SUPPLEMENTAL MATERIAL

Reference Standard

Traditional and advanced approaches were tested against a reference standard for physician encounters. The reference standard consisted of an independent review, with manual annotation of relevant HF-specific features, including 19 unique HF-specific concepts. For each encounter, two independent clinical annotators labeled each concept and all metadata for that concept. For example, an annotator might mark the text "DOE over last month" as dyspnea on exertion, experienced = true, current = true, relative date = 1 month. Concept occurrence was defined as the sum of all concept occurrences, allowing for multiple occurrences per encounter. Encounter occurrence was defined as the number of encounters with at least one occurrence of the concept.

Given that many concepts, such as LVEF are specific to a point in time, concepts were tested at the encounter level. For example, if a patient had an LVEF of 30% in an encounter, the data extraction would only be annotated as correct if it identified "LVEF 30%" in that specific encounter. This reference standard was used to determine accuracy of automated extracted data and structured data. Specifically, this reference standard was used to calculate recall and precision for these individual features for traditional and advanced approaches.