FDA BI Research Weekly Report

Yage Wang, Zheng Gong 20170809

1. **What can we learn from sample list which the current method doesn’t catch.**

The sample list which Asiyah shared us has 100 report numbers that the current method doesn’t catch. We compare the sample list with all mdrfoi tables.

Among 100 report numbers, only 4 could be matched in mdrfoi tables. Considering all our MAUDE files are updated on 8/9/2017 to keep latest records, we conjecture that **for those report numbers that our method doesn’t catch, it’s highly likely that these report numbers are not included in mdrfoi tables.** In other word, MAUDE database may be just a subset of FDA internal database.

1. **What can we learn from 4 matched report number?**

|  |  |  |
| --- | --- | --- |
| MDR\_REPORT\_KEY | REPORT\_NUMBER | DATE\_RECEIVED |
| 6271188 | 9617229-2017-00071 | 1/23/2017 |
| 6271147 | 9617229-2017-00070 | 1/23/2017 |
| 10426 | MW1000048 | 12/9/1993 |
| 222164 | 1645337-1999-00099 | 5/4/1999 |

* 1. According to previous discussion, we set the time filter as 01/01/1994 – 02/01/2017, so it’s possible that some records before 1994 are filtered.
  2. Since MAUDE database is continuously updating, some records after year 2017 will be added.

Based on above two findings, we remove the time filter in our code and we found 27,268 distinct report numbers, 1624 new records are found comparing last time.

1. **Do we need set time filter for records?**

Question: As we were told, FDA internal BI list has a time filter from 1994 to 2017. When looking at mdrfoi table, we found there are two columns related with date: DATE\_RECEIVED and DATE\_REPORT. So, which one should be used for time filter?

Here is the statistic report of DATE\_RECEIVED and DATE\_REPORT column:

We use DATE\_REPORT – DATE\_RECEIVED

|  |  |
| --- | --- |
| mean | -27 days |
| std | 57 days |
| min | -744 days |
| 25% | -19 days |
| 50% | -3 days |
| 75% | -1 days |
| max | 69 days |

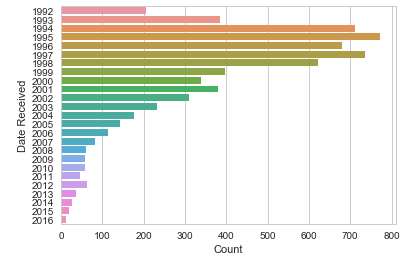
It could be seen from table that for most cases these two columns are different and date report is usually earlier than date received.

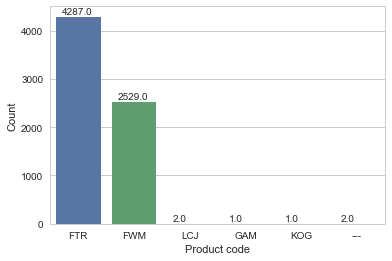
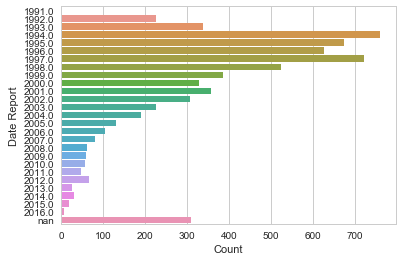
1. **What can we learn from false positives?**

False positives mean those records which are in our list but not in FDA internal list. Based on previous feedback, we have 6660 false positives.

It’s hard to see there is a trend among all these false positives, especially from product code. Almost all false positive has the product code FWM & FTR.

Question: How FDA identify BI records? If it’s by product code, then all records found by product code should be considered as BI records. We think there might be some other conditions to identify BI records other than product code, or MAUDE database does not give all records.





1. **Suggested action item for next week:**

**Yage & Zheng**

* 1. Send column report to Helen.
  2. Field: Use FWM & FTR as new data source (20170809)
  3. Target: extract info from text, make statistic report, focus on algorithm and method rather than specific number of records. Summary of data.
  4. Content:

**Asiyah**

1. Send Yage & Zheng the list of information to extract (column list, keywords list)
2. Generate a new sample which the new method didn’t catch (the new sample should be mostly different from the previous sample) We need to make sure if all internal report numbers are included in mdrfoi tables.