Data cleaning summary for test dataset v1

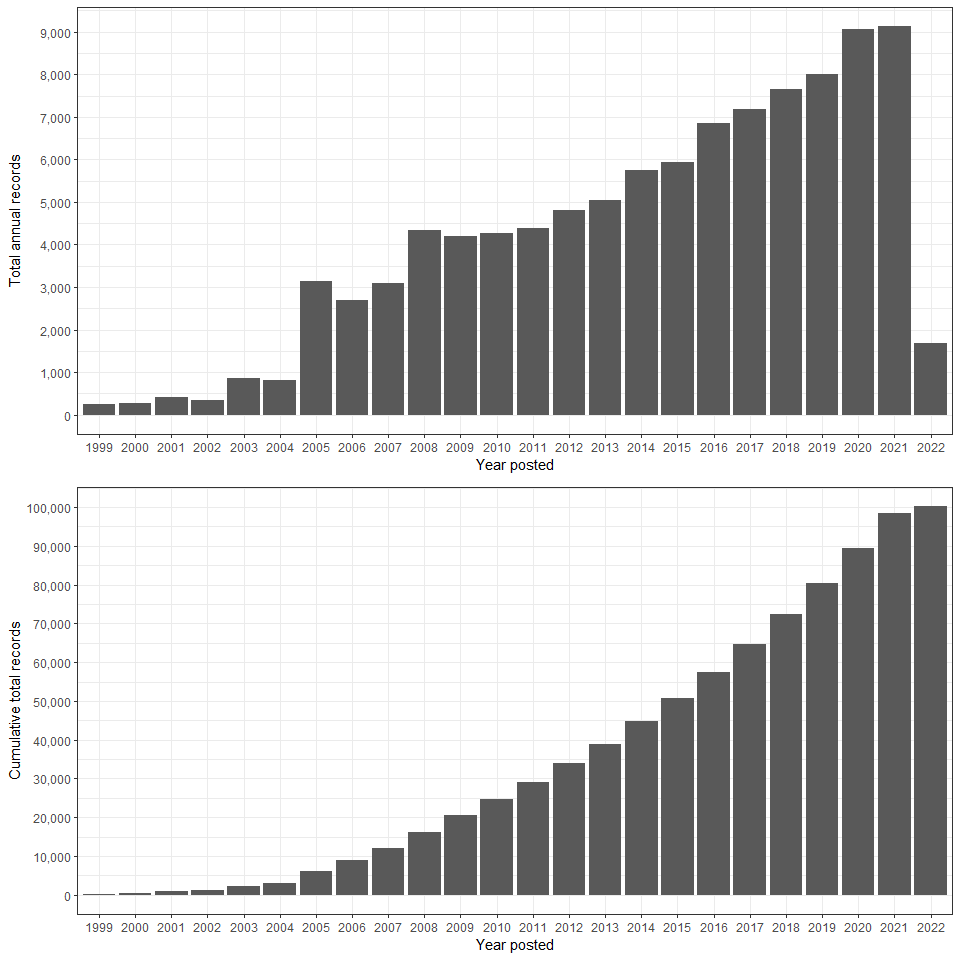
Nicole White

11/03/2022

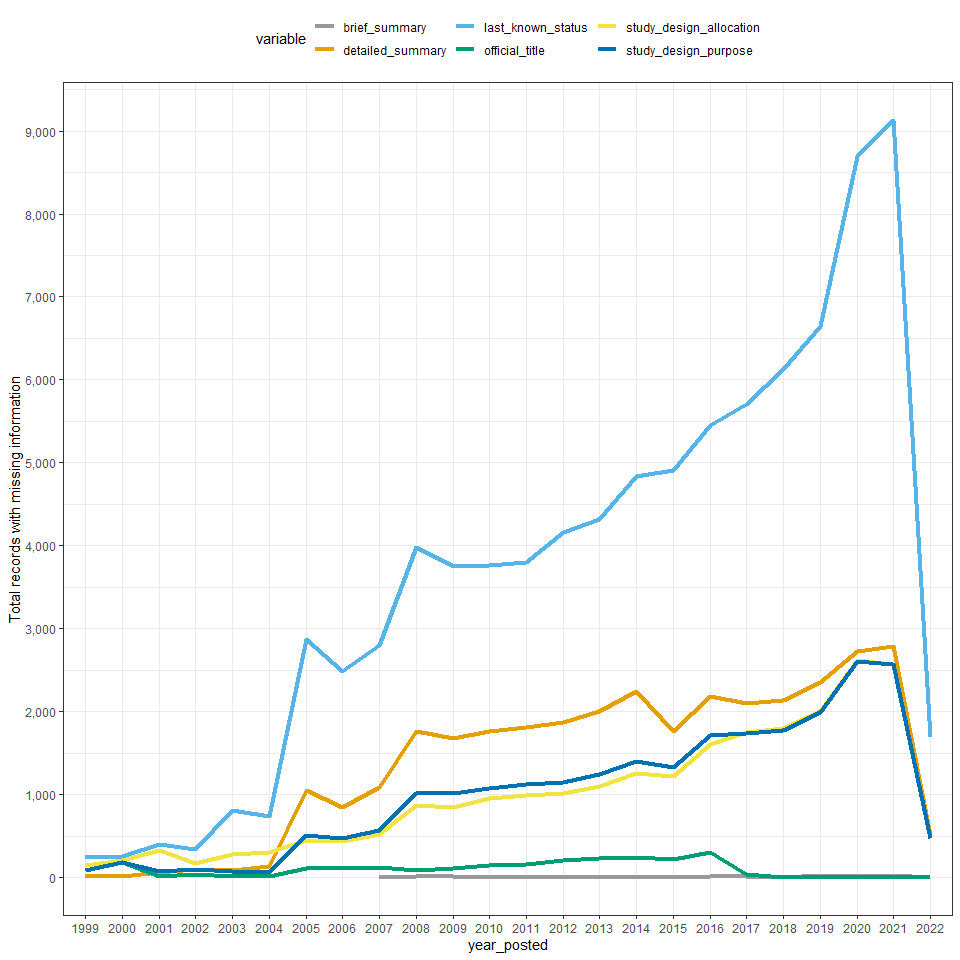
## Fields extracted from XML records

| Field |
| --- |
| Id |
| Official Title |
| Brief Title |
| Brief Summary |
| Detailed Summary |
| Overall Status |
| Last Known Status |
| Submitted |
| Posted |
| Updated |
| Study Type |
| Study Design Allocation |
| Study Design Purpose |
| Primary Outcome Measures |
| Primary Outcome Description |
| Secondary Outcome Measures |
| Secondary Outcome Description |
| Mesh Terms |
| Keywords |

## Total and cumulative records posted over time



## Missing study fields

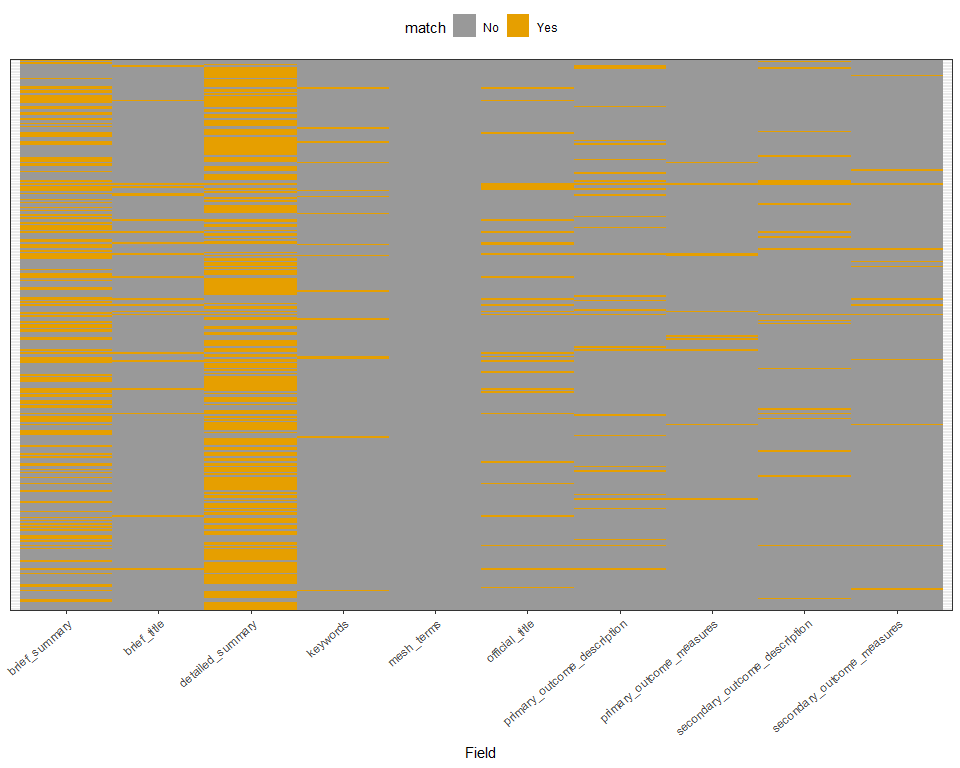


## Targeted search terms

I used the following set of search terms to identify relevant studies on prediction modelling:

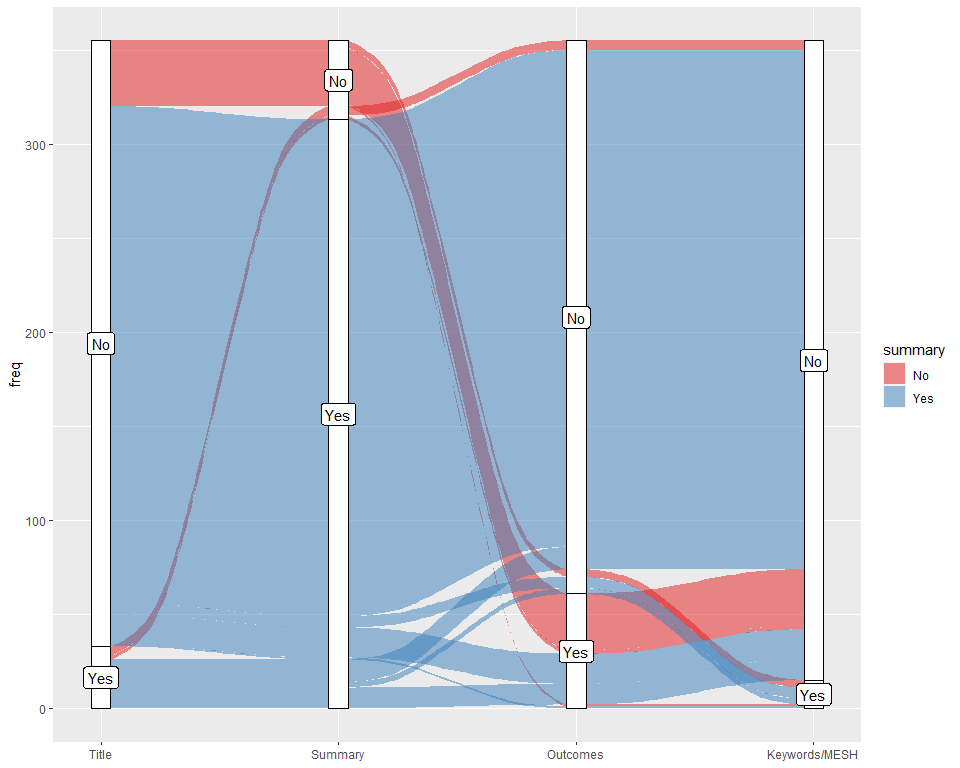
* (machine learning|artificial intelligence|deep learning|prediction model|predictive model|prediction score|predictive score|warning score|risk score|risk prediction) AND (prognosis|prognostic|diagnose|diagnosis|diagnostic)

Overall, 355 matching records were found. A visual summary of matches is given below. Most matches were found in study summaries.



Matching study records

## Warning in to\_lodes\_form(data = data, axes = axis\_ind, discern =  
## params$discern): Some strata appear at multiple axes.  
  
## Warning in to\_lodes\_form(data = data, axes = axis\_ind, discern =  
## params$discern): Some strata appear at multiple axes.  
  
## Warning in to\_lodes\_form(data = data, axes = axis\_ind, discern =  
## params$discern): Some strata appear at multiple axes.



Summary of matches by study field