
- 6: UMLS concept id for MedDRA term
- 7: MedDRA concept name

All side effects found on the labels are given as LLT. Additionally, the PT is shown. There is at least one

PT for every LLT, but sometimes the PT is the same as the LLT.

meddra_all_label_indications.tsv.gz and meddra_all_label_se.tsv.gz

These files contain the same data as the indications/se files, but with an additional first column showing the source label.

meddra.tsv

-
- 1: UMLS concept id
 - 2: MedDRA id
 - 3: kind of term (from MedDRA e.g. PT = preferred term)
 - 4: name of side effect