Received: 11 March 2021 | Revised: 6 May 2021 | Accepted: 12 May 2021

DOI: 10.1111/iwj.13623

ORIGINAL ARTICLE

IWJ WILEY

Exploring prevalence of wound infections and related patient characteristics in homecare using natural language processing

Kyungmi Woo[1] | **Jiyoun Song**[2] | **Victoria Adams**[3] | **Lorraine J. Block**[4] | **Leanne M. Currie**[4] | **Jingjing Shang**[2] | **Maxim Topaz**[2,3,5]

[1]College of Nursing, Seoul National University, Seoul, South Korea

[2]School of Nursing, Columbia University, New York City, New York

[3]Visiting Nurse Service of New York, New York City, New York

[4]School of Nursing, University of British Columbia, Vancouver, British Columbia, Canada

[5]Data Science Institute, Columbia University, New York City, New York

Correspondence

Kyungmi Woo, PhD, RN, The Research Institute of Nursing Science, College of Nursing, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul, South Korea 03080.
Email: woo2020@snu.ac.kr

Funding information

Intramural Pilot Grant from Columbia University School of Nursing; National Institute of Nursing Research, Grant/Award Number: T32 NR014205; Eugenie and Joseph Doyle Research Partnership Fund from Visiting Nurses Service of New York

Abstract

We aimed to create and validate a natural language processing algorithm to extract wound infection-related information from nursing notes. We also estimated wound infection prevalence in homecare settings and described related patient characteristics. In this retrospective cohort study, a natural language processing algorithm was developed and validated against a gold standard testing set. Cases with wound infection were identified using the algorithm and linked to Outcome and Assessment Information Set data to identify related patient characteristics. The final version of the natural language processing vocabulary contained 3914 terms and expressions related to the presence of wound infection. The natural language processing algorithm achieved overall good performance (F-measure = 0.88). The presence of wound infection was documented for 1.03% (n = 602) of patients without wounds, for 5.95% (n = 3232) of patients with wounds, and 19.19% (n = 152) of patients with wound-related hospitalisation or emergency department visits. Diabetes, peripheral vascular disease, and skin ulcer were significantly associated with wound infection among homecare patients. Our findings suggest that nurses frequently document wound infection-related information. The use of natural language processing demonstrated that valuable information can be extracted from nursing notes which can be used to improve our understanding of the care needs of people receiving homecare. By linking findings from clinical nursing notes with additional structured data, we can analyse related patients' characteristics and use them to develop a tailored intervention that may potentially lead to reduced wound infection-related hospitalizations.

KEYWORDS

natural language processing (NLP), wound infection, home health care, nursing notes, Outcome and Assessment Information Set (OASIS)

Key Messages

- detailed information on wounds and wound infections is commonly documented in narrative clinical notes than structured electronic health record data
- nurses frequently document wound infection-related information. The NLP algorithm achieved overall good performance (F-measure = 0.88). The presence of wound infection was documented for 19.19% (n = 152) of patients with wound-related hospitalisation or emergency department visits. Diabetes, peripheral vascular disease, and skin ulcer were significantly associated with wound infection among homecare patients
- the use of natural language processing demonstrated that valuable information can be extracted from nursing notes which can be used to improve our understanding of the care needs of people receiving homecare

1 | INTRODUCTION

Outpatient settings are becoming increasingly important as healthcare systems in the world aim to reduce costs and improve the quality of care. In 2017, 3.4 million Medicare beneficiaries received homecare services from more than 12 000 agencies nationwide.[1] In 2014, almost 10% (5.5 million) of Medicare beneficiaries reported a diagnosis of wound infection, with treatment costs reaching approximately $30 billion in Medicare spending.[2] In homecare, wound infections are one of the main causes of hospitalisation, and their treatment are costly and burdensome not only for patients but also their families.[3] Previous research found that wound infections were one of the top five reasons for hospitalisation or emergency department (ED) visits among homecare patients.[4] Despite this, a comprehensive picture of the prevalence of wound infection among larger homecare populations, its associated indicators and the characteristics of patients with wound infection remains largely unknown.

Opportunity to examine the prevalence of wound infections exists through evaluating the growing amount of electronic patient data collected by homecare agencies. However, one of the challenges in using electronic health record (EHR) data is that up to 50% of this content is stored in an unstructured format (eg, narrative free text charting).[5] This is further complicated as most post-acute care records contain no structured wound assessment content.[6] For example, routine assessments with standardised tools are performed only at admission, transfer, and discharge from service in the United States.[4] More commonly, narrative clinical notes are recorded during the home visits which include detailed information on wounds and wound infections. Given the reporting capabilities of many organisations and agencies (ie, the use of codified structured data), these unstructured free-text notes remain largely underutilised.

Innovative data science methods, such as natural language processing (NLP), can be leveraged to extract valuable data from narrative clinical notes. NLP refers to a set of computer algorithms or systems that process human languages. Although NLP has been increasingly applied in the healthcare domain,[7,8] to our knowledge, our team has been the only group to use NLP to analyse wound information.[5] In our previous work, we used NLP to extract wound-related information from inpatient clinical notes.[5] The study found that over half of the wounds documented in the clinical notes (55%) did not have a wound-related diagnosis code in the structured data. A similar trend of explicating additional critical information with NLP has been demonstrated in other studies examining mental health,[9-11] oncology,[12,13] and other health domains such as depression, food allergy, and family representation.[11,14-18] Given the high prevalence of wounds in homecare, a method to address the wound infection information gap is warranted.

2 | AIMS

This study aimed to bridge this research gap by using an innovative data science method. Specifically, the aims were (1) to create and validate an NLP algorithm to extract wound infection-related information from clinical notes, (2) to estimate the prevalence of wound infections, and (3) to describe related patient characteristics by linking the NLP identified wound infections to structured data in homecare.

3 | METHODS

This was a retrospective cohort study that used the clinical nursing notes and Outcome and Assessment

Information Set (OASIS) data. We integrated and analysed ~2 million nursing notes for 89 459 patients who had documentation of wound infections. For the purpose of this work, we defined a wound as an injury to the body (as from accident or surgery) that typically involves laceration or breaking of the skin and damage to underlying tissues.[19] Furthermore, we specified a wound infection to be an infection at a local wound site or an infection related to the wound. The institutional review boards of Columbia University (reference number IRB-AAAS5545) and the Visiting Nurse Service of New York (E19-004) approved the study protocol.

3.1 | Data sets

We used information obtained from two data sets.

3.1.1 | Clinical notes

Homecare nursing visit notes (n = 1 149 586) and care coordination notes (n = 1 461 171) for 89 459 patients were extracted from the largest non-profit homecare agency in the United States between 1 January 2014 and 31 December 2014. The dataset included 112 789 unique episodes of care for all patients (during a period of time of up to 60 days from admission to the end of homecare services). The average visit note length was 150 words and the care coordination notes had an average of 50 words.

3.1.2 | Outcome and Assessment Information Set (OASIS)

OASIS is a standardised homecare patient assessment that tracks nearly 100 patient characteristics in the domains of socio-demographics, medical history, health status, environmental status, support system, functional status, and health service utilisation.[4] OASIS is the only standardised homecare patient assessment required by the Center for Medicare and Medicaid Services for all Medicare-certified homecare agencies on the national level. OASIS assessments are performed upon admission and the end of a care episode. The version used in our analysis, OASIS-C, was released in 2009. We extracted homecare admission OASIS data for all patients included in the study sample. This included socio-demographic characteristics and clinical status. In addition, homecare nurses are required to document patients' wound status at admission and a reason for hospital or ED admission from homecare. We used this structured data field in OASIS (for wound status: OASIS dataset items M1306, M1330, M1340, and M1350, and for hospitalisation or ED visits: M2430 and M2310) to identify patients with wound and wound infection-related hospitalisation. We only used OASIS items for wound infection-related hospitalisation or ED visits.

3.2 | NLP algorithm creation

We synthesised the literature on wound infection collected from various health research databases and applied our clinical expertise of homecare to generate a list of candidate wound infection related categories. Next, we used a large standardised health terminology database (Unified Medical Language System [UMLS])[20] to identify a preliminary list of terms for each wound infection category. We also extracted lists of UMLS synonyms to enhance the information schema. Two of our team members (a Certified Wound, Ostomy and Continence Nurse [CWOCN] and a Nursing PhD student) reviewed the preliminary list independently then validated and finalised a list of specific and non-specific wound infection symptoms and treatment categories (n = 9) (Table 1).

The NLP algorithm used in the study was developed using multiple methods. Specifically, (1) initial concept identification using literature synthesis and clinical expertise, (2) face validation of concepts with clinical expert and concept validation and expansion using standardised health terminology database, (3) interactive rapid vocabulary explorer, (4) label assignment and review, and finally (5) algorithm testing. Table S in the supplementary material provides more information about each stage.

3.3 | Detailed description of step 3: Interactive rapid vocabulary explorer

The first stage in developing the algorithm is creating a language model. Language models are statistical representations of a certain body of text. To create our language model in NimbleMiner, we identified a large corpus of clinical notes and used a specific type of language model called word embedding models.[21] A word embedding model enables us to identify similar terms in the clinical notes and build a vocabulary based upon our topic of interest.

The next stage is aimed at helping users to rapidly discover large vocabularies of relevant terms and expressions. In this study, interactive rapid vocabulary explorer was implemented by two nurses, who are experts in homecare. The user enters a query term of interest (for example 'wound infection'), and the system returns a list

TABLE 1 Example words and expressions identified in each category of wound infection

Category	Example words and expressions				
Wound type	Open blister	Venous ulcer	Surgical wound	wd	Pressure ulcer
Wound infection	Inflamed ulcer	Local infection of wound	Cellulitis	Surgical site infection	Incision infection
Exudate	Scant purulent drainage	Seropurulent	Draining large amts	White slough	Serous drainage
Foul odour	Bad odour	Bad smell	foul odour	Malodor	Offensive odour
Periwound skin	Swollen wound	Edematous	Granulated slough	Redness	erythema noted
Wound bed tissue	Hyper-granulated tissue	Non-granulating	New necrotic tissue	Bridging	Tunnelling
Spreading systemic signs	Vomiting	Confused	Feverish	So exhausted	Disoriented lethargic
Possible wound infection name	Gangrene	Folliculitis	Necrotizing fasciitis	Skin necrosis	Erysipelas
Possible wound infection treatment	iv abx	Antibiotic ointment	Apply silvadene cream	Medihoney	Surgical debridement

Abbreviations: iv abx, intravenous antibiotics; wd, wound.

of similar terms it identified as relevant (eg, 'infected ulcer', 'infected wounds', and 'wd infect'). In our case, we pre-populated lists of similar expressions for each of the wound infection-related information categories extracted from the UMLS. The user selects and saves the relevant terms by clicking on them in the interactive vocabulary explorer screen. Negated terms or other irrelevant terms not selected by the user are also saved in the system for further tasks, such as negation detection.

In the final stage, the system uses similar terms discovered by the user during stage 2 to assign labels to clinical notes (while excluding notes with negations and other irrelevant terms). Assigning a positive label means that a concept of interest is present in the clinical note. When needed, the user reviews and updates the list of similar terms and negated terms. The user reviews the clinical notes with assigned labels for accuracy. This weakly supervised rapid labelling approach is based on a positive label learning framework validated in previous research.[22,23]

Further details about NimbleMiner's architecture (Figure 1) are described in detail elsewhere[24] and the system can be downloaded from http://github.com/mtopaz/NimbleMiner under General Public License v3.0.

3.4 | NLP algorithm testing

To test the accuracy of our NLP algorithm, we created a gold standard testing set of clinical notes using a high likelihood sampling approach as follows. First, we identified a subset of patients admitted to a hospital for a wound infection during a homecare episode, as indicated in the structured data. Among these patients, we extracted a random subset of 200 clinical notes (50% visit notes and 50% care coordination notes). Each note was annotated by two expert reviewers for the presence of one or more of the nine wound infection-related information categories. The interrater reliability was relatively high (kappa = 0.72), indicating good agreement between reviewers.[25] All disagreements were discussed until a final consensus was reached.

Next, we applied our NLP system to the gold standard testing set and for each category calculated precision (defined as the number of true positives out of the total number of predicted positives), recall (the number of true positives out of actual number of positives), and F-score (the weighted harmonic mean of the precision and recall).

4 | DATA ANALYSIS

We linked patient data identified via NLP with OASIS data (socio-demographic characteristics and clinical status) at a homecare episode level. We used the merged dataset to characterise the patients with wound infections and compared patient characteristics between patients with and without wound infection information.

4.1 | Comparisons based on structured data

We identified cases with any type of wounds at admission to homecare based on the OASIS items (N = 54 316). We

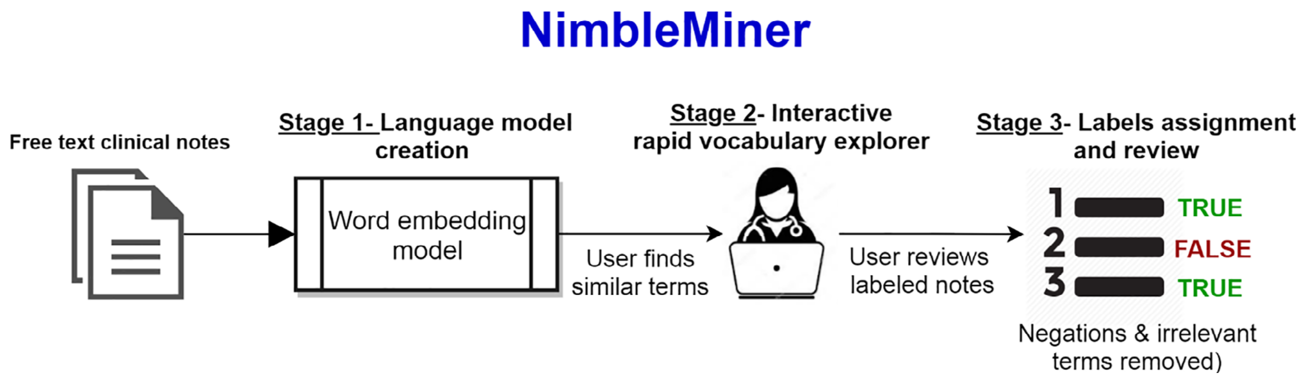

FIGURE 1 NimbleMiner architecture

then compared the frequency of wound infection-related information in the clinical notes identified by the NLP algorithm among patients who were admitted to homecare with wounds and those without wounds as identified by OASIS. We further identified patients who were hospitalised or admitted to the ED for wound infection during a homecare episode for frequency analysis of wound infection-related information from NLP.

Among patients with a wound at homecare admission (N = 54 316) identified by OASIS, differences between patients with and without documentation of wound infection identified by the NLP algorithm were evaluated for significance using t tests for the continuous variables (eg, age) and Pearson's chi-square test for the categorical or binary variables.[26,27] Statistical significance was set at $P < 0.05$. Given the large sample size, we had sufficient power to detect statistically significant differences between the groups.[26]

5 | RESULTS

5.1 | NLP system performance

The final version of the NLP lexicon contained 3914 terms and expressions related to the presence of wound infection, including a number of variations and misspellings. Table 1 displays examples of words and expressions identified in each of the nine categories of wound infection.

Below are some sample sentences that include some of these words or expressions in context:

> Example Note 1: … Wd consult recommended usage of iodosorb to debrid the wd and for antoimicrobial also reported ss of cellulitis: redness, and edema ant calf area shiny. MD will call in Rx for antibiotics to the XXX pharmacie number given, and also MD agreable on wd consult recom usage of iodosorb TIW and DSD. MD told VN that pt has a h/o refusing certain Tx …he refused to use Medihoney for debridment of wd before due to pain after application of Medihoney. Both open wds have slough…

> Example Note 2: …wheel chair bound has home MD program visit monthfully due to not able to go out. Pt's son called to Dr. XXX that noted pts RT lower leg with redness, swelling, and pt c/o pain since this Tuesday. Case referred to XXX for nurse to assess pts skin and ss of cellulitis. Pt's RT ankle and instep has Tr 1+ edema RT lower leg warm to touch. Pt has 4 cm x 3 cm dry skin scab on dorsal aspect of the foot… sometime has serous drainage noted … SN called to Dr. XXX, Np. XXX and left message that pt needs abx for RT lower leg cellulitis…

The NLP algorithm achieved very good overall performance (F-score = 0.88) with high precision (0.87) and recall (0.91). The highest F-score and recall among categories were for 'Foul odour' and 'Possible wound infection treatment'. Highest precision was achieved for the categories 'Foul odour' and 'Wound bed tissue' (Table 2).

5.2 | NLP results

Table 3 presents the prevalence of wound infection-related information among different patient populations. There is a clear trend of increased prevalence among patients who were hospitalised or admitted to ED for wound infection compared with other patient populations. For example, the presence of wound infection was documented for 1.03% (n = 602) of patients without wounds, for 5.95% (n = 3232) of patients with wounds, and 19.19% (n = 152) of patients with wound-related hospitalisation or ED visit.

TABLE 2 Natural language processing (NLP) system performance

Category of wound infection-related information	Recall	Precision	F-measure
Wound type	0.93	0.92	0.89
Wound infection	0.85	0.97	0.89
Exudate	0.92	0.73	0.79
Foul odour	1.00	1.00	1.00
Periwound skin	0.84	0.84	0.84
Wound bed tissue	0.83	0.99	0.90
Spreading systemic signs	0.73	0.98	0.81
Possible wound infection name	0.83	0.83	0.83
Possible wound infection treatment	0.93	0.95	0.94
Overall	0.87	0.91	0.88

TABLE 3 Wound infection prevalence from natural language processing

Categories	Patients without wound (N = 58 472)	Patients with wound (N = 54 316)	Patients with wound-related hospitalisation or ED visits (N = 792)
# of mentions from NLP			
Wound type, n (%)	9134 (15.62%)	42 425 (78.11%)	763 (96.34%)
Wound infection, n (%)	602 (1.03%)	3232 (5.95%)	152 (19.19%)
Exudate, n (%)	1174 (2.01%)	11 675 (21.49%)	380 (47.98%)
Foul odour, n (%)	886 (1.52%)	1767 (3.25%)	129 (16.29%)
Periwound skin, n (%)	9131 (15.62%)	15 883 (29.24%)	373 (47.1%)
Wound bed tissue, n (%)	475 (0.81%)	2645 (4.87%)	75 (9.47%)
Spreading systemic signs, n (%)	17 682 (30.24%)	14 406 (26.52%)	311 (39.27%)
Possible wound infection name, n (%)	114 (0.19%)	1342 (2.47%)	82 (10.35%)
Possible wound infection treatment, n (%)	6702 (11.46%)	18 735 (34.49%)	555 (70.08%)

All differences were statistically significant ($P < 0.05$) using a chi-square test.

Figure 2 describes the appearance of wound infection-related information over time during homecare episodes that resulted in wound infection-related hospitalisation or ED visits. The figure suggests that the frequency of wound infection-related information documentation increased close to wound infection-related hospitalisation or ED visits, peaking within a few days before the event.

5.3 | Wound infection and related patient characteristics

Table 4 presents a descriptive analysis of the study patient sample with a documented wound at homecare admission (N = 54 316 cases), comparing the patients with and without documentation of wound infection in the nursing notes. A total of 3232 patients (5.95%) were identified from NLP as having a wound infection. We found that patients with wound infection were slightly younger (66.36 vs 67.04 years) and more likely to be White (51.13% vs 48.68%) or Hispanic (22.65% vs 20.43%) compared with patients without wound infection. For clinical characteristics, patients with wound infection were more likely to have no inpatient stay 14 days prior to the homecare admission (17.61% vs 16.26%) and have a urine incontinence (17.2% vs 15.18%) and intractable pain (16.99% vs 14.65%) compared with those without wound infection. Diabetes, skin ulcer, and peripheral vascular disease (PVD) were more frequently reported among patients with wound infection.

Patients with wound infection were less likely to have had private health maintenance organisation (HMO) insurance (21.32% vs 25.01%) and a long-term care (such

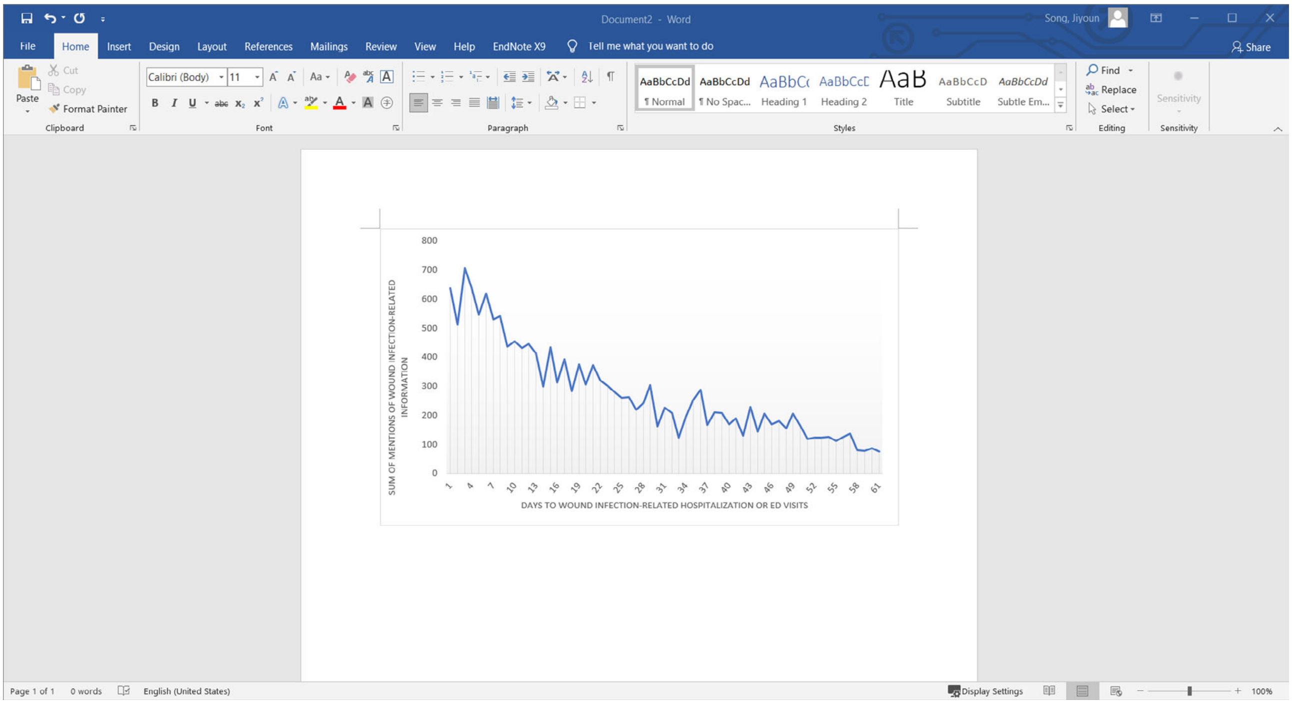

FIGURE 2 Frequency of wound infection information documentation. This figure describes the appearance of wound infection-related information over time during homecare episodes that resulted in wound infection-related hospitalisation or ED visits. The figure suggests that the frequency of wound infection-related information documentation increased close to wound infection-related hospitalisation or ED visits, peaking within a few days before the event. This finding shows the potential of NLP in identifying important wound infection-related information before wound infection-related hospitalisation or ED visits

as skilled nursing facility, long-term nursing home, or long-term care hospital) stay 14 days prior to the homecare admission (8.57% vs 9.86%). Cancer, cerebral degeneration, and dementia were less common in these groups compared with patients without wound infection.

6 | DISCUSSION

We found a significantly higher prevalence of wound infection information co-documentation among patients admitted to ED or hospitalised for wound infection during a homecare episode. About one in five patients admitted to ED or hospitalised for wound infection had documentation of wound infection in clinical notes, whereas only 1% of patients in the sample without wound at homecare admission and about 6% of patients in the sample with a wound at homecare admission. The frequency of wound infection information documentation in clinical notes increased close to wound infection-related hospitalisation or ED admission, peaking a few days before the ED admission event. This finding is consistent with the evidence reported in inpatient settings that there is an association between increased nursing documentation and negative health outcomes.[28,29] This finding may help in the early detection of infection signs and symptoms for timely interventions.

The findings also showed a clear trend of increased identification of wound infection information for patients with wounds and wound-related hospitalisation or ED visits. The differences between patients with and without wound infection identified with NLP were most notable in the diagnosis category. In this study, race and insurance were significantly associated with wound infection. White- and Hispanic-identifying patients were more likely to develop wound infection compared with Black- or Asian/Pacific Island-identifying patients. For insurance, patients with private HMO coverage were less likely to have a description of wound infection when compared with patients who did not have private HMO coverage. We could not find literature specifically exploring race or insurance relating to the development of wound infection in homecare; however, previous research in other settings, for example, in-patient, has reported no significant association for both factors. Two studies found that for surgical site wounds, race and insurance were not contributing factors to postoperative wound infection.[30,31]

Wound infection documentation was more common among patients with diabetes, PVD, and skin ulcer.

TABLE 4 Comparison of patients' characteristics identified as having a wound infection from natural language processing (NLP) among patients with wound at homecare admission (N = 54 316)

	Patients without documentation of wound infection (n = 51 084) n (%)	Patients with documentation of wound infection (n = 3232)
Demographics		
Age		
Age (mean (SD))	67.04 (16.57)	66.36 (16.94)*
Sex		
Female (%)	n (56.69)	57.36
Male (%)	43.31	42.64
Race		
Asian or PI (%)	6.95	4.05*
Black (%)	23.59	21.53*
Hispanic (%)	20.43	22.65*
White (%)	48.68	52.13*
Payer		
Medicare FFS (%)	n (41.51)	43.25
Medicare HMO (%)	17.37	18.16
Medicaid FFS (%)	3.62	3.81
Medicaid HMO (%)	15.35	16.37
Dual eligible (%)	5.9	5.82
Private insurance HMO (%)	25.01	21.32*
Other (%)	4.56	4.18
Language		
English (%)	83.61	84.31
Spanish (%)	11.99	12.78
Previous history and diagnosis		
Inpatient stay 14 days prior to home care admission		
Short-stay acute hospital (%)	69.57	69.83
Long-term care (skilled nursing facility, long-term nursing home, long-term care hospital)	9.86	8.57*
Others (rehab/psych/other) (%)	6.21	5.32*
Not applicable (%)	16.26	17.61*
Prior condition		
Urinary incontinence (%)	15.18	17.2*
Indwelling/suprapubic catheter (%)	1.54	1.18
Intractable pain (%)	14.65	16.99*
Decision (%)	6.72	6.87
Behaviour (%)	0.54	0.46
Memory (%)	4.33	3.34*
None (%)	60.65	58.29*
Diagnosis		
Acute myocardial infarction (%)	14.37	13.27

TABLE 4 (Continued)

Acquired immunodeficiency syndrome (AIDS) (%)	2.12	2.17
Cancer (%)	6.67	4.24*
Cardiac dysrhythmias (%)	8.86	8.51
Cerebral degeneration (%)	1.93	1.11*
Dementia (%)	5.74	4.36*
Depression (%)	9.41	9.19
Diabetes (%)	31.03	36.57*
Heart failure (%)	9.79	9.84
Hypertension (%)	56.54	56.75
Neurological disorder (%)	3.96	3.71
Pulmonary disease (%)	12.22	13.03
Peripheral vascular disease (%)	4.15	7.8*
Renal (%)	10.19	7.95*
Skin ulcer (%)	18.82	25.74*
Stroke (%)	4.71	3.19*
Overall status*		
Stable (%)	11.77	10.19
Likely to be stable (%)	74.62	76.46
Fragile (%)	12.08	12.33
Serious/unknown (%)	0.96	0.68

*$P < 0.05$, t-test or chi-square test or Fisher's exact test, as appropriate.

These three diagnoses have been reported as risk factors for the development of wound infection.[32] For example, diabetes has been found to be a risk factor for surgical wound infection after cardiac surgery.[33] On the other hand, the present study identified that documentation of wound infections was less frequent among patients with cancer, cerebral degeneration, or dementia. However, because these differences are quite small we are cautious of over-interpreting them. In previous studies, cancer patients were found to be at higher risk for wound infection mostly due to immunosuppression from chemotherapy.[34] Dementia and neurodegenerative disorders were reported to be associated with comorbidities of pressure ulcers among older adults in previous studies.[32,35] As our present study did not consider whether cancer patients are in active treatment or had surgery, further research is needed to gain greater insight into these potential relationships.

Future research is also required on identifying risk factors for wound infection in homecare with more targeted populations, such as patients with particular diseases such as diabetes or skin ulcer, to help create better tailored interventions. In doing so, applying advanced technology, such as NLP, may be used to guide the development tailored interventions. In addition, further efforts might be focused on creating a comprehensive wound infection documentation standard that can help to standardise wound infection information in the clinical documentation. Such standardisation can enhance the quality of the note content, leading to clearer communication between healthcare professionals in directly treating patients and facilitating clinical research using more advanced data science tools. The overall performance of the NLP algorithm could be improved in future applications by testing and refining it with the use of diverse data from multiple agencies.

This study demonstrates the actionable implications of data science in clinical practice. First, the NLP algorithm can be applied in current EHR systems to auto-detect high risk patients and support clinical decision-making. Second, with high risk patients thus identified, nursing care can be appropriately prioritised with close observation of these patients. Third, informed by future research findings using NLP algorithms, tailored interventions can be developed to advance effective management of homecare patients.

7 | LIMITATIONS

This study has several notable limitations. First, the wound infection information is based on nurses' reports in clinical notes without any other clinical validation, such as lab tests, and therefore may not be comprehensive. Moreover, due to partial availability of other patient variables such as some socio-demographic characteristics, selected variables previously found to be more closely associated with wound infection such as diagnosis and patient condition were included in our analysis. Second, we only used OASIS data to identify wound infection related hospitalisation or ED visits. Linking outcome data with hospital codes, such as ICD 10 codes if possible, would have yielded more accurate outcomes. In addition, the study was conducted using data from one homecare agency over a 1-year period, so findings might be not widely generalizable. Finally, this study did not exclude patients who might have returned back to homecare after hospitalisation and this subpopulation of patients might need to be examined further.

8 | CONCLUSION

To our knowledge, this study is the first to use NLP to extract wound infection information from nursing notes in homecare. Our findings suggest that nurses document wound infection-related information relatively frequently. Advanced technologies such as NLP can be used to extract valuable information from nurses' notes to improve our understanding of patient care needs in the community setting. Utilising existing clinical notes can have many benefits, such as designing decision support for case management to provide a tailored intervention, which might lead to a reduction in related hospitalizations. Considering the patient characteristics related to wound infection would also help nurses to be alert and more closely observe those patients for potential infections.

ACKNOWLEDGEMENT

Both institutional review boards of our institute and the homecare agency approved the study protocol. This study is funded by the Eugenie and Joseph Doyle Research Partnership Fund from Visiting Nurses Service of New York and the Intramural Pilot Grant from Columbia University School of Nursing. At the time of data analysis and manuscript development, Kyungmi Woo was supported by the Comparative and Cost-Effectiveness Research (T32 NR014205) grant through the National Institute of Nursing Research. Jiyoun Song is funded by Jonas Center for Nursing and Veterans Healthcare.

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Kyungmi Woo https://orcid.org/0000-0002-8710-2696
Jiyoun Song https://orcid.org/0000-0003-0362-0670
Victoria Adams https://orcid.org/0000-0002-5960-7291
Lorraine J. Block https://orcid.org/0000-0002-9496-3208
Leanne M. Currie https://orcid.org/0000-0002-8232-2809
Jingjing Shang https://orcid.org/0000-0003-1815-5556
Maxim Topaz https://orcid.org/0000-0002-2358-9837

REFERENCES

1. Medicare Payment Advisory Commission. March 2019 Report to the Congress: Medicare Payment Policy; 2019. http://www.medpac.gov/-documents-/reports
2. Avalere Health LLC. Home Health Chartbook; 2018. http://www.ahhqi.org/research/home-health-chartbook
3. Perencevich EN, Sands KE, Cosgrove SE, Guadagnoli E, Meara E, Platt R. Health and economic impact of surgical site infections diagnosed after hospital discharge. *Emerg Infect Dis*. 2003;9(2):196-203. https://doi.org/10.3201/eid0902.020232.
4. Shang J, Larson E, Liu J, Stone P. Infection in home health care: results from national outcome and assessment information set data. *Am J Infect Control*. 2015;43(5):454-459. https://doi.org/10.1016/j.ajic.2014.12.017.
5. Topaz M, Lai K, Dowding D, et al. Automated identification of wound information in clinical notes of patients with heart diseases: developing and validating a natural language processing application. *Int J Nurs Stud*. 2016;64:25-31. https://doi.org/10.1016/j.ijnurstu.2016.09.013.
6. Coleman S, Nelson EA, Vowden P, et al. Development of a generic wound care assessment minimum data set. *J Tissue Viability*. 2017;26(4):226-240. https://doi.org/10.1016/j.jtv.2017.09.007.
7. Chapman WW, Nadkarni PM, Hirschman L, D'Avolio LW, Savova GK, Uzuner O. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. *J Am Med Inform Assoc*. 2011;18(5):540-543. https://doi.org/10.1136/amiajnl-2011-000465.
8. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform*. 2009;42(5):760-772. https://doi.org/10.1016/j.jbi.2009.08.007.
9. Althoff T, Clark K, Leskovec J. Large-scale analysis of counseling conversations: an application of natural language processing to mental health. *Trans Assoc Comput Linguist*. 2016;4:463-476.
10. Le DV, Montgomery J, Kirkby KC, Scanlan J. Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting. *J Biomed Inform*. 2018;86:49-58. https://doi.org/10.1016/j.jbi.2018.08.007.
11. Zhou L, Baughman AW, Lei VJ, et al. Identifying patients with depression using free-text clinical documents. *Stud Health Technol Inform*. 2015;216:629-633. https://doi.org/10.3233/978-1-61499-564-7-629.

12. Kreimeyer K, Foster M, Pandey A, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *J Biomed Inform*. 2017;73:14-29. https://doi.org/10.1016/j.jbi.2017.07.012.
13. Yim WW, Yetisgen M, Harris WP, Kwan SW. Natural language processing in oncology: a review. *JAMA Oncol*. 2016;2(6):797-804. https://doi.org/10.1001/jamaoncol.2016.0213.
14. Acker WW, Plasek JM, Blumenthal KG, et al. Prevalence of food allergies and intolerances documented in electronic health records. *J Allergy Clin Immunol*. 2017;140(6):1587-1591.e1581. https://doi.org/10.1016/j.jaci.2017.04.006.
15. Assad WA, Topaz M, Tu J, Zhou L. The application of machine learning to evaluate the adequacy of information in radiology orders. Paper presented at: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), November 13-16, 2017.
16. Goss FR, Plasek JM, Lau JJ, Seger DL, Chang FY, Zhou L. An evaluation of a natural language processing tool for identifying and encoding allergy information in emergency department clinical notes. *AMIA Annu Symp Proc*. 2014;2014:580-588. https://pubmed.ncbi.nlm.nih.gov/25954363.
17. Plasek JM, Goss FR, Lai KH, et al. Food entries in a large allergy data repository. *J Am Med Inform Assoc*. 2016;23(e1):e79-e87. https://doi.org/10.1093/jamia/ocv128.
18. Zhou L, Lu Y, Vitale CJ, et al. Representation of information about family relatives as structured data in electronic health records. *Appl Clin Inform*. 2014;5(2):349-367. https://doi.org/10.4338/aci-2013-10-ra-0080.
19. Merriam-Webster. (n.d.). *Wound*. https://www.merriam-webster.com/dictionary/wound
20. Sinha P, Sunder G, Bendale P, Mantri M, Dande A, eds. Unified Medical Language System. In *Electronic Health Record*. New York, NY: The Institute of Electrical and Electronics Engineers, Inc.; 2012. https://doi.org/10.1002/9781118479612.ch16.
21. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. *Adv Neural Inf Process Syst*. 2013;2:3111-3119. http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf.
22. Elkan C, Noto K. Learning classifiers from only positive and unlabeled data. Paper presented at: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Las Vegas, Nevada, USA; 2008. https://doi.org/10.1145/1401890.1401920
23. Halpern Y, Horng S, Choi Y, Sontag D. Electronic medical record phenotyping using the anchor and learn framework. *J Am Med Inform Assoc*. 2016;23(4):731-740. https://doi.org/10.1093/jamia/ocw011.
24. Topaz M, Murga L, Bar-Bachar O, McDonald M, Bowles K. NimbleMiner: an open-source nursing-sensitive natural language processing system based on word embedding. *Comput Inform Nurs*. 2019;37(11):583-590. https://doi.org/10.1097/cin.0000000000000557.
25. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276-282.
26. Altman DG. *Practical Statistics for Medical Research*. London, England: Chapman and Hall; 1990.
27. Parab S, Bhalerao S. Choosing statistical test. *Int J Ayurveda Res*. 2010;1(3):187-191. https://doi.org/10.4103/0974-7788.72494.
28. Collins SA, Cato K, Albers D, et al. Relationship between nursing documentation and patients' mortality. *Am J Crit Care*. 2013;22(4):306-313. https://doi.org/10.4037/ajcc2013426.
29. Waudby-Smith IER, Tran N, Dubin JA, Lee J. Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients. *PLoS One*. 2018;13(6):e0198687. https://doi.org/10.1371/journal.pone.0198687.
30. Dotters-Katz SK, Feldman C, Puechl A, Grotegut CA, Heine RP. Risk factors for post-operative wound infection in the setting of chorioamnionitis and cesarean delivery. *J Matern Fetal Neonatal Med*. 2016;29(10):1541-1545. https://doi.org/10.3109/14767058.2015.1058773.
31. Takoudes TC, Weitzen S, Slocum J, Malee M. Risk of cesarean wound complications in diabetic gestations. *Am J Obstet Gynecol*. 2004;191(3):958-963. https://doi.org/10.1016/j.ajog.2004.05.063.
32. Jaul E, Barron J, Rosenzweig JP, Menczel J. An overview of comorbidities and the development of pressure ulcers among older adults. *BMC Geriatr*. 2018;18(1):305. https://doi.org/10.1186/s12877-018-0997-7.
33. Cimochowski GE, Harostock MD, Brown R, Bernardi M, Alonzo N, Coyle K. Intranasal mupirocin reduces sternal wound infection after open heart surgery in diabetics and nondiabetics. *Ann Thorac Surg*. 2001;71(5):1572-1578; discussion 1578-1579. https://doi.org/10.1016/s0003-4975(01)02519-x.
34. Payne WG, Naidu DK, Wheeler CK, et al. Wound healing in patients with cancer. *Eplasty*. 2008;8:e9. https://pubmed.ncbi.nlm.nih.gov/18264518.
35. Jaul E, Meiron O. Dementia and pressure ulcers: is there a close pathophysiological interrelation? *J Alzheimers Dis*. 2017;56(3):861-866. https://doi.org/10.3233/jad-161134.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Woo K, Song J, Adams V, et al. Exploring prevalence of wound infections and related patient characteristics in homecare using natural language processing. *Int Wound J*. 2022;19(1):211–221. https://doi.org/10.1111/iwj.13623