N1: *Are there previous conclusive reports on this reaction? (+1 for yes, 0 for no, and 0 for do-not-know).* We have built semi-automatically a knowledge resource of drug–AE pairs from external knowledge resources (http://www.aderepository.org).

N2: *Did the adverse event occur after the suspected drug was administered? (+2 for yes, -1 for no, and 0 for do-notknow).* Our discourse parsing will capture temporal relations between entities.

N3: *Did the adverse reaction improve when the drug was discontinued or a specific antagonist was administered? (+1 for yes, 0 for no, and 0 for do-not-know).* We will build a comprehensive list of relevant antidotes, antagonists and toxicity-specific supportive measures relevant to the drug-AE pairs. We will automatically detect the actions of improving and getting worse with manually curated rules. For example, if a physiological function goes from normal to abnormally high or from normal to abnormally low, it may be represented linguistically as getting worse. Our discourse parsing of temporal relation identification will help capture the condition after a drug is discontinued. We will explore the use of polarity words (e.g., bad, worsening, and improving) and automatically capture polarity with our previously developed approaches[76], a most-cited seminal work in the field of subjective language. We expect that semantic ambiguity can nearly always be resolved by referencing the normal range of physiological function or therapeutic range of drug concentration as represented in the structured data in the EHR. In addition, we will extract related information from package inserts on the National Institutes of Health website DailyMed, which contains over 10,000 drug inserts, their warnings, precautions, and adverse events. We will develop surface features (for example, detecting cue terms withdrawal, terminated, rise, and increase, etc), pattern matching (mapping 1200/mm^3 to the CTCAE 4.03 grade level), and rules to extract information to answer this question.

N4: Did the adverse reaction reappear when the drug was readministered? (+2 for yes, -1 for no, and 0 for do-notknow). We will associate AE and medication events, and will apply discourse parsing to identify temporal relations between events.

N5: Are there alternative causes (other than the drug) that could on their own have caused the reaction? (-1 for yes, +2 for no, and 0 for do-not-know). We will examine patient's longitudinal DSs to identifying whether AEs are introduced by other means.

N6: Did the reaction reappear when a placebo was given? (-1 for yes, +1 for no, and 0 for do-not-know). This Naranjo item would not be relevant in this context and will be omitted.

N7: Was the drug detected in the blood (or other fluids) in concentrations known to be toxic? (+1 for yes, 0 for no, and 0 for do-not-know). We will detect this information from the structured EHR data, focusing on the lab results.

N8: Was the reaction more severe when the dose was increased, or less severe when the dose was decreased? (+1 for yes, 0 for no, and 0 for do-not-know). Our NLP approaches capture events and temporal and causal relations among the events and we will develop additional rule-based approaches and regular expressions for detecting answers to this question.

N9: Did the patient have a similar reaction to the same or similar drugs in any previous exposure? (+1 for yes, 0 for no, and 0 for do-not-know). We will mine patient's longitudinal EHRs to extract temporal ADEs. We will use the World Health Organization's Anatomical Therapeutic Chemical Classification System to identify drug classes. Specifically, we will use the fourth level of the classification system to identify those drugs in the same class as the target drug

N10: Was the adverse event confirmed by any objective evidence? (+1 for yes, 0 for no, and 0 for do-not-know). Eventually, objective evidence will include objective clinical assessments, laboratory results, imaging studies, and other diagnostic tests. This work will focus only on detecting objective evidence from DSs, and leave other EHR data for future work. We will identify objective statements (e.g., "drop in blood pressure") and separate them from subjective statements (e.g., "seems better") with our HedgeScope system, which performed with an F1-score of 0.93 and 0.90, respectively, in detecting hedge cue phrases and their scope in clinical notes [47]