

Supplementary Table 2. Numeric results from Figure 4: Comparison of medication extraction NLP systems for entity-level precision, recall, and F1 performance measures on test sets.

		Test Set Performance								
		Tacrolimus			Lamotrigine			Allopurinol		
		Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
DrugName	medExtractR	.96 [.92, 1.00]	1.00 [1.00, 1.00]	.98 [.96, 1.00]	1.00 [1.00, 1.00]	.97 [.94, 1.00]	.99 [.97, 1.00]	.98 [.94, 1.00]	.99 [.97, 1.00]	.98 [.96, 1.00]
	MedXN	1.00 [1.00, 1.00]	.96 [.92, 1.00]	.98 [.96, 1.00]	.82 [.69, .93]	.79 [.64, .92]	.81 [.67, .92]	1.00 [1.00, 1.00]	.98 [.94, 1.00]	.99 [.97, 1.00]
	MedEx	.91 [.85, .97]	.99 [.96, 1.00]	.95 [.91, .98]	.99 [.95, 1.00]	.95 [.85, 1.00]	.97 [.91, 1.00]	.98 [.95, 1.00]	.97 [.92, 1.00]	.98 [.94, 1.00]
	CLAMP	.97 [.93, 1.00]	.95 [.90, .99]	.96 [.93, .99]	1.00 [1.00, 1.00]	.94 [.84, 1.00]	.97 [.91, 1.00]	.98 [.95, 1.00]	.85 [.79, .91]	.91 [.87, .95]
Strength	medExtractR	.97 [.93, 1.00]	1.00 [1.00, 1.00]	.99 [.96, 1.00]	.97 [.92, 1.00]	.96 [.90, 1.00]	.97 [.93, 1.00]	.99 [.97, 1.00]	.97 [.93, 1.00]	.98 [.96, 1.00]
	MedXN	.94 [.89, .99]	.97 [.92, 1.00]	.96 [.92, .98]	.99 [.96, 1.00]	.87 [.74, .99]	.93 [.85, .99]	.96 [.90, 1.00]	.96 [.90, 1.00]	.96 [.91, 1.00]
	MedEx	.82 [.73, .91]	.98 [.95, 1.00]	.90 [.84, .95]	.93 [.86, .99]	.85 [.72, .97]	.89 [.80, .96]	.93 [.87, .99]	.91 [.77, 1.00]	.92 [.84, .98]
	CLAMP	.89 [.80, .97]	.82 [.72, .90]	.85 [.77, .92]	.98 [.93, 1.00]	.61 [.46, .78]	.75 [.63, .87]	.97 [.91, 1.00]	.80 [.72, .88]	.87 [.82, .93]
DoseAmount	medExtractR	.97 [.92, 1.00]	1.00 [1.00, 1.00]	.99 [.96, 1.00]	.93 [.84, 1.00]	1.00 [1.00, 1.00]	.96 [.91, 1.00]	.93 [.85, .99]	1.00 [1.00, 1.00]	.96 [.92, .99]
	MedXN	.93 [.85, 1.00]	.97 [.91, 1.00]	.95 [.90, .99]	.97 [.91, 1.00]	.93 [.83, 1.00]	.95 [.89, 1.00]	.97 [.93, 1.00]	.97 [.93, 1.00]	.97 [.95, .99]
	MedEx	.93 [.80, 1.00]	.34 [.19, .51]	.50 [.31, .66]	.97 [.91, 1.00]	.90 [.80, 1.00]	.94 [.86, 1.00]	.99 [.96, 1.00]	.99 [.96, 1.00]	.99 [.97, 1.00]
	CLAMP	.70 [.38, 1.00]	.18 [.07, .32]	.29 [.12, .46]	.93 [.81, 1.00]	.68 [.51, .86]	.78 [.65, .90]	1.00 [1.00, 1.00]	.87 [.79, .94]	.93 [.89, .97]
Frequency	medExtractR	.98 [.96, 1.00]	.98 [.95, 1.00]	.98 [.96, 1.00]	.96 [.91, 1.00]	.91 [.83, .98]	.93 [.89, .97]	.94 [.88, .98]	.97 [.93, 1.00]	.96 [.92, .99]
	MedXN	.94 [.88, .98]	.92 [.85, .98]	.93 [.88, .97]	.95 [.89, 1.00]	.77 [.62, .90]	.85 [.74, .93]	.94 [.87, .99]	.96 [.90, 1.00]	.95 [.91, .99]
	MedEx	.51 [.40, .63]	.53 [.41, .64]	.52 [.41, .64]	.79 [.66, .90]	.63 [.48, .76]	.70 [.58, .80]	.83 [.75, .92]	.86 [.78, .94]	.85 [.78, .92]
	CLAMP	.60 [.46, .74]	.41 [.30, .52]	.49 [.37, .60]	.79 [.64, .90]	.42 [.31, .54]	.55 [.42, .66]	.90 [.82, .96]	.71 [.61, .81]	.79 [.72, .87]

F1 represents F-measure. The test sets consisted of 50 notes each for tacrolimus and lamotrigine, and 110 notes for allopurinol. The drug entities presented here reflect a restricted list of entities which have been standardized across all four natural language processing (NLP) systems to ensure comparability. MedExtractR was developed on clinical notes for tacrolimus and lamotrigine, with no training notes for allopurinol.