Clinical Tractor is a natural language understanding system for clinical practice guidelines. Clinical practice guidelines, or CPGS, encompass the most recent evidence-based information available to doctors when caring for patients and contain recommendations for standard approaches to diagnosing and treating patients. Compliance with CPGs has been shown to improve patient care significantly. Compliance improves with the use of clinical decision support systems and electronic healthcare records, which require computer interpretable guidelines, or CIGs. Currently, CPG distribution is not well suited to integration with these systems, and converting CPGs to CIGs requires laborious manual effort. Clinical tractor aims to aid in automating the process of converting CPGs to CIGs.

Clinical Tractor consists of a natural language processing pipeline that ultimately converts the textual content and structural elements of CPGs, into a representation of the semantic knowledge contained in a CPG. The first step in the pipeline is to manually create computer readable XML representations of CPGs for downstream processing. Textual regions of the guideline are represented in JATS, the Journal Article Tag Suite. The visual diagramatic components of the guidelines are transcribed into a CSV file which is the basis of the final graphs. Then a script produces graphs utilizing GraphML in order to provide a standard schema to ensure proper storage of data and ease of parsing. After the guideline has been converted to XML, initial text processing is done with GATE (the General Architecture for Text Engineering). GATE is used to perform tokenization and sentence splitting, part-of-speech tagging and dependency parsing, morphological analysis, named entity recognition, and coreference resolution.

Once initial processing is complete, the syntactic information is converted by the *propositionalizer* to CSNePS assertions to create a syntactic knowledge base. CSNePS is a knowledge representation and reasoning system that is simultaneously logic-based, frame-based, and graph-based. CSNePS uses the logic of arbitrary and indefinite objects for representation. That is, guidelines are represented as knowledge graphs where relationships between entities have logical structure. The syntactic knowledge base created by the propositionalizer is then aligned with background information from various biomedical ontologies. This associates medical terms with logical axioms representing the meaning of the terms.
Background knowledge alignment results in an enhanced syntactic knowledge base, which is then used as input to a process which builds a semantic knowledge base. This is done by the Syntax-Semantics Mapper. The process of converting syntax to semantics ultimately is aimed at building representations of entities like patients (individuals and groups), medical conditions, treatments, etc., and their participation in actions prescribed by the guideline recommendations.

As we continue to develop Clinical Tractor, we also perform regular evaluation. We have a development set that is separate from the test data we will use to evaluate the project upon completion. Evaluation is a continuous and semi-automatic process in which new CSNePS assertions are checked against a gold standard evaluation. If the new assertion is present in the gold standard, then it is assigned the matching rating of goodness or badness. If an assertion is not present, the rating is determined manually by reviewers who hate their lives. Also included in the semi-automatic evaluation is data corresponding to the rule-firings that were lost from the gold standard evaluation, and the new rule-firings that were not present in the gold standard. An

initial calculation of precision and recall relative to the gold standard evaluation is presented as well.

We continue to develop the syntax-semantics mapper, update and improve off-the-shelf tools used in the NLP pipeline, and tackle more complex representational issues.