eC **A commentary by Gwo-Chin Lee, MD,** is linked to the online version of this article at jbjs.org.

Use of Natural Language Processing Algorithms to Identify Common Data Elements in Operative Notes for Total Hip Arthroplasty

Cody C. Wyles, MD, Meagan E. Tibbo, MD, Sunyang Fu, MHI, Yanshan Wang, PhD, Sunghwan Sohn, PhD, Walter K. Kremers, PhD, Daniel J. Berry, MD, David G. Lewallen, MD, and Hilal Maradit-Kremers, MD, MSc

Investigation performed at the Mayo Clinic, Rochester, Minnesota

Background: Manual chart review is labor-intensive and requires specialized knowledge possessed by highly trained medical professionals. Natural language processing (NLP) tools are distinctive in their ability to extract critical information from raw text in electronic health records (EHRs). As a proof of concept for the potential application of this technology, we examined the ability of NLP to correctly identify common elements described by surgeons in operative notes for total hip arthroplasty (THA).

Methods: We evaluated primary THAs that had been performed at a single academic institution from 2000 to 2015. A training sample of operative reports was randomly selected to develop prototype NLP algorithms, and additional operative reports were randomly selected as the test sample. Three separate algorithms were created with rules aimed at capturing (1) the operative approach, (2) the fixation method, and (3) the bearing surface category. The algorithms were applied to operative notes to evaluate the language used by 29 different surgeons at our center and were applied to EHR data from outside facilities to determine external validity. Accuracy statistics were calculated with use of manual chart review as the gold standard.

Results: The operative approach algorithm demonstrated an accuracy of 99.2% (95% confidence interval [CI], 97.1% to 99.9%). The fixation technique algorithm demonstrated an accuracy of 90.7% (95% CI, 86.8% to 93.8%). The bearing surface algorithm demonstrated an accuracy of 95.8% (95% CI, 92.7% to 97.8%). Additionally, the NLP algorithms applied to operative reports from other institutions yielded comparable performance, demonstrating external validity.

Conclusions: NLP-enabled algorithms are a promising alternative to the current gold standard of manual chart review for identifying common data elements from orthopaedic operative notes. The present study provides a proof of concept for use of NLP techniques in clinical research studies and registry-development endeavors to reliably extract data of interest in an expeditious and cost-effective manner.

Total hip arthroplasty (THA) is one of the most common inpatient surgical procedures. Over 500,000 THA and hip hemiarthroplasty procedures are performed each year in the United States, and approximately 2.5 million Americans are currently living with THA implants[1]. Growing demand for improved mobility and quality of life is expected to result in further increases in THA procedures in the coming decades[2].

Lack of high-quality, real-world data is a critical barrier to THA research, policy, and surveillance efforts. In the absence of detailed information, quality-improvement efforts are often restricted to imperfect superficial administrative data for THA classification and risk-stratification[3,4]. Furthermore, delays in data availability make it virtually impossible to perform real-time surveillance of THA implants and outcomes.

THA-specific data elements remain embedded in the unstructured text of electronic health records (EHRs). Manual collection of this information from charts is labor-intensive and requires specialized knowledge possessed by highly trained

Disclosure: The authors indicated support from the National Institutes of Health (NIH), grant R01 AR73147. On the **Disclosure of Potential Conflicts of Interest** forms, *which are provided with the online version of the article*, one or more of the authors checked "yes" to indicate that the author had a relevant financial relationship in the biomedical arena outside the submitted work and "yes" to indicate that the author had a patent and/or copyright, planned, pending, or issued, broadly relevant to this work (http://links.lww.com/JBJS/F493).

TABLE I Number of THA Procedures Included for Training and Testing of NLP Algorithms

	Operative Approach	Fixation	Bearing Surface
Training data set			
No. of procedures	250	467	300
Age* *(yr)*	65	69	51
Percent female	67%	56%	53%
Test data set			
No. of procedures	250	291	284
Age* *(yr)*	63	69	53
Percent female	63%	58%	51%

*The values are given as the mean.

medical professionals. The cost and infrastructure challenges required to implement manual data collection are prohibitive for many hospitals and research teams. Therefore, detailed THA data are limited to a handful of institutions with the resources to implement labor-intensive manual data collection.

Natural language processing (NLP) methods offer an opportunity to efficiently extract THA-specific data elements from the unstructured text of EHRs, and they are increasingly used in research[5]. In partnership with trained orthopaedic surgeons and informatics specialists, we therefore developed a series of NLP-based algorithms for ascertainment of 3 common THA-specific data elements from operative notes: operative approach, fixation technique, and bearing surface. The purpose of the present study was to evaluate the accuracy of THA-specific NLP algorithms and to compare their accuracy against the gold standard of manual chart review by trained registry specialists.

Materials and Methods

Study Setting

Following institutional review board approval, we evaluated all primary THA procedures that had been performed at a single academic institution between 2000 and 2015. The institution had an annual volume of 800 to 1,300 primary THAs performed by 35 orthopaedic surgeons during that time period. As part of the Mayo Total Joint Registry, data collection was performed through manual chart review by trained joint registry personnel with use of standardized definitions for THA-specific data elements, yielding readily available gold-standard data for validation.

Study Design

We focused on 3 THA-specific data elements recorded in operative notes: (1) operative approach, (2) fixation technique, and (3) bearing surface. Within the large cohort of all primary THA procedures between 2000 and 2015, we identified training and test data sets using stratified random sampling to develop and test the NLP algorithms separately for each of the 3 data elements. Data from the Mayo Total Joint Registry were used as the gold standard to assess the performance of the NLP algorithms.

The numbers of THA procedures for training and testing the NLP algorithms are shown in Table I. For operative approach, the training and test data sets comprised 250 THA procedures, stratified into 3 categories (anterolateral, direct anterior, and posterior). The average ages of the patients in the training and test data sets were 65 and 63 years, respectively, and women comprised 67% and 63% of the patients in these data sets, respectively. For fixation technique, the training and test data sets comprised 467 and 291 THA procedures, respectively, stratified into 4 categories (uncemented, cemented, hybrid, and reverse hybrid). The average ages in the training and test data sets were 69 and 69 years, and women comprised 56% and 58% of the patients in these groups, respectively. For bearing surface, the training and test data sets comprised 300 and 284 THA procedures, respectively, stratified into 4 categories (metal-on-polyethylene, ceramic-on-polyethylene, metal-on-metal, and ceramic-on-ceramic). The average ages in the training and test data sets were 51 and 53 years, respectively, and women comprised 53% and 51% of the patients in these groups, respectively.

NLP Algorithm Development

The NLP algorithms were developed in 3 steps: (1) prototype system development based on manual chart review and expert knowledge from orthopaedic surgeons, (2) formative system development using a training data set, and (3) final system evaluation using a test data set. Our NLP system was based on expert rules defined by 2 orthopaedic surgeons after review of operative notes that established target "textual markers" (i.e., keywords related to approaches, fixation technique, or bearing surface). The NLP system has 3 main components: text processing, concept extraction, and classification (Fig. 1).

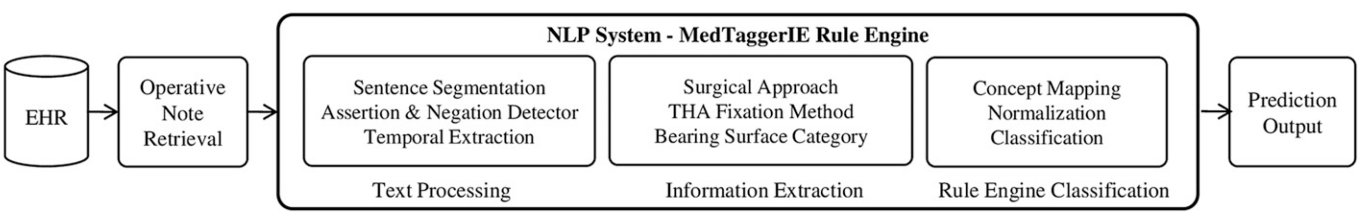

Fig. 1
Workflow of the NLP system for extracting 3 THA-specific data elements from operative notes.

Approach	**Rule for Direct Mentions** Concepts of Direct Lateral Approach, Direct Anterior Approach, and Posterior Approach	
Fixation	**Rule for Direct Mentions** Concepts of Cement, Uncement, Hybrid, Reverse- Hybrid	**Rule for Indirect Mentions** **Cemented:** [Cement Concept] AND ([Shell Concept] AND [Stem Concept]) **Hybrid:** “hybrid” OR ([Stem Concept] AND [No Shell Concept]) **Reverse Hybrid:** “reverse-hybrid” OR ([Shell Concept] AND [No Stem Concept]) **Uncemented:** “uncement” OR ([No Shell Concept] AND [No Stem Concept])
Bearing Surface	**Rule for Direct Mentions** Concepts of Metal-on-poly, Ceramic-on-poly, Metal-on-metal, Ceramic-on-ceramic	**Rule for Indirect Mentions** **Metal-on-poly:** [Metal Concept] AND [No Ceramic Concept] AND [Poly Concept] **Ceramic-on-poly:** ([Poly Concept] AND [No Metal Concept] AND [Ceramic Concept] AND [Head Concept]) OR ([Polybrand Concept] AND [No Metal Concept] AND [Ceramic Concept] AND [Head Concept]) **Metal-on-metal:** [No Poly Concept] AND [Metal Concept] **Ceramic-on-ceramic:** [No Poly Concept] AND [Ceramic Concept]

Fig. 2

Rules for THA operative approach, fixation, and bearing-surface classification.

The key components of the text processing pipeline were sentence segmentation, assertion identification, and temporal extraction. The assertion of each concept includes certainty (i.e., positive, negative, and possible) along with the person who experienced the event (i.e., the patient or someone else, such as husband, child, etc.), whereas temporality identifies the

TABLE II THA Operative Approach-Related Keywords in Operative Notes

Direct mention for anterolateral approach	
	Anterolateral approach; direct lateral approach; Hardinge; lateral approach; gluteus medius detached; gluteus medius reflected; gluteus medius split; gluteus medius trochanter; gluteus medius incised; gluteus minimus detached; gluteus minimus reflected; gluteus minimus split; gluteus minimus trochanter; gluteus minimus incised; abductors detached; abductors reflected; abductors split; abductors trochanter; abductors incised; anterior detached; anterior reflected; anterior split; anterior trochanter; anterior incised; incision directly laterally*
Direct mention for direct anterior approach	
	Direct anterior approach; TFL; tensor fasciae latae; tensor fascia latae; lateral femoral cutaneous nerve; LFCN; lateral circumflex femoral artery; lateral femoral circumflex artery; anterior superior iliac spine; anterior-superior iliac spine; ASIS; Hana table; Smith-Petersen approach; sartorius
Direct mention for posterior approach	
	External rotators; short external rotators; short rotators; conjoined tendon; obturator internus; posterior approach; posterolateral approach; posterior capsule; posterior greater trochanter; quadratus; quadratus femoris; Southern approach; Moore approach; Kocher approach; Kocher-Langenbeck approach; Kocher-Langenbeck; gluteus maximus split; gluteus maximus divided; glut max split; glut max divided; maximus split; maximus divided; gluteus maximus muscle split; gluteus maximus muscle divided; incision posterolateral*
Exclusion keywords	
	Posterior capsule released; piriformis released; obturator released; short external rotators preserve; obturator internus released; external rotators preserve

*Additional refinement from the external validation.

TABLE III THA Fixation-Related Keywords in Operative Notes

Indirect mention for cement concept
Cement; cemented; methyl methacrylate; methacrylate; methyl; vacuum
Indirect mention for shell concept
Shell; Implex shell; PINNACLE shell; cup; acetabular; acetabulum component; acetabulum; socket
Indirect mention for stem concept
Femoral stem; stem; HFx-stem; femoral component; femoral component/stem; permanent prosthesis; stem fem cemented
Direct mention for cement
Cement total hip arthroplasty; cement total hip replacement; total hip arthroplasty, cement; total hip replacement, cement
Direct mention for uncement
Uncemented total hip arthroplasty; uncement total hip replacement; total hip arthroplasty, uncemented; total hip replacement, uncemented; uncemented total arthroplasty*
Direct mention for hybrid
Hybrid total hip arthroplasty; hybrid total hip replacement; total hip arthroplasty, hybrid; total hip replacement, hybrid; arthroplasty hybrid cemented*
Direct mention for reverse hybrid
Reverse hybrid total hip arthroplasty; reverse hybrid total hip replacement; total hip arthroplasty, reverse hybrid; total hip replacement, reverse hybrid
Exclusion keywords
Implant name: cemented [indirect mention for stem concept]; implant name: cemented [indirect mention for stem concept]; liner [indirect mention for cement concept]; components used*; stem SUMMIT cemented*

*Additional refinement from the external validation.

timing of an event (i.e., historical or present). For example, from a passage that reads, "The cement compressor was used. A 50-mm acetabular component was placed in 45 degrees of abduction and 15 degrees of anteversion," "cement" would be extracted as a procedure concept, and "acetabular component" would be extracted as a material concept along with the corresponding assertion status ("positive"), temporality ("present"), and the person who experienced the event ("associated with the patient"). Concept extraction is a knowledge-driven annotation and indexing process that is used to identify phrases

TABLE IV THA Bearing Surface-Related Keywords in Operative Notes

Indirect mention for polyethylene (poly) concept
Polyethylene; poly; plastic; ultra-high molecular weight polyethylene; highly crosslinked poly; highly crosslinked polyethylene; ply
Indirect mention for stem concept
Femoral stem; stem; HFx-stem; femoral component; femoral component/stem; permanent prosthesis; stem fem cemented
Indirect mention for metal concept
Metal; cobalt; chrome; cobalt-chromium; stainless steel; steel; polished
Indirect mention for ceramic concept
Ceramic; delta ceramic; zirconia; zirconia ceramic; BIOLOX delta; BIOLOX
Indirect mention for head concept
Femoral stem; stem; HFx-stem; femoral component; femoral component/stem; permanent prosthesis; stem fem cemented
Indirect mention for polyethylene brand concept
MARATHON; Longevity; Crossfire; Cross; Durasul; X3; ALTRX; Prolong; Vivacit-E; AOX; E1; ArCom
Direct mention for metal-on-poly
(Indirect mention for metal concept) on (indirect mention for polyethylene [poly] concept)
Direct mention for ceramic-on-poly
(Indirect mention for ceramic concept) on (indirect mention for polyethylene [poly] concept)
Direct mention for metal-on-metal
(Indirect mention for metal concept) on (indirect mention for metal concept)
Direct mention for ceramic-on-ceramic
(Indirect mention for ceramic concept) on (indirect mention for ceramic concept)

TABLE V Performance of the Operative Approach Algorithm

	Anterolateral (Gold Standard)	Direct Anterior (Gold Standard)	Posterior (Gold Standard)
Approach predicted by algorithm*			
Anterolateral	79	0	2
Direct anterior	0	35	0
Posterior	0	0	134
Total	79	35	136

*The algorithm had an accuracy of 99.2% (95% confidence interval, 97.1% to 99.9%).

referring to concepts of interests in an unstructured text[6]. After concepts are extracted from operative reports, they are normalized to specific categories. For instance, the concept "acetabular component" will be mapped to the category "finding—material." All keywords and phrases are applied to the sentence level and are categorized by direct mention and indirect mention. The direct mention is defined as the mention of a specific concept directly indicating the final classification category without requiring additional information (e.g., mentions of "anterolateral approach"). The indirect mention is defined as a mention that cannot directly determine the final class (e.g., mentions of "femoral component"). Once direct and indirect concepts are extracted from a document, the rule engine will determine the final class. For example, if the report contains both mentions of "cemented stem" and "cemented shell," this case will be classified as cement fixation (Fig. 2). In addition, exclusion keywords were used to avoid potential false-positive cases. The full list of concepts, keywords, modifiers, and disease categories for the operative approach, fixation methods, and bearing surface categories are listed in Tables II, III, and IV, respectively.

The rule engine independently processes all 3 tasks (approach, fixation technique, and bearing surface) and outputs the classification results (Fig. 2). First, the engine processes all direct mentions. The reports with no direct mentions subsequently have the rules on indirect mentions applied. The rules for indirect mentions contain a series of conditional clauses including "and," "or," and "not." For example, patients with non-negated findings of "stem" and "cement" and without mentions of "cup" are classified as having hybrid fixation.

The infrastructure for the rule-based NLP system was implemented with use of the open-source NLP pipeline MedTaggerIE[7], a resource-driven open-source Unstructured Information Management Architecture (UIMA)[8]-based inference engine framework. The system separates domain-specific NLP knowledge engineering from the generic NLP process, which enables words and phrases containing clinical information to be directly coded by subject matter experts. This functionality allows efficient translation of NLP tools for use by different institutions. The tool has been utilized in the Electronic Medical Records & Genomics (eMERGE) consortium to develop NLP-based phenotyping algorithms[9].

External Validation

We retrieved and annotated 422 operative notes from a different hospital setting. The data set was split into refinement (n = 242) and test (n = 180) data sets. The refinement data set was used to adjust the operative notes suitable to NLP algorithms (i.e., format adjustment) and to refine NLP algorithms (i.e., adding keywords) for further performance improvement. The final performance was measured on the test data set.

TABLE VI Performance of the Fixation Algorithm

	Uncemented (Gold Standard)	Cemented (Gold Standard)	Hybrid (Gold Standard)	Reverse Hybrid (Gold Standard)
Fixation predicted by algorithm*				
Uncemented	90	2	10	0
Cemented	0	72	15	0
Hybrid	0	0	102	0
Reverse hybrid	0	0	0	0
Total	90	74	127	0

*The algorithm had an accuracy of 90.7% (95% confidence interval, 86.8% to 93.8%).

TABLE VII Performance of the Bearing Surface Algorithm

	Metal-on-Polyethylene (Gold Standard)	Ceramic-on-Polyethylene (Gold Standard)	Metal-on-Metal (Gold Standard)	Ceramic-on-Ceramic (Gold Standard)
Bearing surface predicted by algorithm				
Metal-on-polyethylene	72	0	1	1
Ceramic-on-polyethylene	0	66	0	4
Metal-on-metal	0	0	67	1
Ceramic-on-ceramic	0	5	0	67
Total	72	71	68	73

*The algorithm had an accuracy of 95.8% (95% confidence interval, 92.7% to 97.8%).

Statistical Analysis

The performance of each NLP algorithm was assessed with use of the gold-standard manually abstracted data from the Mayo Total Joint Registry and the manually abstracted data from the external-setting EHR. Performance was assessed through calculation of accuracy (i.e., the sum of the correct classifications divided by the total). Because of the multi-class organization and imbalance across groups, we used the micro-averages to assess accuracy of the algorithms[6].

Results

Operative Approach

The THA operative approach algorithm (which classified the operative approach into 3 categories as anterolateral, direct anterior, and posterior) had an accuracy of 99.2% (Table V). During the training phase of algorithm development, 45% of patients had conflicted findings, with ≥2 operative approach classes being detected in the operative reports. We therefore added a header detector to identify procedure-related descriptions. The majority of expressions in the header section of the operative notes were direct mentions. Therefore, prioritizing findings in the header section helped to resolve most cases with conflicting findings. One of the 2 cases in which an anterolateral procedure was falsely classified as posterior was due to a confusing statement by the surgeon that he "had planned for a posterolateral approach, but upon exposure, it was clear that she had a fairly significant chronic degenerative tear of the abductor tendons." The other false classification was due to a lack of direct and indirect identifiers for the anterolateral approach, with the algorithm choosing the keyword *external rotators* as the final prediction for classifying the procedure as having a posterior approach.

Fixation

The THA fixation algorithm (which classified fixation into 4 categories as uncemented, cemented, hybrid, and reverse hybrid) had an accuracy of 90.7% (Table VI). All 90 uncemented THAs were correctly classified. Of the 74 cemented THAs, 2 were falsely classified as uncemented, and in both cases the procedure had been performed with cemented all-polyethylene acetabular components. The highest number of false-negative results was observed in association with hybrid THA. Of the 127 hybrid THAs, 15 were falsely classified as cemented because the procedures had been performed with cemented liners. Furthermore, 10 hybrid THAs were falsely classified as uncemented because the component was not explicitly mentioned in the operative notes. Although our NLP algorithm included terms for reverse hybrid procedures, no such procedures were randomly selected for inclusion in the test data set.

Bearing Surface

The THA bearing surface algorithm (which classified the bearing surface into 4 categories as metal-on-polyethylene, ceramic-on-polyethylene, metal-on-metal, and ceramic-on-ceramic) had an accuracy of 95.8% (Table VII). In 5 cases, ceramic-on-polyethylene surfaces were falsely classified as ceramic-on-ceramic surfaces because of a lack of documentation of polyethylene-related concepts. In 1 case, a metal-on-metal surface was falsely classified as a metal-on-polyethylene surface because of the mention of the term "polyethylene metal liner." In 6 cases, ceramic-on-ceramic surfaces were falsely classified because of incorrect mentions of a liner. In other words, dictation and transcription error was the main reason for the falsely classified cases.

External Validation

We achieved an accuracy of 94.4% for operative approach, 95.6% for fixation technique, and 98.0% for bearing surface. These estimates were comparable with performance with Mayo Total Joint Registry data.

Discussion

Total joint arthroplasty research, policy, and surveillance efforts have been bolstered by large-scale data, primarily in the form of national and institutional registries. These repositories are rich sources of longitudinal information; nevertheless, they are usually the product of labor-intensive and cost-intensive manual abstraction and coding efforts from individuals with specialized knowledge. NLP tools are distinctive in their ability to extract critical information from raw text in EHRs. NLP algorithms offer a more sophisticated way of handling free text than simple keyword searches. First, an NLP algorithm can handle information context such as assertion (e.g., confirmed, negation, possible, hypothetical), status (e.g., present, historical), and experiencer (e.g., associated with patients or others/family members). Second, specific sections of clinical text (e.g., the postoperative diagnosis in the operative notes) can be identified to capture the data element under the desired section. Third, an NLP algorithm is able to identify time descriptions (date, time, duration, frequency, before-after), thereby incorporating temporal information. Last, an NLP algorithm allows development of user-specific rules with combinations of various conditions, enabling the implementation of complex criteria. As a proof of concept for the potential application of this technique, we examined its ability to correctly identify common elements in operative notes for THA. We also demonstrated that NLP algorithms that are developed internally can be applied to other hospital settings with comparable performance, showing external validity and adaptability[10,11].

NLP technology has been used in various industries as a means of efficiently abstracting data from large repositories of text. The medical field is perhaps ideally suited to leverage informatics technology as clinical notes remain the standard of communication and documentation. Notes generated by clinicians are typically free text, creating a challenging environment for automated data abstraction. Therefore, complete and accurate data are often reliant on manual abstraction by trained professionals, which is time-consuming and costly. NLP can automate data abstraction and can achieve high levels of accuracy, allowing efficient mining of data for broad applications. Importantly, many questions remain unasked simply because they are too cumbersome to answer if not pursued in a prospective fashion. NLP has the potential to obviate that obstacle in appropriate circumstances. Murff et al. recently evaluated the ability of NLP to identify postoperative complications in the EHRs of 2,974 patients[12]. They noted that NLP had higher sensitivity, but lower specificity, compared with patient-safety indicators based on discharge coding. At the very least, this finding demonstrates the ability of NLP to serve as a screening tool for queries of large data sets. In contrast, in all 3 of our algorithms, we achieved higher specificity than sensitivity. This finding is due to the fact that our outcomes were restricted to 3 or 4 discrete results. Furthermore, we developed the algorithms through an iterative process with continuous modification of key search terms and rules based on areas identified as consistent sources of algorithm failure. Specifically as it applies to orthopaedic registry data, an iterative process of this nature against a limited subset of manually abstracted data would be critical to obtain high-fidelity information in the absence of a gold standard for comparison across the entire cohort.

Although NLP techniques were applied to a narrow area (i.e., THA operative reports), the applications of this technology throughout orthopaedics, and medicine in general, are wide-ranging. NLP has been shown to have the ability to abstract medical data from EHRs in a variety of settings[13-18]. Furthermore, NLP has been successfully tested as a tool for enhancing real-time clinical decision-making[19,20]. The implications for research, policy, and patient care are vast and will continue to become more apparent and viable as this technology matures. In an era in which data are plentiful, but the ability to extract and interpret data is limited, NLP offers one potential solution to mitigate this critical barrier. With the widespread availability of EHRs and open-source NLP-based tools, data collection can potentially be more comprehensive and efficient in the future. The tools and algorithms described in the present report also will be provided on an open-source basis through Open Health Natural Language Processing (OHNLP) (https://github.com/ohnlp) to facilitate further development and application.

The present study must be interpreted in light of important limitations. First, we had the benefit of a well-developed and reliable institutional total joint arthroplasty registry to serve as the gold standard. Developing algorithms without a preestablished gold standard would be difficult and time-consuming for institutions without this resource. What seems most logical is to have algorithms created at centers with gold-standard registries and then disseminated for refinement to institutions interested in applying the technology to their own operative reports. A second limitation is that the algorithms initially were developed after review of operative reports at a single institution, with the structure of the reports being tailored to a specific EHR system. Different surgeons are apt to describe a procedure in a distinctive fashion; therefore, surveillance of reports from surgeons at our institution may create an inherent bias toward their manner of description. Although we have successfully demonstrated the external validity on EHRs in another hospital setting, additional work is necessary to demonstrate the transportability of algorithms. The present study leveraged the operative reports of 29 different surgeons. The algorithms will certainly improve as efforts are undertaken to apply them to operative reports from other centers as algorithm refinement is an ongoing iterative process. Finally, the success of NLP algorithms depends on the level of detail and accuracy of the medical records and operative notes.

NLP-enabled algorithms are a promising alternative to the current gold standard of manual chart review for data collection in orthopaedics. NLP algorithms demonstrated excellent accuracy in delineating common elements that are typically described in THA operative notes and showed a capacity to translate with high fidelity across multiple practice facilities. The present study provides a proof of concept for the use of NLP technology in research and registry development endeavors, as it reliably obtained data of interest in an expeditious and cost-effective manner. ■

Cody C. Wyles, MD[1]
Meagan E. Tibbo, MD[1]
Sunyang Fu, MHI[1]
Yanshan Wang, PhD[1]
Sunghwan Sohn, PhD[1]
Walter K. Kremers, PhD[1]

Daniel J. Berry, MD[1]
David G. Lewallen, MD[1]
Hilal Maradit-Kremers, MD, MSc[1]

[1]Departments of Orthopedic Surgery (C.C.W., M.E.T., D.J.B., D.G.L., and H.M.-K.) and Health Sciences Research (S.F., Y.W., S.S., W.K.K., and H.M.-K.), Mayo Clinic, Rochester, Minnesota

Email address for H. Maradit-Kremers: maradit@mayo.edu

ORCID iD for C.C. Wyles: 0000-0002-8629-7567
ORCID iD for M.E. Tibbo: 0000-0003-2756-1353
ORCID iD for S. Fu: 0000-0003-1691-5179
ORCID iD for Y. Wang: 0000-0003-4433-7839
ORCID iD for S. Sohn: 0000-0001-8256-2602
ORCID iD for W.K. Kremers: 0000-0001-5714-3473
ORCID iD for D.J. Berry: 0000-0003-0103-8003
ORCID iD for D.G. Lewallen: 0000-0003-3101-9129
ORCID iD for H. Maradit-Kremers: 0000-0003-3882-602X

References

1. Maradit Kremers H, Larson DR, Crowson CS, Kremers WK, Washington RE, Steiner CA, Jiranek WA, Berry DJ. Prevalence of total hip and knee replacement in the United States. J Bone Joint Surg Am. 2015 Sep 2;97(17):1386-97.

2. Kurtz SM, Ong KL, Lau E, Bozic KJ. Impact of the economic downturn on total joint replacement demand in the United States: updated projections to 2021. J Bone Joint Surg Am. 2014 Apr 16;96(8):624-30.

3. Berríos-Torres SI, Umscheid CA, Bratzler DW, Leas B, Stone EC, Kelz RR, Reinke CE, Morgan S, Solomkin JS, Mazuski JE, Dellinger EP, Itani KMF, Berbari EF, Segreti J, Parvizi J, Blanchard J, Allen G, Kluytmans JAJW, Donlan R, Schecter WP; Healthcare Infection Control Practices Advisory Committee. Centers for Disease Control and Prevention guideline for the prevention of surgical site infection, 2017. JAMA Surg. 2017 Aug 1;152(8):784-91.

4. Grosso LM, Curtis JP, Lin Z, Geary LL, Vellanky S, Oladele C. Hospital-level risk-standardized complication rate following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA): measure methodology report. 2012. https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/measure-methodology.html. Accessed 2019 Jul 25.

5. The Electronic Medical Records and Genomics (eMERGE) Network. 2019. https://emerge.mc.vanderbilt.edu/. Accessed 2019 Jul 25.

6. Manning CD, Schütze H. Foundations of statistical natural language processing. MIT Press; 1999.

7. Liu H, Bielinski SJ, Sohn S, Murphy S, Wagholikar KB, Jonnalagadda SR, Ravikumar KE, Wu ST, Kullo IJ, Chute CG. An information extraction framework for cohort identification using electronic health records. AMIA Jt Summits Transl Sci Proc. 2013 Mar 18;2013:149-53.

8. Ferrucci D, Lally A. UIMA: an architectural approach to unstructured information processing in the corporate research environment. J Nat Lang Eng. 2004;10(3-4):327-48.

9. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010 May 1;26(9):1205-10. Epub 2010 Mar 24.

10. Wi CI, Sohn S, Ali M, Krusemark E, Ryu E, Liu H, Juhn YJ. Natural language processing for asthma ascertainment in different practice settings. J Allergy Clin Immunol Pract. 2018 Jan - Feb;6(1):126-31. Epub 2017 Jun 19.

11. Sohn S, Wang Y, Wi CI, Krusemark EA, Ryu E, Ali MH, Juhn YJ, Liu H. Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions. J Am Med Inform Assoc. 2017 Nov 30;25(3):353-9. Epub 2017 Nov 30.

12. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, Dittus RS, Rosen AK, Elkin PL, Brown SH, Speroff T. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA. 2011 Aug 24;306(8):848-55.

13. Li Y, Salmasian H, Harpaz R, Chase H, Friedman C. Determining the reasons for medication prescriptions in the EHR using knowledge and natural language processing. AMIA Annu Symp Proc. 2011;2011:768-76. Epub 2011 Oct 22.

14. Salmasian H, Freedberg DE, Friedman C. Deriving comorbidities from medical records using natural language processing. J Am Med Inform Assoc. 2013 Dec;20(e2):e239-42. Epub 2013 Oct 31.

15. Byrd RJ, Steinhubl SR, Sun J, Ebadollahi S, Stewart WF. Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. Int J Med Inform. 2014 Dec;83(12):983-92. Epub 2013 Jan 11.

16. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, Liu S, Zeng Y, Mehrabi S, Sohn S, Liu H. Clinical information extraction applications: a literature review. J Biomed Inform. 2018 Jan;77:34-49. Epub 2017 Nov 21.

17. Sohn S, Clark C, Halgrim SR, Murphy SP, Chute CG, Liu H. MedXN: an open source medication extraction and normalization tool for clinical text. J Am Med Inform Assoc. 2014 Sep-Oct;21(5):858-65. Epub 2014 Mar 17.

18. Sohn S, Ye Z, Liu H, Chute CG, Kullo IJ. Identifying abdominal aortic aneurysm cases and controls using natural language processing of radiology reports. AMIA Jt Summits Transl Sci Proc. 2013 Mar 18;2013:249-53.

19. Sung SF, Chen K, Wu DP, Hung LC, Su YH, Hu YH. Applying natural language processing techniques to develop a task-specific EMR interface for timely stroke thrombolysis: a feasibility study. Int J Med Inform. 2018 Apr;112:149-57. Epub 2018 Feb 8.

20. Wagholikar KB, MacLaughlin KL, Henry MR, Greenes RA, Hankey RA, Liu H, Chaudhry R. Clinical decision support with automated text processing for cervical cancer screening. J Am Med Inform Assoc. 2012 Sep-Oct;19(5):833-9. Epub 2012 Apr 29.