International Journal of Nursing Studies 64 (2016) 25–31

Contents lists available at ScienceDirect

International Journal of Nursing Studies

journal homepage: www.elsevier.com/ijns

Automated identification of wound information in clinical notes of patients with heart diseases: Developing and validating a natural language processing application

Maxim Topaz[a,b,*], Kenneth Lai[a], Dawn Dowding[c,d], Victor J. Lei[a], Anna Zisberg[e], Kathryn H. Bowles[c,f], Li Zhou[a,b]

[a] Brigham and Women's Hospital, Boston, MA, USA
[b] Harvard Medical School, Boston, MA, USA
[c] Visiting Nurse Service of New York, NY, USA
[d] School of Nursing, Columbia University, NY, USA
[e] The Cheryl Spencer Department of Nursing, Haifa University, Israel
[f] School of Nursing, University of Pennsylvania, PA, USA

ARTICLE INFO

Article history:
Received 6 April 2016
Received in revised form 13 September 2016
Accepted 18 September 2016

Keywords:
Nursing informatics
Medical informatics
Natural language processing
Wound
Electronic health records
Pressure ulcer

ABSTRACT

Background: Electronic health records are being increasingly used by nurses with up to 80% of the health data recorded as free text. However, only a few studies have developed nursing-relevant tools that help busy clinicians to identify information they need at the point of care.
Objective: This study developed and validated one of the first automated natural language processing applications to extract wound information (wound type, pressure ulcer stage, wound size, anatomic location, and wound treatment) from free text clinical notes.
Methods and design: First, two human annotators manually reviewed a purposeful training sample (n = 360) and random test sample (n = 1100) of clinical notes (including 50% discharge summaries and 50% outpatient notes), identified wound cases, and created a gold standard dataset. We then trained and tested our natural language processing system (known as MTERMS) to process the wound information. Finally, we assessed our automated approach by comparing system-generated findings against the gold standard. We also compared the prevalence of wound cases identified from free-text data with coded diagnoses in the structured data.
Results: The testing dataset included 101 notes (9.2%) with wound information. The overall system performance was good (F-measure is a compiled measure of system's accuracy = 92.7%), with best results for wound treatment (F-measure = 95.7%) and poorest results for wound size (F-measure = 81.9%). Only 46.5% of wound notes had a structured code for a wound diagnosis.
Conclusions: The natural language processing system achieved good performance on a subset of randomly selected discharge summaries and outpatient notes. In more than half of the wound notes, there were no coded wound diagnoses, which highlight the significance of using natural language processing to enrich clinical decision making. Our future steps will include expansion of the application's information coverage to other relevant wound factors and validation of the model with external data.

What is already known about the topic

- There were significant advances in the field of medical natural language processing (or automated information extraction from narrative clinical notes) over the past two decades.
- To date, no studies have explicitly focused on the feasibility of using natural language processing methods to process wound information.

What this paper adds

- We developed and validated one of the first automated natural language processing applications to extract wound information (wound type, pressure ulcer stage, wound size, anatomic location, and wound treatment) from free text clinical notes.

* Corresponding author at: 93 Worcester St., Wellesley Gateway, Suite 2030I, Wellesley, MA, 02481 USA.
E-mail address: mtopaz80@gmail.com (M. Topaz).

http://dx.doi.org/10.1016/j.ijnurstu.2016.09.013
0020-7489/

- In more than half of the wound notes, there were no coded wound diagnoses, which highlight the significance of using natural language processing to enrich clinical decision making.

1. Introduction and background

Wounds, defined as injuries to soft tissues that can vary from minor tears to severe crushing wounds, and can be acute or chronic in nature (O'Connell Smeltzer et al., 2010). In general, acute wounds (e.g., skin graft donor sites, partial thickness burns, and posttraumatic or surgical wounds) heal in an orderly and timely fashion. On the other hand, chronic wounds (e.g., leg ulcers, pressure ulcers and diabetic foot ulcers) take longer to heal as a result of impaired tissue repair due to malnutrition, infection, or poor oxygenation (Chaby et al., 2007). Wounds are a key issue in terms of morbidity and quality of life. It has been estimated that in the United States, Medicare spent \$2.4 billion on hospital stays for infections of surgical and traumatic wounds and disruption of surgical wounds in 2005 (Schwien and Lang, 2008), and that the cost of treating chronic wounds exceeds \$50 billion per year (Fife et al., 2012). International studies support the high prevalence and increasing costs of wound treatments (Hurd and Posnett, 2009; Posnett and Franks, 2016; Vanderwee et al., 2007). Wounds are more common among several clinical populations, such as patients with cardiac diseases (this study's patient population) (Sen et al., 2009). A significant proportion of wound patients require hospitalization or emergent care within 30 days of their hospital admission; in a recent study, having a skin ulcer or wound was the largest predictor of a patient being re-hospitalized (Westra et al., 2013). With the increase in age of the world's population, together with a rise in comorbidities (e.g., obesity, diabetes and venous insufficiency), the number of individuals with chronic wounds is estimated to rise in the near future (Werdin et al., 2009). Thus, identifying patients with wounds to ensure timely interventions and appropriate management is of key importance in health care settings.

In general, health practitioners tend to record up to 80% of health information as free text (Murdoch and Detsky, 2013). Free text documentation is very common in nursing, in part, due to poor usability and lack of standards in electronic health record systems and also because it is hard to capture all the nursing-sensitive information as structured data (Cho et al., 2016; Farber et al., 2007). Although large databases of health and wound specific information exist on local and national levels, only a fraction of this data is captured in a standardized structured format (Ross et al., 2014; Roth et al., 2009). Free-text narratives can help clinicians make a better sense of the patient's wound status (Rosenbloom et al., 2011), using unstructured data may limit the quality and safety of care by increasing the time required to find the relevant data or reducing the ability to use computer-aided applications, such as clinical decision support. Also, our ability to conduct much-needed wound research (i.e. identifying optimal treatments for different types of wounds) with free-text data is limited.

Recently, novel informatics methods have been increasingly applied to large health data sources for data standardization and extraction (Ross et al., 2014). One promising technique- an automated natural language processing (NLP)- can potentially assist in processing free-text wound information in clinical notes. In previous studies, NLP has been successfully used to extract family history data (Zhou et al., 2014), medication information (Zhou et al., 2011), radiology findings (Pham et al., 2014), and other health information from different types of clinical notes (Demner-Fushman et al., 2009; Meystre et al., 2008). To our knowledge, no studies have explicitly focused on the feasibility of using natural language processing methods to process nursing-sensitive data, such as wound information in clinical notes.

The goal of this study was to develop and validate an NLP-based approach to automatically extract wound information (including wound type, pressure ulcer stage if applicable, wound size, anatomic location, and wound treatment) from free-text clinical narratives. We also assessed the performance of the NLP application designed to summarize wound-related information on a corpus of discharge summaries and outpatient notes.

2. Methods

For the purposes of this study, a wound was defined as any skin lesion, regardless of external (e.g., trauma) or internal (e.g., venous hypertension with its secondary consequences to skin integrity) etiology. Several preparatory steps for this study included data collection and building a wound minimum-dataset information model. Our NLP system development and testing methods are summarized in Fig. 1. We first manually reviewed a purposeful training sample and random test sample of clinical notes (including discharge summaries and outpatient notes), identified wound cases, and created a gold standard dataset with the help of domain experts (Steps 1–2). We then trained our NLP system (known as MTERMS) (Zhou et al., 2015, 2014, 2011) to process the wound information and applied our system on a randomly selected sample of clinical notes (Steps 3–4). Finally, we assessed our automated approach by comparing system-generated findings against the gold standard (Step 5). We also identified coded wound diagnoses and compared wound cases identified from free-text data with coded diagnoses in the structured data (Steps 6–7).

2.1. Natural language processing engine description

In this study, we utilized our natural language processing system called Medical Text Extraction, Reasoning and Mapping System (MTERMS). MTERMS applies advanced methodologies and technologies in computer science, artificial intelligence, computational linguistics, and biomedical informatics to conduct text analytics. MTERMS' natural language processing engine consists of both computational and knowledge components. It uses a modular, pipeline approach to conducting linguistic analyses at different linguistic structure levels (i.e., words, phrases, paragraphs, sections and whole notes) while handling abbreviations and lexical variations. MTERMS' engine generates structured output in a standard, interoperable documentation format that can be used for subsequent applications. The system was previously validated for identifying clinical terms within narrative health records in order to extract medications, clinical problems (e.g., presence of depression), family history, and so forth, with high accuracy metrics (Zhou et al., 2015, 2014, 2011).

2.2. Preparatory steps

2.2.1. Data collection

We used a retrospective cohort of patients from an ongoing study focusing on identifying factors associated with readmissions. Patients in this cohort had a history of ischemic heart disease and were hospitalized between 01/01/2011 and 12/31/2013 at different hospitals in Partners Healthcare System, a large integrated healthcare network in Boston, Massachusetts, United States. The database included about 120,000 distinct patients and 3 million notes generated over 2 years of the study period. This study was approved by Partners Institutional Review Board (IRB) at Partners Healthcare System.

Because of a relatively low general distribution of wound information in clinical notes (about 10% of patients), we first

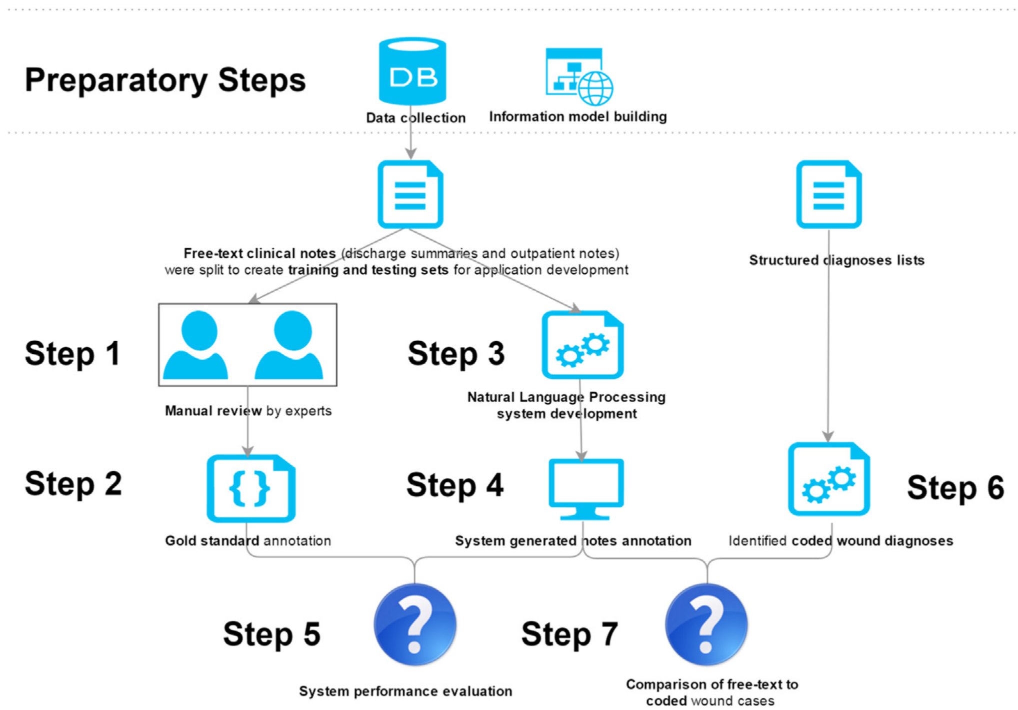

Fig. 1. Automated identification of wound information in clinical notes: system development and testing methods.
Several preparatory steps for this study included data collection and building a wound minimum-dataset information model. We then manually reviewed a purposeful training sample and random test sample of clinical notes (including discharge summaries and outpatient notes), identified wound cases, and created a gold standard dataset with the help of domain experts (Steps 1–2). We trained our NLP system (known as MTERMS) to process the wound information and applied our system on a randomly selected sample of clinical notes (Steps 3–4). Finally, we assessed our automated approach by comparing system-generated findings against the gold standard (Step 5). We also identified coded wound diagnoses and compared wound cases identified from free-text data with coded diagnoses in the structured data (Steps 6–7).

conducted a manual review of clinical notes to select a purposeful sample of notes. This sample comprised of 50% discharge summaries and 50% outpatient clinical notes was used as a training set. A discharge summary in our institution typically consists of a complete copy of the admission note as well as standard discharge summary information including a detailed hospital course, discharge medications and a plan of care. Outpatient notes included primary care physician visit notes, homecare nursing notes, and clinical nurse specialist notes. In natural language processing studies (similarly to qualitative studies), training set sample size is usually determined using maximum variation approach where positive cases are continuously reviewed and annotated until no new concepts/vocabulary/ lexical variants are found in the text. In our study, we reached saturation at 120 notes identified with keyword search (i.e.; looking for regular expressions related to wound or wound treatment); random notes review and notes with structured wound codes on the patient's problem list. We also then examined 240 additional notes without wound information to test our algorithm's ability to identify negative cases and potential negated expressions.

Based on the sample size calculation, we needed about 1100 cases for the testing set to have adequate power when the expected algorithm specificity = 0.8, sensitivity >0.9, outcome prevalence = 0.1 (based on the background literature of wound prevalence), and the desired precision is 0.05 (Malhotra and Indrayan, 2010). Thus, for the testing set, we randomly selected 1100 additional clinical notes (550 discharge summaries and 550 outpatient notes) that were not used for training. Other studies developing NLP systems use a similar approach and number of cases for system performance evaluation (Pakhomov et al., 2007; Pons et al., 2016).

2.2.2. Building a wound minimum-dataset information model

In this study, we aimed to explore the feasibility of using NLP for automated wound information processing. Rather than creating an extensive algorithm to extract all wound-related information, we created a wound minimum-dataset information model with central wound-related factors. In general, information models are used to describe a phenomenon of interest (e.g., health problem, sign/symptom, allergy, etc.) in informatics for further computer processing. For example, a generic health problem can have such attributes as an anatomic location (e.g., right hand, left lower lung lobe, etc.), status (e.g., acute or chronic), organ or tissue involved (heart, skin, liver, etc.), among others.

To specify attributes of a wound minimum-dataset information model for the purposes of this study, we first reviewed the commonly used standard representations and information models for wound related information, such as:

- Relevant concepts and their relationships (e.g., "Disorder of Skin and/or Subcutaneous Tissue") in the Systematic Nomenclature of Medical Terms Clinical Terms (SNOMED CT) (IHTSDO, 2014). SNOMED CT is one of the most commonly used international standard medical ontologies for representing clinical problems and findings.
- The "Open Wound" model in the openEHR (openEHR, 2014). openEHR is open-source standard for health information that is being used in several European countries.
- The "Pressure Ulcer" model defined within the Consolidated Continuity of Care Document (C-CDA) (Health Level 7, 2014). C-CDA is a government-required standard in the U.S. for all health information exchange.

We described the models graphically and in a table format and shared those findings with our study experts (two PhD prepared nurses with clinical and research expertise in wound management). These expanded wound information models had different numbers of attributes, for example wound type, wound size, wound bed characteristics, etc. In general, the models differed by the level of specificity; some models were more generic, offering multiple attributes that can be used to describe wound information, while other models were more problem-oriented, focusing on specific types of wounds. For example, SNOMED-CT offered an extensive "Disorder of Skin and/or Subcutaneous Tissue" hierarchy (IHTSDO, 2014) with possible multiple hierarchies for individual problems (e.g. pressure ulcer) while C-CDA (Health Level 7, 2014) only had a specific "Pressure Ulcer" model. Our team (including nurse and physician informaticians, a doctoral student in pharmacy, a computational linguist and a software developer) and the wound experts reviewed the models and arrived at an agreed wound minimum-dataset information model that included five factors: 1) wound type; 2) pressure ulcer stage, if available; 3) wound size (length/depth/width), when available; 4) wound anatomic location; and 5) wound treatment. We decided to include the pressure ulcer stage (factor #2) in this generic information model since it is a very nursing-sensitive model component that is likely to be documented and is associated with wound treatment, prognosis, etc.

2.3. Automated identification of wound information in clinical notes: system development and testing methods

2.3.1. Manual review and gold standard generation (Steps 1–2)

All training (n = 360) and testing (n = 1100) notes were independently reviewed by a doctorally prepared nurse informatician and a doctoral student in pharmacy. Each note was classified as either having or not having wound-related information and further annotated with the five elements of our wound minimum-dataset information model, when applicable. We used Knowtator, a general-purpose Protégé-integrated text annotation tool for annotations (Ogren, 2009). Annotations were compared and discussed until full consensus between the annotators was achieved. We also calculated Cohen's Kappa to measure the inter-annotator agreement. Fig. 2 presents an example of an annotated wound note. We also calculated the number of attributes of each type and an average number of instances per note for each note type. For example, we quantified the prevalence of different wound-related attributes (e.g., wound type, anatomic location, etc.) in the discharge summaries versus other types of notes. These findings were then used by the research group to modify our automated classification algorithms introduced below.

2.3.2. Natural language processing wound module development (Steps 3–4)

We created a specialized wound module to process wound information. To create the wound type lexicon, we used a combination of terms from standard terminologies, namely the SNOMED-CT (September 2014 Release), the International Statistical Classification of Diseases v.9 (ICD-9) (WHO, 2014) and Logical Observation Identifiers Names and Codes (LOINC) (LOINC, 2014), and narrative terms identified in the training dataset. We also used those sources to identify any pressure ulcer stage terms (i.e. "stage I" or "unstageable"). For wound size, a separate regular expression-based module was created with possible wound size formats from the narrative clinical descriptions in the notes (i.e. "3 cm × 1cm" or "Length: 3.5 cm Width: 2 cm Depth: 1 cm"). For body locations, we used an existing previously-created MTERMS module based on SNOMED-CT anatomic attributes. For wound treatments, we created an additional specific subset that included frequently used wound treatment equipment (e.g., "wound V.A.C.", "Unna boot" etc.), dressings (e.g., "Aquacel", "DuoDerm" etc.), and procedure types (e.g., "wet to dry dressing", "wound cleansing with saline", etc.) from the training set. We also used several wound specific treatments from the literature (e.g., "Silvadene"). Finally, we used an existing MTERMS lexicon to identify negation terms (e.g. "no", "denies", and "absent").

We then used MTERMS to annotate each clinical note with the five elements of the wound minimum dataset information model. MTERMS have also identified whether the wound related term was negated. For example, the software recognized that a sentence indicating " . . . there were no wounds or other ulcers . . . " does not include wound information (as suggested by the expression "there were no").

2.3.3. System performance evaluation (Step 5)

We used standard metrics to evaluate the system performance, which include precision (p- the percentage of wound attributes that were true wound attributes), recall (r- the percentage of true wound attributes that were captured by our automated wound module), and F-measure (F- the harmonic mean of the precision and recall and represents their combined quality). For each category, let TP, TN, FP, and FN be the number of true positives, true negatives, false positives, and false negatives, respectively. Then $p = TP/(TP + FP)$, $r = TP/(TP + FN)$, and $F = 2pr/(p + r)$.

Statistical analyses and attribute prevalence comparisons in different note types were conducted with STATA v.11 (StataCorp, 2009). Statistical procedures including *t*-tests and chi-square tests were used, when appropriate.

2.3.4. Comparison of coded vs. free-text identification of wounds (Steps 6–7)

In order to measure how the NLP-based approach, in addition to using structured data, can enhance the detection of patients suffering from wounds, we calculated the percentage of cases with wounds identified by our annotators that also contained coded diagnosis information. The goal of this step is to estimate the number of wound cases that are not indicated in patients' problem lists but are only captured in the free text. When present, some of the coded diagnoses indicated definite wound presence (i.e. International Classification of Diseases- ICD-9 code 707.00- "PRESSURE ULCER, UNSPECIFIED SITE") while others indicated potential wound presence (ICD-9 code E885.9- "FALL FROM OTHER SLIPPING, TRIPPING, OR STUMBLING"). We considered notes with at least one *definite* wound code as having a *definite* wound code and notes with at least one *potential* wound code (and no definite wound codes) as having a *potential* wound code. Code identification was done with structured diagnosis data which was retrieved from inpatient discharge diagnoses (for discharge summaries) as well as patient problem lists from our outpatient electronic health record system (for outpatient records).

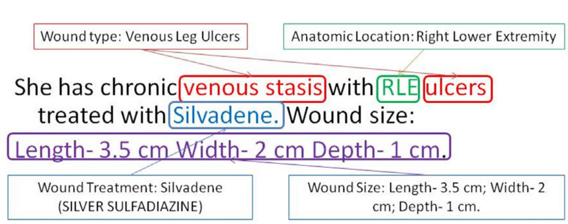

Fig. 2. Wound information annotation example.

Table 1
Proportion and number of instances per note for wound-related attributes in 49 discharge summaries and 52 outpatient notes (testing set notes).

Measure	Note type	Overall n	Wound type n (%)	Anatomic location n (%)	Wound size n (%)	Pressure ulcer stage n (%)	Wound treatment n (%)
Number of instances	Discharge summaries	591	275 (46.5)	129 (21.8)	22 (3.7)	5 (0.8)	160 (27.1)
	Outpatient notes	314	155 (49.4)	67 (21.3)	20 (6.4)	2 (0.6)	70 (22.3)
	Total	905	430 (47.5)	196 (21.7)	42 (4.6)	7 (0.8)	230 (25.4)
Average instances per note	Discharge summaries	12.1*	5.6	2.6	0.4	0.1	3.3
	Outpatient notes	6*	3	1.3	0.4	0.1	1.3
	Total	9	4.3	1.9	0.4	0.1	2.3

* $p < 0.001$.

3. Results

3.1. Manual review and gold standard generation (Steps 1–2)

We identified 101 notes with wound information in the testing dataset (9.2% out of 1100 notes); 49 discharge summaries and 52 outpatient notes. These notes represented unique patients in the dataset. Inter-rater agreement between human annotators was high (Cohen's Kappa [K] = 0.91) and full consensus was achieved on all wound cases between the two annotators. Generally, discharge summaries and outpatient notes included a similar fraction of information focused on each of the wound attributes (Table 1). For example, about half of the included information was a description of a wound type (47.5% total, 46.5% of information in the discharge notes and 49.4% of information in the outpatient notes), one-fifth of the attributes were related to anatomic location (21.7% total, 21.8% vs. 21.3%, respectively), and only a small fraction of information was about wound size or pressure ulcer stage. We found about 5% more information about wound treatment in the discharge summaries (27.1%) than in the outpatient notes (22.3%), although the difference was not statistically significant ($p = 0.32$).

In terms of frequencies, there were a total of 905 instances of wound-related attributes in the notes (Table 1). Each note included 9 wound-related information attributes on average. Overall, discharge summaries included significantly more wound information instances (12.1 attributes per note) than outpatient notes (6 attributes per note, $p < 001$). The common wound attributes were more frequently mentioned in the discharge summaries; for instance, wound type was documented 5.6 times in a discharge note versus 3 times in an outpatient note, similar to the wound treatment (3.3 vs. 1.3) and anatomic location (2.6 vs. 1.3).

The major wound types presented were "Wound" (n = 185, 43%), followed by "Ulcer [unspecified]" (n = 70, 16.3%), "Incision" (n = 58, 13.5%), "Pressure ulcer" (n = 15, 3.5%) and "Blister" (n = 14, 3.3%) (Fig. 3). For half of the pressure ulcers, no stage information was present (7 out of 15, 46.6% pressure ulcers included stage). The remaining one-fifth of the wound types included a diverse range of terms, such as "Skin tears", "Sores", "Venous stasis ulcers", etc.

3.2. System performance evaluation (Step 5)

The overall (micro-averaged) F-measure of the NLP method was 92.7%, with a precision of 95.3% and a recall of 90.3% (Table 2). Among the five factors, best results were obtained for wound treatment (precision = 98.6% and recall = 93%), and poorest results for wound size (precision = 82.9% and recall = 81%). Low recall was due to non-supported patterns (i.e. "7 mm"), terms/expressions not in our lexicon (i.e. "broken skin"), or large lexical distance between a wound type and other attributes (i.e. pressure ulcer stage). There were also several instances of false positive attributes found by MTERMS, mainly because of the close lexical proximity to the wound type. Table 3 presents several examples of false positive and false negative instances- these examples identify areas for further system development.

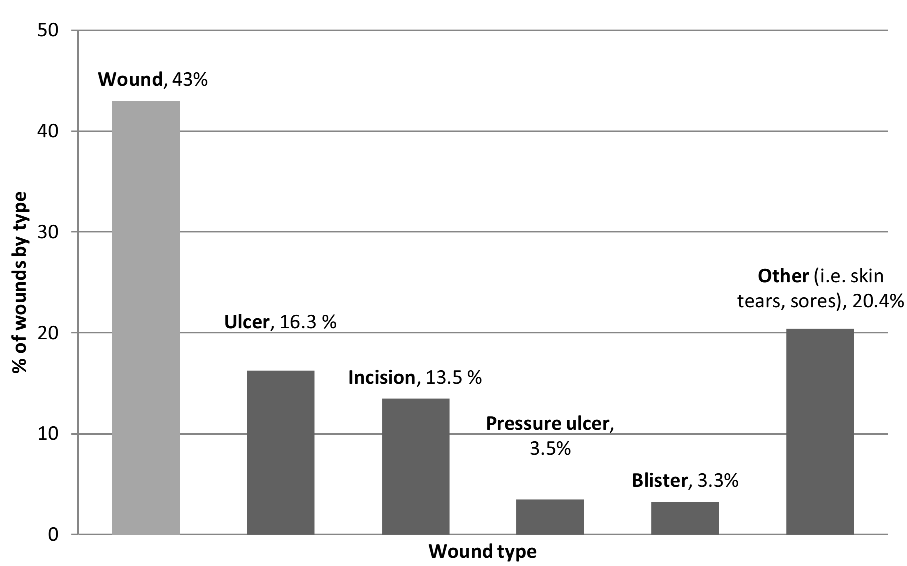

Fig. 3. Distribution of wound type information in the clinical notes.

Table 2
MTERMS system performance on processing wound information in clinical notes.

	Overall	Wound type	Wound treatment	Anatomic location	Press ulcer stage	Wound size
Precision (%)	95.3	98.3	98.6	88	100	82.9
Recall (%)	90.3	91.9	93	86.2	71.4	81
F-measure (%)	92.7	95	95.7	87.1	83.3	81.9

Table 3
Issues with our natural language processing system MTERMS- false negatives and false positives.

Issue	Examples
False Negatives	
Not in lexicon (Wound Type)	"broken skin" | "traumatic blistering"
Not in lexicon (Anatomic Location)	"UIQ" [upper inner quadrant] | "LUE" [left upper extremity]
Not in lexicon (Wound Treatment)	"cleansed with normal saline" | "wet-to-dry packing changes"
No Wound Type near mention (Pressure Ulcer Stage)	"stage 3 on heel"
Pattern not supported (Wound Size)	"7 mm" | "partial-thickness"
False Positives	
Anatomic location near wound type	" . . . drainage at the incision site; chest pain . . . " | "laceration to the right pinna and some jaw pain"
Wound size near wound type	"(inferior to the segment involved in the calf ulcer), diffuse small vessel disease in the foot"

3.3. Comparison of coded vs. free-text identification of wounds (Steps 6–7)

For about half of the notes (53.5%) with wound information in free text, there were no coded diagnoses of wound present. Only one out of five notes (19.8%) had a *definite* wound code and in the remainder of notes (26.7%), a *potential* wound code was present.

4. Discussion

This is one of the first studies that developed an NLP application focused on nursing-sensitive data. We modified the general purpose natural language processing system, MTERMS, to explore the feasibility of automated processing of wound information in free-text clinical notes. First, we reviewed the common wound information representations and models described in the literature (preparatory work). Despite some differences in levels of information models specificity (some models were more generic while others described a specific type of wound), five common attributes used by most of the models emerged.

When testing our NLP approach, we found that free-text wound information appeared in about one out of ten clinical notes in our sample. Discharge summaries included slightly more information about wound treatment than in the outpatient notes. This difference is reasonable, because discharge notes often include recommendations for future treatments or a summary of past treatments while outpatient notes in our sample (including nurses' and primary care providers' visit notes) were more focused on current problems and briefly mentioned wounds as a patient's clinical context. Discharge summaries also tended to be longer than outpatient notes; hence they had a higher frequency of wound information instances per note.

Although wound size documentation is of high importance for clinical follow-up and healing status progress assessment (Flanagan, 2003; Gethin, 2006), it was one of the least-documented wound attributes. In addition, slightly more than half of the pressure ulcers were presented with no stage information. These findings are in line with other studies indicating low documentation quality for chronic wounds (Jones et al., 2007; Stremitzer et al., 2007), including the pressure ulcers (Gunningberg and Ehrenberg, 2004; Gunningberg et al., 2000). Although more in-depth analysis of the wound size documentation is needed, these results suggest that there is certainly room for improvement of wound documentation to provide a more accurate and complete wound description. We recommend emphasizing the importance of wound characteristics and wound treatment documentation at all levels of nursing education, especially undergraduate programs and nurses' in-service training. Another forthcoming recommendation is to examine wound documentation modules in electronic health records to ensure that necessary wound information model attributes are covered.

An additional concerning finding was the lack of a coded wound diagnosis in almost half of the patient notes where free-text wound information was recorded (Steps 6–7). Only one out of five notes had a structured wound code, and in a quarter of notes, a potential wound code was present. Thus, the real prevalence of wound presence remains largely underestimated when no free-text analysis is conducted (which is the case in many clinical studies based on the electronic health record structured data or claims data). Absence of structured wound information reduces the quality of medical billing and impacts our ability for effective risk adjustment. Also, busy clinicians can miss the needed wound documentation in the large volume of free text documentation, which in turn can impact the quality and timeliness of care. This finding underlines the critical need for NLP analysis of clinical notes and further verification of this trend with a larger sample of patient notes.

Using the wound module developed in this study, we were able to identify the five wound information model elements with relatively high performance. Our best results were obtained for wound treatment and wound type. Currently, we are working on improving the algorithm's accuracy for wound size detection. Also, we will expand the coverage of the lexicon to include additional wound related terms from the testing sample. In general, our system can be applied to augment clinical research (i.e. by retrospectively identifying the most effective wound treatments) or practice (i.e. by proactively flagging patients with wounds for specialist care).

4.1. Limitations and future directions

The study is limited by the sample used to generate the algorithm, as it only included patients with heart disease from one healthcare system. We intend to address this in future research by applying our wound module on data from a broader range of patients across healthcare systems. Also, the tool is intended to be used by 'generalist' nurses and other providers rather than wound care specialists. For this feasibility study, we compiled five common elements into a generic wound minimum-dataset information model. Future work should explore the applicability of natural language processing to other wound attributes, such as wound bed description (i.e. epithelialization or granulation), skin characteristics (color, turgor etc.) and other relevant factors. Finally, in this study we used dictionary-based NLP approach and plan to extend it to include statistical techniques (e.g., text classification) in the future algorithm development.

5. Conclusions

This study developed and validated one of the first automated NLP applications to extract wound information from free text clinical notes. The system achieved good performance on a subset of randomly selected discharge summaries and outpatient notes. In more than half of the notes describing wounds in a free-text clinical narrative, there were no coded wound diagnoses, which highlight the significance of using NLP to enrich clinical decision making. International researchers can examine the quality of the wound documentation in their practices and, if necessary, follow the steps we propose here to develop NLP solutions in other languages.

Conflict of interest

None.

Funding

None.

Ethical approval

IRB approval given by the Partners Healthcare System IRB.

References

Chaby, G., Senet, P., Vaneau, M., Martel, P., Guillaume, J.-C., Meaume, S., Téot, L., Debure, C., Dompmartin, A., Bachelet, H., Carsin, H., Matz, V., Richard, J.L., Rochet, J.M., Sales-Aussias, N., Zagnoli, A., Denis, C., Guillot, B., Chosidow, O., 2007. Dressings for acute and chronic wounds: a systematic review. Arch. Dermatol. 143, 1297–12304. doi:http://dx.doi.org/10.1001/archderm.143.10.1297.

Cho, I., Kim, E., Cho, W., Staggers, N., 2016. Comparing usability outcomes and functions of six nursing record systems. Int. J. Med. Inf. 78–85.

Demner-Fushman, D., Chapman, W.W., McDonald, C.J., 2009. What can natural language processing do for clinical decision support? J. Biomed. Inf. 42, 760–772. doi:http://dx.doi.org/10.1016/j.jbi.2009.08.007.

Farber, J., Siu, A., Bloom, P., 2007. How much time do physicians spend providing care outside of office visits? Ann. Intern. Med. 147, 693–698.

Fife, C., Carter, M., Walker, D., Thomson, B., 2012. Wound care outcomes and associated cost among patients treated in US outpatient wound centers: data from the US wound registry. Wounds 24, 10–17.

Flanagan, M., 2003. Wound measurement: can it help us to monitor progression to healing? J. Wound Care 12, 189–194. doi:http://dx.doi.org/10.12968/jowc.2003.12.5.26493.

Gethin, G., 2006. The importance of continuous wound measuring. Wounds 2.

Gunningberg, L., Ehrenberg, A., 2004. Accuracy and quality in the nursing documentation of pressure ulcers: a comparison of record content and patient examination. J. Wound. Ostomy Continence Nurs. 31, 328–335.

Gunningberg, L., Lindholm, C., Carlsson, M., Sjödén, P.O., 2000. The development of pressure ulcers in patients with hip fractures: inadequate nursing documentation is still a problem. J. Adv. Nurs. 31, 1155–1164.

Health Level 7, 2014. HL7 Standards Product Brief – CDA. http://www.hl7.org/implement/standards/product_brief.cfm?product_id = 7 (accessed 3.30.15).

Hurd, T., Posnett, J., 2009. Point prevalence of wounds in a sample of acute hospitals in Canada. Int. Wound J. 6, 287–293. doi:http://dx.doi.org/10.1111/j.1742-481X.2009.00615.x.

IHTSDO, 2014. SNOMED Clinical Terms. URL http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html (accessed 2.23.15).

Jones, K.R., Fennie, K., Lenihan, A., 2007. Evidence-based management of chronic wounds. Adv. skin Wound Care 20, 591–600. doi:http://dx.doi.org/10.1097/01.ASW.0000284936.32707.8d.

LOINC [WWW Document]. https://loinc.org/ (accessed 2.23.15.).

Malhotra, R.K., Indrayan, A., 2010. A simple nomogram for sample size for estimating sensitivity and specificity of medical tests. Indian J. Ophthalmol. 58, 519–522. doi:http://dx.doi.org/10.4103/0301-4738.71699.

Meystre, S.M., Savova, G.K., Kipper-Schuler, K.C., Hurdle, J.F., 2008. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb. Med. Inf. 128–144.

Murdoch, T.B., Detsky, A.S., 2013. The inevitable application of big data to health care. JAMA 309, 1351–1352. doi:http://dx.doi.org/10.1001/jama.2013.393.

O'Connell Smeltzer, S., Bare, B., Hinkle, L., Cheever, K., 2010. Brunner & Suddarth's Textbook of Medical-surgical Nursing. Lippincott Williams & Wilkins, Philadelphia, PA.

Ogren, P., 2009. Knowtator [WWW Document]. http://knowtator.sourceforge.net/index.shtml (accessed 3.30.15).

Pakhomov, S., Weston, S.A., Jacobsen, S.J., Chute, C.G., Meverden, R., Roger, V.L., 2007. Electronic medical records for clinical research: application to the identification of heart failure. Am. J. Manage. Care 13, 281–288.

Pham, A.-D., Névéol, A., Lavergne, T., Yasunaga, D., Clément, O., Meyer, G., Morello, R., Burgun, A., 2014. Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings. BMC Bioinf. 15, 266. doi:http://dx.doi.org/10.1186/1471-2105-15-266.

Pons, E., Braun, L.M.M., Hunink, M.G.M., Kors, J.A., 2016. Natural language processing in radiology: a systematic review. Radiology 279, 329–343. doi:http://dx.doi.org/10.1148/radiol.16142770.

Posnett, J., Franks, P.J., 2016. The burden of chronic wounds in the UK. Nurs. Times 104, 44–45.

Rosenbloom, S.T., Denny, J.C., Xu, H., Lorenzi, N., Stead, W.W., Johnson, K.B., 2011. Data from clinical notes: a perspective on the tension between structure and flexible documentation. J. Am. Med. Inf. Assoc. 18, 181–186. doi:http://dx.doi.org/10.1136/jamia.2010.007237.

Ross, M.K., Wei, W., Ohno-Machado, L., 2014. Big data and the electronic health record. Yearb. Med. Inf. 9, 97–104. doi:http://dx.doi.org/10.15265/IY-2014-0003.

Roth, C.P., Lim, Y.-W., Pevnick, J.M., Asch, S.M., McGlynn, E.A., 2009. The challenge of measuring quality of care from the electronic health record. Am. J. Med. Qual. 24, 385–394. doi:http://dx.doi.org/10.1177/1062860609336627.

Schwien, T., Lang, C., 2008. Changes in Wound Care Outcomes Analysis – New Home Health Compare Measures. OCS, Seattle, WA.

Sen, C.K., Gordillo, G.M., Roy, S., Kirsner, R., Lambert, L., Hunt, T.K., Gottrup, F., Gurtner, G.C., Longaker, M.T., 2009. Human skin wounds: a major and snowballing threat to public health and the economy. Wound Repair Regen. 17, 763–771. doi:http://dx.doi.org/10.1111/j.1524-475X.2009.00543.x.

StataCorp, 2009. STATA 11. College-Park, TX.

Stremitzer, S., Wild, T., Hoelzenbein, T., 2007. How precise is the evaluation of chronic wounds by health care professionals? Int. Wound J. 4, 156–161. doi:http://dx.doi.org/10.1111/j.1742-481X.2007.00334.x.

Vanderwee, K., Clark, M., Dealey, C., Gunningberg, L., Defloor, T., 2007. Pressure ulcer prevalence in Europe: a pilot study. J. Eval. Clin. Pract. 13, 227–235. doi:http://dx.doi.org/10.1111/j.1365-2753.2006.00684.x.

WHO, 2014. International Classification of Diseases (ICD) [WWW Document]. http://www.who.int/classifications/icd/en/ (accessed 2.23.15).

Werdin, F., Tennenhaus, M., Schaller, H.-E., Rennekampff, H.-O., 2009. Evidence-based management strategies for treatment of chronic wounds. Eplasty 9, e19.

Westra, B.L., Bliss, D.Z., Savik, K., Hou, Y., Borchert, A., 2013. Effectiveness of wound, ostomy, and continence nurses on agency-level wound and incontinence outcomes in home care. J. Wound. Ostomy Continence Nurs. 40, 25–53. doi:http://dx.doi.org/10.1097/WON.0b013e31827bcc4f.

Zhou, L., Plasek, J.M., Mahoney, L.M., Karipineni, N., Chang, F., Yan, X., Chang, F., Dimaggio, D., Goldman, D.S., Rocha, R.A., 2011. Using Medical Text Extraction, Reasoning and Mapping System (MTERMS) to process medication information in outpatient clinical notes. AMIA Annu. Symp. Proc. 2011, 1639–1648.

Zhou, L., Lu, Y., Vitale, C.J., Mar, P.L., Chang, F., Dhopeshwarkar, N., Rocha, R.A., 2014. Representation of information about family relatives as structured data in electronic health records. Appl. Clin. Inf. 5, 349–367. doi:http://dx.doi.org/10.4338/ACI-2013-10-RA-0080.

Zhou, L., Baughman, A.W., Lei, V.J., Lai, K.H., Navathe, A.S., Chang, F., Sordo, M., Topaz, M., Zhong, F., Murrali, M., Navathe, S., Rocha, R.A., 2015. Identifying patients with depression using free-text clinical documents. Stud. Health Technol. Inf. 216, 629–633.

openEHR, 2014. openEHR: An Open Domain-driven Platform for Developing Flexible E-Health Systems [WWW Document]. http://www.openehr.org/ (accessed 2.23.15).