



HHS Public Access

Author manuscript

Comput Inform Nurs. Author manuscript; available in PMC 2015 April 30.

Published in final edited form as:

Comput Inform Nurs. 2009 ; 27(4): 215–225. doi:10.1097/NCN.0b013e3181a91b58.

Exploring the Ability of Natural Language Processing to Extract Data from Nursing Narratives

Sookyung Hyun, RN, DNSc^a, Stephen B. Johnson, PhD^b, and Suzanne Bakken, RN, DNSc^{a,b}

^aSchool of Nursing, Columbia University, New York, New York, USA

^bDepartment of Biomedical Informatics, Columbia University, New York, New York, USA

Abstract

Natural Language Processing (NLP) offers an approach for capturing data from narratives and creating structured reports for further computer processing. We explored the ability of a NLP system, Medical Language Extraction and Encoding (MedLEE), on nursing narratives. MedLEE extracted 490 concepts from narrative text in a sample of 553 oncology nursing progress notes. The most frequently monitored and recorded signs and symptoms were related to chemotherapy care, such as adverse reactions, shortness of breath, nausea, pain, bleeding. In terms of nursing interventions, chemotherapy, blood culture, medication, and blood transfusion were commonly recorded in free text. NLP may provide a feasible approach to extract data related to patient safety/quality measures and nursing outcomes by capturing nursing concepts that are not recorded through structured data entry. For better NLP performance in the domain of nursing, additional nursing terms and abbreviations must be added to MedLEE's lexicon.

Introduction

Electronic Health Record (EHR) systems approach the issue of data capture from clinicians in one or both of two general ways. The first consists of using specific user interfaces for capturing data in a structured and possibly coded format. The second consists of collecting and storing clinical data in computer systems in free text format.

In order to use clinical data for additional purposes such as quality assurance, outcomes research, or public health reporting, the data must be structured and coded [1]. However, using only structured, coded approaches for data entry may result in the loss of significant clinical information typically contained in narratives (free text data). For instance, the medication list may be captured only in structured format, but many of the important nuances of the medication regimen, the history of past regimens, patient reactions to the regimen, and adherence behavior, are typically captured in narrative formats [2]. Studies of medical errors have often used retrospective chart review to measure adverse event rates. Although collecting data in this manner produces important clinical information, it is more costly, and manual chart review detects only documented adverse events [3]. A nursing

Address for Correspondence: Sookyung Hyun, RN, DNSc, 630 W. 168th Street, Maibox 6, New York, NY 10032, sh2138@columbia.edu, Voice: (212) 342-6845, Fax: (212) 305-6937.

documentation system that has a combination of structured data entry and unstructured (i.e., free text) data entry assisted by natural language processing (NLP) may better support the acquisition and use of nursing data by placing data in a meaningful context based on the original text, allowing users the freedom of narrative documentation, and ensuring that the free text data are available for reuse.

NLP offers an approach for capturing data from narratives and creating structured reports for further computer processing [4]. Previous research has demonstrated that NLP is an applicable method for capturing clinical information from free text data [5]. However, fewer studies have been conducted to determine if NLP can be used to extract data from nursing narratives to support data reuse for purposes such as decision support, measurement of patient safety and quality of care.

Nursing narratives contain descriptive information about the patient, specific interventions that have been completed, and patient responses to the interventions (e.g., patient adherence or adverse drug event). For example, in one study, about 14% of adverse drug events were detected from electronic nursing free text data [6]. Therefore, it is desirable to investigate NLP performance on nursing narratives. MedLEE (**M**edical **L**anguage **E**xtraction and **E**ncoding), an NLP system in use at New York Presbyterian Hospital, has been extensively evaluated with different types of clinical free text data, such as x-ray reports, discharge summaries, and sign-out notes [4]. However, it has not been evaluated with nursing data.

The purpose of this study was to explore the ability of NLP for capturing nursing concepts so that we could determine opportunities for supporting nursing documentation and data reuse. Specific research questions are:

- What types of data can be extracted by NLP from oncology nursing narratives?
- What terms of relevance to patient safety and quality of care measures can be extracted from oncology nursing narratives using NLP?

Background

Previous research has demonstrated that NLP is an effective method of accurately identifying and reusing data from clinical notes in several domains. Fisman and Haug [7] demonstrated that NLP systems supported real-time decision support for community-acquired pneumonia by extracting specific radiology findings. In another study, a lexically-based NLP system showed promise as a method for detecting adverse events in outpatient visit notes [8].

MedLEE was originally designed for decision support applications in the domain of chest x-ray reports. MedLEE showed a high accuracy, sensitivity, and specificity in extracting specific clinical information from discharge summaries and x-ray reports when it was compared to a reference standard obtained manually by an expert [9]. MedLEE effectively identified findings suspicious for breast cancer from mammogram reports [10] and suspected tuberculosis patients from chest x-ray reports [11]. In addition, MedLEE demonstrated better precision than an ad-hoc approach and acceptable recall for its intended use in ophthalmology visit notes [12].

Nursing narratives are different from those of physicians. Nursing documentation is more like a picture that describes a patient's status illustratively, whereas physicians' documentation is more like a headline due to focus on problem-oriented summarization and abstraction [13]. Nursing narratives describe aspects of the patient's condition that are not addressed in the flowsheet or other structured data, such as change in status, nursing interventions, and patient responses [14]. Consequently, the ability of NLP to extract data from nursing narratives may differ from performance on reports and physicians' notes.

Little research on NLP has been published in the nursing domain. One study identified potential challenges associated with using NLP for HIV/AIDS clinic notes [2]. A further study reported that HIV/AIDS ambulatory progress notes had a more diverse vocabulary and the language structures were different from radiology reports [15]. In the study, notable semantics were categorized for determining automated strategies to deal with the issues. Hsieh et al. demonstrated the potential for automatic extraction of the linguistic meaning of the terms patients use in their electronic mail messages using NLP [16]. Bakken et al. examined the applicability of NLP for nursing narratives by comparing the semantic categories of the ISO (International Organization for Standardization) reference terminology models for nursing diagnoses and nursing actions with the semantic categories of MedLEE [17] and indicated that the current semantic structure of MedLEE was not sufficient for processing verb-rich nursing narratives, and proposed further research regarding NLP semantic categories for better processing of nursing narratives [18].

Methods

Sample and Setting

The study setting was an oncology unit at Weill Cornell Medical Center campus of New York Presbyterian Hospital in New York City. The sample comprised nursing narratives (free text data) from a corpus of de-identified nursing progress notes written by oncology nurses using the Eclipsys Sunrise Critical Care (SCC) documentation system between April 1 and July 31, 2006. The study was approved by Columbia University Medical Center's Institutional Review Board.

Procedure

The study procedures included: pre-processing, natural language processing, and analysis (Figure 1). The analysis focused on two areas: 1) extraction of data in general and 2) extraction of data for quality and safety purposes.

Pre-processing—Perl [19], a computer programming language that facilitates manipulation of a large volume of text data, was used to prepare nursing free text for NLP. The pre-processing involved 1) the addition of a colon (:) to identify the end of each section name and 2) the addition of a period (.) to identify the end of each sentence. This routine pre-processing was a required technical step for MedLEE processing.

Previous research demonstrated that words undefined by the NLP tool, such as abbreviations, symbols, or specialized vocabulary, might affect NLP performance on extracting terms since the words do not yet exist in its internal dictionary [2]. In regard to

this, we ran the first NLP session in order to purposely collect undefined words by MedLEE in our data. The undefined words identified from the output of the NLP were mostly abbreviations (e.g., D/C [discontinue], premed [premeditated]) and non-medical words (e.g., endorse, team). The undefined words occurred since they were not in MedLEE's current lexicon. We replaced the abbreviations with full names to reduce the number of undefined terms from MedLEE's parsing of nursing narratives. After these replacements, we ran the second NLP session (see extraction of data in following section) and analyzed the output to examine which nursing concepts were extracted from the nursing narratives. Although our primary interest was the nursing concepts extracted by MedLEE, we were also interested in the abbreviations since they were commonly used in nursing notes. Thus, we collected the unknown abbreviations (Table 1).

Extraction of Data—MedLEE extracted concepts from the nursing narratives. MedLEE labeled the extracted terms with various tags, such as problem, device, procedure, bodymeas (body measure), labtest, status, med (medication), normalfinding (normal finding), timeper (time period), bodyfunc (body function), and finding. The extracted data were summarized using descriptive statistics.

Use for Quality/Safety—In order to determine what terms of relevance to patient safety and quality of care measures can be extracted by MedLEE from the nursing free text data, two measures (chemotherapy side effects and pain management) were selected by investigators by referring to the American Nurses Association (ANA)'s patient safety initiatives, National Quality Forum (NQF)'s quality measures, and a patient safety report [20].

Gold standards were needed to compare against the terms extracted through NLP. Clinical practice guidelines (CPGs) from the National Guideline Clearinghouse's Web site [21] were selected as a source of terms of relevance to the measures. Two guidelines, Assessment and Management of Chronic Pain and Assessment and Management of Acute Pain, published by Institute for Clinical Systems Improvement [22, 23], were used for *pain management*. For *chemotherapy side effects*, Chemotherapy and Biotherapy Guidelines and Recommendations for Practice, published by Oncology Nursing Society [24], was used as the source of terms of relevance to the measures. These processes are summarized in Figure 2.

The CPGs consisted of several typical categories, such as Scope, Methodology, Recommendations, Evidence Supporting the Recommendations, Benefits/Harms of Implementing the Guideline Recommendations, Contraindications, etc. The recommendation section was used as the source for terms of relevance. Six categories¹ of acute pain management and eleven categories² of chronic pain management were listed under 'Recommendations.' Each category had a one- or two-level hierarchy either listed in

¹1) Somatic Pain Treatment, 2) Visceral Pain Treatment, 3) Neuropathic Pain Treatment, 4) Prevention/Intervention, 5) Pharmacological Therapy, and 6) Patient Controlled Analgesia (PCA).

²1) General Management; 2) Physical Rehabilitation and Psychosocial Management; 3) Psychosocial Management; 4) Pharmacologic Management; 5) Intervention Management; 6) Complementary Management; 7) Level I Treatment: Neuropathic Pain; 8) Level I Treatment: Muscle Pain; 9) Level I Treatment: Inflammatory Pain; 10) Level I Treatment: Mechanical/Compressive Pain; and 11) Level II Treatment: Interdisciplinary Team Referral, Plus a Pain Medicine Specialist or Pain Medicine Specialty Clinic.

bulleted or paragraph form; these compromised terms of relevance to *pain management* in this study. For instance, ‘acetaminophen,’ ‘cold packs,’ and ‘corticosteroids’ were components of Somatic Pain Treatment.

For the chemotherapy CPG, two sections, Immediate Complications of Cytotoxic Therapy (e.g., extravasation, hypersensitivity) and Side Effects of Cancer Therapy (e.g., myelosuppression, gastrointestinal and mucosal side effects) under ‘Recommendations’, were used to obtain terms of relevance to *chemotherapy side effects* for this study. For instance, Neutropenia, Anemia, and Thrombocytopenia were obtained from Myelosuppression (a complication); and Nausea, Vomiting, Diarrhea, Mucositis, and Anorexia were obtained from Gastrointestinal and mucosal side effects.

MedLEE tagged terms were compared to the gold standards to examine whether the terms of relevance to *pain management* and *chemotherapy side effects* could be identified from the free text data through NLP. When the term was a drug name, the drug information database from MedlinePlus [25], was used to look up synonyms of the particular drug for accurate matching. The results were summarized using descriptive statistics.

Results

Extraction of Data

Five hundred and fifty-three nursing progress notes associated with 22 patients were processed by MedLEE. Table 2 shows an example of sample text and its associated output from MedLEE. Words underlined in the output from NLP represent the terms captured with a tag.

We classified the abbreviations collected from the first NLP into three categories: 1) Abbreviations in a predefined hospital abbreviation list, but currently not in the MedLEE lexicon, 2) Abbreviations not in the hospital abbreviation list, but common in the nursing free texts, and 3) Abbreviations that commonly appeared in the nursing free texts could be interpreted with more than one meaning.

Examples of the first category were VSS (vital signs stable), NS (normal saline), BM (bowel movement), PICC (peripheral insertion central catheter), RN (registered nurse), FFP (fresh frozen plasma), OOB (out of bed), RUE (right upper extremity), D&C (dilatation and curettage), N&V (nausea and vomiting), and C&S (Culture and Sensitivity). Some examples of the second category were cont (continue), app (approximately), premed/premeds (premeditated), C+S (Culture and Sensitivity), and plt/plts (platelets). Examples of the last category were DL (Direct Laryngoscopy; Double Lumen), Cx (Cervix; culture – meaning blood culture), n (nausea; night), IVF (in vitro fertilization; intravenous fluid), and D/C (discharge; discontinue).

After the second NLP session, 3929 terms were captured by MedLEE. Some terms appeared in more than one narrative; therefore, the total number of nonredundant terms was 490. Selected examples of terms with associated MedLEE tags are shown in Table 3.

We made a list of undefined words so that the list could be added to the MedLEE lexicon for better NLP on nursing narratives in the future. Table 4 displays undefined words with frequencies equal to or greater than 10.

Use for quality/safety

Fifty-five terms relevant to *pain management* (e.g., tylenol, heat, massage) and 35 terms relevant to *chemotherapy side effects* (e.g., nausea, constipation) were selected from the CPGs (Tables 5 & 6).

The terms extracted through NLP were compared with the CPG-based gold standards. In several cases, the concepts from the CPGs were broader than the concepts extracted through NLP; for instance, 'Cardiac Toxicity' (a term from the CPGs) and 'Palpitations' (a term extracted through NLP). These cases were regarded as a difference in granularity of the terms and considered a non-match. About 18% and 43% of terms extracted using MedLEE were matched with the terms of relevance to *pain management* and *chemotherapy side effects*, respectively (Table 7 and 8).

Discussion

We could identify terms that were frequently documented in our sample. They were reaction, pain, nausea, adverse reaction, distress, emesis, shortness of breath, chill, discomfort, bleeding, swelling, febrile, headache, cough, and vomiting. Since the data were from the oncology nursing progress notes, these terms appear to be most frequently monitored and recorded by the oncology nurses in free text.

Terms that frequently appeared in nursing free text, but currently not in the MedLEE lexicon included *ordered, aware, premedicated, informed, sensitivity, flowsheet* and *incident*. In order for MedLEE to perform better on nursing free text data, these undefined words need to be added to the MedLEE lexicon.

The output of the first NLP session revealed many abbreviations frequently used in nursing narratives in oncology nursing progress notes. Some abbreviations truncated at the end of a word were commonly found in our sample, such as *premed/premeds* (premeditated) and *cont* (continue). Other abbreviations frequently found in the data, but not on the hospital abbreviation list included *C+S* (Culture and Sensitivity) and *plt/plts* (platelet). Other abbreviations had more than one meaning, such as *Cx* (cervix; blood culture). Some abbreviations were on the hospital abbreviation list indicating that they are legitimate abbreviations, but were absent in the MedLEE lexicon, such as *VSS* (vital signs stable). While using abbreviations in patient records is a convenient and efficient way for nurses to document since many medical terms are very long, the use of abbreviations may be a detriment to patient safety. One of the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) National Patient Safety Goals is to improve the effectiveness of communication among caregivers [26]. To meet this goal, the hospital wishes to designate a list of abbreviations and acronyms that are not to be used because their use increases the risk of medical error. In addition, the use of any unauthorized abbreviations is not allowed by the

JCAHO requirements. In this context, NLP may support the identification of unauthorized clinical abbreviations from free text in electronic records.

Nurses directly affect the lives of patients [27] and measuring the performance of care that patients receive is indispensable to evaluating the effectiveness of nursing care. In an attempt to capture quality/safety measures from nursing free text, a new method, NLP, was explored in this study. Matching between terms of relevance from the gold standards and MedLEE output is a function of MedLEE performance and the free text in the corpus of the oncology nursing narratives. Terms that are not in the corpus cannot be found by MedLEE; therefore, the percent matching (i.e., 18% matching for *pain management*, 43% matching for *chemotherapy side effects*) should not be viewed as a measure of MedLEE performance. Although, further study is desired to validate the findings, the results from this exploratory study demonstrate that NLP has the potential to capture data related to safety/quality measures in the domain of nursing oncology notes. It may assist nurses in evaluating patient progress and determining which interventions are effective and which are ineffective. In addition, NLP may provide a method to assess nursing outcomes that are not captured through structured data entry.

NLP may contribute significantly to the availability of nursing records in EHRs for reuse in clinical decision support and patient outcomes research and, ultimately, the quality of care. Strategies are needed to handle special features of nursing records and vocabulary that are unique to the domain of nursing.

The generalizability of the findings of this study is limited since the free texts were oncology nursing narratives and only from one institution. Further research is needed to examine the use of NLP on nursing narratives from various nursing specialties and across institutions which may vary in their documentation procedures.

Conclusion

We explored the ability of NLP on nursing narratives. NLP may provide a feasible approach to extract data related to patient safety/quality measures and nursing outcomes by capturing nursing concepts that are not documented through structured data entry. For better NLP performance in the domain of nursing, additional nursing terms and abbreviations must be added to MedLEE's lexicon.

Acknowledgments

This study was supported by 1R01LM07593 (S. Johnson, Principal Investigator) from the National Library of Medicine and P20NR007799 (S. Bakken, Principal Investigator) from the National Institute of Nursing Research, USA. We would like to thank Drs. Carol Friedman and Leanne Currie for their assistance and support.

References

1. Barrows RC, Busuioc M, Friedman C. Limited parsing of notational text visit notes: Ad-hoc vs. NLP approaches. Proc AMIA Symp. 2000:51–5. [PubMed: 11079843]
2. Hyun S, Bakken S, Friedman C, Johnson S. Natural language processing challenges in HIV/AIDS clinic notes. AMIA Annu Symp Proc. 2003:872. [PubMed: 14728377]

3. Jha A, Kuperman G, Teich J, Leape L, Shea B, Rittenberg E, et al. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *J Am Med Inform Assoc.* 1998; 5(3):305–14. [PubMed: 9609500]
4. Friedman C. A broad coverage Natural Language processing system. *Proc AMIA Symp.* 2000:270–4. [PubMed: 11079887]
5. Honigman BA. A computerized method for identifying incidents associated with adverse drug events in outpatients. *International Journal of Medical Informatics.* 2001; 61(1):21–32. [PubMed: 11248601]
6. Hurdle J, Weir C, Roth B, Hoffman J, Nebeker J. Critical gaps in the world's largest electronic medical record: Ad Hoc nursing narratives and invisible adverse drug events. *AMIA Annu Symp Proc;* 2003. 2003:309–12.
7. Fiszman M, Haug P. Using medical language processing to support real-time evaluation of pneumonia guideline. *Proc AMIA Symp.* 2000:235–9. [PubMed: 11079880]
8. Honigman B. A computerized method for identifying incidents associated with adverse drug events in outpatients. *International Journal of Medical Informatics.* 2001; 61(1):21–32. [PubMed: 11248601]
9. Friedman C, Knirsch CA, Shagina L, Hripcsak G. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. *Proc AMIA Symp.* 1999:256–60. [PubMed: 10566360]
10. Jain NL, Friedman C. Identification of findings suspicious for breast cancer based on natural language processing of mammogram reports. *Proc AMIA Symp.* 1997:829–33.
11. Jain NL, Knirsch CA, Friedman C, Hripcsak G. Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports. *J Am Med Inform Assoc.* 1996; 3(Suppl):542–6.
12. Barrows RC, Busuioic M, Friedman C. Limited parsing of notational text visit notes: ad-hoc vs. NLP approach. *Proc AMIA Symp.* 2000:51–5. [PubMed: 11079843]
13. Scalise D. Clinical communication and patient safety. *Hospitals & Health Networks.* 2006; 80(8): 50. [PubMed: 16838807]
14. Scoates G, Fishman M, McAdam B. Health care focus documentation--more efficient charting. *Nurs Manage.* 1996; 27(8):30–2. [PubMed: 8850969]
15. Hyun, S.; Bakken, S.; Friedman, C.; Johnson, S. Natural language processing challenges in HIV/AIDS clinic notes (preliminary evaluation). In: Fieschi, Marius; Coiera, Enrico; Li, J., editors. *Medinfo;* 2004. San Francisco: 2004. p. 1655
16. Hsieh Y, Hardardottir G, Brennan P. Linguistic analysis: terms and phrases used by patients in e-mail messages to nurses. *Medinfo;* 2004. 2004:511–5.
17. Bakken, S.; Hyun, S.; Friedman, C.; Johnson, S. *Medinfo;* 2004. Vol. 11. San Francisco: 2004. A comparison of semantic categories of the ISO reference terminology models for nursing and the MedLEE natural language processing system; p. 472-6.
18. Bakken S, Hyun S, Friedman C, Johnson S. ISO reference terminology models for nursing: applicability for natural language processing of nursing narratives. *Int J Med Inform.* 2005; 74(7–8):615–22. [PubMed: 16043086]
19. The Source for Perl. Documentation. Apr 21. 2005 [cited 2005 April, 22]; Available from: <http://perl.oreilly.com/>
20. Committee on Data Standards for Patient Safety. *Patient Safety: Achieving a New Standard for Care.* 2004.
21. the National Guideline Clearinghouse 2006. Oct 9. 2006 [cited 2006 August, 3]; Available from: <http://www.guideline.gov/>
22. Institute for Clinical Systems Improvement Assessment and management of chronic pain. 2005. [cited 2006; Available from: http://www.guideline.gov/summary/summary.aspx?doc_id=8363&nbr=4684
23. Institute for Clinical Systems Improvement. Assessment and management of acute pain. 2006. [cited 2006; Available from: http://www.guideline.gov/summary/summary.aspx?doc_id=9009&nbr=004884&string=Assessment+and+%22management+of+acute+pain%22

24. Oncology Nursing Society. Chemotherapy and biotherapy guidelines and recommendations for practice. 2005. [cited 2006; 2nd ed.]; Available from: http://www.guideline.gov/summary/summary.aspx?doc_id=8337&nbr=004665&string=chemotherapy+AND+side+AND+effects
25. U.S. National Library of Medicine. MedlinePlus. 2006. [cited; Available from: <http://www.nlm.nih.gov/medlineplus/druginformation.html>
26. Joint Commission. 2007 National Patient Safety Goals. 2006. [cited; Available from: <http://www.jointcommission.org/PatientSafety/NationalPatientSafetyGoals/>
27. National Quality Forum. Nursing Care Quality at NQF. 2006

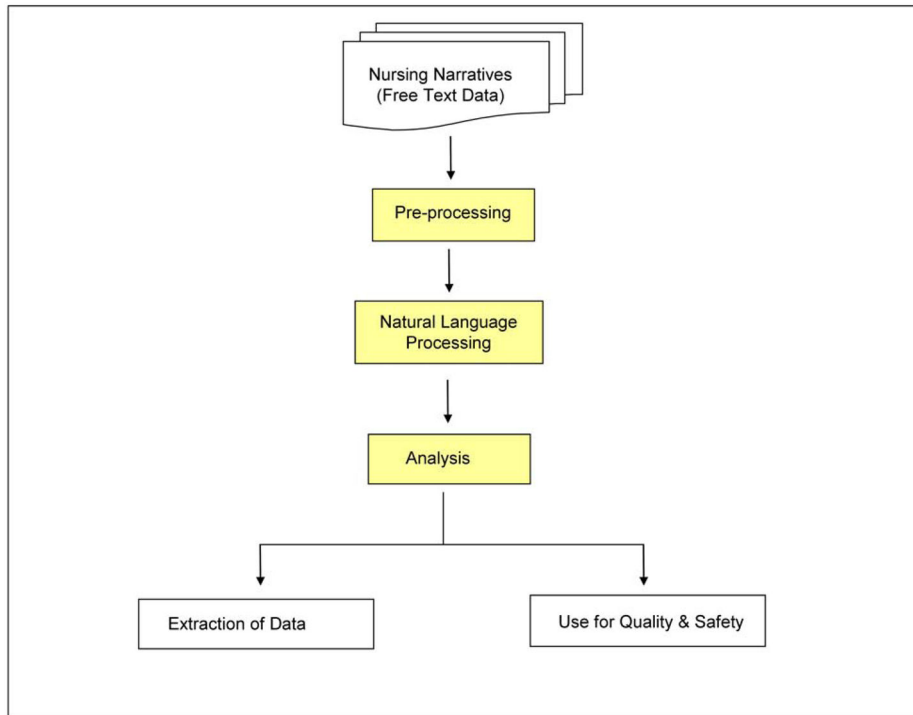


Figure 1.
Procedure

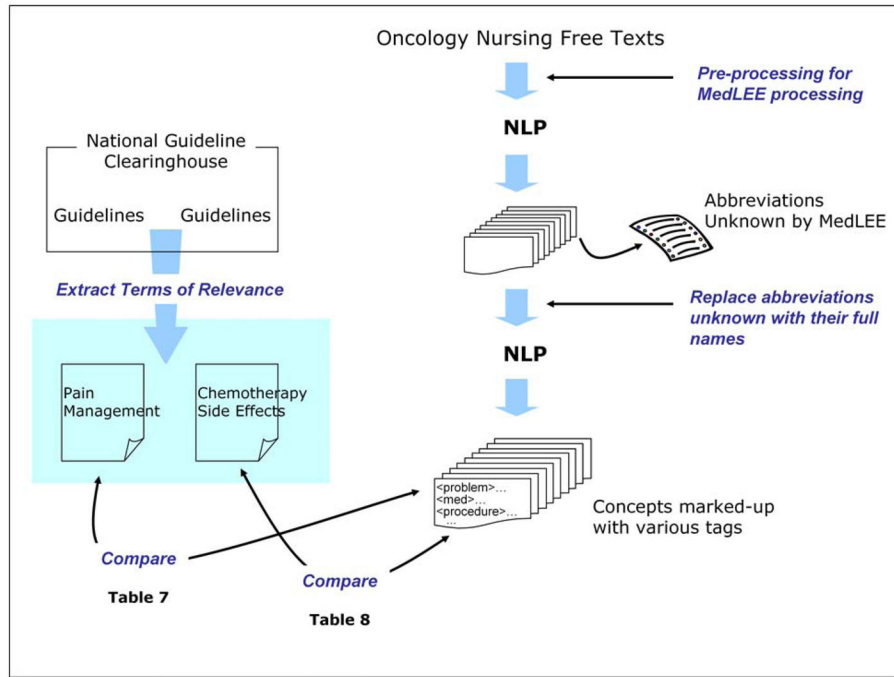


Figure 2.
Procedure – Use for quality/safety

Table 1

Examples of abbreviations unknown by MedLEE

Examples

Temp 38.2C at 10pm.

Pt voided 30 min post Foley D/C.

Stool cultures sent as ordered at app 945.

Pt with small BM at app 1400 per pt report.

Blood cultures x 2 from RUE DL PICC collected and sent to lab as ordered.

Pt continues on IVF D5 1/2 NS with 75 meq sodium bicarb @ 150 ml/hr.

Hgb = 7.8. 2 U of PRBCs were administered w/o reaction.

Pt is presently receiving plts and well thus far, will cont to monitor.

C/o abdominal tenderness also cont.

Pt's right forearm 20g PIV discontinued d/t redness at insertion site and pt c/o slight pain at site.

Note. Shadowed word: undefined word in MedLEE

Table 2

Original text and its corresponding MedLEE output

Original text	Output from NLP
<i>Platelets 45,000 from am labs.</i>	<p><u>labtest:platelets</u></p> <p>idref>> 11 measure>> [45000, [idref, 13]] parsemode>> mode2 sectname>> report summary item sid>> 1</p> <p><u>procedure:laboratory test</u></p> <p>idref>> 19 parsemode>> mode5 sectname>> report summary item sid>> 1 code>> UMLS:C0022885_laboratory procedure idref>> [19]</p>
<i>Pt with <u>redness</u> to sclera of right eye.</i>	<p><u>finding:redness</u></p> <p>certainty>> high certainty idref>> 28 bodyloc>> sclera idref>> 34 bodyloc>> eye idref>> 40 region>> right idref>> 38 code>> UMLS:C0229089_right eye structure idref>> [38, 40] code>> UMLS:C0929516_right sclera idref>> [34, 38] idref>> 30 parsemode>> model sectname>> report summary item sid>> 2 code>> UMLS:C0235267_redness of eye idref>> [30, 40]</p>
<i>Patient medicated with <u>Tylenol</u> 650 mg PO and <u>Benadryl</u> 25 mg IV as ordered.</i>	<p><u>med:tylenol</u></p> <p>dose>> [650, [idref, 55], mg, [idref, 57]] idref>> 53 manner>> po idref>> 59 parsemode>> mode2 sectname>> report summary item sid>> 3 code>> UMLS:C0699142_Tylenol idref>> [53]</p> <p><u>med:benadryl</u></p> <p>dose>> [25, [idref, 65], mg, [idref, 67]] idref>> 63 manner>> intravenous</p>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Original text	Output from NLP
<p><i>Transfused with <u>platelets</u> 1 unit as ordered.</i></p>	<pre> idref>> 69 parsemode>> mode4 sectname>> report summary item sid>> 3 code>> UMLS:C0700899_Benadryl idref>> [63] <u>procedure:transfusion</u> idref>> 80 parsemode>> mode2 sectname>> report summary item sid>> 4 code>> UMLS:C0199960_transfusion - action idref>> [80] <u>labtest:platelets</u> idref>> 84 measure>> [1, [idref, 86]] parsemode>> mode4 sectname>> report summary item sid>> 4 </pre>
<p><i>No <u>adverse reaction</u> noted.</i></p>	<pre> <u>problem:adverse reaction</u> certainty>> no idref>> 99 idref>> 101 parsemode>> mode1 sectname>> report summary item sid>> 5 code>> UMLS:C0879626_adverse effect idref>> [101] </pre>

Platelets 45,000 from am labs. Pt with redness to sclera of right eye.

Patient medicated with Tylenol 650 mg PO and Benadryl 25 mg IV as ordered.

Transfused with platelets 1 unit as ordered.

No adverse reaction noted...

Table 3

Examples of extracted terms with MedLEE tags

Term	Tag
vital signs	bodymeas
heart rate	bodymeas
oxygen sat	bodymeas
nebulizer	device
monitor	device
Foley catheter	device
face mask	device
chest tube	device
patient-controlled analgesia	device
potassium	labtest
erythrocyte sedimentation rate	labtest
hemoglobin	labtest
platelets	labtest
complete blood count	labtest
Tylenol	med
Benadryl	med
Rituxan	med
packed red blood cells	med
oxygen	med
ara-C	med
gentamycin	med
vancomycin	med
computerized axial tomography	procedure
telemetry	procedure
chemotherapy	procedure
assessment	procedure
liver function test	procedure
magnetic resonance imaging	procedure
intervention	procedure
biopsy	procedure
blood culture	procedure
laboratory test	procedure
sputum culture	procedure
transfusion	procedure
nursing therapy	procedure
pain measurement	procedure
assessment	procedure

Term	Tag
x-ray	procedure
clothing assistance	procedure
supportive care	procedure
prevention	procedure
toilet procedure	procedure
skin care	procedure
pain scale	procedure
adverse reaction	problem
emesis	problem
shortness of breath	problem
chill	problem
palpable	problem
discomfort	problem
bleeding	problem
febrile	problem
chest pain	problem
wheezing	problem
abdominal pain	problem
dizzy	problem
redness	finding
good	finding
distended	finding
elevated	finding
no change	finding
asymptomatic	finding
streaky	finding
floating	finding
flushing	finding
intact	finding
non-infectious	finding
watery	finding
alert	finding
withdrawn	finding
active disease	normalfinding
afebrile	normalfinding
alert	normalfinding
comfortable	normalfinding
intact	normalfinding
within normal limits	normalfinding

Term	Tag
alert and oriented times 3	normalfinding
doing well	normalfinding
pupils equal	normalfinding
complaint	status
stable	status
pending	status
previous	status
ability	status
again	status
resident	status

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Undefined words in MedLEE

Undefined word	Frequency
ordered	310
aware	235
sent	87
order	82
unit	74
adverse	59
flowsheet	57
endorse	51
premedicated	49
well	47
notified	35
team	35
orders	33
voiced	29
Protocol/protocol	29
escort	27
slept	25
awaiting	24
reviewed	22
medicated	20
sensitivity	20
sheet	20
infusing	19
port	19
states	18
endorsed	17
incident	17
results	17
bag	16
collected	15
consent	15
implemented	15
informed	14
see	14
call	12
care	11
closely	11

Undefined word	Frequency
due	10
transparent	10

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5Terms of relevance to *pain management* from CPGs

Term
Acetaminophen
Acupuncture
Adverse drug reactions
Amitriptyline (Elavil)
Analgesia
Anticonvulsant
Antidepressants
Antiepileptic drugs
Aquatic therapy
Art therapy
Balance activity
Biofeedback
Capsaicin
Carbamazepine (Atretol;Carbatrol;Epilex;Equetro;Tegretol;Tegretol-XR)
Cold packs
Consultation
Coping
Corticosteroid
Cyclobenzaprine (Flexeril)
Desipramine (Norpramin;Pertofrane)
Duloxetine (Cymbalta)
Endurance activities
Epidural corticosteroid injections
Exercise
Flexibility
Heat
Imipramine (Tofranil)
Immobilization
Ketamine
Kyphoplasty
Lidoderm (Lidocaine)
Massage
Methadone
Music therapy
Nerve blocks
Neurotomy
Nonsteroidal anti-inflammatory drugs (nsaids)
Non-Tricyclic Anti-depressants
Nortriptyline (Aventyl;Pamelor)

Term
Opioids
Patient controlled analgesia (PCA)
Play therapy
Referral
Relaxation
Rest
Self-management
Strengthening
Surgery
Topical Agents
Topical lidocaine
Tramadol
Transcutaneous electrical nerve stimulation
Tricyclic Anti-depressants (tcas)
Trigger point injections
Vertebroplasty

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6Terms of relevance to *chemotherapy side effects* from CPG

Term
Alopecia
Alterations in sexuality and reproductive function
Anaphylaxis
Anemia
Anorexia
Bleeding
Blood in stool or urine
Bruises
Cardiac toxicity
Constipation
Cutaneous toxicity
Diarrhea
Ecchymoses
Extravasation
Fatigue
Flare reaction
Headaches
Hemorrhagic cystitis
Hepatotoxicity
Hypersensitivity
Mucositis
Myelosuppression
Nausea
Nephrotoxicity
Neurotoxicity
Neutropenia
Nosebleeds
Ocular toxicity
Pancreatitis
Petechiae
Prolonged menstruation
Pulmonary toxicity
Secondary malignancies
Thrombocytopenia
Vomiting

Table 7Match between terms of relevance to *pain management* and terms extracted through NLP

Pain Management	Number of Matched Terms (%)
Yes	10 (18.18)
No	40 (72.73)
Different Granularity	5 (9.09)
Total	55 (100)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 8

Match between terms of relevance to *chemotherapy side effects* and terms extracted through NLP

Chemotherapy Side Effects	Number of Matched Terms (%)
Yes	15 (42.86)
No	12 (34.29)
Different Granularity	8 (22.86)
Total	35 (100)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript