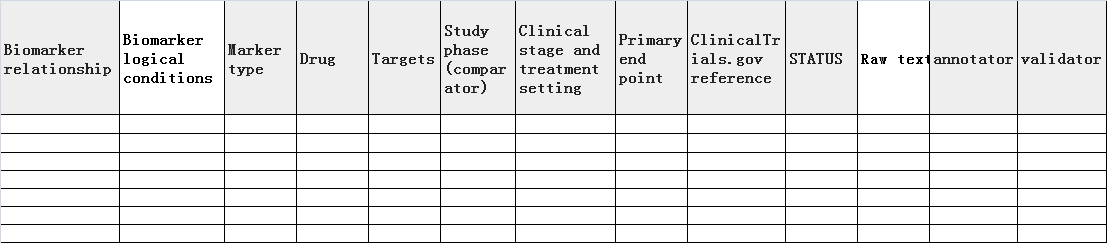
**Breast Cancer clinical trial data annotation**

1. Setting Up the table in Excel.
   1. Open Excel;
   2. Create a table:



1. Obtain Raw Data.

2.1. Go to [www.ClinicalTrials.gov](http://www.ClinicalTrials.gov);

2.2. At “Find a study”, “Condition or disease”, enter “breast cancer”;

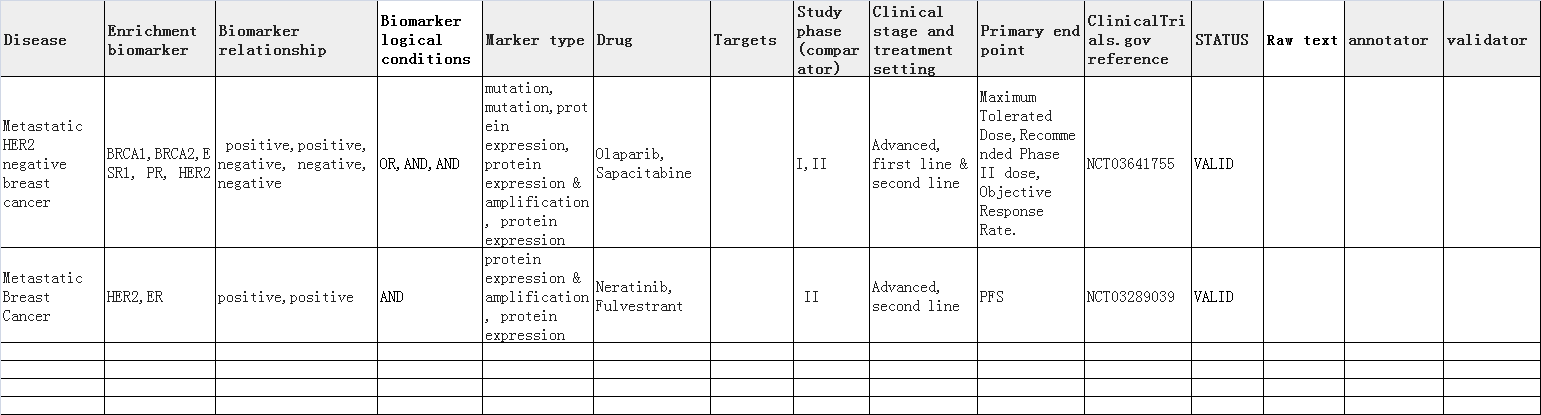
2.3. Search;

2.4. 9709 clinical trials were retrieved on March 2, 2020. Go to filters, status, recruitment; select all except “Suspended” and “withdraw”;

2.5. Apply the filter change;

2.6. 9350 clinical trials were retrieved;

2.7. Fill table with required information from the trials, for example:



1. Edit and standardize the information in the table

3.1. Use the following info to standardize biomarker type:

Biomarker types:

1. polymorphism

2. mutation

3. amplification

4. translocation

5. deletion

6. gene fusion

7. genetic others

8. methylation

9. epigenetic others

10. protein expression

11. protein others: e.g., phosphorylation

12. gene expression

13. gene others

14. metabolite abundance

15. metabolite others

16. others

3.2 Use NCBI entrez name to standardize biomarker names.

3.2.1. find NCBI entrez name for each biomarker;

3.2.2. replace biomarker names with NCBI entrez names. For example: enter “estrogen receptor” into

<https://www.ncbi.nlm.nih.gov/gene/>

found the NCBI entrez name of “estrogen receptor” is “ESR1”; replace all “ER” with “ESR1” in the table.

1. Validate the info in the table and correct the mistakes.