

An Exploration of the Terminology of Clinical Cognition and Reasoning

James J. Cimino, MD;^{1,2} Ziran Li, MD;² Chunhua Weng, PhD²

¹Informatics Institute, University of Alabama at Birmingham, Birmingham, Alabama

²Department of Biomedical Informatics, Columbia University, New York, New York

Abstract

We hypothesize that the functionality of electronic health records could be improved with the addition of formal representations of clinicians' cognitive processes, including such things as the interpretation and synthesis of patient findings and the rationale for diagnostic and therapeutic decisions. We carried out a four-phase analysis of clinical case studies to characterize how such processes are represented through relationships between clinical terms. The result is a terminology of 26 relationships that were validated against published clinical cases with 85.4% interrater reliability. We believe that capturing patient-specific information with these relationships can lead to improvements in clinical decision support systems, information retrieval and learning health systems.

Introduction

It seems that no day goes by without publication of a complaint about electronic health records (EHRs) in the scientific and other professional literature. Practitioners and informaticians call repeatedly for improved access to patient information, reduced requirements for data entry by clinicians, and relief from information overload and needless alerts. A recent Viewpoint in the *Journal of the American Medical Association* captures much of this sentiment [1] and a popular book written by a recent AMIA Symposium keynote speaker provides extensive detail with illustrative anecdotes on where the EHR falls short.[2]

While acknowledging that EHRs do some things well (e.g., billing, privacy, availability, legibility, and coordination with ancillary departments), current complaints generally center on poor user interface design, lack of interoperability, and overzealous alerts and reminders. The American Medical Association has called for redesign of EHRs[3], while Rudin et al. have recently called for acceleration in innovation in EHR development and outlined the necessary mechanisms for doing so.[4] However, improved programming alone may be insufficient to overcome the current shortcomings.

Consider, for example, automated alerts related to computer-based prescriber order entry. A recent study by Bryant and colleagues showed that alerts are overridden 93% of the time.[5] When a reason for override is given, it is often based on something the clinician knew that was not considered by the alert logic, either because the alert authors did not consider it, its representation in the EHR was not usable by the alert, or it was missing entirely.[6,7]

We hypothesize that what is missing is a clear representation of the clinical reasoning (by clinicians and patients) regarding the interpretation of information, as well as the application of that information and medical knowledge to decision-making processes. For example, if the EHR contains information about why a medication has been discontinued (cure, intolerance, noncompliance, insurance coverage, etc.), decision support logic would be better able to make appropriate recommendations about alternative therapy (or remain silent, if the reason is "cure"). In addition, the user interface would be better able to retrieve data from the patient's record relevant to selection and course of the medication's use, while the reuse of these data for research purposes would be better able to inform a "learning healthcare system".[8]

It is generally the case that EHR functions such as billing, data display, decision support, and infobuttons, to name a few, are more successful when they can make use of structured data represented with a controlled vocabulary (i.e., coded data), rather than relying on medical language processing. If we want a decision support system to take into account our clinical reasoning, for example to decide whether to alert us about a potentially inappropriate repeat chest x-ray or to retrieve information from the patient's record relevant to a particular problem, it will be better able to do so if it has some knowledge about the reasoning behind our requests. It follows that a controlled vocabulary for representing this information will be needed.

Any clinician can easily provide a list of some of the kinds of information that can be used to describe his or her clinical reasoning: "I think this physical finding explains this symptom", "my differential diagnosis for this finding includes these diseases", "I am going to order this test to verify my hypotheses about a potential diagnosis", "I am going to use this medication to treat this disease", and so on. It strikes us that much of this information is in the form

of relationships between familiar classes of medical concepts, such as symptoms, findings, tests, diseases, medications, and procedures. Terminologies such as SNOMED-CT, LOINC, and RxNorm provide good coverage of the medical concepts, but the relations they provide (a list of symptoms or treatments of a disease, for example) is different than the assertions about a particular situation for a particular patient (that a patient's symptom was ascribed to a particular disease or that a particular treatment was chosen to treat that particular disease).

To the best of our knowledge, there is no terminology that covers such relations. We therefore undertook an initial empirical study to begin to identify the concepts and terms involved in clinical cognition and reasoning with the intent of developing an terminology to support richer EHR functionality. This paper describes our initial experience.

Methods

In developing our approach, we considered a number of possible sources of examples of clinical cognition and reasoning: patient records, expert clinicians, and case reports. We chose not to use patient records in this initial study due to their incomplete nature, which was an original motivation for this research. Simply asking medical experts how they think might seem the most expedient route, but we believe this would be subject to recall bias and that the inherently tacit nature of clinical reasoning [9] might hinder discovery.

We believe that case reports, as published in the peer-reviewed medical literature, expose the kind of clinical reasoning that underlies what is taking place, or should be taking place, in actual patient care. As a result, published case reports provide a potentially richer source of explicit statements about what clinicians think is going on with their patients and why they choose the diagnostic and therapeutic actions that they do. We therefore chose an approach that combined case reports and clinical expertise, involving clinical informaticians who (by nature of their training) we thought might have better insight into the tacit nature of their own reasoning processes. We carried out a four-phase process to develop and evaluate a terminology that could be used to build a terminology of clinical reasoning and cognition (see Figure 1).

As an initial data set, we chose the three most recent clinical problem-solving cases published in the *New England Journal of Medicine*[10-12] (as of October, 2016) because (a) they are freely available for use in such research, (b) they provide high quality case descriptions, (c) they cover the clinical progress of the patient and the evolution of his or her situation over time, and (d) they are annotated with interpretations of observations and recommendations provided by an expert. The cases were obtained by searching PubMed with the strategy “clinical problem solving and nejm”, obtaining links to the full text at www.nejm.org, and downloading them.

The text of each article was processed manually to remove header and footer information, figures, tables, and the “Commentary” sections (as they have a much more erudite discussion of general medical knowledge than one would expect to see in EHR records). The text was then parsed into individual sentences and loaded into a spreadsheet, with one sentence per row. Particularly complex sentences were divided into multiple rows.

We asked fellows of the American College of Medical Informatics (ACMI) with clinical backgrounds (that is, nurses, dentists and physicians) and who attended the 2017 ACMI Symposium to annotate random subsets of these sentences. Their task was to annotate each sentence with one or more tuples, consisting of two concepts and some relationship between them (e.g., “patient” – “has symptom” – “dyspnea”). One of the authors (JJC) annotated the entire set of sentences and then reviewed all of the expert results to identify common relationship themes. This set of relationships formed the *initial terminology*.

In Phase 2, we sought a preliminary validation of the initial terminology. One of the authors (ZL), applied the initial terminology in a blinded fashion to a random sample of the sentences. These annotations were compared to the previous annotations by JJC and interrater reliability (IRR) was measured using joint probability of agreement. Discrepancies were discussed and resolved either by agreeing which term was appropriate, by merging terms as overlapping, or splitting terms determined to be ambiguous to produce a *revised terminology*.

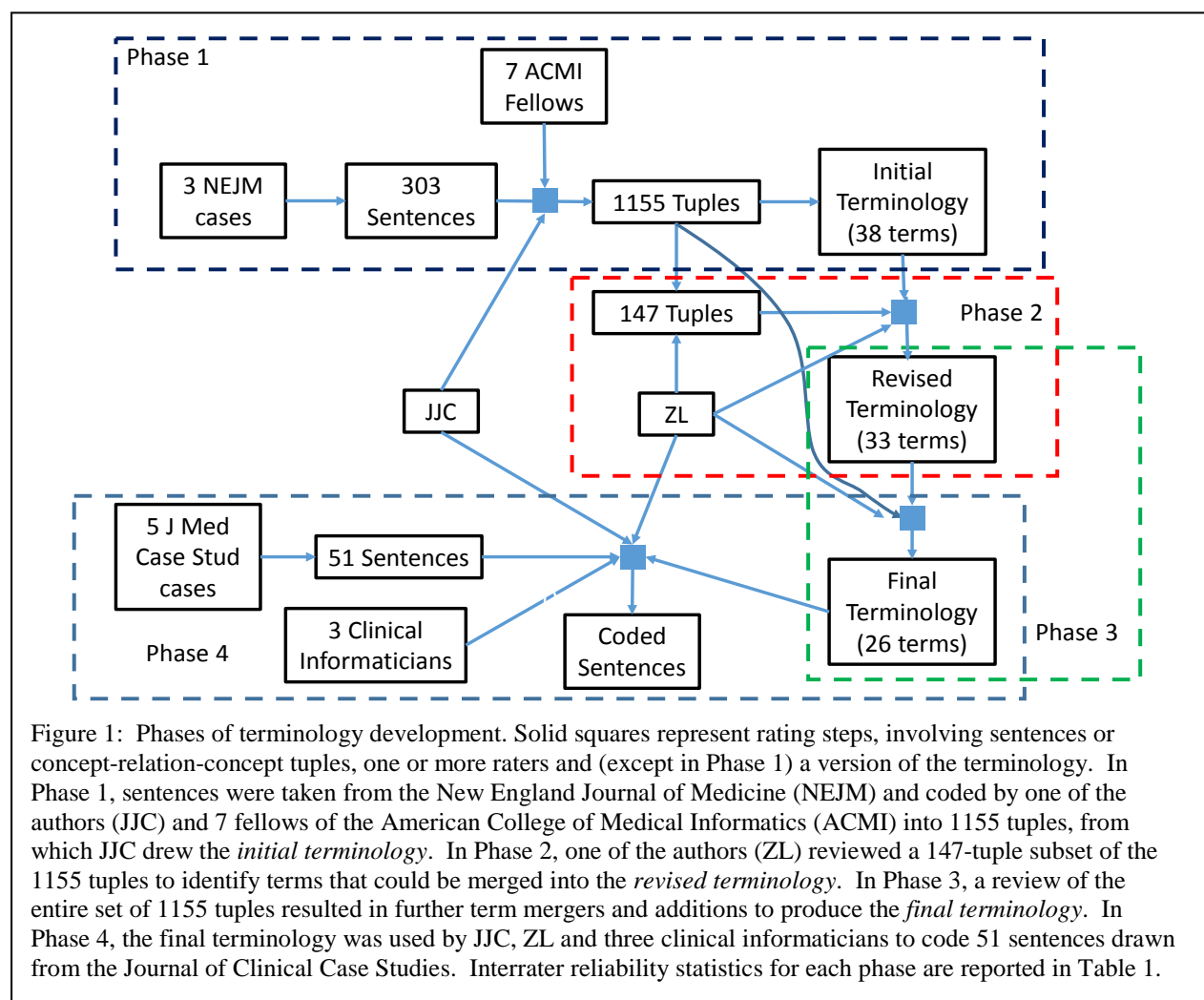
In Phase 3, all of the sentences were further annotated (by ZL) using the revised terminology, and again the IRR (joint probability of agreement with JJC's original annotations, now updated with the revised terminology) was measured. Discrepancies were discussed to determine if additional term merging was appropriate, resulting in our *final terminology*.

In Phase 4, we validated the final terminology by using it to annotate a new set of sentences that were drawn from the “Clinical Presentation” sections of five case reports in the *Journal of Clinical Case Reports*. [13-17] Each sentence from these sections was placed in a spreadsheet along with three columns for data entry, using pull-down lists that contained the final terminology. Copies of the spreadsheet were given to three clinical informaticians

(specializing in family practice, pediatric hematology/oncology and pathology) to select one to three terms (e.g., “has symptom”, one per data entry column) for each sentence. Two of us (JJC and ZL) also coded the sentences.

We examined the degree of agreement among all the raters using Fleiss’s Kappa statistics[18] to describe the IRR between all five raters assigning categorical ratings to a number of sentences. In order to better understand the IRR, the raw data were reshaped and re-interpreted such that for each sentence, raters decided whether the sentence contains each of the relationship concepts in our final terminology. IRRs for each relationship concept were calculated respectively. For each sentence, the most common relationship among all raters was considered the “dominant” relationship. Only the most common chosen relationship was counted for each rater’s choice and the overall IRR among the relationships for all five raters across 51 sentences was assessed.

Unweighted kappa statistics (Fleiss’s Kappa for all five raters) were employed to measure the IRR for the overall agreement among raters. Recommended guidelines for interpretation of kappa value (k) are as follows[19]: $k \leq 0$ Poor agreement; $0 < k \leq 0.2$ Slight agreement; $0.2 < k \leq 0.4$ Fair agreement; $0.4 < k \leq 0.6$ Moderate agreement; $0.6 < k \leq 0.8$ Substantial agreement; $0.8 < k \leq 1.0$ Almost perfect agreement. The calculation was performed using the “irr” package in R.[20]



Results

A total of 303 sentences or sentence fragments were extracted from the three *New England Journal of Medicine* papers. Twenty-five ACMI members were contacted. Of those who agreed to conduct some analysis, seven (six physicians and one nurse) returned their spreadsheets with some sentences coded. Taken together, experts (including

JJC provided 1155 tuples for the 303 sentences (272, or 90.4%, of sentences has results from two or experts). The consolidation of the 1155 tuples by JJC resulted in 395 instances of 38 relations in the *initial terminology*.

Initial remapping of 147 tuples (from 27 sentences) of the total 1155 tuples yielded 127 agreements (86%). Discussion of the 20 discrepancies led to the following decisions to change relationships in the initial terminology:

- The term “finding due to therapy” covers the scope of “adverse effect” and so these two were merged into "finding/condition due to therapy/intervention "
- The term "findings co-occur" is synonymous with “co-occurrence” and so these two were merged into "findings co-occur”
- Because the term “history” does not distinguish between something that has resolved and something that continues, it was merged into “finding/history of”
- The terms "natural history of finding” and “disease course” overlap significantly and were merged into "natural history of finding/condition"
- The term “condition suggests finding” was more appropriately covered by “finding suggests condition” and was removed

These revisions reduced the 38 relationships to 33 (the *revised terminology*) and resulted in agreement on 138 of 147 tuples (94%).

Reannotation of the full 1155 tuples with the revised terminology resulted in agreement on 998 (86.4%), with disagreement on 157. Discussion of discrepancies yielded the following observations:

- Findings (especially physical examination) were sometimes difficult to characterize as positive or negative; e.g., presence of a normal finding implied absence of a negative finding; the presence and absence terms were therefore merged into “finding/history of” (75 instances)
- There were similar results related to test results; these terms were also merged into “finding/history of” (29 instances)
- The terms “description of test” and “natural history of finding/condition”, both involving general medical knowledge, as opposed to patient-specific information were deemed to be overlapping and were merged into “general knowledge/physician experience” (9 instances)
- A new term, “progress and prognosis” was suggested and accepted (15 instances)

These changes resulted in a *final terminology* of 26 terms and retroactively reduced the number of disagreements to 4 (97.3% agreement) in the initial 147 tuples and to 29 (97.5% agreement) in the total set of 1155.

The five Phase 4 reviewers identified a total of 594 relationships in the 51-sentence sample from the Journal of Medical Case Studies. The kappa values for the use of each 26 final terms (plus “other”) across all five reviewers ranged from -0.03 to 0.81, with an average of 0.31 and 11 having moderate or better agreement ($k \geq 0.4$). When only the most common term assigned by the five reviewers was considered for each sentence, Fleiss’s Kappa IRR was 0.70. Table 1 lists summary statistics for IRR measurements in all phases of the study. Table 2 shows details for each of the terms in the final terminology with individual kappa values.

All five phase 4 reviewers provided a total of 58 suggestions. These are currently undergoing analysis to obtain consensus on how the terminology should change to accommodate them.

Table 1: Summary IRRs for all evaluations in the study

| Data Set | Initial Terminology | Revised Terminology | Final Terminology |
|---|---------------------|---------------------|-------------------|
| 147 tuples from the New England Journal of Medicine | 86%* | 94%* | 97%* |
| 1155 tuples from the New England Journal of Medicine | N/A | 86.4%* | 97.5%* |
| 51 sentences from the Journal of Medical Case Studies | N/A | N/A | 0.70** |

*Joint probability of agreement between ZL and JJC

**Fleiss’s Kappa IRR among all 5 raters

Table 2: Twenty-six terms in the final terminology, with definitions, examples and Fleiss's Kappa from Phase 4 evaluation. Example words in italics indicate clinical concepts associated with each other through the relationship.

| Relationship concepts | Definition | Example | kappa |
|---|--|--|-------|
| finding/history of | Any simple statement of fact about a patient, including a past or current diagnosis, a symptom, physical finding, test result, etc.; no inference or interpretation involved | The <i>patient</i> had <i>slurred speech</i> | 0.39 |
| absence of finding | Any simple statement of fact about something absent, currently or in the past; no inference or interpretation involved | <i>She</i> had no history of <i>easy bleeding</i> or bruising | 0.65 |
| finding altered by activity | An observation (by clinician or patient) that a finding is altered (increased, relieved, etc.) by some action (physical activity, medication, etc.) | <i>She</i> noted <i>shortness of breath on exertion</i> | 0.06 |
| findings co-occurrence | Findings are related to each other in some temporal or topographical way | The <i>bruising</i> began shortly after she had had <i>sore throat, coryza, and malaise</i> | 0.24 |
| finding suggests condition | A condition is inferred by presence of a finding, or a finding can be explained by condition | The <i>pink sputum</i> suggests <i>hemoptysis</i> | 0.69 |
| finding suggests condition is not present | Finding rules out or argues against the condition | These <i>results</i> essentially rule out a <i>platelet disorder</i> | 0.27 |
| finding does not suggest condition | Finding does not support the condition, but cannot rule out the condition | <i>Elevation of blood sugar</i> in a patient receiving intravenous glucose does not imply the presence of <i>diabetes</i> . | -0.01 |
| finding absence argues against condition | Absence of finding rules out or argues against the condition | The <i>normal blood pressure</i> argues against the <i>HELLP syndrome</i> | 0.31 |
| finding/condition due to intervention | Adverse effects, side effect or complication of a treatment or procedure | Medications are typically thought to trigger this disorder; exposure to antibiotics such as <i>cephalosporins</i> , which this patient received, commonly precedes the <i>rash</i> | 0.55 |
| finding/condition suggests risk of some other condition | Inference that a condition might be present based on the presence of some fact about the patient | <i>HIV-positive patients</i> are at risk for a variety of <i>lesions in the nervous system</i> | 0.00 |
| differential diagnosis | Consideration of multiple explanatory conditions for the presence of some observed fact or inferred condition | Other causes of an <i>isolated elevated aPTT</i> include <i>lupus anticoagulants, factor deficiencies, factor inhibitors, and severe cases of von Willebrand's disease</i> | 0.70 |
| therapy | A decision to initiate or continue a therapeutic activity | The <i>patient</i> began to receive <i>levofloxacin</i> | 0.49 |
| discontinue therapy | A decision to cease a therapeutic activity | When the Gram's staining and culture from <i>his</i> biopsy were negative, all <i>antibiotic treatment</i> was discontinued | 0.81 |
| therapy indicated/considered | A recommendation that a therapeutic activity should be initiated (or continued) based on clinical assessment | A trial of <i>glucocorticoids</i> could be considered for relief of symptoms if the patient's initial <i>headache</i> persists | 0.40 |

| | | | |
|---|--|---|-------|
| therapy not indicated | A recommendation that a therapeutic activity should be not initiated (or should be discontinued) based on clinical assessment | Inadvertent use of <i>quinolone</i> monotherapy to treat <i>tuberculosis</i> facilitates the emergence of resistant strains and may delay a diagnosis by decreasing the sensitivity of sputum culture | -0.01 |
| therapy not available | Statement that a therapeutic activity, whether recommended or not, is not an available or appropriate option | With regard to the patient's dysarthria, her age, smoking, hypertension, hyperlipidemia, and family history all arouse concern about stroke, although she is beyond the window for any <i>therapy to prevent acute stroke</i> | 0.16 |
| progress and prognosis | An assessment of current or future change in a clinical condition | During this time, he had <i>complete resolution</i> of his <i>skin abnormalities</i> , with no recurrence after discontinuation of prednisone treatment | 0.41 |
| diagnostic/assessment procedure indicated | A recommendation that an assessment activity should be initiated (or continued) based on clinical assessment | In this case, given the history that is suggestive of <i>hemoptysis</i> , <i>imaging studies</i> are indicated | 0.43 |
| intention of intervention | The intended reason for carrying out an intervention to assist with an inference or alteration of a patient condition | In a patient with <i>acute cough</i> , the history and physical examination are key to determining whether <i>imaging</i> is needed | 0.01 |
| procedure performed | A statement that some intentional activity occurred | <i>A lumbar puncture</i> was performed [on patient] | 0.36 |
| activity description | Report of patient performing some activity | The <i>patient</i> reported <i>strict adherence to medications</i> | -0.03 |
| description of knowledge/experience | A statement of medical knowledge or clinical experience that may be relevant to the patient's findings, conditions or decision-making | Many <i>acquired bleeding diatheses</i> have no obvious <i>hereditary component</i> | 0.40 |
| has feature | A statement of some (nontopographical, nontemporal) attribute of a finding | <i>Chest radiography</i> showed a <i>wedge-shaped opacity</i> in the right lower lobe with air bronchograms | 0.19 |
| location | A statement about a topographical aspect of a finding | During the week before presentation, she also noted <i>pain</i> in her <i>right elbow and both knees</i> | 0.03 |
| time | A statement about a temporal aspect of a finding | She recalled that she first noticed <i>difficulty speaking 4 days</i> before presentation | 0.66 |
| causes | A statement of medical knowledge causal relationships between findings /conditions or an inference that such as relationship is present in the patient | If <i>systemic release of toxins</i> occurs, a diffuse pustular variant of the <i>toxic shock syndrome</i> can also occur | 0.20 |
| other relation (please explain) | Statements that cannot be characterized by one of the above | N/A | 0.05 |

Discussion

We carried out a four-phase analysis of clinical case reports to characterize how clinicians' cognitive processes are represented through relationships between clinical terms. We also manually extracted and harmonized frequently seen relationships among clinical concepts and achieved a promising interrater agreement for a terminology defining these relationships. The resulting terminology contains 26 unique relationships connecting clinical concepts for symptoms, signs, findings, treatments, and so on. The terminology was validated against published clinical cases with an overall interrater reliability of 85.4%. Examination of individual annotations of clinical case statements shows great variability in the consistency of their use. This is likely due in part to a need for more rigorous and systematic annotator training. However, based on initial review of annotator comments, there is likely room in the terminology for additional relationships. The fact that the IRR of the Journal of Clinical Case Studies was more than 10 percentage points lower than the IRR with New England Journal of Medicine cases suggests that the

terminology needs to be adapted to additional types of reports before saturation is reached. Not the least important of these types will be real-world data from electronic health records.

This is the initial pilot study intended to create a “starter set” of relationships that appear to be relevant to data captured in current EHRs. The data set was drawn from published annotated cases. These cases are probably not typical of what appears in patient records today, but that is exactly the point. The cases identify things that *should* be going through clinicians’ minds and should therefore be captured somehow in the EHR, either explicitly by clinicians or implicitly through inferences performed behind the scene by artificial intelligence, in order to better communicate with others caring for the patient or to improve reuse of the data (whether for immediate decision support or retrospective research).

This work presents a much-needed first step to catalyze a larger study to explore which of the 26 relationships found here can be reliably detected through automated means and which are worth having clinicians (and patients) enter explicitly into EHRs, what additional relationships might be found in other clinical sources, and which ones are not worth further effort. Developing non-disruptive methods for capturing these data (see [21], for example), finding corresponding ways to reduce other workload and educating clinicians and patients on the importance of making their insights and plans explicit are all challenges that will require the breadth of scientific disciplines comprising biomedical informatics. Once they are there, we can focus our efforts on making productive use of contextual information to support cognitive computing needed by clinicians.

This study has several limitations. First, we analyzed only a sample of case reports and clinical statements. Our current set of 26 relationships is unlikely to be sufficient for representing all clinical reasoning. This work is merely the first step in a principled, empiric approach to developing an ontology for this domain. Second, our recruitment strategy for human annotators was based on convenience samples. Now that an initial set of concepts has been established, development of a more rigorous training process for annotators should be feasible. Third, much work remains to clarify the relationships and the terms to which they can be applied.

Despite these limitations, we believe that our “starter set” of clinical cognition and reasoning can form the basis for representing patient data in ways that, to our knowledge, cannot be captured with current ontologies (such as SNOMED [22]) or data models (such as the HL7 Reference Information Model [23]). And while medical knowledge bases have long included relationships such as “Symptom X can be due to Disease Y” and “Disease Y can be treated with Drug Z”, there is currently no controlled terminology or data model for capturing patient-specific clinical reasoning, such as “I believe that Ms. Jones’s Symptom X is explained by her Physical Finding Y” or “I am considering Diseases A, B and C to explain Ms. Jones’s Physical Finding Y”.

While data mining of noisy, incomplete narrative text records may be useful for discovering patterns in patient records that can help us care for future patients, we believe that explicit, formally represented statements by clinicians regarding their inferences and decisions about Ms. Jones will be better sources of knowledge to caring for Ms. Jones herself. For example, by knowing the clinician’s differential diagnosis, an expert system can help rule out some diseases, suggest additional diseases to consider, and offer the best evidence for choosing diagnostics tests. By knowing how the clinician believes signs and symptoms relate to diagnoses, decision support systems can help monitor the care plan to propose alternative strategies when the treatment is “right” but the patient is not improving.

We also believe that a formal representation of clinicians’ interpretations and decisions can unlock new approaches to addressing the criticisms currently leveled against EHRs. Knowing what patient attributes clinicians consider important can drive data retrieval from the patient’s historical record and can silence irrelevant alerts. Information about the patient’s state and the clinician’s thinking can lead to more informative clinical documentation. Finally, the “learning health system” will be better able to actually learn from health records if their authors say explicitly what they are thinking.

Conclusions

Representing the patient’s true situation and clinician’s cognitive processes has always been a goal of good health care but has not been pursued diligently. Adding such information to the chart, in a structured, controlled manner has the potential to lead to the next generation of sophisticated EHRs that are better equipped to provide cognitive decision support. The work presented here can serve as an initial step towards a larger effort to characterize such valuable situational information with a broad range of applications to improve healthcare quality and efficiency.

Acknowledgments

This work was supported by research funds from the Informatics Institute and the Center for Clinical and Translational Science of the University of Alabama at Birmingham, funded under grant **1TL1TR001418-01** from the National Center for the Advancement of Translational Science (NCATS). This work would not have been possible without the assistance of the raters: Christopher Chute (Johns Hopkins University), Patricia Dykes (Brigham and Women's Hospital), Pavithra Dissanayake (University of Alabama at Birmingham), Kevin Johnson (Vanderbilt University), Wayne Liang (University of Alabama at Birmingham), Thomas Payne (University of Washington), and Marc Overhage (Cerner Corporation), Charles Safran (Beth Israel-Deaconess Medical Center), Marc Williams (Geisinger Health System), and Amy Wang (University of Alabama at Birmingham).

References

1. Zulman DM, Shah NH, Verghese A. Evolutionary pressures on the electronic health record – caring for complexity. *JAMA*. 2016 Sep 6;316(9):923-4.
2. Wachter RM. *The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age*. McGraw-Hill, New York; 2015.
3. Monegane B. AMA demands EHR overhaul, calls them 'poorly designed and implemented'. *Healthcare IT News*, September 12, 2017. <http://www.healthcareitnews.com/news/ama-demands-ehr-overhaul-calls-them-poorly-designed-and-implemented> (last visited July 5, 2018)
4. Rudin RS, Bates DW, MacRae C. Accelerating Innovation in Health IT. *N Engl J Med*. 2016 Sep 1;375(9):815-7
5. Bryant AD, Fletcher GS, Payne TH. Drug interaction alert override rates in the Meaningful Use era: no evidence of progress. *Appl Clin Inform*. 2014 Sep 3;5(3):802-13.
6. Her QL, Amato MG, Seger DL, Beeler PE, Slight SP, Dalleur O, Dykes PC, Gilmore JF, Fanikos J, Fiskio JM, Bates DW. The frequency of inappropriate nonformulary medication alert overrides in the inpatient setting. *J Am Med Inform Assoc*. 2016 Sep;23(5):924-33.
7. Baysari MT, Tariq A, Day RO, Westbrook JI. Alert override as a habitual behavior - a new perspective on a persistent problem. *J Am Med Inform Assoc*. 2016 Oct;94:67-74
8. Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, Gunter C, Musen M, Platt R, Stead W, Sullivan K, Van Houweling D. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc*. 2015 Jan;22(1):43-50.
9. Patel VL, Arocha JF, Kaufman DR. Expertise and Tacit Knowledge in Medicine. In, Sternberg RJ and Horvath JA, eds. *Tacit knowledge in professional practice*. Hillsdale, NJ: Lawrence Erlbaum Associates (Dec,1998).
10. Lampson BL, Gray SW, Cibas ES, Levy BD, Loscalzo J. CLINICAL PROBLEM-SOLVING: Tip of the Tongue. *N Engl J Med*. 2016 Sep 1;375(9):880-6.
11. Aday AW, Saavedra AP, Levy BD, Loscalzo J. CLINICAL PROBLEM-SOLVING: Prevention as Precipitant. *N Engl J Med*. 2016 Aug 4;375(5):471-5.
12. Gibson CJ, Berliner N, Miller AL, Loscalzo J. CLINICAL PROBLEM-SOLVING: A Bruising Loss. *N Engl J Med*. 2016 Jul 7;375(1):76-81.
13. Dalugama C. Mesalazine-induced myocarditis: a case report. *J Med Case Rep*. 2018 Feb 22;12(1):44.
14. Munakomi S. Aggressive fibromatosis in the infratemporal fossa presenting as trismus: a case report. *J Med Case Rep*. 2018 Feb 19;12(1):41.
15. Takemori N, Imai G, Hoshino K, Ooi A, Kojima M. A novel combination of bortezomib, lenalidomide, and clarithromycin produced stringent complete response in refractory multiple myeloma complicated with diabetes mellitus - clinical significance and possible mechanisms: a case report. *J Med Case Rep*. 2018 Feb 18;12(1):40.
16. D'Amico F, Bertacco A, Cesari M, Mescoli C, Caturegli G, Gondolesi G, Cillo U. Extraordinary disease-free survival in a rare malignant extrarenal rhabdoid tumor: a case report and review of the literature. *J Med Case Rep*. 2018 Feb 17;12(1):39. doi: 10.1186/s13256-017-1554-2.
17. Ferreira-Hermosillo A, Casados-V R, Paúl-Gaytán P, Mendoza-Zubieta V. Utility of rituximab treatment for exophthalmos, myxedema, and osteoarthropathy syndrome resistant to corticosteroids due to Graves' disease: a case report. *J Med Case Rep*. 2018 Feb 16;12(1):38. doi: 10.1186/s13256-018-1571-9.
18. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*; 1960; 20(1):37-46.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.

20. Gamer M, Lemon J, Fellows I, Singh P. <puspendra.pusp22@gmail.com> (2012). irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84. <https://CRAN.R-project.org/package=irr>
21. Payne TH, Alonso WD, Markiel JA, Lybarger K, White AA. Using voice to create hospital progress notes: Description of a mobile application and supporting system integrated with a commercial electronic health record. *J Biomed Inform.* 2018 Jan;77:91-96.
22. Wang AY, Sable JH, Spackman KA. The SNOMED clinical terms development process: refinement and analysis of content. *Proc AMIA Symp.* 2002:845-9.
23. Shakir AM. HL7 Reference Information Model. More robust and stable standards. *Healthc Inform.* 1997 Jul;14(7):68.