FDA BI Research Weekly Report

Yage Wang, Zheng Gong 20170802

1. **Include ‘anaplastic’ and ‘lymphoma’ to ALCL word list to find ALCL records**

Please see file:

1\_Total\_ALCL\_Report\_Number\_List (union of file 2 and 3).txt

2\_ALCL\_Report\_Number\_List\_Keyword[anaplastic].txt

3\_ALCL\_Report\_Number\_List\_Keyword[lymphoma].txt

4\_Total\_ALCL\_Intersect\_with\_BI\_Records.txt

We use ‘anaplastic’ and ‘lymphoma’ as keyword to search through all MDR text data.

Method: Use text mining to identify

Data source: MAUDE database [foitext.zip, foitextadd.zip, foitextchange.zip, foitextthru1995.zip, foitext(1996-2016).zip] (25 files in total), use column “FOI\_TEXT”

Word List: [‘anaplastic’], [‘lymphoma’] (separated)

If one record contains at least one word in Word List, it would be considered as ALCL related record.

For keyword ‘anaplastic’, 619 records are found with 380 distinct report keys. 361 records have matched report number with 249 distinct report number.

For keyword ‘lymphoma’, 1648 records are found with 1059 distinct report keys. 1168 records have matched report number with 819 distinct report number.

The union of two keywords result has 827 distinct report number. Among these, 411 report numbers are included in BI records.

According to previous ALCA text mining result, we got 106 distinct report number. The intersection of previous result and result from this week is 412 report numbers.

**Conjecture from research:**

* 1. Most of mdr reports don’t use the keyword ‘ALCL’ in text to describe the adverse event.
  2. Keyword ‘anaplastic’ seems to have stronger relation with ALCL than ‘lymphoma’.

1. **Go through duplicated data and give summary of the duplicated type**

We focus on duplicated data in foitext files, which means for a specific report key, there might be several records with different text content.

For foitext2016, it has 2,280,753 records with 965,211 distinct report number. 57.68% of records are duplicated.

NOTE: when we are talking “duplicated”, it means there are several records share a same report number but these records could be different.

**Possible duplicate type:**

* 1. The mdr report is imported from some other resources and when imported the text content is divided into several pieces by specific format or character with the same report key.
  2. When user report their adverse event on FDA website, there is a limitation to text content length. So, a single record may be divided to several records to store.
  3. When user return to FDA website to give a follow-up report, they may use their previous record number to add new report. The system will store it as a new record but with same record number.

NOTE: For type c, there is a column called MDR\_TEXT\_KEY to identify different report with same record number.

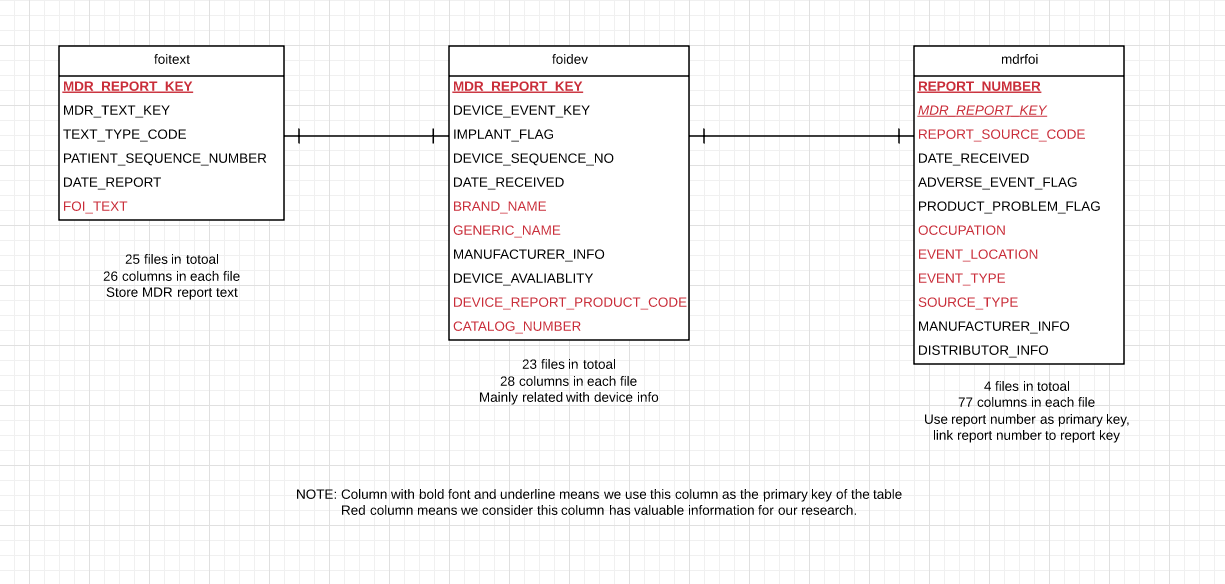
1. **Full value list of following table columns.**

Please see file:

5\_foidev\_column\_full\_list.xlsx

6\_mdrfoi\_column\_full\_list.xlsx

1. **Give a diagram of MAUDE database tables showing**



NOTE: MDR\_REPORT\_KEY is identical in these three kinds of tables. But it doesn’t mean MDR\_REPORT\_KEY is foreign key for each table. Some report key may only appear in one table but missing in other tables.