

Original Paper

# Natural Language Processing–Enabled and Conventional Data Capture Methods for Input to Electronic Health Records: A Comparative Usability Study

David R Kaufman<sup>1</sup>, PhD; Barbara Sheehan<sup>2</sup>, NP, PhD; Peter Stetson<sup>3</sup>, MA, MD; Ashish R Bhatt<sup>4</sup>, MBA; Adele I Field<sup>4</sup>, MFA; Chirag Patel<sup>5</sup>, MD, PhD; James Mark Maisel<sup>4</sup>, MD

<sup>1</sup>Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ, United States

<sup>2</sup>Health Strategy and Solutions, Intel Corp, Santa Clara, CA, United States

<sup>3</sup>Internal Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, United States

<sup>4</sup>ZyDoc Medical Transcription LLC, Islandia, NY, United States

<sup>5</sup>Department of Neurology & Neurological Sciences, Stanford School of Medicine, Stanford University, Palo Alto, CA, United States

**Corresponding Author:**

David R Kaufman, PhD

Department of Biomedical Informatics

Arizona State University

13212 East Shea

Scottsdale, AZ, 85260

United States

Phone: 1 4808840250

Fax: 1 480 884 0239

Email: [dave.kaufman@asu.edu](mailto:dave.kaufman@asu.edu)

## Abstract

**Background:** The process of documentation in electronic health records (EHRs) is known to be time consuming, inefficient, and cumbersome. The use of dictation coupled with manual transcription has become an increasingly common practice. In recent years, natural language processing (NLP)–enabled data capture has become a viable alternative for data entry. It enables the clinician to maintain control of the process and potentially reduce the documentation burden. The question remains how this NLP-enabled workflow will impact EHR usability and whether it can meet the structured data and other EHR requirements while enhancing the user's experience.

**Objective:** The objective of this study is evaluate the comparative effectiveness of an NLP-enabled data capture method using dictation and data extraction from transcribed documents (NLP Entry) in terms of documentation time, documentation quality, and usability versus standard EHR keyboard-and-mouse data entry.

**Methods:** This formative study investigated the results of using 4 combinations of NLP Entry and Standard Entry methods ("protocols") of EHR data capture. We compared a novel dictation-based protocol using MediSapien NLP (NLP-NLP) for structured data capture against a standard structured data capture protocol (Standard-Standard) as well as 2 novel hybrid protocols (NLP-Standard and Standard-NLP). The 31 participants included neurologists, cardiologists, and nephrologists. Participants generated 4 consultation or admission notes using 4 documentation protocols. We recorded the time on task, documentation quality (using the Physician Documentation Quality Instrument, PDQI-9), and usability of the documentation processes.

**Results:** A total of 118 notes were documented across the 3 subject areas. The NLP-NLP protocol required a median of 5.2 minutes per cardiology note, 7.3 minutes per nephrology note, and 8.5 minutes per neurology note compared with 16.9, 20.7, and 21.2 minutes, respectively, using the Standard-Standard protocol and 13.8, 21.3, and 18.7 minutes using the Standard-NLP protocol (1 of 2 hybrid methods). Using 8 out of 9 characteristics measured by the PDQI-9 instrument, the NLP-NLP protocol received a median quality score sum of 24.5; the Standard-Standard protocol received a median sum of 29; and the Standard-NLP protocol received a median sum of 29.5. The mean total score of the usability measure was 36.7 when the participants used the NLP-NLP protocol compared with 30.3 when they used the Standard-Standard protocol.

**Conclusions:** In this study, the feasibility of an approach to EHR data capture involving the application of NLP to transcribed dictation was demonstrated. This novel dictation-based approach has the potential to reduce the time required for documentation

and improve usability while maintaining documentation quality. Future research will evaluate the NLP-based EHR data capture approach in a clinical setting. It is reasonable to assert that EHRs will increasingly use NLP-enabled data entry tools such as MediSapien NLP because they hold promise for enhancing the documentation process and end-user experience.

(*JMIR Med Inform* 2016;4(4):e35) doi:[10.2196/medinform.5544](https://doi.org/10.2196/medinform.5544)

## KEYWORDS

electronic health records; natural language processing; medical transcription; user-computer interface

## Introduction

### Electronic Health Records and Data Entry

Electronic health records (EHRs) permeate most medical practices in the United States [1]. A promising feature of EHRs is that they provide machine-readable structured data that can be stored electronically, so that patient-centered information can be reviewed, retrieved, reported, and shared in real time to facilitate patient care. Although narrative data entry supports a measure of flexibility, structured data entry confers a number of advantages such as ready access to clinical decision support and interoperability between EHRs and health information exchanges. To achieve the full complement of these benefits, health care providers must generate clinical notes and reports in both human-readable and machine-readable formats. This adds effort to the documentation workflow [2,3] and requires new computer skills of physicians. The additional work required has led to a growing number of data entry alternatives [4], which is the subject of this paper.

Since their inception, EHRs “have been proposed as a means for improving availability, legibility, and completeness of patient information” [5]. The potential benefits of EHRs as instruments of patient care are widely recognized. Spurred by the 2009 US Health Information Technology for Economic and Clinical Health (HITECH) Act and accompanying incentives for providers to use EHRs, advancements in EHR technologies and implementation in the United States have grown rapidly. Approximately 78% of office-based physicians reported using some form of EHR in 2013 [6]. The role of EHRs is now considered integral to achieving federal health care goals, as expressed in the Meaningful Use mandate, for example. This has compelled physicians to adapt to new methods of documentation with concomitant changes to clinical workflow. This has resulted in great uncertainty about the impact of these requirements on the effective application of EHR systems [7].

As EHR implementations continue, physicians frequently express dissatisfaction with EHR documentation methods and usability [8]. Problems associated with usability impact not only the quality of patient records but can even contribute to compromised patient safety [9,10]. EHR documentation places ever-increasing demands on clinicians’ time, which contributes further to diminished quality of documents (eg, replete with irrelevant, redundant, and erroneous information) and physician dissatisfaction. EHR usability is a complex problem involving a multitude of factors [11-13], many of which are beyond the scope of this study. The focus in this paper is on the usability of data capture methods designed to enhance and potentially alleviate some of the burden resulting from manual input methods.

### Natural Language Processing–Based Solutions

Natural language processing (NLP) has emerged as a viable solution for clinical data capture. Many challenges remain for keyboard-and-mouse entry, namely, having to type text and negotiate the often unwieldy EHR interface to record information in structured fields. This is exacerbated by the fact that much of the EHR content continues to be unstructured [3,14]. A 2015 American Medical Informatics Association report identified time-consuming data entry as a problem with EHRs and recommended improving the EHR interface by allowing “multiple modes of data entry to accommodate provider preferences, including voice, typing, clicking, and handwriting recognition” [15].

Although most clinical information in EHRs is stored as unstructured data, such as clinical narrative, its electronic capture or retrieval has been challenging [16]. NLP has the potential to enable the clinician to reduce the documentation burden with the advantages of dictation—efficiency, usability, and quality—and also satisfy the needs for both machine-readable structured data and human-usable rich text in the EHR.

Problems associated with the time required for documentation and usability are well established. However, there is also evidence to suggest that quality of EHR documents (eg, progress reports) is problematic [17,18]. Physician documents often contain redundant, extraneous information or missing and inaccurate patient data [17]. EHR notes are not optimally used to either facilitate clinical communication or enhance patient care [8]. The measure of the quality and completeness of data in the EHR represents a challenging issue [19,20]. Stetson and colleagues [21] developed a tool for quantifying documentation quality, the Physician Documentation Quality Instrument (PDQI), and demonstrated its construct validity and internal consistency reliability. The initial tool consisted of 22 items and was subsequently reduced to a 9-item tool in order to facilitate its real-world application. The instrument can be used to assess the output quality of EHR note modules, as well as explain the components of document areas in need of improvement. Stetson and colleagues [21] assessed the interrater reliability using the intraclass correlation for consistency of average measures on the PDQI-9 total scores and found it to be 0.83 (CI 0.72-0.91). The tool can reliably be used to compare documentation methods and changes in quality resulting from a change in such methods.

In this study, we were focally concerned with testing NLP-enabled dictation-based data capture as a potential solution for relieving the increased burden of documentation. The benchmarks of performance include measures of time, data quality, and usability. According to Cimino [5], “Improvements

in the documentation process hold promise for more than simply reduced data entry effort and more readable notes. If impressions and plans can be captured as explicit data elements, using standard terminology, rather than being buried in the narrative text of a note, EHRs could use this information to better support clinical work flow.” As a result of physicians capturing explicit data elements, their clinical reasoning can be made more transparent and more easily available to colleagues caring for the same patient via electronic access to their EHR or data exchange.

### Data Capture Methods

A variety of modalities have been used for creating clinical documentation for EHR data capture or extraction to generate structured, actionable data (ie, data that are consumable, usable, reusable by a computer, and exchangeable with other computer systems in an efficient manner). These modalities include paper-based records transfer; verbal communication; direct entry or direct entry with macros; electronic templates; “Smart Forms”; dictation using speech recognition, sometimes known as voice recognition or continuous speech recognition; transcription or transcription with manual error correction; patient-recorded data (various methods); and hybrids, with or without NLP data capture, also termed “text processing” [22]. Rosenbloom et al observe that in spite of a “profusion of computer-based documentation (CBD) systems that promote real-time structured documentation,” it is a challenge “integrating clinical documentation into workflows that contain EHR systems.” They further note that health care providers prefer the ability to achieve a certain balance by both using a standardized note structure and having the flexibility to use expressive narrative text, facilitated by speech recognition. NLP systems afford that expressivity in developing a patient narrative as well as offering the capability to encode structured notes in a range of clinical document types and forms [22,23].

Figure 1 illustrates 5 alternative dictation-based EHR data capture methods. The NLP Entry method used in the study is shown in the center of the figure (labeled as 3), with bold arrows and boxes. Methods 1 and 4 show speech recognition and transcription, respectively, converting dictation to text that is inserted into the EHR. Methods 2 and 3 show NLP being applied to the speech-recognized and transcribed text, respectively, to generate structured data that are inserted into the EHR alongside the text. Method 5 shows a human scribe manually entering text and structured data into the EHR immediately in live time as the physician dictates or at a later time from recorded dictation.

NLP encompasses a family of methods for processing text. These methods have been used for a range of EHR applications [24] including information extraction [25], information retrieval [26], question answering, and text summarization [27]. NLP has also been used for the automatic encoding of narrative text into EHRs [4,28]. NLP and associated technologies used in conjunction with dictation for capturing and structuring medical data have advanced considerably in recent years [4,29].

Whereas relatively few NLP systems for structured clinical data capture are implemented outside academic medical centers [22], NLP is gaining more traction as a viable commercial technology for populating EHRs. In addition to the MediSapien NLP

(ZyDoc Medical Transcription LLC) application used in this study (the user interface for which is shown in Figure 2), there are a limited number of other NLP products being marketed or developed for use with EHR systems to enable NLP Entry (ie, free-text data capture, structuring, and EHR population). For example, both M\*Modal [30] and Nuance [31] offer dictation products with voice recognition that structure some data for EHRs. Certain EHR vendors, such as Allscripts, Greenway, and Cerner among others, have integrated the M\*Modal or Nuance technologies into their EHRs [32]. Other NLP-based research studies (including one on the interpretation of free-text Papanicolaou test reports for clinical decision support [33] and another on the use of “cognitive analytic tools to gain insight from all types of healthcare information,” including “knowledge-driven decision support” and “data-driven decision support” [34]) demonstrate the increasing importance of NLP in generating and analyzing structured health data.

The NLP engine used by the MediSapien NLP data capture application is the Medical Language Extraction and Encoding System (MedLEE), which was developed at Columbia University in the Department of Biomedical Informatics. MedLEE accepts unstructured clinical text inputs and outputs structured clinical information in a variety of formats [35]. Utilizing clinical lexicons, it is able to normalize clinical concepts in the text to conform to various standard terminologies. It is also able to identify, among other attributes, negation, degrees of certainty, temporal data, and results associated with the identified clinical concepts. MedLEE has been used for a number of data extraction purposes but was not specifically optimized for generating clinical documentation [36]. A commercially available version of MedLEE is now licensed and maintained by Health Fidelity under the product name REVEAL.

Developed by ZyDoc Medical Transcription, the MediSapien NLP data capture application allows doctors to use unstructured dictation to capture structured data in the EHR. MediSapien NLP preprocesses documents, leverages the MedLEE NLP engine, and postprocesses the NLP output using patent-pending processes that augment the NLP engine’s output. It also enables a workflow by which (1) the physician dictates, (2) the dictation is transcribed or subjected to speech recognition, (3) MediSapien NLP generates structured data from the transcription, and (4) the structured data and text are inserted into the EHR, although we simulated the EHR interface in the study.

Figure 2 shows part of a screen from the MediSapien NLP application in which source text is displayed on the left, with medical concepts highlighted, and structured data generated from the source text are displayed on the right. The document included in Figure 2 was selected to illustrate the volume of structured data generated by the MediSapien NLP application; the text was not produced as part of the study presented in this paper. As an indication of the volume of data generated by MediSapien NLP, the average number of clinical concepts and corresponding modifiers identified in a sample of notes from the study using the NLP-NLP protocol was 392. These concepts are coded in various standard terminologies—including ICD-10-CM (International Classification of Diseases, Tenth Revision, Clinical Modification); ICD-9-CM (International

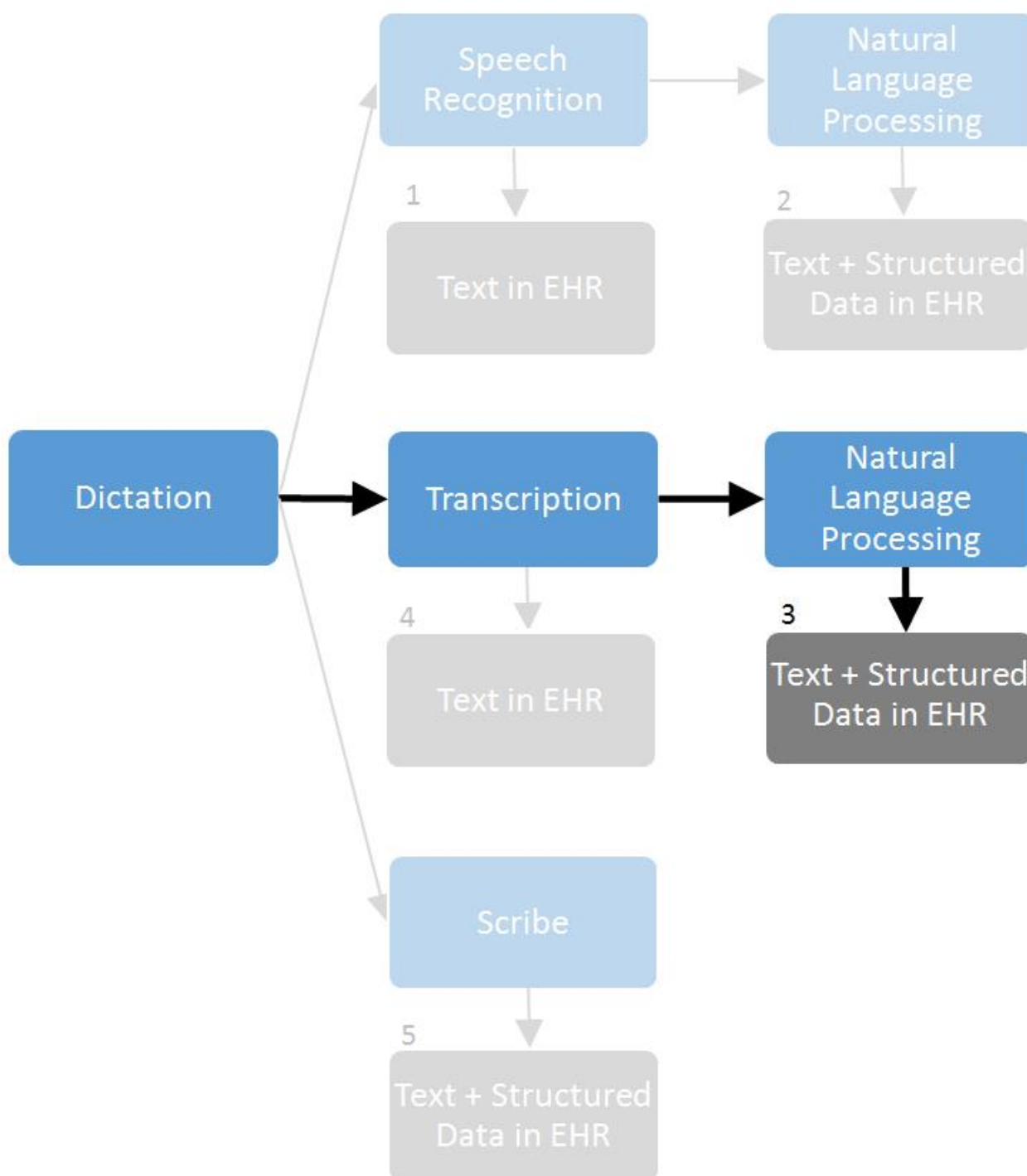
Classification of Diseases, Ninth Revision, Clinical Modification); SNOMED-CT (Systematized Nomenclature of Medicine, Clinical Terms); RxNorm; LOINC (Logical Observation Identifiers Names and Codes); and CPT (Current Procedural Terminology)—depending on the type of concept identified. Modifiers are structured data elements that provide additional properties related to a clinical concept. Examples of modifiers include body location, status, and dose (as shown on the right side of Figure 2).

It should be noted that this was a formative study designed to investigate the comparative effects of data capture methods

enabled by the NLP system. The focus of the analysis was on characterizing interactive behavior and system usability rather than the NLP method. Future studies will investigate the efficacy of the NLP processes used by the system.

The objectives of this study were to (1) measure the effects, relative to using Standard Entry only, of using 3 NLP-based documentation protocols on EHR documentation time and quality and (2) measure the effects of an NLP Entry-based protocol and a Standard Entry-based protocol on the usability of the documentation process.

Figure 1. Five dictation-based electronic health record (EHR) data capture methods.



**Figure 2.** MediSapien NLP application user interface, illustrating the volume of structured data generated by MediSapien NLP. NLP: natural language processing.

The screenshot displays the MediSapien NLP application interface. At the top, patient information is shown: Job Id: 47508, Date of Service: 01/01/1754, Patient Last: [redacted], Patient First: [redacted], Patient MI: [redacted], Patient DOB: 01/01/1754. A filter menu on the right includes checkboxes for 'Prob', 'Proc', 'Med', 'Phy', 'Findings', 'Lab', 'Demo', 'Allergy', 'All', 'ICD-9', 'ICD-10', 'LOINC', 'RXNORM', 'SNOMED', 'CPT-4', and 'Expand All Concept'. The main content area shows a clinical note with several sections highlighted in yellow: 'Weakness, fatigue, and tiredness...', 'PHYSICAL EXAM:', 'DIAGNOSTIC STUDIES:', 'ASSESSMENT:', and 'TREATMENT PLAN:'. The 'ASSESSMENT:' section lists: '1. Atrial Fibrillation with rapid ventricular rate 2. Pneumonia 3. Congestive heart failure'. The 'TREATMENT PLAN:' section lists: '1. We will start her on a low dose of beta-blocker for rate control and antibiotic for pneumonia. Once, if she is stable, we will consider further cardiac workup. 2. We will also obtain an echocardiogram to assess the prior valve surgery. 3. Discussed with the patient's husband regarding plan of care, fully understood, and consents for the same. All the questions answered in detail.' On the right side, a structured data panel titled 'assessment' and 'assessment and plan' lists various medical concepts with their corresponding codes (ICD-9, ICD-10, SNOMED) and checkboxes, such as 'atrial fibrillation', 'ventricular rate', 'pneumonia', and 'congestive heart failure'.

## Methods

### Study Overview

The study evaluated an NLP-enabled solution for documentation. Specifically, we focused on three problem areas related to EHR data capture: (1) efficiency, including time required for data capture; (2) effectiveness, encompassing documentation quality; and (3) physician satisfaction, based on usability. We compared a novel dictation-based protocol using MediSapien NLP for structured data capture (“NLP-NLP”) against a standard, keyboard-and-mouse structured data capture protocol where the study participant was instructed to generate EHR documentation as in normal clinical practice (“Standard-Standard”) as well as 2 novel hybrid protocols (“NLP-Standard” and “Standard-NLP”) to determine which protocols provided better results in terms of data capture time, documentation quality, and physician satisfaction. The hybrid protocols were included because we anticipated that mixed forms and modalities of interaction may serve as realistic alternatives to a one-dimensional NLP approach or standard data entry. For example, certain parts of clinical notes may be better served by one modality of entry or the other; a note’s assessment and plan sections are often more given to free text and may therefore be suited to a dictation-based modality, whereas a note’s history and physical examination sections are less so and therefore may be better suited to a different modality. A hybrid approach may offer greater flexibility and can be adapted to the preferences of individual users. The study

presented here is formative work that focused more directly on the nature of user interaction and the user experience rather than the efficacy or precision of the NLP system or the system for insertion of data in the EHR. These will be addressed in the next phase of research.

### Study Design

This study contrasted 4 conditions involving combinations of NLP Entry and Standard Entry (referred to in this paper as documentation protocols) on the following measures: documentation time, documentation quality, and usability of the documentation process.

The Standard Entry method (ie, how physicians typically use an EHR to document) entailed using a keyboard and mouse for typing text and negotiating the graphical user interface (eg, drop-down menus, check-boxes) to record information in structured fields.

In the NLP Entry method, the participants dictated the content of the documents. They did not enter any documentation using the keyboard or mouse. Their dictation was transcribed, and the transcription was inputted into the MediSapien NLP application. That application outputted a document containing structured data (an example of which is shown in Figure 2) generated from the transcription. Finally, following precise instructions, study assistants entered the transcribed text and part of the generated structured data into a Microsoft Word document to produce a final note.

**Table 1.** Documentation methods used for each documentation protocol.

Documentation protocols	Documentation method for history and physical examination	Documentation method for assessment and plan
NLP <sup>a</sup> -NLP	NLP Entry	NLP Entry
NLP-Standard	NLP Entry	Standard Entry
Standard-NLP	Standard Entry	NLP Entry
Standard-Standard (control)	Standard Entry	Standard Entry

<sup>a</sup>NLP: natural language processing.

In the study, each physician was asked to document 4 notes using 4 methods including 1 control method (Standard-Standard protocol) and 3 experimental protocols consisting of combinations of NLP Entry and Standard Entry for documenting different parts of the note, as presented in [Table 1](#). The Standard-NLP protocol involved using Standard Entry to generate the history and physical examination sections and NLP Entry to generate the assessment and plan sections. The NLP-Standard protocol involved using NLP Entry for the history and physical examination sections and Standard Entry for the assessment and plan sections. The NLP-NLP protocol involved using NLP Entry for the entire note. The order in which the protocols were used was randomized for each participant.

## Participants

Physician participants were recruited through referrals. Two of the coauthors (BS and PS) referred us to several physicians who in turn made additional referrals. The inclusion criteria for the participants were as follows: (1) must be a neurologist, cardiologist, or nephrologist, the 3 specialties included in the study; (2) must be a senior resident, fellow, or attending; and (3) must be a current user of the Columbia University Medical Center's (CUMC) Crown Allscripts EHR (Chicago, IL). The participants were each compensated US \$500 for their efforts.

## Setting

This study was conducted at CUMC. The test protocol was administered with physician participants at their offices. Fictitious patients were created for the study, and the participants documented their cases in a test environment of the Crown Allscripts EHR, which was the same EHR in which the participants documented during normal clinical practice. Participants were all experienced users of the system. The Crown Allscripts EHR had been in use in excess of 5 years at CUMC as of the time of the study. This test environment contained the same custom templates that participants used during normal clinical practice. As a result, the Standard Entry method simulated documentation during normal clinical practice as closely as reasonably possible.

## Test Scripts

The test scripts were based on anonymized transcription documents that were modified by 4 expert clinicians (2 fellows and 2 attending physicians). These clinicians were not participants in the study. The test scripts consisted of history

and physical examination sections but excluded assessment and plan sections. After reviewing test scripts that described cases of the fictitious patients, the participants generated 4 multisectional consultation or admission notes using 4 documentation protocols ([Table 1](#)).

## Procedure

First, each participant read the instructions for generating consultation or admission notes based on the 4 provided test scripts, an example of which is shown in [Figure 3](#). The instructions indicated that the participant must generate documentation without copying verbatim any part of the test script and that the assessment and plan sections of these notes would be generated based on the participant's medical judgment.

Second, the participant was asked to review the test scripts and to generate 4 notes from 4 test scripts, 4 examples of which are shown in [Figures 4-7](#).

Third, for documentation in which NLP Entry was used, the participant's dictation was transcribed; the transcription was processed by MediSapien NLP; and the transcription, structured data generated, and any documentation generated for the note by Standard Entry (if applicable) were combined to create the final note. A simulated interface and simulated note were used for NLP Entry: following a protocol, study assistants copied the generated unstructured and structured data into a Microsoft Word document to generate the final note. For Standard Entry, an actual EHR interface was used.

Finally, after reviewing their final notes, the participants completed 2 System Usability Scale (SUS) surveys [37] to evaluate the usability of the NLP-NLP protocol and the Standard-Standard protocol. Given the limited availability of clinicians' time, we determined this would be the most important contrast to include in the study. The SUS is a widely used and reliable tool. It consists of 10 Likert items measured on a 5-point scale (ranging from "completely agree" to "completely disagree") [38]. Half of the items are framed as positive questions (eg, "easy to use") and half are negative (eg, "unnecessarily complex"). The scores were tabulated accordingly. The surveys were made available in SurveyMonkey, a Web-based survey application. The SUS was slightly modified for language and context. The questions with tabulated responses are presented in [Table 2](#).

**Table 2.** Summary of usability scores (mean, SD) and paired *t* test comparisons between use of the Standard-Standard and the NLP-NLP protocols (n=23 cases); scores have been normalized such that higher scores indicate greater usability.

Usability question	Standard-Standard, mean (SD)	NLP <sup>a</sup> -NLP, mean (SD)	<i>P</i> value
I think that I would like to use this method frequently for admitting notes.	2.9 (0.9)	3.3 (0.8)	.21
I found this method unnecessarily complex.	2.5 (1.4)	3.8 (0.8)	.003
I thought this method was easy to use.	2.8 (1)	4.2 (0.6)	<.001
I think that I would need assistance to be able to use this method.	3.3 (1.1)	3.6 (0.9)	.24
I found the various functions in the processes of the method were well integrated.	2.6 (0.9)	3.2 (1)	.05
I would imagine that most people would learn to use this method very quickly.	3.0 (0.9)	3.8 (0.7)	.01
I found this method very cumbersome/awkward to use.	2.6 (1.1)	3.7 (0.9)	.004
I felt very confident using this method.	3.6 (0.8)	3.4 (0.8)	.43
I would need to learn a lot of things before I could get going with this method.	3.6 (1)	3.8 (0.8)	.40
I feel the method would fit well in my existing workflow.	2.8 (0.9)	3.4 (0.9)	.08

<sup>a</sup>NLP: natural language processing.

**Figure 3.** Example of a neurology test script.**Documentation Method:****Test Subject ID:****Patient ID:****MRN:** 11111**DOS:** November 1, 2007**Job Type:** Admission Note**BACKGROUND INFORMATION:**

You were contacted by the neurologist caring for the patient who sent the patient for an elective admission with possible adult hydrocephalus. You are asked to write an admitting note.

**CHIEF COMPLAINT:**

Imbalance

**HISTORY OF PRESENT ILLNESS:**

This is a 74 year old right handed woman who was generally well until about 6 months ago. At that time, her husband noticed she had some difficulty with her balance. She saw her internist, Dr. Smith, who followed her and observed on serial visits that the balance was deteriorating slowly. Accordingly, Dr. Smith obtained a non-contrast CT scan of the head that was suggestive of ventriculomegaly. She was then seen in consultation in the office of a neurologist, Dr. Jones, at the request of Dr. Smith. Dr. Jones' office notes are available separately. However, she continued to worsen such that she now has difficulties with activities of daily living. With respect to her walking and balance, she states "I think I walk funny." Her husband has noticed over the last six months or so that she has broadened her base and become more stooped in her posture. Her balance has also gradually declined such that she frequently touches walls and furniture to stabilize herself. She has difficulty stepping up on to things like a scale because of this imbalance. She does not festinate. Her husband has noticed some slowing of her speed. She does not need to use an assistive device. She has occasional difficulty getting in and out of a car. Recently she has had more frequent falls. In March of 2007, she fell when she was walking to the bedroom and broke her wrist. Since that time, she has not had any emergency room trips, but she has had other falls. She denies any tremor. She denies any lateralizing incoordination. She has no difficulty with fine motor movements such as buttons. She has no bowel problems. She has no urinary frequency or urgency. The patient does not have headaches. She denies photophobia, phonophobia, diplopia, dysarthria, dysphagia, weakness, paresthesias, numbness, or seizures. With respect to thinking and memory, she states she is still able to pay the bills, but over the last few months she states, "I do not feel as smart as I used to be." She feels that her thinking has slowed down. Her husband states that he has noticed, she will occasionally start a sentence and then not know what words to use as she is continuing. The patient has not had trouble with syncope. She has had past episodes of vertigo, but not recently. Of note, she has not had any vertigo since the balance difficulty started. She also denies tinnitus or changes in hearing.

**PAST MEDICAL HISTORY:**

1. Significant for hypertension diagnosed in 2006, reflux in 2000, insomnia, but no snoring or apnea. She has been on Ambien, which is no longer been helpful.

2. She has had arthritis since year 2000, hypothyroidism diagnosed in 1968, a hysterectomy in 1986, and a right wrist operation after her fall in 2007 with a titanium plate and eight screws.
3. She has two normal vaginal deliveries.

**FAMILY HISTORY:**

Her father died with heart disease in his 60s and her mother died of colon cancer. She has a sister who she believes is probably healthy. She has had two sons one who died of a blood clot after having been a heavy smoker and another who is healthy.

**SOCIAL HISTORY:**

She lives with her husband. She is a nonsmoker and no history of drug or alcohol abuse. She does drink two to three drinks daily. She completed 12th grade. Danish is her native language, but she has been in the United States for many, many years and speaks fluent English, as does her husband. She has a Living Will and if unable to make decisions for herself, she would want her husband, Vilhelm to make decisions for her.

**ALLERGIES:**

Codeine and sulfa medications.

**CURRENT MEDS:**

Premarin 0.625 mg p.o. q.o.d., Aciphex 20 mg p.o. q. daily, Metoprolol 50 mg p.o. q. daily, Norvasc 5 mg p.o. q. daily, multivitamin, B-complex vitamins and vitamin C daily.

**PHYSICAL EXAM:**

On examination today, this is a pleasant and healthy appearing woman. Blood pressure 154/72, heart rate 87, and weight 153 pounds. Pain is 0/10. Head is normocephalic and atraumatic. Head circumference is 54 cm, which is in the 10-25th percentile for a woman who is 5 foot and 6 inches tall. Spine is straight and nontender. Spinous processes are easily palpable. She has very mild kyphosis, but no scoliosis. There are no neurocutaneous stigmata. Regular rate and rhythm. No carotid bruits. No edema. No murmur. Peripheral pulses are good. Lungs are clear. Assessed for recent and remote memory, attention span, concentration, and fund of knowledge. She scored 30/30 on the MMSE when attention was tested with either spelling or calculations. She had no difficulty with visual structures. Pupils are equal. Extraocular movements are intact without nystagmus. Face is symmetric. Tongue and palate are midline. Jaw muscles strong. Cough is normal. SCM and shrug 5 and 5. Visual fields intact. Dix-Halpike maneuver did not produce nystagmus or vertigo. Normal for bulk, strength, and tone. There was no drift or tremor. Intact for pinprick and proprioception. Normal for finger-to-nose and no disidiadokokinesis. Are 2+ throughout. No Myerson. Assessed using the Tinetti assessment tool. She was fairly quick, but had some unsteadiness and a widened base. She did not need an assistive device. I gave her a score of 13/16 for balance and 9/12 for gait for a total score of 22/28. No postural instability.

**DIAGNOSTIC RESULTS:**

MRI was reviewed from June 26, 2008. It shows mild ventriculomegaly with a trace expansion into the temporal horns. The frontal horn span at the level of foramen of Munro is 3.8 cm with a flat 3rd ventricular contour and a 3rd ventricular span of 11 mm. The Sylvian aqueduct is patent. There is no pulsation artifact. Her corpus callosum is bowed and effaced. She has a couple of small T2 signal abnormalities, but no significant periventricular signal change. There is transependymal absorption of

**Figure 4.** Example of part of the history and physical examination section of a neurology consultation note generated using the Standard-NLP protocol, illustrating the part of the note that was generated by Standard Entry. NLP: natural language processing.

**Patient:** AMIE2 TEST      **MRN:**      **DOB:**  
**Date of Service:** 05/24/2013

**Test Subject ID:**  
**Patient ID:**  
**MRN:**  
**DOS:**

#### Neuro

#### History of Present Illness

Patient presented in late June 2006 with sudden onset of blurred vision, diplopia, weakness (L arm > R), and L eye ptosis after a C. jejuni GI infection. She was admitted to the hospital and lumbar puncture showed increased protein, an EMG/NCS showed early signs of AIDP. She was treated with IVIg and had some improvement of her symptoms. Her vital capacities were normal during the hospitalization. She was then transferred to rehab and was discharged on July 20, 2006. her walking is better but she still has some weakness, blurry vision due abnormal eye movement and some tightness and pain in her mid-back.

#### Social History

Never A Smoker |  
Never Drank Alcohol  
Occupation: Retired

#### Current Meds

Fluticasone Propionate 50 MCG/ACT Nasal Suspension; Therapy: (Recorded:24May2013) to  
Gabapentin 100 MG Oral Capsule; Therapy: (Recorded:24May2013) to  
GliPIZIDE 10 MG Oral Tablet; Therapy: (Recorded:24May2013) to  
Insulin Purified NPH (Pork) SUSP; Therapy: (Recorded:24May2013) to  
Lasix 20 MG Oral Tablet; TAKE 1 TABLET TWICE DAILY; Therapy: (Recorded:24May2013) to  
Norvasc 10 MG Oral Tablet; Therapy: (Recorded:24May2013) to  
Percocet 5-325 MG Oral Tablet; Therapy: (Recorded:24May2013) to  
Protonix 40 MG Oral Tablet Delayed Release; Therapy: (Recorded:24May2013) to  
Toprol XL 50 MG Oral Tablet Extended Release 24 Hour; Therapy: (Recorded:24May2013) to  
Zocor 10 MG Oral Tablet; Therapy: (Recorded:24May2013) to

#### Allergies

Acnotex LOTN  
Bactrim DS TABS



**Figure 5.** Example of part of the history and physical examination section of a neurology consultation note generated using the NLP-Standard protocol, illustrating the part of the note that was generated by NLP Entry. NLP: natural language processing.

**Test Subject ID:**  
**Patient ID:**  
**MRN:**  
**DOS:**

### Neuro

#### Chief Complaint

Follow-up hospital discharge for Guillain-Barre syndrome.

#### History of Present Illness

This is a 62 year old right-handed woman with history of hypertension, diabetes, and history of stroke, right basal ganglia stroke in June 2006 with residual weakness. In late June 2006, she developed acute onset of blurry vision, diplopia, and weakness of the arms, right arm greater than left arm and left-sided ptosis, following diarrhea which she was diagnosed with *Campylobacter jejuni*. During the hospital admission, she was worked up with MRI, which showed old right basal ganglia infarct. She also had lumbar punctures done, which showed elevated protein. EMG and nerve conduction study was done, confirmed diagnosis of AIDP. She was treated with IVIG and subsequently showed mild improvement in her symptoms. She was discharged to rehab on 07/12/06. Since rehab discharge, she has made some progress in which she felt her walking has improved; however, she has not noticed improvement of her eye movement or blurry vision, and she also still have occasional tightness and pain in her back.

- Hypertension: History
- Diabetes: History
- Stroke: History, basal ganglia, Right, 2006, in
- Blurred vision: Acute, new
- Diplopia
- Weak: Arm
- Ptosis: Arm, left
- Diarrhea

#### Past Medical History

- Hypertension
- Diabetes: Type 2
- Stroke: Basal ganglia, right
- Guillain Barre syndrome: 2006, in

Hypertension.  
 Diabetes type 2.  
 Right basal ganglia stroke.

**Figure 6.** Example of part of the assessment and plan section of a neurology consultation note generated using the Standard-NLP protocol, illustrating the part of the note that was generated by NLP Entry. NLP: natural language processing.

#### Assessment

This patient had acute inflammatory demyelinating polyneuropathy after *C. jejuni* diarrheal infection. She is now recovering and just was discharged from rehab. She continues to have persistent bilateral ophthalmoparesis, which seems to be her main problem at this time.

- Polyneuropathy: Demyelination, inflammation, acute
- Infection: Diarrhea, recover
- Ophthalmoparesis: Bilateral, continue, persistent

#### Plan

- Physical therapy: Outpatient, continuous, future

We will refer her for continuous outpatient physical therapy. We will see her back in a month.

**Figure 7.** Example of part of the assessment and plan section of a neurology consultation note generated using the NLP-Standard protocol, illustrating the part of the note that was generated by Standard Entry. NLP: natural language processing.

#### Assessment

62 year-old right-handed woman with h/o R basal ganglia stroke and GBS.  
 1. GBS

#### Discussion

Much improved since last hospitalization, still has some residual weakness from stroke, and left CN3 palsy-questionable from DM

#### Plan

-will refer pt to see ophthalmologist for full evaluation and treatment  
 -PT/OT for weakness and gait problem  
 -RTC in 6 months

**Measures of Analysis**

***Time, Documentation Quality, and Usability***

The time required for documentation was measured using a stopwatch.

The 4 expert clinicians (2 fellows and 2 attending physicians) were not participants and were blind to the protocols used to generate the documentation they evaluated. They were provided with gold standard versions of the test documentation they were asked to evaluate and told that the gold standard versions represented “high quality notes.” They were then instructed to measure documentation quality by comparing participants’ final test documentation against the gold standard versions of that documentation using the PDQI-9 tool [21], shown in Figure 8. The expert clinicians were given minimal background on the purpose for the study. They were independent and highly trained in their specialties (and they only graded documents within their own domains). In addition, they were provided with clear instructions and had a sound understanding of the PDQI-9 tool, which is known to be a reliable instrument [21]. Unfortunately,

it was not practical to test interrater reliability. We used only 8 of the 9 PDQI-9 prompts (items) because one of the prompts required a judgment of whether the documentation was up-to-date and this was not meaningful in this particular context.

The gold standard versions of the documentation were generated by the expert clinicians in Microsoft Word. They produced these documents from clinical notes and modified them so that they were consistent with the clinical profile of the patient (ie, the patient’s assessment and treatment were consistent and derivable from history and physical examination findings). The expert clinicians were instructed to ensure that all elements of the documents were internally consistent and that they truly reflected a gold standard. The expert clinicians were compensated at a rate of US \$125 per hour for their efforts. We did not have access to the interim work product of the expert clinicians. We were only provided with the expert clinicians’ grades.

The usability of the documentation processes was assessed using a modified version of the SUS [38]. Each participant completed 1 SUS questionnaire for each of the NLP-NLP and Standard-Standard protocols.

**Figure 8.** Physician Documentation Quality Instrument (PDQI-9) tool.

**Please review the gold standard note before assessing the test subject's note. Then rate the test subject's note on each of the the following attributes:**

	1 (Not at all)	2	3	4	5 (Extremely)
1) ACCURATE. The note is true. It is free of incorrect information.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2) THOROUGH. The note is complete and documents all of the issues of importance to the patient.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3) USEFUL. The note is extremely relevant, providing valuable information and/or analysis.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4) ORGANIZED. The note is well-formed and structured in a way that helps the reader understand the patient's clinical course.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5) COMPREHENSIBLE. The note is clear, without ambiguity or sections that are difficult to understand.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6) SUCCINCT. The note is brief, to the point, and without redundancy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7) SYNTHESIZED. The note reflects the author's understanding of the patient's status and ability to develop a plan of care.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8) INTERNALLY CONSISTENT. No part of the note ignores or contradicts any other part.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Statistical Analysis

Data were analyzed using Intercooled Stata version 9.2 (StataCorp LP). Demographics were tabulated in regard to participants' years of EHR experience, years of experience with dictation, the number of cases per subject area, the frequency of use of each of the 4 protocols, dictation time, usability scores, and quality scores. The Shapiro-Wilk *W* test was used to determine whether continuous variables were normally distributed. Results are presented as mean (SD) or analysis of variance (ANOVA) results, median (interquartile range), or percentage; respectively, comparisons were made using *t* test, Wilcoxon rank sum analysis, or chi-square analysis.

Pearson correlation was performed on continuous variables. The association of years of EHR experience, years of experience with dictation, the 4 protocols (Standard-Standard, Standard-NLP, NLP-Standard, and NLP-NLP), and the 3 subject areas (cardiology, nephrology, and neurology) with the note dictation time was assessed using ANOVA.

Statistical significance was defined as  $\alpha=0.05$  and Bonferroni correction was used where applicable for multiple comparisons.

### Human Subjects Protection

The study was approved by the Institutional Review Board at Columbia University (#AAAK2458). All participants gave

consent before their participation and were fully briefed on the true objectives of the study. The study protocol adhered to strict standards of confidentiality and privacy.

### Results

A total of 31 unique individuals documented 3.8 (SD 0.7) notes on average. Of these, 28 participants completed all 4 protocols, 2 participants completed 2 protocols, and 1 participant completed 1 protocol. The participants who did not complete all 4 protocols were called away from the study and therefore could not finish the task. These individuals had an average EHR experience of 6.6 (SD 3.4) years (data were available for 30 individuals) and an average dictation experience of 2.8 (SD 5.6) years (data were available for 29 individuals). There was a significant association between years of EHR experience and years of dictation experience ( $r=0.47$ ,  $P=0.01$ ).

A total of 118 notes were documented across the 3 subject areas of cardiology (22/118, 18.6%), nephrology (21/118, 17.8%), and neurology (75/118, 63.6%). The Standard-Standard, Standard-NLP, NLP-Standard, and NLP-NLP protocols were used in 28/118 (23.7%), 28/118 (23.7%), 30/118 (25.4%), and 32/118 (27.1%) documented notes, respectively. The frequency of use of the 4 protocols was balanced across the 3 subject areas (see Table 3).

**Table 3.** Frequency of use of the 4 protocols by subject area for each documented note.

Protocol	Documented cardiology notes, n (%)	Documented nephrology notes, n (%)	Documented neurology notes, n (%)	Total number of documented notes, n (%)
Standard-Standard	5 (23)	5 (24)	18 (24)	28 (23.7)
Standard-NLP <sup>a</sup>	5 (23)	4 (19)	19 (25)	28 (23.7)
NLP-Standard	6 (27)	5 (24)	19 (25)	30 (25.4)
NLP-NLP	6 (27)	7 (33)	19 (25)	32 (27.1)
Total	22	21	75	118

<sup>a</sup>NLP: natural language processing.

**Table 4.** Median documentation time in minutes, with interquartile ranges, by protocol and subject area.

Protocol	Median (IQR <sup>a</sup> ) time to document cardiology note (minutes)	Median (IQR) time to document nephrology note (minutes)	Median (IQR) time to document neurology note (minutes)
Standard-Standard	16.9 (16.5-19.7)	20.7 (18.6-23.2)	21.2 (17.6-29.9)
Standard-NLP <sup>b</sup>	13.8 (13.0-17.2)	21.3 (14.5-29.8)	18.7 (16.0-22.9)
NLP-Standard	7.5 (7.1-9.1)	12.1 (10.7-12.2)	11.0 (8.5-14.6)
NLP-NLP	5.2 (4.7-8.0)	7.3 (6.6-9.1)	8.5 (6.4-11.4)

<sup>a</sup>IQR: interquartile range.

<sup>b</sup>NLP: natural language processing.

**Table 5.** Interprotocol comparisons (Wilcoxon rank sum analysis).

Interprotocol comparisons	Statistical analysis of time difference ( <i>P</i> value <sup>a</sup> )		
	Cardiology notes	Nephrology notes	Neurology notes
Standard-Standard vs Standard-NLP <sup>b</sup>	.60	.81	.20
Standard-Standard vs NLP-Standard	.01	.03	<.001
Standard-Standard vs NLP-NLP	.006	.005	<.001
Standard-NLP vs NLP-Standard	.006	.05	.001
Standard-NLP vs NLP-NLP	.006	.008	<.001
NLP-Standard vs NLP-NLP	.11	.02	.02

<sup>a</sup>Statistical significance level: alpha=.0083 after Bonferroni correction.

<sup>b</sup>NLP: natural language processing.

**Table 6.** Document quality for each protocol (median values are presented).

Protocols and statistical comparisons	Document quality metrics <sup>a</sup>								
	A	T	U	O	C	S	Sy	I	Sum
<b>Protocol, median score</b>									
Standard-Standard (n=24)	3.5	3	4	4	4	4	4	4	29
Standard-NLP <sup>b</sup> (n=24)	4	4	4	3.5	4	2.5	4	4	29.5
NLP-Standard (n=27)	4	3	3	3	3	3	4	4	26
NLP-NLP (n=30)	4	4	3	3	3	2	3	4	24.5
<b>Interprotocol comparisons,</b>									
<b><i>P</i> value<sup>c</sup></b>									
Standard-Standard vs Standard-NLP		.04		.03			<.001		
Standard-Standard vs NLP-Standard					.04	.006			
Standard-Standard vs NLP-NLP				.002	.02	<.001	.03		
Standard-NLP vs NLP-Standard						.005			
Standard-NLP vs NLP-NLP							.02		
NLP-Standard vs NLP-NLP				.03		.001			

<sup>a</sup>The 8 document quality metrics are as follows: Accurate, Thorough, Useful, Organized, Comprehensible, Succinct, Synthesized, and Internally Consistent.

<sup>b</sup>NLP: natural language processing.

<sup>c</sup>Statistical significance level: alpha=.0083 after Bonferroni correction.

The documentation times were not normally distributed ( $z=4.6$ ); thus, comparison of documentation times was performed using Wilcoxon rank sum analysis of medians. Table 4 summarizes the median time and interquartile range for the documentation of each note by protocol and subject area. Table 5 presents statistical analysis of interprotocol times. In each subject area, the NLP-NLP protocol required significantly less documentation time compared with either the Standard-Standard or Standard-NLP protocol. Compared with the Standard-Standard protocol, the NLP-Standard protocol required significantly less documentation time for the neurology subject area. Compared with the Standard-NLP protocol, the NLP-Standard protocol required significantly less documentation time for the cardiology and neurology subject areas.

On the basis of the ANOVA of documentation time, the model was statistically significant (adjusted  $R^2=.54$ ,  $P<.001$ ). This indicates that, taken together, the input variables used in the ANOVA model (EHR experience; years of experience with dictation; the 4 protocols, Standard-Standard, Standard-NLP, NLP-Standard, and NLP-NLP; and the 3 subject areas, cardiology, nephrology, and neurology) accounted for 54% of the variance in documentation time, the outcome variable. The factors significantly associated with documentation time included the protocol method ( $P<.001$ ), subject area ( $P=.009$ ), and the number of years of EHR experience ( $P=.047$ ) but not the number of years of dictation experience ( $P=.77$ ).

Document quality was assessed using 8 observed PDQI-9 metrics (Figure 8). Median values of the document quality metrics are presented for each protocol, in Table 6. Statistical

comparisons across the protocols are also presented in [Table 6](#). The significant differences among the protocols occurred within the “Organized” metric (Standard-Standard vs NLP-NLP, 4 vs 3,  $P=.002$ ) and the “Succinct” metric (Standard-Standard vs Standard-NLP, 4 vs 2.5,  $P<.001$ ; Standard-Standard vs NLP-Standard, 4 vs 3,  $P=.006$ ; Standard-Standard vs NLP-NLP, 4 vs 2,  $P<.001$ ; Standard-NLP vs NLP-Standard, 2.5 vs 3,  $P=.005$ ; and NLP-Standard vs NLP-NLP, 3 vs 2,  $P=.001$ ).

The usability data were analyzed using a paired  $t$  test in a subset of 23 cases ( $n=5$  cardiology,  $n=5$  nephrology, and  $n=13$  neurology) in which the same participant documented a case with both Standard-Standard and NLP-NLP protocols. The average duration of EHR experience for these 23 individuals was 6.3 (SD 2.8) years. The total score of the 10-component SUS measure was significantly higher when the participants used the NLP-NLP protocol compared with when they used the Standard-Standard protocol (mean 36.7, SD 5.4, compared with mean 30.3, SD 7.7;  $P=.007$ ). [Table 2](#) summarizes the usability scores and paired comparisons between the 2 protocols. Responses to 4 of the 10 usability questions (complexity, ease of use, learning the method very quickly, and cumbersome or awkwardness of use) significantly favored the NLP-NLP protocol over the Standard-Standard protocol.

## Discussion

### Findings

This formative study sought to assess the feasibility of using an EHR documentation method based on dictation and NLP by evaluating the effect of the method on documentation time, documentation quality, and usability. We found that a pure protocol of NLP Entry as well as hybrid protocols (involving both NLP Entry and Standard Entry) showed promise for EHR documentation, relative to Standard Entry alone (Standard-Standard Entry). It is our opinion that different parts of the note should be documented differently, but reaching a conclusion on the optimal method of documentation for each part of the note will require further study.

The finding that NLP-NLP Entry and NLP-Standard Entry required significantly less time than Standard-Standard Entry can be explained by the faster speed of dictation relative to that of entering data using the keyboard and mouse, rather than by the involvement of NLP.

No statistically significant difference was found between the overall documentation quality (measured using the PDQI-9 tool) of Standard-Standard Entry and that of any of the other 3 documentation protocols. The succinctness of Standard-Standard Entry documentation was found to be significantly greater than that of the other 3 protocols. This suggests that the note was judged to be more to the point and with less redundancy. In addition, documentation from Standard-Standard Entry was found to be more organized than that from NLP-NLP Entry, indicating that it was structured in a way that the reader could better understand the patient’s clinical course. When the participant used the Standard-Standard protocol, they used Standard Entry for history and physical examination sections as well as assessment and plan sections. When they used the

Standard-NLP protocol, they used Standard Entry for history and physical examination sections and NLP Entry for assessment and plan sections. In the former (Standard-Standard), the participants tended to type shorter paragraphs for the assessment and plan sections. In the latter (Standard-NLP), they dictated the assessment and plan resulting in a larger volume of text. This difference warrants future scrutiny. On the basis of the results of the modified SUS, the participants’ usability ratings for NLP-NLP Entry were significantly higher than for Standard-Standard Entry. These findings suggest that, pending further study, EHR documentation methods using a combination of dictation and NLP show potential for reducing documentation time and increasing usability while maintaining documentation quality, relative to EHR documentation via standard keyboard-and-mouse entry.

Documentation methods using dictation and NLP have the potential to reduce some of the most egregious “pain points” for EHR data entry. These methods can facilitate capture and insertion of both structured data and transcribed text into the appropriate EHR sections, affording the user of the note the option of using one or both types of information. The structured data are ideal for interoperability and coding and may prove to be useful for analytics.

Opinion is divided regarding the relative advantages of narrative fields and structured fields in clinical documentation and in which contexts each excels or is preferable [39,40]. The flexibility to allow providers to enter text in narrative fields clashes with the desire to produce structured data to facilitate reuse of this information in EHRs [22]. It is this quest for achieving a balance of expressivity, richness, and completeness of detail in the patient health story, with reusable, and thus actionable, structured data in the patient EHR that motivates this field of inquiry. Improving the quality of structured notes generated within an EHR from the structured output of NLP Entry and transcribed text is an area requiring further study and development work.

### Future Research

In future research, for the purpose of achieving documentation quality using dictation and NLP that, in all respects, is comparable to or better than documentation quality resulting from Standard Entry, certain changes to the NLP Entry process will be evaluated. We will assess the effects of requiring participants to use dictation under the constraints of a structured template on improving the organization, comprehensibility, succinctness, and synthesis of notes produced from NLP Entry. The templates would reflect the structure of the participant’s EHR. In addition, we will aim to improve the procedures by which NLP output data are translated into and transferred to the clinical note. We also plan to more systematically scrutinize data capture differences pertaining to documenting in different sections of the EHR note. This will enable us to fine-tune hybrid methods of data entry.

In a subsequent study, we will measure the time required for, and documentation quality and usability of, NLP Entry in live clinical use. This will require developing automated interfaces for sending the participants’ dictation to MediSapien NLP and for sending structured data and free text from MediSapien NLP

into the EHR, during which process the participant will be able to modify the documentation. This process will be facilitated by the emergence and widespread adoption of interoperability standards and messages that can carry rich structured data.

### Limitations

This study has several limitations. One limitation is that the simulated interface used in this controlled experiment is somewhat lacking in ecological validity. In a real-world live setting, the structured data and the transcribed text data would both be inserted into the EHR via an automated interface. In addition, the physician would be able to review or modify the documentation before it was finalized. For the purposes of this formative study, this process was simplified. Therefore, an interface to automatically insert the structured data and text into the EHR or allow the physician to review the documentation before finalization was not used for this study. Instead, the insertion process was simulated by manually generating a note in Microsoft Word resembling one that might have been generated by the automated insertion process. Time required for generating the note was not included in the study's time measurements. To ensure that the manually generated note could have been produced by an automated process, it was produced following strict predetermined rules and without any reliance on human discretion.

Second, physicians generated documentation for the study based on test scripts about fictitious patient encounters. Test scripts included history and physical examination sections and were formatted as transcription notes. The assessment and treatment plan sections were excluded from the test scripts. Participants were instructed not to dictate or type verbatim what was written in the test scripts, but to understand what was written and document it in their own way. In addition to being instructed to generate history and physical examination sections, they were instructed to generate their own assessment and treatment plan sections, because those sections were excluded from the test script.

The sample size for cardiology and nephrology was rather small owing to recruiting challenges. This affected the power for determining differences for related contrasts. Clearly, a larger sample size would have enabled us to detect more subtle group differences.

A limitation of this method of generating test documentation was that because it presented medical information in a free-text

format, it may have favored documentation methods requiring the physician to generate free text. NLP Entry requires documentation via dictation exclusively, and Standard Entry entails only some documentation via typing, with the rest entered by pointing and clicking using a mouse. Consequently, the results for time required to complete documentation may be biased toward free text and therefore toward NLP Entry. Nevertheless, we perceive a value in measuring the temporal differences and think that such differences may be consequential in real-world use of this system.

### Conclusions

Current standard methods of EHR documentation have been shown to be extremely time consuming and are judged to have suboptimal usability. In this formative study, the feasibility of an approach to EHR data capture involving applying NLP to transcribed dictation was demonstrated. This approach was shown to have the potential to reduce the time required for documentation and improve usability while maintaining documentation quality in several respects. Future research will evaluate the NLP-based EHR data capture approach in a live clinical setting where generated structured data and transcribed text for real patients are inserted into the EHR via an automated interface.

The past decade has witnessed a dramatic increase in the adoption of EHRs as central instruments in medical practice. However, these systems have not yet proven to be reliable tools for facilitating clinical workflow or enhancing patient care. Recent advances in usability have led to the development of frameworks, new methods, and robust assessment tools that can be used to more precisely delineate the source of the problems associated with an interface [41,42]. In addition, novel approaches to design have provided new EHR approaches that better support flexibility and expressivity [43]. We anticipate that NLP-enabled data entry tools will form an important part of the solution space and will serve to enhance the user's experience.

There is ample evidence that clinicians spend many hours documenting patient records and sometimes at the expense of time that could be devoted to patient care. Dictation is a familiar method of data entry to most clinicians. The proposed solution leverages that familiarity and has the potential to produce a quality document or patient note in less time along with highly structured machine-readable codes.

---

### Acknowledgments

The authors thank the clinicians who participated in the study. The research was supported by the Center for Advanced Information Management of New York State (#13240001) and the National Library of Medicine of the National Institutes of Health under Award Number R43LM011165. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

---

### Conflicts of Interest

JM, AB, and AF are employed by ZyDoc Medical Transcription LLC, whose MediSapien NLP software product was used in this study.

---

## References

1. Healthit. 2012. EHR Adoption & Meaningful Use Progress is Assessed in New Data Briefs URL: <http://www.healthit.gov/buzz-blog/meaningful-use/ehr-adoption-meaningful-use-progress-assessed-data-briefs/> [accessed 2015-08-23] [WebCite Cache ID 6b0B1Iih]
2. Fiks AG, Alessandrini EA, Forrest CB, Khan S, Localio AR, Gerber A. Electronic medical record use in pediatric primary care. *J Am Med Inform Assoc* 2011;18(1):38-44 [FREE Full text] [doi: [10.1136/jamia.2010.004135](https://doi.org/10.1136/jamia.2010.004135)] [Medline: [21134975](https://pubmed.ncbi.nlm.nih.gov/21134975/)]
3. Füchtbauer LM, Nørgaard B, Mogensen CB. Emergency department physicians spend only 25% of their working time on direct patient care. *Dan Med J* 2013 Jan;60(1):A4558. [Medline: [23340186](https://pubmed.ncbi.nlm.nih.gov/23340186/)]
4. McDonald C, Tang P, Hripcsak G. Electronic health record systems. In: Shortliffe EH, Cimino JJ, editors. *Biomedical Informatics*. London: Springer; 2014:255-284.
5. Cimino JJ. Improving the electronic health record—are clinicians getting what they wished for? *JAMA* 2013 Mar 13;309(10):991-992 [FREE Full text] [doi: [10.1001/jama.2013.890](https://doi.org/10.1001/jama.2013.890)] [Medline: [23483171](https://pubmed.ncbi.nlm.nih.gov/23483171/)]
6. Hsiao C, Hing E. 2014. Use and Characteristics of Electronic Health Record Systems Among Office-based Physician Practices: United States, 2001–2013. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Atlanta URL: <http://www.cdc.gov/nchs/data/databriefs/db143.htm> [accessed 2015-08-23] [WebCite Cache ID 6b0OxO6Wy]
7. Slight SP, Berner ES, Galanter W, Huff S, Lambert BL, Lannon C, et al. Meaningful use of electronic health records: experiences from the field and future opportunities. *JMIR Med Inform* 2015;3(3):e30 [FREE Full text] [doi: [10.2196/medinform.4457](https://doi.org/10.2196/medinform.4457)] [Medline: [26385598](https://pubmed.ncbi.nlm.nih.gov/26385598/)]
8. Friedberg M, Chen P, Van BK, Aunon F, Pham C, Caloyeras J, et al. Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy. Santa Monica, CA: Rand Corporation; 2013. URL: [http://www.rand.org/content/dam/rand/pubs/research\\_reports/RR400/RR439/RAND\\_RR439.pdf](http://www.rand.org/content/dam/rand/pubs/research_reports/RR400/RR439/RAND_RR439.pdf) [WebCite Cache ID 6ds7RJ7In]
9. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, et al. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. *J Am Med Inform Assoc* 2013 Jun;20(e1):e2-e8 [FREE Full text] [doi: [10.1136/amiajnl-2012-001458](https://doi.org/10.1136/amiajnl-2012-001458)] [Medline: [23355463](https://pubmed.ncbi.nlm.nih.gov/23355463/)]
10. Westbrook JI, Baysari MT, Li L, Burke R, Richardson KL, Day RO. The safety of electronic prescribing: manifestations, mechanisms, and rates of system-related errors associated with two commercial systems in hospitals. *J Am Med Inform Assoc* 2013;20(6):1159-1167 [FREE Full text] [doi: [10.1136/amiajnl-2013-001745](https://doi.org/10.1136/amiajnl-2013-001745)] [Medline: [23721982](https://pubmed.ncbi.nlm.nih.gov/23721982/)]
11. Schumacher R, Lowry S. Customized common industry format template for electronic health record usability testing. National Institute of Standards and Technology, US Department of Commerce. URL: [http://www.nist.gov/healthcare/usability/upload/LowryNISTIR-7742Customized\\_CIF\\_Template\\_for\\_EHR\\_Usability\\_Testing\\_Publication\\_Verion.doc.pdf](http://www.nist.gov/healthcare/usability/upload/LowryNISTIR-7742Customized_CIF_Template_for_EHR_Usability_Testing_Publication_Verion.doc.pdf) [accessed 2015-12-10] [WebCite Cache ID 6dfslEml]
12. Schumacher R, Lowry S. IST guide to the processes approach for improving the usability of electronic health records. National Institute of Standards and Technology, US Department of Commerce; 2010 Nov 1. URL: [http://www.nist.gov/healthcare/usability/upload/Guide\\_Final\\_Publication\\_Version.pdf](http://www.nist.gov/healthcare/usability/upload/Guide_Final_Publication_Version.pdf) [WebCite Cache ID 6dfslYqDT]
13. Armijo D, McDonnell C, Werner K. Electronic health record usability: evaluation and use case framework. Healthit: Agency for Healthcare Research and Quality; 2009. URL: [https://healthit.ahrq.gov/EHR\\_evaluation\\_and\\_use\\_case\\_framework](https://healthit.ahrq.gov/EHR_evaluation_and_use_case_framework) [WebCite Cache ID 6dfDJ6zr]
14. Ohno-Machado L. Electronic health records and computer-based clinical decision support: are we there yet? *J Am Med Inform Assoc* 2011 Mar 01;18(2):109. [doi: [10.1136/amiajnl-2011-000141](https://doi.org/10.1136/amiajnl-2011-000141)]
15. Payne TH, Corley S, Cullen TA, Gandhi TK, Harrington L, Kuperman GJ, et al. Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. *J Am Med Inform Assoc* 2015 Sep;22(5):1102-1110. [doi: [10.1093/jamia/ocv066](https://doi.org/10.1093/jamia/ocv066)] [Medline: [26024883](https://pubmed.ncbi.nlm.nih.gov/26024883/)]
16. Kimia AA, Savova G, Landschaft A, Harper MB. An introduction to natural language processing: how you can get more from those electronic notes you are generating. *Pediatr Emerg Care* 2015 Jul;31(7):536-541. [doi: [10.1097/PEC.0000000000000484](https://doi.org/10.1097/PEC.0000000000000484)] [Medline: [26148107](https://pubmed.ncbi.nlm.nih.gov/26148107/)]
17. Bernat JL. Ethical and quality pitfalls in electronic health records. *Neurology* 2013 Mar 12;80(11):1057-1061. [doi: [10.1212/WNL.0b013e318287288c](https://doi.org/10.1212/WNL.0b013e318287288c)] [Medline: [23479465](https://pubmed.ncbi.nlm.nih.gov/23479465/)]
18. Cusack CM, Hripcsak G, Bloomrosen M, Rosenbloom ST, Weaver CA, Wright A, et al. The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. *J Am Med Inform Assoc* 2013 Jan 1;20(1):134-140 [FREE Full text] [doi: [10.1136/amiajnl-2012-001093](https://doi.org/10.1136/amiajnl-2012-001093)] [Medline: [22962195](https://pubmed.ncbi.nlm.nih.gov/22962195/)]
19. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013 Jan 1;20(1):144-151 [FREE Full text] [doi: [10.1136/amiajnl-2011-000681](https://doi.org/10.1136/amiajnl-2011-000681)] [Medline: [22733976](https://pubmed.ncbi.nlm.nih.gov/22733976/)]
20. Office of the National Coordinator of Health Information Technology. Capturing high quality electronic health records data to support performance improvement: a learning guide. 2013. URL: <http://www.healthit.gov/sites/default/files/onc-beacon-lg3-ehr-data-quality-and-perform-impvt.pdf> [accessed 2015-08-23] [WebCite Cache ID 6b0HF2qpV]

21. Stetson PD, Bakken S, Wrenn JO, Siegler EL. Assessing electronic note quality using the Physician Documentation Quality Instrument (PDQI-9). *Appl Clin Inform* 2012;3(2):164-174 [[FREE Full text](#)] [Medline: [22577483](#)]
22. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *J Am Med Inform Assoc* 2011;18(2):181-186 [[FREE Full text](#)] [doi: [10.1136/jamia.2010.007237](#)] [Medline: [21233086](#)]
23. Suominen H, Zhou L, Hanlen L, Ferraro G. Benchmarking clinical speech recognition and information extraction: new data, methods, and evaluations. *JMIR Med Inform* 2015;3(2):e19 [[FREE Full text](#)] [doi: [10.2196/medinform.4321](#)] [Medline: [25917752](#)]
24. Friedman C, Elhadad N. Natural language processing in health care and biomedicine. In: Shortliffe EH, Cimino JJ, editors. *Biomedical Informatics*. 4th edition. London: Springer; 2014:255-284.
25. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet* 2012 Jun;13(6):395-405. [doi: [10.1038/nrg3208](#)] [Medline: [22549152](#)]
26. Wu S, Zhu D, Hersh W, Liu H. Clinical information retrieval with split-layer language models. Presented at: Proceedings of the Association for Computing Machinery SIGIR workshop on health search and discovery (HSD); August 1, 2013; Dublin, Ireland. URL: [https://www.eecis.udel.edu/~zhu/uploads/hsd2013\\_paper.pdf](https://www.eecis.udel.edu/~zhu/uploads/hsd2013_paper.pdf)
27. Demner-Fushman D, Lin J. Answering clinical questions with knowledge-based and statistical techniques.. *Computational Linguistics* 2007 Mar;33(1):63-103. [doi: [10.1162/coli.2007.33.1.63](#)]
28. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc* 2011 Oct;18(5):544-551 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000464](#)] [Medline: [21846786](#)]
29. Liu G, Weng C, Yu H. Natural language processing, electronic health records, and clinical research. In: Richesson RL, Andrews JR, editors. *Clinical Research Informatics*. London: Springer; 2012:293-310.
30. MModal. M\*Modal's cloud-based artificial intelligence is widely adopted by physicians to improve patient care. Press release. URL: <https://mmodal.com/resources/press/mmodals-cloud-based-artificial-intelligence-is-widely-adopted-by-physicians-to-improve-patient-care/> [accessed 2016-06-14] [[WebCite Cache ID 6iGbdZbuh](#)]
31. Nuance. Meaningful use: preserving the health story, providing structure for the EHR. URL: [http://www.nuance.com/ucmprod/groups/healthcare/@web-enus/documents/collateral/nd\\_003313.pdf](http://www.nuance.com/ucmprod/groups/healthcare/@web-enus/documents/collateral/nd_003313.pdf) [accessed 2016-06-15] [[WebCite Cache ID 6iHuujbOD](#)]
32. Sun L. Sep 10. The rise of speech recognition technology in EHRs. *The Motley Fool*; 2013 Sep 10. URL: <http://www.fool.com/investing/general/2013/09/10/a-closer-look-at-speech-recognition-technology-in.aspx> [accessed 2016-10-23] [[WebCite Cache ID 6b0KeiRV6](#)]
33. Waghlikar KB, MacLaughlin KL, Henry MR, Greenes RA, Hankey RA, Liu H, et al. Clinical decision support with automated text processing for cervical cancer screening. *J Am Med Inform Assoc* 2012;19(5):833-839 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2012-000820](#)] [Medline: [22542812](#)]
34. Kohn MS, Sun J, Knoop S, Shabo A, Carmeli B, Sow D, et al. IBM's Health Analytics and clinical decision support.. *Yearb Med Inform* 2014;9:154-162 [[FREE Full text](#)] [doi: [10.15265/IY-2014-0002](#)] [Medline: [25123736](#)]
35. Wu Y, Denny JC, Rosenbloom ST, Miller RA, Giuse DA, Xu H. A comparative study of current clinical natural language processing systems on handling abbreviations in discharge summaries. *AMIA Annu Symp Proc* 2012;2012:997-1003 [[FREE Full text](#)] [Medline: [23304375](#)]
36. 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-e242 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2013-001889](#)] [Medline: [24177145](#)]
37. Brooke J. SUS: A quickdirty usability scale. In: Jordan PW, Thomas B, Lyall I, editors. *Usability Evaluation in Industry*. London: Taylor & Francis, Ltd; 1996:189-194.
38. US Department of Health & Human Services. System Usability Scale (SUS). URL: <http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> [accessed 2015-08-23] [[WebCite Cache ID 6b0M4XSKc](#)]
39. Jacobs L. Interview with Lawrence Weed, MD—the father of the problem-oriented medical record looks ahead.. *Perm J* 2009;13(3):84-89 [[FREE Full text](#)] [Medline: [20740095](#)]
40. Johnson SB, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, et al. An electronic health record based on structured narrative. *J Am Med Inform Assoc* 2008;15(1):54-64 [[FREE Full text](#)] [doi: [10.1197/jamia.M2131](#)] [Medline: [17947628](#)]
41. Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. *J Biomed Inform* 2011 Dec;44(6):1056-1067 [[FREE Full text](#)] [doi: [10.1016/j.jbi.2011.08.005](#)] [Medline: [21867774](#)]
42. Zheng K, Haftel HM, Hirschl RB, O'Reilly M, Hanauer DA. Quantifying the impact of health IT implementations on clinical workflow: a new methodological perspective. *J Am Med Inform Assoc* 2010;17(4):454-461 [[FREE Full text](#)] [doi: [10.1136/jamia.2010.004440](#)] [Medline: [20595314](#)]
43. Senathirajah Y, Bakken S, Kaufman D. The clinician in the driver's seat: part 1 – a drag/drop user-composable electronic health record platform. *J Biomed Inform* 2014 Dec;52:165-176 [[FREE Full text](#)] [doi: [10.1016/j.jbi.2014.09.002](#)] [Medline: [25240253](#)]



## Abbreviations

**ANOVA:** analysis of variance

**CBD:** computer-based documentation

**CUMC:** Columbia University Medical Center

**EHR:** electronic health record

**HITECH:** Health Information Technology for Economic and Clinical Health

**MedLEE:** Medical Language Extraction and Encoding System

**NLP:** natural language processing

**PDQI:** Physician Documentation Quality Instrument

**SUS:** System Usability Scale

*Edited by G Eysenbach; submitted 20.01.16; peer-reviewed by PY Yen, C Lovis; comments to author 22.02.16; revised version received 21.06.16; accepted 15.09.16; published 28.10.16*

*Please cite as:*

*Kaufman DR, Sheehan B, Stetson P, Bhatt AR, Field AI, Patel C, Maisel JM*

*Natural Language Processing-Enabled and Conventional Data Capture Methods for Input to Electronic Health Records: A Comparative Usability Study*

*JMIR Med Inform 2016;4(4):e35*

*URL: <http://medinform.jmir.org/2016/4/e35/>*

*doi: [10.2196/medinform.5544](https://doi.org/10.2196/medinform.5544)*

*PMID: [27793791](https://pubmed.ncbi.nlm.nih.gov/27793791/)*

©David R. Kaufman, Barbara Sheehan, Peter Stetson, Ashish R. Bhatt, Adele I. Field, Chirag Patel, James Mark Maisel. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org>), 28.10.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.