

**Table 1**

Intrinsic Evaluation: Gold Standard Agreement Measures for NLP Engine, Compared with Human Coders and Upper Bounds.

	<b>CPT F</b>	<b>ICD F</b>	<b>ICD-1</b>
<b>Engine</b>	0.838	0.435	0.486
<b>Coders (Paper)</b>	0.872	0.540	0.516
<b>Upper Bound</b>	0.910	0.592	0.572

Figures are aggregated over all engine confidence levels.