

Exclude articles not relevant to EHR-based automated phenotyping

Siyue Yang, Jessica Gronsbell

06/15/2022

After merged articles of PubMed and Web of Sciences by `1-merging.rmd`, there are 850 papers.

Table 1: Number of papers included after merging papers from two database

| Source | n |
|--------|-----|
| Both | 510 |
| PubMed | 205 |
| WoS | 135 |
| Total | 850 |

We annotated based on the `annotations050822.csv` file. Our analysis consists of 3 steps:

- **Screen for title and abstract** and exclude the papers by our inclusion or exclusion criteria.
- **Screen for full-text** and exclude the papers by the same inclusion and exclusion criteria.
- **Include relevant papers** that are cited by papers in our pool or we encountered in our previous work.

Table 2: Number of papers excluded

| Primary reason for exclusion | Title and abstract review | Full-text review | Total |
|------------------------------|---------------------------|------------------|-------|
| Not phenotyping | 545 | 63 | 608 |
| Not using EHR | 37 | 5 | 42 |
| Not using ML | 6 | 40 | 46 |
| Review/Perspective/Editorial | 54 | 0 | 54 |
| Total | 642 | 108 | 750 |

We also report the number of papers removed in Table 3.

Table 3: Number of paper removed and added

| Process | Number of papers removed | Number of papers added | Number of papers remained |
|---------------------------|--------------------------|------------------------|---------------------------|
| Before screening | 0 | 0 | 850 |
| Title and abstract review | 642 | 0 | 208 |
| Full-text review | 108 | 0 | 100 |