Type of Paper: Original research

Title:

Novel Approach to Inpatient Fall Risk Prediction and its Cross-site Validation Using Time-Variant Data

Insook Cho, PhD, RN\textsuperscript{1,4}, Eun-Hee Boo, PhD, RN\textsuperscript{2}, Eun-Ja Chung, MS, RN\textsuperscript{3}, David W. Bates, PhD, MD\textsuperscript{4,5}, Patricia C. Dykes, PhD, RN\textsuperscript{4,5}

Author affiliations:
\textsuperscript{1} Nursing Department, Inha University, Incheon, Republic of Korea
\textsuperscript{2} Department of Nursing, National Health Insurance Service Ilsan Hospital, Gyeonggi-do, Republic of Korea
\textsuperscript{3} Department of Nursing, Bundang Seoul National University Hospital, Gyeonggi-do, Republic of Korea
\textsuperscript{4} The Center for Patient Safety Research and Practice, Division of General Internal Medicine, Brigham and Women’s Hospital, Boston, MA, USA
\textsuperscript{5} Medicine, Harvard Medical School, Boston, MA, USA

Correspondence to:
Insook Cho, RN, PhD.
Nursing Department, Inha University, Republic of Korea
Inharo 100, Namgu, Incheon 402-751,
E-mail: insook.cho@inha.ac.kr
Tel.: +82-32-8608201, Fax: +82-32-8748201

Abstract word count: 278, Text word count: 4,190

Number of references: 33, Number of tables: 2, Number of figures: 4

Keywords: inpatient falls, predictive model, electronic medical records, across sites validation, nursing data set
ABSTRACT

**Objectives:** Electronic medical records (EMRs) contain a considerable amount of information about patients. We investigated whether the readily available longitudinal EMRs data including nursing records could be utilized to compute the risk of inpatient falls and its accuracy when compared to existing fall risk assessment tools.

**Methods:** Two study cohorts from two tertiary hospitals with different EMR systems and located near Seoul, South Korea were used. The modeling cohort included 14,307 admissions (122,179 hospital-days) and the validation cohort comprised 21,172 admissions (175,592 hospital-days) from each of 6 nursing units. A probabilistic Bayesian network model was used, and patient data were divided into windows with a length of 24 hours. Data on existing fall risk assessment tools, nursing processes, Korean Patient Classification System groups, medications, and administration data were used as model parameters. Model evaluation metrics were averaged using 10-fold cross validation.

**Results:** The initial model showed an error rate of 11.7% and a spherical payoff of 0.91 with a c-statistic of 0.96, which represent far superior performance compared to that for the existing fall risk assessment tool (c-statistic = 0.69). The cross-site validation revealed an error rate of 9.3% and a spherical payoff of 0.92 with a c-statistic of 0.87, compared to a c-statistic of 0.65 for the existing fall risk assessment tool. The calibration curves for the model displayed more reliable results than the risk assessment tools. Nursing intervention data showed potential contributions to reducing the variance in the fall rate as did the risk factors of individual patients.
**Conclusion:** A risk prediction model that considers longitudinal EMR data including nursing interventions can improve the ability to identify individual patients at a high risk of falling.
INTRODUCTION

There is a considerable body of literature on fall prevention and reduction, yet despite many attempts by hospitals to reduce fall rates, significant and sustained reductions have proved elusive.[1] Risk assessment tools have been developed for several decades, and many risk factor identification studies have been published. St. Thomas’ Risk Assessment Tool in Falling Elderly Inpatients (STRATIFY), Hendrich II, and Morse are examples consisting of five or six subscales.[2-4] However, most predate broad use of electronic medical records (EMRs), and the tools were largely developed using limited data collected by researchers. The Cochrane reviews of Cameron et al.[5] and Gillespie et al. [6] imply that there is a significant lack of evidence on the efficacy of tools used to assess the risk of falling.

EMRs contain a considerable amount of information about patient histories and patient information conveyed both for discrete events and in narratives such as nursing notes. The increasing adoption of EMRs makes such clinical documentation a potentially rich and underutilized source of information for supporting nursing decisions.[7] Two types of data in the EMRs in particular present an opportunity for automated risk prediction: (i) structured longitudinal data and (ii) semi-structured or narrative data of nursing statements conveyed in clinical notes. Nursing assessment data are usually recorded in a structured form or using a predefined template. Nursing notes contain rich nursing-process information about identified nursing problems, provided interventions, and how patients responded to them. Several studies have investigated inpatient prediction models using EMR data. One study[8] used physician orders, nursing assessments and care plans,
progress notes, and the intensity of nursing care needs to predict inpatient falls. Another study[9] that was conducted at 13 nursing homes used a Minimum Data Set and structured data from EMRs, such as medications and nursing problems at one week after admission and one week after a room change to predict resident falls. In addition, Tescher et al.[10] and Giles et al.[11] used EMR data to identify risk factors for the development of pressure ulcers and inpatient falls, respectively. These studies are limited by their use of summary metrics rather than time-varying variables, and they did not consider the nursing interventions provided in attempt+ to prevent falls.

The rapid adoption of EMRs and the integration of nursing data into clinical repositories have made large quantities of clinical data available for both clinical practice and research.[7] The aims of the present study were to incorporate longitudinal EMR nursing-process data as a novel feature in calculating the risk of falls, and to validate the findings at an external site. We believe that intended nursing activities contribute to decreasing the risk, and thus controlling for this will facilitate the ability to predict the risk at a specific point in time. In addition, external validation is important for generalizability and discrimination when a model is applied at other sites or when using other EMR systems. [12] We note several points that Goldstein et al.[12] addressed in their systematic review of EMR-based prediction models: (i) it is easier to predict the short-term risk of events, since the data are observed more frequently, (ii) patient populations included in EMRs may be more reflective of the real world than the data collected for research purposes, and (iii) prediction models based on EMR data can often be implemented more easily
than traditional algorithms that need to be translated before being applied in a clinical setting.

In this study we investigated the following research questions: (i) How can longitudinal data from nursing records be incorporated into fall risk modeling? (ii) How can electronic EMR data be incorporated into a fall risk modeling paradigm, focusing on two types of data elements of the EMR (structured data and semi-structured data of nursing notes)? (iii) Does the fall risk model developed at a particular site or using a particular EMR system environment work at another site with a different EMR system?

We cast the problem of risk modeling as a probabilistic Bayesian network, which has several advantages for capturing and reasoning with uncertainty.[13] These methods are capable of producing two valuable outcomes: (i) an interpretable set of concept variables associated with the risk of falling at the population level and (ii) an actionable model to estimate the risk of falling for individual patients.

METHODS

Study Sites and Modeling Strategy

The two study cohorts were derived from the clinical data repositories of two institutions. One tertiary hospital was the “development site,” and the other tertiary hospital was the “validation site;” both are located near Seoul, South Korea. Both hospitals have approximately 1,000 beds and have used EMR systems for more than 10 years. The
development site had 24,000 coded nursing statements mapped to the International Classification for Nursing Practice (ICNP) terminology. These statements are used for documenting nursing notes with free-text entries. The validation site has coded nursing statements represented by 3N (North American Nursing Diagnosis Association, Nursing Intervention Classification, and Nursing Outcome Classification). The two study sites have different EMR systems with two different terminology standards.

Our research team employed the following principles to enable prediction model translation into practice; (i) based on existing nursing knowledge or clinical guidelines, (ii) interpretable to users, and (iii) parameterized to be adjusted and refined based on target population’s characteristics changing over time and/or sites. At the development-site, we first constructed a concept model and then mapped the concept variables to local data elements, which followed by training with local cohort data. The same concept model was then applied to the validation site and the model parameters were trained and tested by the local cohort.

**Research Cohorts**

The development cohort consisted of inpatients who were admitted to six nursing units with high fall rates during the 1-year period from September 2014 to August 2015. The patients were mainly registered in cardiovascular, hematology-oncology, and neurology medical departments. Inclusion criteria were adults admitted for at least 24 hours. Exclusion criteria were being admitted to a psychiatric, emergency, or pediatric medical department. Patients who died or had received resuscitation treatment were excluded. We
identified 14,307 admissions of 122,179 hospital days that conformed with these criteria. Two hundred and twenty events were identified by analyzing the hospital’s event-reporting system, and an additional 18 cases were found through chart reviews conducted after pre-filtering the free-text entries.

The validation cohort included 21,172 admissions from 6 medical-surgical nursing units. The same eligibility criteria applied to the development cohort were applied to the validation cohort. The fall rate on nursing units was estimated to be lower in the validation site, so we extended data collection to a 2-year period from June 2014 to May 2016. This resulted in a total of 172,592 hospital days and 292 falls being identified after analyzing the reporting system and chart reviews. We adopted the NDNQI® operational definition of falls and level of injury.[14]

Each cohort was divided randomly into model training and testing sets. For both training and testing, the patient stays were divided into windows with a length of 24 hours because nurses’ fall-risk assessments are conducted on a daily basis in practice at both participating hospitals. For example, a patient hospitalized for 4 days can have a maximum 4 fall-risk assessments performed and documented in the EMR. A sliding-window approach was used to generate multiple windows covering a patient’s data during their hospital stay by shifting the window to consecutive fall events. For fallers, only data that applied to within 24 hours before a fall were considered; data obtained prior to this were eliminated since it is unclear whether they should receive a positive or negative label. For non-fallers, all of their data were included and labelled as negative. Samples
were split into the training and testing sets while including samples from a given patient only in one of these sets. This approach was used to mirror the end-use situation more closely, where the system is evaluated on patients who are different from those on whom the model was trained. The imbalance between positive and negative labels was removed by oversampling the positives based on the ratio of positive-to-negative examples. According to a study [15] on machine learning using imbalanced data, the oversampling method is better than intelligent sampling techniques such as SMOTE (synthetic minority oversampling technique) and borderline SMOTE. For assuring the model validation. We applied the 10-fold validation method.

The retrospective study was reviewed and approved by the institutional review boards at the two hospitals, and the need for patient informed consent was waived because the study involved the collection of de-identified data.

**Variables**

Variables were selected based on a literature review focusing on clinical guidelines published within past five years (2012-2017). We adopted the following seven fall prevention guidelines recommended by the Joint Commission\(^1\) including the guideline of the Korean Hospital Nurses Association[16]: Agency for Healthcare Research and Quality,[17] ECRI Institute,[18] Institute for Clinical Systems Improvement,[19] Institute for Healthcare Improvement,[20] Joint Commission Center for Transforming Healthcare, [21] Veterans Affairs,[22] and Veterans Affairs National Center for Patient Safety.[23]
Table 1 lists the concepts identified according to category and care components. We used the concepts to build a predictive Bayesian network structure for falls. We standardized the concepts by mapping to standard nursing terminologies in the current releases of the LOINC (Logical Observation Identifiers Names and Codes) and the ICNP. Then the standard concepts were semantically mapped to local data elements of each individual EMR environment. Nursing assessment, diagnosis, and intervention items were extracted from structured nursing records and nursing notes. The mapping process was conducted by a project team consisting of six experts in the following relevant domains: nursing informatics, terminology, quality management, and patient safety. We have previously described the mapping process.[23-24]

Table 1. Concepts derived from the literature review and local data elements mapped to concept variables in the prediction models.

<table>
<thead>
<tr>
<th>Category</th>
<th>Care component</th>
<th>Model concept</th>
<th>EMR data element in development-site</th>
<th>EMR data element in validation-site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient characteristics</td>
<td>Demographics</td>
<td>Age</td>
<td>Age</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary and secondary dx,</td>
<td>Medical dx. (ICD code), dates of surgical operation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>surgical operation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discharge unit, medical</td>
<td>Discharge unit, medical department, length of stay</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>department, hospital days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contributing factors:</td>
<td>Physiological/disease-related factors</td>
<td>Visual and hearing impairment, elimination impairment, gait,</td>
<td>Nursing assessment and dx.; physiologic evaluation and problem (e.g. impaired mobility, incontinence, etc), KPCSa</td>
<td></td>
</tr>
<tr>
<td>patient</td>
<td></td>
<td>mobility impairment, use of walking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Description</td>
<td>Action/Intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive</td>
<td>Dementia, delirium, disorientation, level of consciousness, fear, irritability, noncompliance</td>
<td>Nursing assessment or dx.; cognitive function (e.g. acute confusion, disorientation, noncompliance, etc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral</td>
<td>Fall history, sleep impairment</td>
<td>Presence of past falls, nursing dx. related to sleep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Therapeutics</td>
<td>Medications, adverse reaction to medications, catheter (IV line, tube, Foley), use of restraints</td>
<td>Medication list by class (sedatives, antidepressant, anti-emetics, antipsychotics, antianxiety drugs, diuretics, antiepileptics, antihypertensives, analgesics, antiarrhythmics and NSAIDs), Physician order of fluid injection, tube, Foley and restraints.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actions to reduce risk</td>
<td>Universal fall precautions Fall precautions on admission, regular rounds</td>
<td>Nursing interventions; safety education on admission, rounds per 2 hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>Patient and caregiver education, presence of bedsitter, use of visual indicators, communicating fall risk status to care team</td>
<td>Nursing interventions; fall prevention education, presence of bedsitter, use of visual indicators, and activities communicating fall risk status to care team</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation and surveillance</td>
<td>Fall-risk assessment tool</td>
<td>Hendrich II score STRATIFY score and subscores[2], and subscores[1],</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk-target intervention</td>
<td>Cognitive and mental function</td>
<td>Nursing interventions: repeatedly provision of orientation, hourly rounding, assigning room close to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>Nursing interventions:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toileting problem</td>
<td>provision toilet scheduling, assist toileting, provision comodo/ bed-pan, etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impaired mobility</td>
<td>provision of mobility devices, walking aids, and assistance, etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medication review</td>
<td>rearranging medication time, provision side-effect precaution, etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep disturbance</td>
<td>attention to night movement and noise, inducing sleep pattern changes, etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental intervention</td>
<td>keeping paths clear, inspect furniture, equipment, lighting, floor, room arrangement</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*KPCS: Korean Patient Classification System,*[26]  
*NSAIDs: nonsteroidal anti-inflammatory agents: dx. means diagnoses.*

Figure 1 shows the four steps used in this study to develop and validate the fall risk prediction model; (i) review of guidelines and literature, (ii) represent the concepts in standardized terminology, (iii) train and evaluate the model, and (iv) cross validate and adjust the model.
Figure 1. The 4 steps of building a predictive Bayesian network model.

**Bayesian Network**

The Bayesian network model was specified as follows: A Bayesian network or probability network $B = (Pr, G)$ is a model of a multivariate probability distribution over
a set of selected concept variables, and consists of a graphical structure $G$ and an
associated distribution $\Pr$. The graphical structure takes the form of a directed acyclic
graph $G = (V(G), A(G))$ with nodes $V(G) = \{v_1, v_2, \ldots, v_n\}$ and arcs $A(G) \subseteq V(G) \times V(G)$, where $G$ represents a random variable that takes one of a finite set of values. The arcs in the graph present the probabilistic influences between the variables.

To build the Bayesian network model structure, we identified relationships between the concepts selected from the eight fall prevention guidelines. The direction of the arcs in the network graph were determined based on physiological, chronological, and logical processes. The network structure was iteratively constructed followed by several refinements based on the availability of data elements in the development-site’s EMR system. The local conditional probability distributions $\Pr(V_i \mid \pi(V_i))$ (we call it parameters) for each variable $V_i$ were obtained from each local (training) data set. For the identified networks, the conditional probability distributions were computed based on the weighted averages of probability estimates from the local data set and a prior Dirichlet distribution; that is, multinomial distributions whose parameters can be interpreted as counts on the data set:

$$
\Pr(V_i \mid \pi(V_i), D) = \frac{n}{n+n_0} \, \hat{\Pr}_D(V_i \mid \pi(V_i)) + \frac{n}{n+n_0} \, \Theta(V_i \mid \pi(V_i))
$$

where $\hat{\Pr}_D$ is the probability distribution estimated from a given data set $D$, $\Theta$ is the Dirichlet prior over the possible values of $V_i$, $n$ is the size of data set $D$, and $n_0$ is the number of past cases on which the contribution of $\Theta$ is based.
Model Evaluation and Cross Site Validation

The model prediction performance was assessed using sensitivity, specificity, receiver operating characteristics (ROC) curves, 10-fold cross validation, and performance indices such as the spherical payoff.[28] The model was also compared with the performance of two fall-risk assessment tools (Hendrich II and STRATIFY) using calibration curves and ROC curves. A calibration curve does not quantitatively measure the reliability of probability predictions, but instead gives a graphical representation to capture the intuitive meaning of the calibration of a given system.[29]

We performed a sensitivity analysis to establish the quality and clinical utility of the fully specified Bayesian network. We observed the output of the network to detect possible inaccuracies in the underlying probability distribution. We determined the degree to which variations in the posterior probability distributions were explained by other variables. The model sensitivity was calculated as the variance reduction with continuous variables and the entropy reduction with ordinal-scale or categorical variables. We used Netica modeling software (version 3.2, Norsys Software Corporation, Vancouver, Canada) to complete the analysis.

Statistical Analysis

Descriptive statistics on population profiles are presented as mean and standard deviation or frequency and percentage values. Each cohort was compared using the chi-square test or the t-test to quantify differences in the population characteristics. Univariate analysis
was used to select the initial features. Statistical analyses were performed using R software (version 3.3, R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Cohort Description

The two cohort populations had some differences in their characteristics (Table 2). The development-site patients were distributed almost equally across the age groups, but they had a longer length of stay than those in the validation-site. Almost 70% of the development-site patients had a neoplasm or circulatory disease, and most of them also had secondary diagnoses. The validation-site patients were older and had more admissions for respiratory and gastrointestinal diseases and surgical procedures. However, there was no significant difference in the frequency of falling: the total falls per 1,000 hospital days were 1.95 and 1.69 at the development and validation sites, respectively; the corresponding rates for injurious falls per 1,000 hospital days were 0.44 and 0.40. The rates of injurious falls were calculated based only on data from the event-reporting system, and so they could have been underestimated due to missing reports. Among the injurious falls, 90.7% and 80% were no injury or minor at the development and validation sites, respectively, and 9.3% and 20% had moderate levels of injury. No major injury was reported.

Table 2. Characteristics of the two cohorts.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Development site</th>
<th>Validation site</th>
<th>$\chi^2$ or $t (P)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

16
<table>
<thead>
<tr>
<th></th>
<th>((n = 14,307))</th>
<th>((n = 21,172))</th>
<th>(332.20)</th>
<th>(&lt; .001)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>6,157 (43.0)</td>
<td>11,199 (52.9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age, years</th>
<th>(629.04)</th>
<th>(629.04)</th>
<th>(629.04)</th>
<th>(629.04)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&lt; 50)</td>
<td>3,165 (22.1)</td>
<td>5,593 (26.4)</td>
<td>(&lt; .001)</td>
<td>(&lt; .001)</td>
</tr>
<tr>
<td>(50–60)</td>
<td>3,251 (22.7)</td>
<td>3,844 (18.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(60–70)</td>
<td>3,356 (23.5)</td>
<td>3,517 (16.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(70–80)</td>
<td>3,281 (22.9)</td>
<td>5,039 (23.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(&gt; 80)</td>
<td>1,254 (8.8)</td>
<td>3,179 (15.0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Length of stay            | 8.54±11.52 | 8.15±11.28 | 3.14 (0.002) | 3.14 (0.002) |

<table>
<thead>
<tr>
<th>Medical diagnosis</th>
<th>11,701.00</th>
<th>11,701.00</th>
<th>11,701.00</th>
<th>11,701.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoplasm</td>
<td>4,639 (32.4)</td>
<td>4,869 (23.0)</td>
<td>(&lt; .001)</td>
<td>(&lt; .001)</td>
</tr>
<tr>
<td>Benign</td>
<td>385 (2.7)</td>
<td>1,066 (5.0)</td>
<td>(&lt; .001)</td>
<td>(&lt; .001)</td>
</tr>
<tr>
<td>Circulatory disorder</td>
<td>5,670 (39.6)</td>
<td>769 (3.6)</td>
<td>(&lt; .001)</td>
<td>(&lt; .001)</td>
</tr>
<tr>
<td>Respiratory and gastrointestinal disorders</td>
<td>655 (4.6)</td>
<td>5,630 (26.6)</td>
<td>(&lt; .001)</td>
<td>(&lt; .001)</td>
</tr>
</tbody>
</table>

| Surgical procedure         | 517 (3.6) | 2,163 (10.2) | \(< .001\) | \(< .001\) |
| Neurology disorder         | 998 (7.0) | 263 (1.2) | \(< .001\) | \(< .001\) |
| Infectious disorder        | 115 (0.8) | 813 (3.8) | \(< .001\) | \(< .001\) |
| Other                      | 1,328 (9.3) | 5,599 (26.5) | \(< .001\) | \(< .001\) |

| Presence of secondary diagnosis | 14,242 (99.6) | 13,421 (63.4) | 6,497.45 | \(< .001\) |

<table>
<thead>
<tr>
<th>KPCS(^a)</th>
<th>52.79</th>
<th>52.79</th>
<th>52.79</th>
<th>52.79</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>227 (1.6)</td>
<td>377 (2.0)</td>
<td>(&lt; .001)</td>
<td>(&lt; .001)</td>
</tr>
<tr>
<td>Group 2</td>
<td>8,197 (57.7)</td>
<td>11,349 (60.7)</td>
<td>(&lt; .001)</td>
<td>(&lt; .001)</td>
</tr>
<tr>
<td>Group 3</td>
<td>3,898 (27.4)</td>
<td>5,630 (30.1)</td>
<td>(&lt; .001)</td>
<td>(&lt; .001)</td>
</tr>
<tr>
<td>Group 4</td>
<td>1,627 (11.5)</td>
<td>1,332 (7.1)</td>
<td>(&lt; .001)</td>
<td>(&lt; .001)</td>
</tr>
<tr>
<td>Groups 5 and 6</td>
<td>262 (1.8)</td>
<td>0</td>
<td>(&lt; .001)</td>
<td>(&lt; .001)</td>
</tr>
</tbody>
</table>

| Number of medications daily | 2.5±6.8 | 18.6±9.9 | \(−1835.04\) | \(< .001\) |

<p>| Total number of medications | 24.4±75.7 | 172.3±317.7 | (−63.07) | (&lt; .001) |</p>
<table>
<thead>
<tr>
<th>Fall events</th>
<th>4.71</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>231 (1.6)</td>
</tr>
<tr>
<td>Multiple</td>
<td>7 (0.05)</td>
</tr>
</tbody>
</table>

Data are n (%) or mean±standard-deviation values.

*KPCS: Korean Patient Classification System. Group 1 has the lowest nursing needs and group 6 has the highest nursing needs.

**Prediction Modeling at the Development-site**

The fall-prediction model identified at the development site consisted of 56 nodes and 82 links. The error rate of the prediction model was 11.7% and the spherical payoff was 0.91. The calibration curves showing the relationship between observed and predicted outcome event rates divided into deciles revealed that the prediction reliability differed between the prediction model and the Hendrich II tool [Figure 2. (A)]. The prediction model was imprecise at the two extreme probability ranges, with high probabilities underestimated and low probabilities overestimated; the Hendrich II tool (for a high-risk score of ≥5) showed a similar pattern.

(A) Curves for the prediction and Hendrich II models at the development site
Curves for the prediction model and STATIFY tool at the validation site

Figure 2. Calibration curves for the two sites. The data are mean and 95% confidence intervals.

The ROC curves created to determine the ability of the model to discriminate between at-risk and no-risk patients are shown on the left side of Figure 3. The area under the ROC curve was 0.96 for the prediction, demonstrating almost perfect discrimination, while it was only 0.69 for the Hendrich II tool.

Figure 3. ROC curves showing the discrimination ability in fall prediction.
In model development site, the sensitivity test showed that Hendrich II data reduced the variance the most (dark-gray bars in Figure 4), followed in order by nursing assessments and diagnoses, nursing interventions, Korean Patient Classification System (KPCS), and medications. The demographics and administrative data made virtually zero contributions. However, for the validation site, medication and KPCS contributed better to the variance reduction than nursing process data.

Figure 4. Results of the sensitivity analysis for subgroup summations of the prediction models. Dark-gray and light-gray bars correspond to the development and validation sites, respectively.

**Cross Site Validation**

The validation model consisted of 48 nodes and 80 links. The error rate was 4.87%. The logarithmic loss and spherical payoff was and 0.96 respectively. These scores indicate the classification abilities of the model.[30] The logarithmic loss is a cross-entropy estimate
which measures the additional penalty for using an approximation instead of the true model. Closer to 0 indicate a lower penalty.\[31\] The spherical payoff is a very useful accuracy measure, with 1 representing best classifier performance. The calibration curves in Figure 2. (B) show that the lowest projected risk decile accounted for only 3% of the observed falls. The proportion of observed falls increased steadily with the projected risk, to reach 85% in the highest risk decile, while the curve for the STRATIFY tool did not exhibit a consistent increasing trend. The prediction model showed a good calibration curve with better precision at extreme probability ranges; the STRATIFY has blunt calibration with a cut-off score of 2.

The area under the ROC curve was 0.99, and slightly higher than for the development site model, which implies that the model performance was over 30% higher than that of the STRATIFY tool. (Right side of Figure 3) The results of the sensitivity analysis (light-gray bars in Figure 4,) showed that the medication and demographics/administrative data had a greater influence on the occurrence of falls at the validation site.

**DISCUSSION**

We found that longitudinal EMR data could be incorporated successfully into a prediction model, which performed better at discriminating at-risk and no-risk patients than did the existing tools for assessing the risk of falling. Furthermore, the model exhibited acceptable performance at the two sites with different EMR systems, patient populations, fall risk assessment tools and nursing terminology standards. In particular, semi-structured EMR data (mostly nursing-process data) were semantically incorporated into a prediction model. These results imply that evidence-based prediction models based on
EMR data may be considered as an alternative to and may be superior to traditional fall-risk assessment tools, especially as they utilize previously unavailable nursing data.

The two sites involved in this study have different patient profiles in terms of age, primary diagnosis, and medication distributions. However, the rates of falling and injurious falls at the two sites were similar. This finding is consistent with a study from the National Institutes of Health in 2013[32] that involved 1,263 hospitals across the US; They found no trend in the rates of falling or injurious falls according to hospital size or staffing level and the differences in fall rates within each organizational characteristically ranged from 0.17 to 0.33 falls per 1,000 hospital days.

We used all the data available in the EMR systems that are known to be relevant to inpatient falls based on clinical guidelines. One of the challenges in EMR-based studies is the presence of missing data.[12, 33] We observed about 63% missing data for the risk-assessment-tool score and 10% of the KPCS score and subscores. The missing risk-assessment-tool score data was due to the hospital’s local policy that specified reassessment period at everyday for all or at two or three times a week for at-risk patients and once a week for no-risk patients as well as changes in status. These local policies varied by each hospital’s structural factors such as staffing level and patient-nurse ratio. In practice, nurses assume that a previous score could be effective by the next reassessment. However, in data aspect, it was not explicit. The Bayesian inference is greatly advantageous for handling missing data and can produce accurate predictions even when complete data are not available.[13] The expectation-maximization algorithm
that we used in the learning network performed automatic inference based on a-priori probabilities.[13]

A key challenge when building predictive models from EMRs is handling nursing interventions. These interventions are confounders in that they can reduce the likelihood of a fall and thereby make it difficult to distinguish between patients who are at risk for falls based on their fall risk assessment score and those who are at risk but their fall risk is mitigated by preventative interventions. Paxton et al.[34] pointed out that not taking this type of masking into account may lead to models that are useless in practice. We therefore adopted a prognostic Bayesian network, and noticed that the occurrence of falling for a specific patient is generally influenced by the sequence of preventive actions performed by nurses, which in turn may depend on the information that is available about the patient before any interventions aimed at preventing falls are implemented. Falls are often also influenced by the underlying condition of a patient. We therefore formally defined a prediction as a probability distribution, \( Pr(\text{falls} | E, T) \), where \( E \) are the available patient data and \( T \) denotes nursing interventions provided by nurses.

The model developed in this study could be used to evaluate the performance and uncertainty of the Bayesian network. The c-statistic values of 0.96 and 0.99 found in the present study were much higher than those found in studies of prediction models for mortality and clinical outcomes based on EMR data (c-statistic = 0.84 and 0.83, respectively).[12] Our c-statistic values were assured through the testing set and 10-fold cross validation, which supports the reliability of the performance of the models. In
addition, the present c-statistic values were much higher than that in the study of Yokota and Ohe\cite{8} (c-statistic = 0.72), which developed a model for predicting the risk of falling based on EMR data. Yokota and Ohe’s study included physician-order items such as treatment directions, laboratory test and imaging findings, therapies, medications, and nursing assessment and plans. However, only items related to the intensity of nursing-care needs with age and sex remained in their final regression model.

Another comparable study is that of Marier and colleagues, who investigated fall prediction using the Minimum Data Set (MDS) and EMR data of 13 nursing-home residents.\cite{6} They compared four regression models and found that the rate of observed falls increased from 28.6% to 32.3% among residents in the highest risk decile when EMR data were added to an MDS-only model. But the report of that study did not include any model performance metrics such as c-statistic values.

The approach adopted in the present study has several advantages over previously proposed methods for estimating the risk of falling. The first advantage relates to external validation, which is uncommon given that almost all studies have validated performance within the same EMR environment.\cite{12} We conducted an external validation of the developed model at a second site with a different EMR system, patient population, fall risk assessment tool, and nursing terminology standard. For fall risk assessment, a substantial number of tools are readily available and several of them (e.g. Morse fall scale, STRATIFY, Hendrich II) are widely used in hospitals. These tools assess many of
the same areas of risk. [35] These findings suggest that our model is highly portable and comprehensive.

Second, we incorporated more than 50 concepts mapped to 70 time-varying data elements, which represents a relatively large number of variable sets. We found only a small number of studies that used longitudinal EMR data, and they did not fully utilize the depth of information on patients available in the nursing records to identify predictor variables.[8, 10, 11] Instead, those studies used summary metrics or opted for smaller predefined lists. Considering the advantage that the size of EMR data is not limited to the number of patients or the number of potential predictor variables, integrating repeated observations over time is a key strength of this study’s use of EMR data.

A third advantage of our approach relates to the incorporation of nursing-process data including the fall prevention interventions provided to the patients. It is difficult to find an EMR-based study that has integrated the nursing activities of assessments, diagnoses, and interventions—this was possible in the present study because the two EMR systems included complete electronic nursing notes consisting of coded and standardized statements using locally developed data dictionaries.[36] In addition, we identified how the nursing activities captured by the EMR system affect the reduction in the variance of fall events. This finding showed that to accurately predict falls, the nursing data in EMR systems are as important as the individual risk factors of the patients. This implies that using readily available data for risk prediction may simplify computation. Early identification and more precise prediction of at-risk patients has the potential to improve
outcomes by facilitating timely initiation of appropriate and targeted attention, interventions, and monitoring.

Finally, using our model, we calculated for each individual patient, the daily estimate of their risk of falling. The estimated probability ranges from 0% to 100%, so users could set a cut-off of risk depending on an appropriate level of sensitivity and specificity. The next steps involve implementing our approach more broadly and performing a prospective evaluation, of the net benefits obtained by providing fall prevention nursing decision support in practice, as well as validating the model at other sites. For example, interventions could be recommended in real-time for patients that are tailored to their individual fall risk factors. We plan to incorporate a tailored intervention guide according to the individual risk factors of at-risk patients. This will be a great opportunity to explore how the algorithms impacts the clinical decision-making of nurses.
CONCLUSIONS

We found that a risk prediction model that utilizes longitudinal EMR data on nursing assessments, diagnoses, and interventions can improve the ability to identify individual patients who are at a high risk of falling. The prediction model has demonstrated portability and reliability, and can therefore be applied across hospitals with different EMR environments. Current EMR systems—even suboptimal ones—can be leveraged for the secondary use of clinical data to prevent patients from falling.

ACKNOWLEDGEMENTS

We thank to Keoung-Hee Choi, Jungmee Han and Won-Hee Park of Medical Information Department of the two hospitals for helping us to retrieve the raw data efficiently. We also appreciate the clinical staffs of Nursing Departments and graduate students involved in concept mapping process of local data elements with standard vocabularies.

COMPETING INTERESTS

None.

FUNDING

This study was supported by a grant of the Korea Healthcare Technology R&D Project, Ministry for Health and Welfare, Republic of Korea (No. HI15C1089).

CONTRIBUTORS

I.C. conceived and designed the study, supervised and contributed to the data analysis, interpreted the results, and drafted and revised the paper. E.B. and E.J.C. contributed to study design, data acquisition, results interpretation, and paper revision. DW.B. and PC. D. made a substantial contribution to data interpretation and also made critical revisions regarding the intellectual content.
REFERENCES


http://www.k-hen.com/Portals/16/Topics/Falls/ANAsNQFspects.pdf. Archived at: http://www.webcitation.org/70hROcY18


http://khna.or.kr/bbs/linkfile/resource/khna_Fcare.pdf. Archived at:
http://www.webcitation.org/70hRn7JeG.


https://www.ecri.org/components/HRC/Pages/SafSec2.aspx. Archived at:
http://www.webcitation.org/70hRw3uHr.

https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_gui


