Results

	Precision	Recall	$F0.5$	$F1$	$F3$	WSS
Baseline	0.010	0.815	0.012	0.018	0.053	0.801
Greedy	0.034 (+236.33%)	0.725 (−11.11%)	0.034 (+184.24%)	0.039 (+114.74%)	0.068 (+27.29%)	0.720 (−10.10%)
Oracle_p	**0.078 (+680.68%)**^{b}	0.521 (−36.06%)^b	0.077 (+537.16%)^b	0.086 (+380.70%)^b	0.146 (+174.71%)^b	0.520 (−35.10%)^b
Oracle_r	0.011 (+8.98%)	**0.938 (+15.04%)**^{b}	0.012 (+0.86%)	0.015 (−15.35%)	0.027 (−49.97%)	0.865 (+7.99%)
$\text{Oracle}_{F0.5}$	0.076 (+663.67%)^b	0.543 (−33.37%)^b	**0.078 (+540.90%)**^{b}	0.088 (+389.45%)^b	0.153 (+188.80%)^b	0.542 (−32.37%)^b
Oracle_{F1}	0.073 (+633.48%)^b	0.545 (−33.13%)^b	0.077 (+534.48%)^b	**0.089 (+394.48%)**^{b}	0.155 (+192.26%)^b	0.544 (−32.12%)^b
Oracle_{F3}	0.062 (+522.25%)^b	0.577 (−29.21%)^b	0.068 (+463.08%)^b	0.084 (+367.42%)^b	**0.159 (+199.48%)**^{b}	0.576 (−28.15%)^b
Oracle_{WSS}	0.023 (+133.45%)	0.925 (+13.41%)	0.024 (+99.43%)	0.027 (+48.91%)	0.047 (−12.10%)	**0.903 (+12.69%)**
Classifier_p	0.025 (+147.72%)^o	0.693 (−15.02%)	0.028 (+129.36%)	0.035 (+95.90%)	0.075 (+40.18%)^b	0.681 (−15.01%)
Classifier_r	*0.035 (+253.09%)*	*0.717 (−12.11%)^o*	*0.037 (+203.61%)*	*0.042 (+132.98%)*	0.073 (+38.25%)	*0.711 (−11.22%)*
$\text{Classifier}_{F0.5}$	0.032 (+224.53%)	0.713 (−12.57%)	0.035 (+185.13%)^o	0.040 (+125.70%)	0.077 (+44.45%)	0.704 (−12.11%)
Classifier_{F1}	0.026 (+157.78%)	0.693 (−14.98%)	0.029 (+136.51%)	0.036 (+99.29%)^o	0.075 (+40.21%)^b	0.682 (−14.85%)
Classifier_{F3}	0.026 (+160.37%)	0.692 (−15.11%)	0.029 (+138.90%)	0.036 (+101.19%)	0.075 (+40.45%)^{bo}	0.681 (−15.00%)
Classifier_{WSS}	0.033 (+229.41%)	0.684 (−16.11%)	0.035 (+188.95%)	0.041 (+129.61%)	*0.080 (+50.85%)^b*	0.671 (−16.22%)^o
Ranker_p	0.046 (+358.47%)	0.653 (−19.90%)^b	0.043 (+252.78%)	0.044 (+147.11%)	0.065 (+22.97%)	0.649 (−18.97%)^b
Ranker_r	0.034 (+237.78%)	*0.753 (−7.70%)^o*	0.033 (+172.19%)	0.036 (+101.95%)	0.058 (+9.60%)	*0.748 (−6.64%)*
$\text{Ranker}_{F0.5}$	0.046 (+362.90%)	0.619 (−24.12%)^b	0.042 (+247.79%)	0.043 (+141.94%)	0.062 (+16.95%)	0.615 (−23.24%)^b
Ranker_{F1}	0.048 (+380.17%)	0.614 (−24.69%)^b	0.043 (+257.07%)	0.045 (+149.91%)	0.067 (+26.71%)	0.610 (−23.83%)^b
Ranker_{F3}	*0.056 (+455.82%)*	0.645 (−20.92%)^b	*0.051 (+319.32%)*	0.053 (+197.77%)	0.076 (+42.90%)	0.641 (−20.00%)^b
Ranker_{WSS}	0.053 (+428.60%)	0.610 (−25.25%)^b	0.050 (+314.01%)	*0.054 (+199.11%)*	*0.081 (+52.34%)*	0.607 (−24.27%)^{bo}

Task 3: Retrieval for Screening

- Retrieve all relevant studies while minimising non-relevant

  - Medical citations from a medical database

  - Using a Boolean query

  - Has the effect of **reducing the workload** for reviewers

- PICO is commonly used to formulate the Boolean query, however it is not used by the underlying search system

  - Should it be?

- This work was published in CIKM 2017 proceedings as short paper