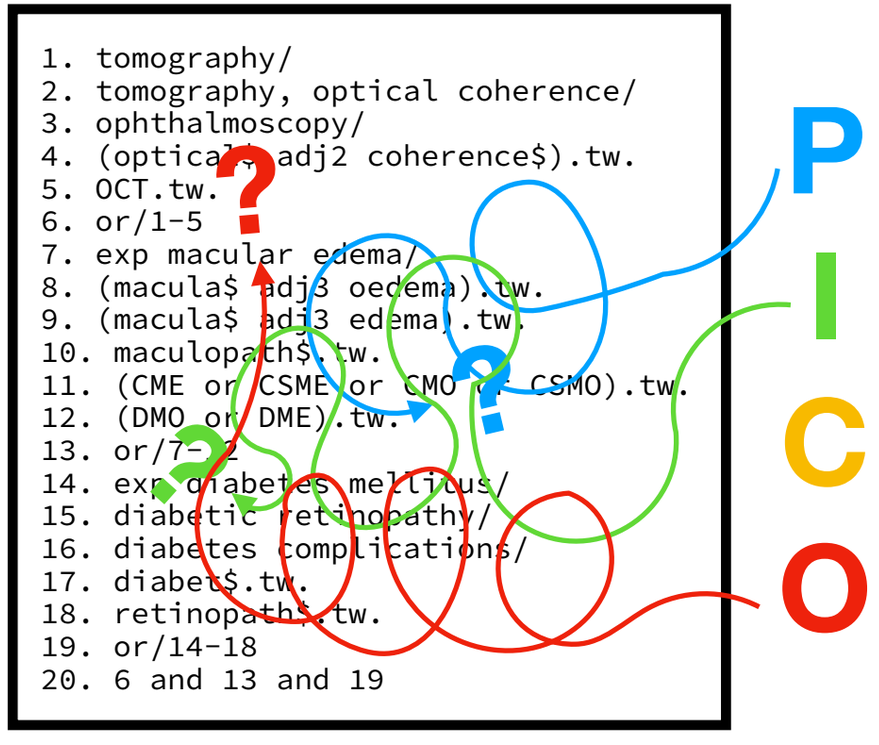
# fP

- We expect the fP approach to be the best, by design
- It shows the upper bounds on effectiveness for this task
- **Can we design a model that can predict which keywords require PICO annotations?**

| | Recall | Precision | F3 | F1 | F0.5 | WSS |
|---|---|---|---|---|---|---|
| B | $\mathbf{0.7553}^{p}$ | $0.0137^{pf}$ | $0.0901^{pf}$ | $0.0255^{pf}$ | $0.0168^{pf}$ | $0.0120^{pf}$ |
| P | $0.6509^{bcf}$ | $0.0215^{bc}$ | $0.1128^{bc}$ | $0.0375^{bc}$ | $0.0258^{bc}$ | $0.0206^{bc}$ |
| cP | $0.7002^{pf}$ | $0.0139^{pf}$ | $0.0886^{pf}$ | $0.0257^{pf}$ | $0.0170^{pf}$ | $0.0124^{pf}$ |
| fP | $\mathbf{0.7553}^{pc}$ | $\mathbf{0.0223}^{bc}$ | $\mathbf{0.1263}^{bc}$ | $\mathbf{0.0400}^{bc}$ | $\mathbf{0.0271}^{bc}$ | $\mathbf{0.0214}^{bc}$ |



# Task 4: Screening Prioritisation

- Trained a learning to rank models using clinical features (PICO) and limited number of QPPs
    - IDFSum, IDFStd, IDFMax, IDFAvg, P-Count,  I-Count, O-Count
- Systematic review queries then transformed manually into Elasticsearch queries
- Train 5 LTR models using these features and evaluate the performance of each
    - MART, AdaRank, Coordinate Ascent, LambdaMART, Random Forests
- Compare the difference between queries with PICO features and non-PICO features