**代码改进**

我仔细看了提取出来的数据，基本达到要求，但是和我想要的数据还有一点差距，恳请作一点小修改，谢谢！

1、功能一修订为：

**检索1、**检索drug2018q1.csv表中drugname包含Risperidone（不区分大小写），且role\_cod等于PS或SS的记录，筛选出该表的caseid。

**检索2、**以检索1生成的caseid检索drug2018q1.csv表，提取其中所有字段。（与之前的区别在于该表除包括了role\_cod为PS或SS的记录，还包括了其他同时使用的药物信息，数据更加全面）

**检索3、**以检索1生成的caseid和drug2018q1.csv表中role\_cod为PS或SS的drug\_seq检索indi2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

**检索4、**以检索1生成的caseid和role\_cod为PS或SS的drug\_seq检索ther2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

**检索5、**以检索1生成的caseid检索demo2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

**检索6、**以检索1生成的caseid检索outc2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

**检索7、**以检索1生成的caseid检索reac2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

**检索8、**以检索1生成的caseid检索rpsr2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

2、功能二没有变化。

3、增加功能三：

**检索1、**条件1：drug2018q1.csv表中drugname包含Risperidone（不区分大小写），且role\_cod等于PS或SS的记录，同时增加条件2：reac2018q1.csv表中pt字段包含Overdose（不区分大小写）的记录，条件1与条件2取交集，即提取同时符合条件1与条件2的记录，提取其caseid。

**检索2、**以检索1生成的caseid检索drug2018q1.csv表，提取其中所有字段。（与之前的区别在于该表除包括了role\_cod为PS或SS的记录，还包括了其他同时使用的药物信息，数据更加全面）

**检索3、**以检索1生成的caseid和drug2018q1.csv表中role\_cod为PS或SS的drug\_seq检索indi2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

**检索4、**以检索1生成的caseid和role\_cod为PS或SS的drug\_seq检索ther2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

**检索5、**以检索1生成的caseid检索demo2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

**检索6、**以检索1生成的caseid检索outc2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

**检索7、**以检索1生成的caseid检索reac2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

**检索8、**以检索1生成的caseid检索rpsr2018q1.csv表中的数据，提取其中所有字段。（这条与之前版本没有差别）

4、能否把功能一、功能二与功能三分开写代码，因为是为了不同的研究目的提取数据，放在一起特别容易搞混，谢谢！

5、能否增加一个数据合并的程序。如我提取各季度的数据后，必产生多个数据表，以demo为例（见下表）。因为同一患者可能在不同时间报告了不良事件，因此必然会产生重复数据，为了清洗和整理数据，必须对数据进行合并整理，才有可能得到更加严谨的结果。需要合并的表包括demo.csv及drug.csv、indi.csv、outc.csv、reac.csv、ther.csv、rpsr.csv表。

|  |  |  |  |
| --- | --- | --- | --- |
| **原始数据** | **提取数据** | 提取数据合并> | 合并为demo.csv |
| demo2004q1.csv | f1\_demo2004q1.csv |
| demo2004q2.csv | f1\_demo2004q2.csv |
| demo2004q3.csv | f1\_demo2004q3.csv |
| demo2004q4.csv | f1\_demo2004q4.csv |
| demo2005q1.csv | f1\_demo2005q1.csv |
| demo2005q2.csv | f1\_demo2005q2.csv |
| demo2005q3.csv | f1\_demo2005q3.csv |
| demo2005q4.csv | f1\_demo2005q4.csv |
| ... | ... |
| demo2017q1.csv | f1\_demo2017q1.csv |
| demo2017q2.csv | f1\_demo2017q2.csv |
| demo2017q3.csv | f1\_demo2017q3.csv |
| demo2017q4.csv | f1\_demo2017q4.csv |
| demo2018q1.csv | f1\_demo2018q1.csv |
| demo2018q2.csv | f1\_demo2018q2.csv |

6、在dignose2中的drug表筛选中使用的条件为：

df = df[(df[**'drugname'**].str.match(**'risperidone'**)) & ((df[**'role\_cod'**].str.match(**'SS'**)) | (df[**'role\_cod'**].str.match(**'PS'**)))]

其中的str.match是否是包含的意思，此处需要特别注意，如果为等于可能会漏掉部分记录，必须是包含该字段。

谢谢，给您添麻烦了！

2018.11.10