HIR
Healthcare Informatics Research

Applying of Decision Tree Analysis to Risk Factors Associated with Pressure Ulcers in Long-Term Care Facilities

Mikyung Moon, PhD, RN[1], Soo-Kyoung Lee, PhD, RN[2]

[1]College of Nursing, the Research Institute of Nursing Science, Kyungpook National University, Daegu, Korea; [2]College of Nursing, Keimyung University, Daegu, Korea

Objectives: The purpose of this study was to use decision tree analysis to explore the factors associated with pressure ulcers (PUs) among elderly people admitted to Korean long-term care facilities. **Methods:** The data were extracted from the 2014 National Inpatient Sample (NIS)—data of Health Insurance Review and Assessment Service (HIRA). A MapReduce-based program was implemented to join and filter 5 tables of the NIS. The outcome predicted by the decision tree model was the prevalence of PUs as defined by the Korean Standard Classification of Disease-7 (KCD-7; code L89*). Using R 3.3.1, a decision tree was generated with the finalized 15,856 cases and 830 variables. **Results:** The decision tree displayed 15 subgroups with 8 variables showing 0.804 accuracy, 0.820 sensitivity, and 0.787 specificity. The most significant primary predictor of PUs was length of stay less than 0.5 day. Other predictors were the presence of an infectious wound dressing, followed by having diagnoses numbering less than 3.5 and the presence of a simple dressing. Among diagnoses, "injuries to the hip and thigh" was the top predictor ranking 5th overall. Total hospital cost exceeding 2,200,000 Korean won (US $2,000) rounded out the top 7. **Conclusions:** These results support previous studies that showed length of stay, comorbidity, and total hospital cost were associated with PUs. Moreover, wound dressings were commonly used to treat PUs. They also show that machine learning, such as a decision tree, could effectively predict PUs using big data.

Keywords: Data Mining, Decision Trees, Long-Term Care, Pressure Ulcer, Risk Factors

I. Introduction

Pressure ulcers (PUs) are defined as "localized injuries of the skin and/or underlying tissue over a bony prominence due to pressure and shear" [1]. Hospital-acquired PUs are the most concerning problems among healthcare professionals because they threaten the patients' quality of life and increase hospital costs [2-4]. In particular, PUs among elderly residents in long-term care facilities are prevalent in comparison with patients admitted to other types of health organizations, and their presence serves as an important indicator of care quality [5,6].

In Korea, long-term care facilities have rapidly increased since the implementation in 2008 of long-term care insur-

ance, a social insurance for Korean older people, which provides physical and social support for older people and helps relieve caregivers' financial burdens [7,8]. The elderly in long-term care facilities do not receive intensive hospital care but need constant care services for geriatric care problems, such as PUs, falls, and cognitive impairment [9,10]. To provide optimal care for them in these facilities, Korean healthcare professionals have been more educated to prevent and treat these problems.

To reduce the occurrence of PUs, prevention is emphasized more than treatment is [1]. The first goal is to identify patients with specific risk factors and then to implement appropriate interventions to prevent PU occurrence [11]. Numerous studies related to risk factors for PUs have been conducted. Some individual physiological factors and organizational characteristics, such as nurse staffing levels, may place adults at risk for PU development [4]. A recent review study showed that a complex interplay of factors could increase the probability of PU development [11]. However, the quality of studies is often limited by the small sample size, analysis methods, and reporting standards [11]. These previous studies have suggested that further research is needed with adequate numbers of PUs to maximize the validity and generalizability of their findings [11].

With the widespread use of electronic health records, big data is digitally collected and stored in healthcare sections [12]. This could be a solution for this issue of generalizability [12]. In particular, the National Patient Samples (NPS)—reimbursement claim data provided by Health Insurance Review and Assessment Service (HIRA)—could be a good source for research data because the sample is representative of the total Korean patient population. Moreover, the data comprise comprehensive information on health services, including diagnoses, treatments, procedures, and administrative information, such as staffing levels [13].

However, because most big data sets, such as the HIRA NPS, are not purposely collected for research purposes and are complex. This makes it difficult to manage common and traditional tools and methods [14]. More advanced data analytics are needed to manage big data that consider a variety of variables and subjects [15].

Data mining is the process of selecting, exploring, and modeling large amounts of data to uncover events and characteristics for accurate predictions of future data [16,17]. This model is non-parametric in nature and does not need the assumptions that are made in traditional statistical techniques [18]. In particular, decision tree analysis is a data mining method, and it shows a global picture of a given pa-

tient's risk. It may aid in better clinical decision-making than individual predictors obtained from a regression model [19].

Several studies have used the data mining approach to identify factors associated with disease prevention and management. Some studies have identified factors associated with postoperative recovery [20], smoking cessation failure [21], or a low-risk population for type 2 diabetes [22] using a decision tree. Raju et al. [18] and Lee et al. [23] explored factors related to PUs using a hospital database. These studies showed that the data mining approach used in a manner similar to decision tree analysis is a simple tool for predicting risk factors [22]. However, there is still a lack of studies that explore factors associated with PUs with a number of variables, including patient characteristics and administrative information.

Therefore, the purpose of this study was to explore the factors associated with PUs among elderly patients admitted to Korean long-term care facilities according to the 2014 Health Insurance Review and Assessment Service National Inpatient Sample (HIRA NIS) using decision tree analysis.

II. Methods

1. Data Source and Study Population
Data were extracted from the 2014 NIS provided by HIRA (HIRA-NIS-2014-0071). HIRA claims data are national data collected from healthcare providers all over the country for reimbursements for healthcare services [13]. The NIS is extracted from the HIRA claim data using a stratified randomized sampling method. Of 5,888,921 patients (estimated population), data from 765,564 patients (13%) were collected [13]. From 2014 HIRA NIS data, the inclusion criteria for this study were patients (1) admitted to Korean long-term care facilities and (2) aged 65 years or more.

2. Data Preparation
We first cleaned and preprocessed the data. An expert data manager who had skills to deal with big data analytics tools, such as Apache Hadoop and machine learning, was involved in this process. Preprocessing included initial filtering based on inclusion criteria and research interests, variable transformation and binarization, and table joining.

The HIRA NIS data consists of 5 tables: general specifications, health services, diagnosis information, outpatient prescription information, and health service provider information [13]. Among the 5 tables of data, the outpatient prescription table was excluded, and only cases admitted in long-term care facilities were included based on the in-

clusion criteria. A MapReduce-based program was implemented to join and filter those 5 tables (Apache Hadoop 2.7.1). Specifically, the billing statement identification code (key ID) was used as the key to join those 5 tables. Thus, we obtained a huge table comprising all information pertaining to the prediction task (Figure 1). MapReduce is a tool for big data analytics in healthcare, which makes it possible to process large datasets with various structures or no structure [15]. The powerful interface of this programing model enables automatic parallelization and distribution of large-scale computation to achieve high performance on large clusters of machines [24].

3. Variable Selection

Non-useful variables were initially excluded after reviewing; only medication/prescription, injection/procedure, and treatment/operation among healthcare services were included. A variable with too raw values without classification was categorized, and each category was dichotomized for a new variable. For example, drug codes in a medication variable were categorized into 34 mid-classes according to Korean drug classification [25], and each class was transformed into a binary variable. Diagnosis codes in a medical diagnosis variable were classified into 264 classes according to the subcategories of Korean Standard Classification of Diseases-7 codes (KCD-7) [26]. Each was transformed into a binary variable.

Initial machine learning was performed considering all selected variables. Based on the analysis results, a few meaningless variables were removed to improve interpretation and performance. Table 1 shows the finalized 830 variables. The final variables largely fell into three categories: demographic characteristics, disease and treatment characteristics, and healthcare provider characteristics. Demographic characteristics included age, gender, insurance type, hospital arrival pathway, length of stay, patient status at discharge, and total hospital cost. Variables related to disease and treatments were surgical status, the number of diagnoses, 264 diagnoses, 34 medications, and 517 procedures. Finally, variables related to healthcare providers were the type of providers, presence of special equipment (CT and MRI), bed grade (number of beds), and number of healthcare providers (nurse and doctor) per 50 beds.

The outcome predicted by the decision tree model was the prevalence of PUs as defined by the KCD-7 (code L89*) [26]. A binary dependent variable representing PUs was created. There were 7,928 (3.5%) PUs in 227,130 cases. The performance of machine learning was threated because the sample was skewed due to the small number of PU cases. To improve the performance, the same number of PU patients was randomly selected in 722,658 non-PU patients by the shuf program in Linux. Finally, 15,856 cases were prepared for machine learning (Figure 2).

4. Data Analysis

A decision tree was generated using R 3.3.1 to visualize associated factors and explore the patients most at risk of pressure. It could help to identify sub-populations with/without pressure through easily interpreted grouping rules [18].

The 10-fold cross-validation method was used to minimize the bias associated with the random sampling of the training. In the 10-fold cross-validation, the data set was divided into 10 parts, and then 9 parts were used for training and 1 set was used for testing. The process was then repeated until all parts were tested. The goal of this process was to determine which data mining algorithm performs best so we could use it to generate our target predictive model [16].

Three performance measures were used to evaluate the models: accuracy, sensitivity, and specificity. Accuracy is the ability to differentiate between patient and healthy cases correctly. Sensitivity is the ability to identify patient cases correctly. Specificity is the ability to identify healthy cases correctly [27]. Finally, we chose the best model with the highest accuracy based on the 10-fold cross-validation results.

Statistical analysis examined associations between PUs and predictors identified by decision tree analysis: chi-square
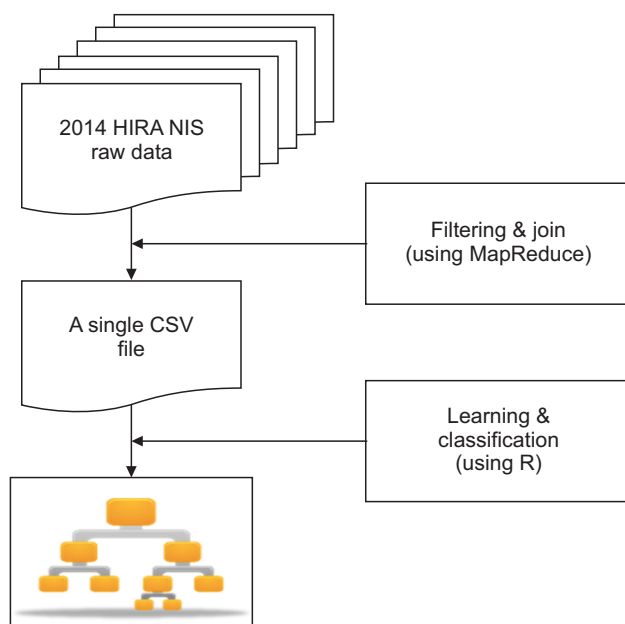

Figure 1. Flow chart of data management.

Table 1. Characteristics of variables

Characteristic	Number of unique values	Description[a]	Value
Demographic			
Gender	2	Women	11,237 (70.9)
Insurance type	2	National Health Insurance subscriber	12,944 (81.6)
Hospital arrival pathway	7	Via OPD	10,811 (68.2)
Patient status at discharge	5	Keep admission	13,855 (87.4)
Age (yr)			81 (65–100)
Length of stay (day)			16.5 (0–31)
Total hospital cost (Korean won)			1,381,190.6 (0–9,175,850)
Disease and treatment			
Surgical status	2	Yes	301 (1.9)
264 medical diagnosis classification[b] [KCD-7 code]			
Organic, including symptomatic, mental disorders [F00-F09]	2	Yes	10,500 (66.2)
Other disorders of the skin and subcutaneous tissue [L80-L99]	2	Yes	7,940 (50.1)
Hypertensive disease [I10-I15]	2	Yes	7,682 (48.4)
34 drug classification[b] [code]			
Central nerve system drug [110]	2	Yes	10,471 (66.0)
Blood and body fluid [330]	2	Yes	2,627 (16.6)
Antibiotics [610]	2	Yes	2,426 (15.3)
517 procedures[b] [billing code]			
Acupuncture (over 2 sites) [40012]	2	Yes	3,874 (24.4)
Acupuncture (Toja method) [40080]	2	Yes	3,333 (21.0)
Fee for medication preparation and guide to medication administration [J2000]	2	Yes	2,514 (15.9)
Number of diagnoses			6.1 (1–30)
Facilities and providers			
Type of provider	8	Private	7,555 (47.6)
Bed grade	14	15 < n ≤ 20	4,347 (27.4)
Computed tomography	2	Yes	199 (1.3)
Magnetic resonance imaging	2	Yes	17 (0.1)
Number of physicians per 50 beds			0.5 (0–3)
Number of nurses per 50 beds			3.5 (0–22.6)

Values are presented as number (%) or mean (range).

KCD-7: Korean Standard Classification of Diseases-7, OPD: outpatient department.

[a]The only top or meaningful value was described. [b]Top 3 variables among the variables were selected and described.

analysis for categorical variables and independent t-test for continuous variables.

5. Ethical Consideration

This study was a secondary analysis using HIRA NIS. The NIS data obtained from HIRA were de-identified and did not contain any patient-specific information. Furthermore,

```
┌─────────────────────────────────────────────────────────────┐
│                    2014 HIRA NIS                              │
│                  n = 36,321,078                               │
└─────────────────────────────────────────────────────────────┘
                            ⇓
┌─────────────────────────────────────────────────────────────┐
│      Only patients admitted in long-term care facilities      │
│                    n = 923,842                                │
└─────────────────────────────────────────────────────────────┘
                            ⇓
┌─────────────────────────────────────────────────────────────┐
│             Patients 65 years old and older                   │
│   n = 365,293; the prevalence of pressure ulcers = 4.9%       │
│                    (18,029 cases)                             │
└─────────────────────────────────────────────────────────────┘
                            ⇓
┌─────────────────────────────────────────────────────────────┐
│   Data filtering (only included cases with procedures and     │
│                    medications)                               │
│   n = 227,130; the prevalence of pressure ulcers = 3.5%       │
│                    (7,928 cases)                              │
└─────────────────────────────────────────────────────────────┘
                            ⇓
┌─────────────────────────────────────────────────────────────┐
│           Apply machine learning algorithm                    │
│   (selected the same number of non-pressure ulcer group)      │
│           Random sampling by shuf program                     │
│ n = 15,856 = 7,928 (pressure ulcer group) + 7,928             │
│              (non-pressure ulcer group)                       │
└─────────────────────────────────────────────────────────────┘
```
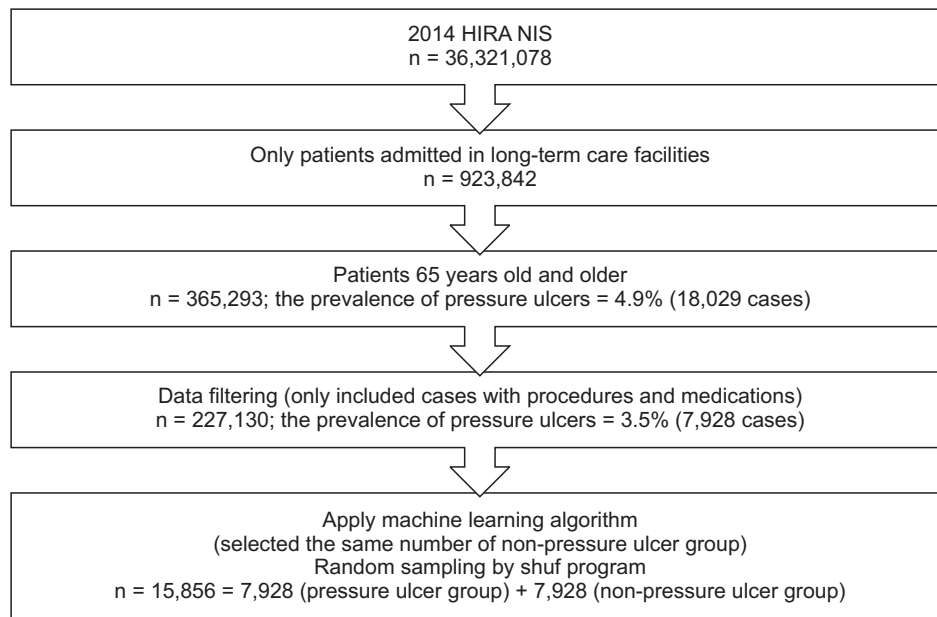
Figure 2. Sample selection process from 2014 Health Insurance Review and Assessment Service National Inpatient Sample (HIRA NIS).

this study was reviewed and exempted by the Kyungpook National University ethical committee (IRB No. 2016-0104).

III. Results

1. Characteristics of Data
Table 1 shows the data characteristics. The majority of patients were female (70.9% of the total population) and National Health Insurance subscribers (81.6%). The sample ranged in age from 65 years to 100 years (mean, 81 years). About 70% of patients were admitted via outpatient departments (OPDs). The patients had stayed for an average of 16.5 days with an average of 6.1 medical diagnoses. The average total hospital cost was US $1,200 (1,381,190.6 Korean won). Of the 264 medical diagnosis classification variables, organic including symptomatic mental disorders (KCD-7 codes, F00-F09) showed the highest frequency among patients (66.2%). In 34 drug classification variables, central nerve system drugs were the most commonly prescribed to the patients (66.0%). The most common bed grade (the number of beds) in the long-term care facilities was between 15 and 20 (15 < n ≤ 20). The facilities had an average of 0.5 physicians and 3.5 nurses per 50 beds.

2. Predictive Performance
Figure 3 shows a box plot of the 10-fold cross-validation results. Accuracy ranged from 0.771 to 0.804, sensitivity ranged from 0.793 to 0.876, and specificity ranged from 0.707 to 0.787. Of the models considered, the model with the highest accuracy was selected as the best model for this
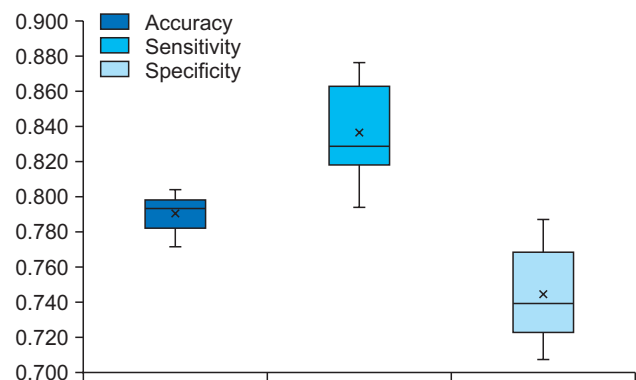

Figure 3. Boxplot of the 10-fold cross-validation results.

study. The model showed 0.804 accuracy, 0.820 sensitivity, and 0.787 specificity. The results indicated that decision tree analysis was the best predictor with 80.4% accuracy.

3. Decision Tree
Figure 4 presents the decision tree derived from the best predictive model in the R output. Decision tree analysis identified 15 subgroups (nodes) and 8 associated factors. A length of stay shorter than 0.5 (a half) days was identified as the top associated factor with the presence of PUs. Next was the presence of an infectious wound dressing followed by a number of diagnoses less than 3.5, and then the presence of a simple dressing. Among diagnoses, "injuries to the hip and thigh" was the top predictor, ranking 5th overall, followed by a number of diagnoses less than 5.5 and episodic and paroxysmal disorders, total hospital cost, and bed grade.
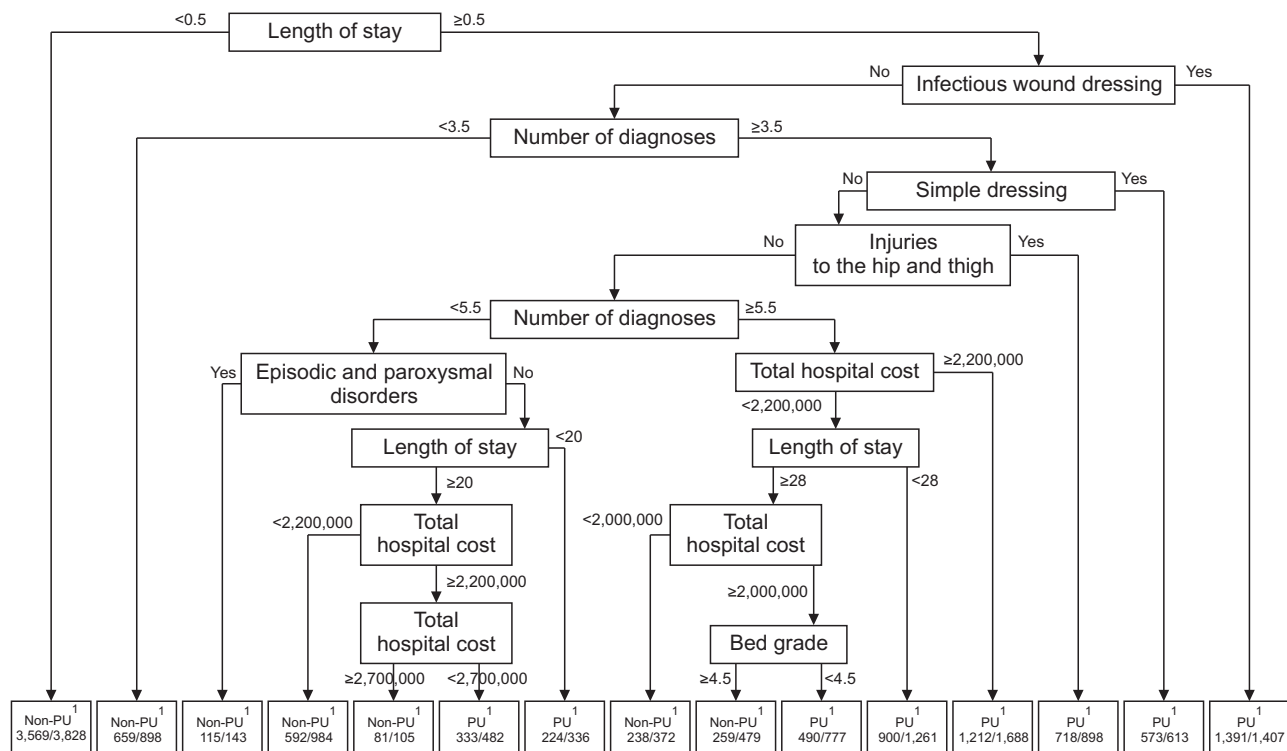The group most likely to have PUs was the 1,391 patients

Figure 4. Decision tree.

who stayed in the hospital for more than or equal to 0.5 days and had infectious wound dressing. The second pressure group included 1,212 patients. Patients who stayed in the hospital for a period longer than or equal to 0.5 days, who did not have infectious wound dressing, who had 3.5 or more than diagnoses, who did not have simple dressing, who did not have injuries the hip and thigh, who had 5.5 or more diagnoses were more likely to have PUs when they had total hospital cost exceeding US $2,000 (2,200,000 Korean won).

4. Variables Associated with PUs from the Statistical Analyses

Table 2 shows the association between PUs and 8 variables identified by a decision tree. There were statistically significant differences in infection wound dressing ($\chi^2 = 1658.0$, $p < 0.001$), simple dressing ($\chi^2 = 985.1$, $p < 0.001$), injuries to the hip and thigh ($\chi^2 = 426.6$, $p < 0.001$), and episodic and paroxysmal disorders ($\chi^2 = 115.0$, $p < 0.001$). Patients with these diseases and procedures had a higher proportion of PUs than patients without them. There were significant differences in length of stay ($t = -33.2$, $p < 0.001$), the number of diagnoses ($t = -69.7$, $p < 0.001$), the hospital cost ($t = -41.3$, $p < 0.001$), and the bed grade ($t = 8.6$, $p < 0.001$) between the PU and non-PU group. The PU group had longer length of stay, more medical diagnoses and hospital cost, and lower bed numbers.

IV. Discussion

This study explored the factors associated with the development of PUs using a data mining approach. The data were extracted from the HIRS NIS. A decision tree was generated with 15,856 cases and 830 variables. The decision tree displayed 15 subgroups with 8 variables showing good prediction performance. First of all, this study highlighted the usefulness of the data mining approach in managing and analyzing healthcare big data, such as the HIRA NIS data. Data mining accurately identified meaningful associations between an outcome and many variables.

The length of stay was the top variable associated with PUs. Moreover, the group with PU had a significantly longer length of stay. These results were similar to those of previous PU studies. PU could be a significant factor that prolongs the length of stay beyond expectations based on diagnosis at admission [28].

Infectious wound dressings and simple dressings were the second- and fourth-most commonly associated variables with PU. The results are quite reasonable because wound dressing is a main component of PU care. Dressings are used to keep a wound bed moist or to keep the periwound dry and prevent maceration to facilitate healing [1]. This study provides limited information regarding types of dressings because the procedure codes for the reimbursement claims

Table 2. Association between 8 variables and pressure ulcers (PU)

Variable	PU	Non-PU	Total	Mean (range)	χ^2	t	p-value
Infectious wound dressing (M0121)[a]					1,658.0		<0.001
Yes	1,550 (98.8)	19 (1.2)	1,569 (9.9)				
No	6,378 (44.6)	7,909 (55.4)	14,287 (90.1)				
Simple dressing (M0111)[a]					985.1		<0.001
Yes	1,067 (95.2)	54 (4.8)	1,121 (7.1)				
No	6,861 (46.6)	7,874 (53.4)	14,735 (92.9)				
Injuries to the hip and thigh [S70-S79][b]					426.6		<0.001
Yes	1,008 (77.4)	294 (22.6)	1,302 (8.2)				
No	6,920 (47.5)	7,634 (52.5)	14,554 (91.8)				
Episodic and paroxysmal disorders [G40-G47][b]					115.0		<0.001
Yes	1,166 (61.6)	728 (38.4)	1,894 (11.9)				
No	6,762 (48.4)	7,200 (51.6)	13,962 (88.1)				
Length of stay (day)	20.2	13.0		16.5 (0–31)		−33.2	<0.001
Number of diagnoses	8.1	4.2		6.1 (1–30)		−69.7	<0.001
Total hospital cost (Korean won)	1,724,438.8	1,037,942.3		1,381,190.6 (0–9,175,850)		−41.3	<0.001
Bed grade	4.6	4.9		4.75 (0–14)		8.6	<0.001
Total	7,928 (50.0)	7,928 (50.0)	15,826 (100)				

Values are presented as number (%) or mean (range).
[a]Billing codes. [b]Korean Standard Classification of Disease 7 (KCD-7) codes.

did not reflect dressing types. Various types of dressings could be applied over time as an ulcer heals or deteriorates.

The number of medical diagnoses was an important variable to predict PUs. The numbers of diagnoses less than 3.5 and 5.5 were ranked 3rd and 6th splits. The number of medical diagnoses could be considered as a comorbidity. The results confirmed that comorbidity is a risk factor for PUs [29]. Of the individual medical diagnoses, "injuries to the hip and thigh" was the fifth most commonly related to PUs in. This diagnosis classification includes diseases such as 'fracture of femur' and 'injury of nerves at hip and thigh level' affecting the mobility of patients. This could be because immobility is common in patients with these diseases, and it increases the risk of developing PUs [9]. This study highlights the need for careful assessment of the elderly with higher comorbidity or these diseases to prevent PUs in long-term care facilities.

We found that total hospital cost was a factor associated with PUs, and this is supported in the literature. Multiple factors, including the prolonged stay, labor of healthcare providers, and treatment material costs can increase hospital costs [2,30]. Today, the increases in hospital costs caused by PUs could receive more attention due to the limited budgets for healthcare. Therefore, PU prevention is vital to reduce the costs related to PUs.

Overall, the results of this study with big data confirmed other previous study results related to PUs; our results appear to be more valid and generalizable than those of previous studies. PUs are associated with length of stay, number of diagnoses, and total hospital costs. Moreover, elderly patients with PUs are more likely to have simple, infected dressings. Eventually, longer length of stay or additional procedures, such as changing the dressing could lead to increased hospital costs for PU patients. Therefore, the importance of PU prevention to alleviate the financial burden of long-term care facilities is highlighted by this study.

Contrary to our expectation, the results did not show any drug associated with PUs even though previous research has identified specific drugs that seem to increase the incidence of PUs [4,11]. Further studies with pre-processing in detail could be conducted to identify other association factors.

Big data continue to impact healthcare; such information can be used to improve patient care and clinical decision-making. However, big data are very large and complex, and they are hard to manage with traditional manipulation methods. This study suggested that the use of data mining in healthcare big data can minimize the time spent manually search while minimizing insignificant studies; it can identify association [16,17]. Furthermore, the decision tree analysis used in this study was used to create a model of association factors and can deal with non-linear relationships to automatically capture multilevel interactions among variables [22]. Therefore, data mining can help manage a variety of healthcare big data.

In conclusion, this study used a decision tree to find factors associated with PUs using the 2014 NIS, which is data from the HIRA. A decision tree was generated with 15,856 cases and 830 variables. The decision tree displayed 15 subgroups with 8 variables showing 0.804 accuracy, 0.820 sensitivity, and 0.787 specificity. These results support those of previous studies that showed length of stay, comorbidity, and total hospital cost were associated with developing PUs. Moreover, wound dressings were commonly used to treat PUs. Finally, this study showed that data mining methods, such as decision tree analysis, could identify outcome variables in a big data set with many variables.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

References

1. National Pressure Ulcer Advisory Panel, European Pressure Ulcer Advisory Panel, Pan Pacific Pressure Injury Alliance. Prevention and treatment of pressure ulcers: quick reference guide. 2nd ed. Perth, Australia: Cambridge Media; 2014.

2. Demarre L, Verhaeghe S, Annemans L, Van Hecke A, Grypdonck M, Beeckman D. The cost of pressure ulcer prevention and treatment in hospitals and nursing homes in Flanders: a cost-of-illness study. Int J Nurs Stud 2015;52(7):1166-79.

3. Gunningberg L, Stotts NA. Tracking quality over time: what do pressure ulcer data show? Int J Qual Health Care 2008;20(4):246-53.

4. Lyder CH, Ayello EA. Pressure ulcers: a patient safety issue. In: Hughes RG, editor. Patient safety and quality: an evidence-based handbook for nurses. Rockville (MD): Agency for Healthcare Research and Quality; 2008.

5. Mueller C, Karon SL. ANA nurse sensitive quality indicators for long-term care facilities. J Nurs Care Qual 2004;19(1):39-47.

6. Keelaghan E, Margolis D, Zhan M, Baumgarten M. Prevalence of pressure ulcers on hospital admission among nursing home residents transferred to the hospital. Wound Repair Regen 2008;16(3):331-6.

7. National Health Insurance Service. Long-term care insurance [Internet]. Wonju, Korea: National Health Insurance Service; 2010 [cited at 2017 Jan 23]. Available from: http://longtermcare.or.kr/npbs/e/e/100/index.web.

8. Korean Health Industry Development Institute. KHISS current state of long term care facilities [Internet]. Cheongju, Korea: Korean Health Industry Development Institute; 2014 [cited at 2017 Jan 23]. Available from: http://www.khiss.go.kr/treeManager.do?task=treeFrm&vw_cd=MT_ITITLE&list_id=358_E&htxt_code=125369725751510746110771881221239.

9. Lahmann NA, Tannen A, Kuntz S, Raeder K, Schmitz G, Dassen T, et al. Mobility is the key! Trends and associations of common care problems in German long-term care facilities from 2008 to 2012. Int J Nurs Stud 2015;52(1):167-74.

10. Shin JH, Hyun TK. Nurse staffing and quality of care of nursing home residents in Korea. J Nurs Scholarsh 2015;47(6):555-64.

11. Coleman S, Gorecki C, Nelson EA, Closs SJ, Defloor T, Halfens R, et al. Patient risk factors for pressure ulcer development: systematic review. Int J Nurs Stud 2013;50(7):974-1003.

12. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA 2013;309(13):1351-2.

13. Kim L, Kim JA, Kim S. A guide for the utilization of Health Insurance Review and Assessment Service National Patient Samples. Epidemiol Health 2014;36:e2014008.

14. Frost & Sullivan. Drowning in big data? Reducing information technology complexities and costs for healthcare organizations [Internet]. Mountain View (CA): Frost & Sullivan; 2011 [cited at 2017 Jan 23]. Available from: https://www.emc.com/collateral/analyst-reports/frost-sullivan-reducing-information-technology-complexities-ar.pdf.

15. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3.

16. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform 2008;77(2):81-97.

17. Pedrassani de Lira C, Toniazzo de Abreu LL, Veiga Silva AC, Mazzuchello LL, Rosa MI, Comunello E, et al. Use of data mining to predict the risk factors associated with osteoporosis and osteopenia in women. Comput Inform Nurs 2016;34(8):369-75.

18. Raju D, Su X, Patrician PA, Loan LA, McCarthy MS. Exploring factors associated with pressure ulcers: a data mining approach. Int J Nurs Stud 2015;52(1):102-11.

19. Kang Y, McHugh MD, Chittams J, Bowles KH. Utilizing home healthcare electronic health records for tele-homecare patients with heart failure: a decision tree approach to detect associations with rehospitalizations. Comput Inform Nurs 2016;34(4):175-82.

20. Michelet D, Andreu-Gallien J, Skhiri A, Bonnard A, Nivoche Y, Dahmani S. Factors affecting recovery of postoperative bowel function after pediatric laparoscopic surgery. J Anaesthesiol Clin Pharmacol 2016;32(3):369-75.

21. Kim SR, Kim HK, Kim JY, Kim HY, Ko SH, Park M. Smoking cessation failure among Korean adolescents. J Sch Nurs 2016;32(3):155-63.

22. Ramezankhani A, Pournik O, Shahrabi J, Khalili D, Azizi F, Hadaegh F. Applying decision tree for identification of a low risk population for type 2 diabetes. Tehran Lipid and Glucose Study. Diabetes Res Clin Pract 2014;105(3):391-8.

23. Lee TT, Lin KC, Mills ME, Kuo YH. Factors related to the prevention and management of pressure ulcers. Comput Inform Nurs 2012;30(9):489-95.

24. Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. Commun ACM 2008;51(1):107-13.

25. Korea Ministry of Food and Drug Society. Regulation on drug classification numbers [Internet]. Cheongju, Korea; Ministry of Food and Drug Society; 2013 [cited at 2017 Jan 4]. Available from: http://www.mfds.go.kr/index.do?mid=1015&pageNo=6&seq=6325&sitecode=2013-04-05&cmd=v.

26. Korean Standard Statistical Classification. The standard of classification of disease and cause of death [Internet]. Daejeon, Korea; Korean Standard Statistical Classification; 2016 [cited at 2017 Jan 2]. Available from: https://kssc.kostat.go.kr:8443/ksscNew_web/link.do?gubun=004.

27. Baratloo A, Hosseini M, Negida A, El Ashal G. Part 1: Simple definition and calculation of accuracy, sensitivity

and specificity. Emerg (Tehran) 2015;3(2):48-9.

28. Graves N, Birrell F, Whitby M. Effect of pressure ulcers on length of hospital stay. Infect Control Hosp Epidemiol 2005;26(3):293-7.

29. Santamaria N, Carville K, Prentice J, Ellis I, Ellis T, Lewin G, et al. Pressure ulcer prevalence and its relationship to comorbidity in nursing home residents: results from phase 1 of the PRIME Trial. Prim Intent Aust J Wound Manag 2005;13(3):107-16.

30. Dealey C, Posnett J, Walker A. The cost of pressure ulcers in the United Kingdom. J Wound Care 2012;21(6): 261-6.