Published in final edited form as:

Proc IEEE Int Congr Big Data. 2015 ; 2015: 409–415. doi:10.1109/BigDataCongress.2015.67.

Predictive Modeling for Comfortable Death Outcome Using Electronic Health Records

Muhammad Kamran Lodhi,

College of Engineering, University of Illinois at Chicago, Chicago, United States

Rashid Ansari,

College of Engineering, University of Illinois at Chicago, Chicago, United States

Yingwei Yao,

College of Engineering, University of Illinois at Chicago, Chicago, United States

Gail M. Keenan,

College of Nursing, University of Florida, Gainesville, United States

Diana J. Wilkie, and

College of Nursing, University of Florida, Gainesville, United States

Ashfaq A. Khokhar

College of Engineering, Illinois Institute of Technology, Chicago, United States

Muhammad Kamran Lodhi: mlodhi3@uic.edu; Rashid Ansari: ransari@uic.edu; Yingwei Yao: yyao@uic.edu; Gail M. Keenan: gkeenan@ufl.edu; Diana J. Wilkie: diwilkie@ufl.edu; Ashfaq A. Khokhar: ashfaq@iit.edu, ashfaq@uic.edu

Abstract

Electronic health record (EHR) systems are used in healthcare industry to observe the progress of patients. With fast growth of the data, EHR data analysis has become a big data problem. Most EHRs are sparse and multi-dimensional datasets and mining them is a challenging task due to a number of reasons. In this paper, we have used a nursing EHR system to build predictive models to determine what factors impact death anxiety, a significant problem for the dying patients. Different existing modeling techniques have been used to develop coarse-grained as well as fine-grained models to predict patient outcomes. The coarse-grained models help in predicting the outcome at the end of each hospitalization, whereas fine-grained models help in predicting the outcome at the end of each shift, therefore providing a trajectory of predicted outcomes. Based on different modeling techniques, our results show significantly accurate predictions, due to relatively noise-free data. These models can help in determining effective treatments, lowering healthcare costs, and improving the quality of end-of-life (EOL) care.

Keywords

component; predictive modeling; end-of-life (EOL); Electronic health record (EHR)

Disclosure Statement

Dr. Gail Keenan is currently the President and CEO of this company and has a current conflict of interest statement of explanation and management plan in place with the University of Illinois at Chicago.

I. Introduction

Sparse and high dimensional datasets pose significant challenges for data mining and machine learning related tasks. The problem is further complicated when the variables involved have non-linear scales and exhibit varying degree of importance from the application perspective. One such dataset stems from electronic health record (EHR) systems, where thousands of variables exists (high dimensionality), but only a few are tracked for any given patient (sparseness), The significance of a particular variable present in a given record may also depend on the context. The use of data mining tools to analyze such data sets can be highly beneficial. The capability to predict patient's condition while the patient is hospitalized is crucial to providing cost effective care, and it is based on diverse factors such as the patient's personal and psychological characteristics and other health issues. In this paper, our focus is on the analysis of nursing care data within EHR systems, which is an important but often ignored component of the EHR. Nurses are the front-line providers of care so the data they enter into EHR systems is extremely imperative to the improvement of patients' care. Mining the nursing data can potentially help guide more effective treatment of patients and thus to help reduce costs and produce better patient outcomes. Unfortunately, prior to now, most of the data that are currently entered by the nurses cannot be analyzed using big data techniques, because in most cases, the data entered are not in a standardized format.

EHRs have traditionally been used for monitoring the progress of patients' condition over time [1, 2]. However, a major initiative is underway in the recent years to discover the hidden knowledge through applying big data techniques to the EHR data [3]. As EHR systems are quite large in size and contain diverse data, they are ideal candidates to study Big Data concerns, including data analytics, storage, retrieval techniques, and decision making tools. Other industries have saved millions of dollars by uncovering multiple cost saving measures using big data science. The healthcare industry on the other hand, has been wasting around a trillion dollars annually, out of which \$88 billion is wasted because of inefficient use of technology [4] and these costs are expected to rise [5]. By using these big data science techniques, a lot of money can be saved. Discovering the hidden knowledge within EHR data for improving patient care offers an important approach to reduce these costs by recognizing at-risk patients who may be aided by targeted interventions and disease prevention treatments [6].

One significant application of predictive modeling is to appropriately recognize the characteristics of different health issues by understanding the patient data present in EHR [7]. In addition to early detection of different diseases, predictive modeling can also help to individualize patient care, by differentiating individuals who can be helped from a specific intervention from those that will be adversely affected by the same intervention [8, 9].

Death anxiety is one of the common problems that patients face at the end of life (EOL). Death Anxiety is defined as a "vague uneasy feeling of discomfort or dread generated by perceptions of a real or imagined threat to one's existence" [10]. The concept of death anxiety has been well studied in the field of psychology [11–13] and nursing [14, 15]. However, most of these papers measure death anxiety of different age groups [13, 16–18]

and compare them based on gender and age. Most of these papers are studies conducted on students, parents, church-goers, or nurses. Only a few studies have been conducted on hospitalized patients measuring their death anxiety [19]. Also, there are only a few papers in literature that talk about prediction of death anxiety in residents of the nursing home [20, 21]. There are some limitations in these papers as well. [20] only considers male residents of the nursing home of age of more than 88 years, whereas [21] built a predictive model using functional, emotional, cognitive, and disease variables found in the Minimum Data Set of older patients (age > 65) and not on nursing metrics. One notable difference between the previous work and our research is that there are no studies that focused on patients with death anxiety using data from the EHRs, possibly because no efficient tools were available to record this diagnosis in a standardized format.

In this paper, we examine the comfortable death outcome of EOL patients suffering from death anxiety. The treatment costs of EOL patients take up a massive share of the healthcare expenditures. About a third of these costs are spent on only 5% of all the beneficiaries who die every year. These costs are high mostly because of out-of-date and ineffective treatments being provided to the patients [22]. These expenditures are anticipated to escalate significantly in upcoming years as 20% of the US population will be older than 65 years of age by year 2030. Therefore, the hidden information within the EHR systems needs to be extracted so that cost-effective care can be delivered to the EOL patients. In this paper, we first examine different associations and then explore different predictor models to determine links that may exist between a variety of patient characteristics and nursing interventions and improved outcomes for EOL patients. These predictive models can then be used to identify the best practices and then to predict expected outcomes in real world settings [23].

II. Data Description

The data for our study was gathered over a period of three years (2005–2008) from nine different hospital units by a nursing care planning system called Hands-On Automatic Nursing Documentation System [23]. The system is specifically designed to record nursing care provided to the patients using standard terms. For each patient in the system, the medical record comprises the various nursing diagnoses for the patient's ailments, nursing outcomes associated with these diagnoses, and nursing interventions directed toward achieving the nursing outcomes. All the nursing diagnoses are based on North American Nursing Diagnosis Association (NANDA-I) [10]. Similarly, nursing outcomes are based on Nursing Outcome Classification (NOC) [24] and nursing interventions are coded using the Nursing Interventions Classifications (NIC) [25].

There were a total of 42,403 episodes (34,927 distinct patients) recorded during the three years of the original study. An episode is defined as a continuous patient stay on a single hospital unit and comprises multiple nurse shifts. Each nurse shift corresponds to a single plan of care (POC) that the nurses document at every formal handoff (admission, shift change, discharge). Whenever a patient is admitted, a new POC is created and updated at every shift change. The POC, along with the different diagnoses, also includes current ratings for each nursing outcome and a rating expected at discharge (the goal for the outcome). The rating scale is 1 to 5 where 5 represents the best outcome. Nursing

interventions (NICs) used to achieve the expected outcome are also stored within the POC. In addition, the dataset also contains demographic information of the patients and nurses. Table 1 provides illustrations of two episodes with multiple POCs that nurses documented.

Out of the 42,403 episodes present in the database, we chose only a subset of episodes that belonged to the EOL category for our experiments. An episode is defined as an EOL episode when any of its POCs contains one or more of the following indicators.

- Discharge Status = “expired” or “hospice home care” or “discharge to hospice medical facility”
- NOC = “Comfortable Death” or “Dignified Life Closure”
- NIC = “Dying care”.

The database consisted of 1,546 unique EOL patients, a few with multiple episodes. For this study, we considered only the last episode for patients that had multiple episodes implying one episode per patient in the chosen dataset. We did not consider earlier episodes for patients with multiple episodes because there were only eight such patients and the potential information gain is minimal by including these episodes and having to account for within patient correlation.

462 patients with multiple POCs were diagnosed with Death Anxiety. Patients with single POCs were ignored. The mean age of these patients was 78.8 years ($SD = 12.8$ years). The set of patients was further decreased by including only those patients that had a NOC: Comfortable Death. The final sample includes 438 unique patient episodes, containing a total of 3928 POCs.

I. Feature Extraction

In this paper, we conduct predictive modeling at two different levels of granularity. First analysis has been done with episode-level data whereas second, more fine-grained analysis has been conducted at the POC level. The target variable for predictive modeling is “NOC met” and is defined as:

NOC met: A NOC is considered “met” if the final NOC outcome rating is same or better than the NOC expected rating, which is set when a NOC was first put on a patient’s plan of care. NOC is assumed to be “not met” otherwise. For episode level modeling, it is computed using the NOC rating set at the time of discharge or death.

The aim of the study is to develop predictive models that determine how different patient and nursing variables impact the outcome. The variables used in our analysis, extracted from the database, include Initial, Expected and Final NOC outcome ratings varying between 1 and 5, with 1 being the worst outcome. Other variables were selected by selecting the most important variables through association mining.

In our dataset of 438 patients, there are 168 NANDA-Is and over 500 NICs. Our dataset also contains patients’ demographics like their age, race, gender and length of stay information. The dataset also includes nurse experience and nurse education. In the dataset, the patient’s

race and gender as well as nurse education were not required fields, therefore most of these values are null. Similarly, there are no patients that had all of the NANDA-Is and NICs within their plan of cares. We therefore have a high dimensional sparse dataset.

To select clinically important features and to reduce dimensions of the dataset, association mining was performed. We chose association mining over other dimension reduction techniques, such as Principal Component Analysis, because conventional techniques only consider frequent data, which may not be important in a particular domain. Association mining revealed a few interesting results on its own.

We observed that younger patients had less chance of meeting the outcome, whereas older patients had a much higher chance of meeting the outcome. A chi-squared test indicated that this difference was statistically significant ($\chi^2(2) = 9.27$, $p = 0.002$). It was also observed that the chances of meeting the NOC improved as the length of patient stay increased. Only 38% of the patients with stay of less than 48 hours met their outcome, whereas 53% of the patients that stayed for 5 days or longer met their outcome. The chi-square test showed that for those patients that had NOC: Comfortable death, their chances of meeting the NOC improved significantly as their stay increased ($\chi^2(2) = 6.47$, $p < 0.04$). Though statistically not significant, we also noticed that patients who received care from more experienced nurses were more likely to meet expected outcomes for Comfortable Death. For this purpose, patient's age, length of stay (LOS) and nurse experience were selected as features. All of these features are continuous variables that have been discretized for our analysis.

Apart from these patient and nursing variables, we selected only those NANDA-Is and NICs that had strong correlations with death anxiety and had appeared a significant number of times. Therefore these NANDA-Is and NICs were also incorporated as predictive variables. To further reduce the dimensionality of our dataset, the NANDA-Is and NICs were grouped into NANDA-I and NIC domains respectively. The groups were based on respective taxonomic structures of these systems [10, 25]. In our dataset for this study, the NANDA-Is belonged 4 of the 12 domains. NANDA domains included were Activity/Rest, Comfort, Perception/Cognition, and Safety/Protection domains. The other eight domains were not studied due to low frequency of episodes that appeared in those domains. In addition, we considered five of the seven NIC domains in our study. The included NIC domains are Behavioral, Family, Physiological: Basic, Physiological: Complex and Safety domain. A particular NIC or NANDA domain was assumed to be either present or absent in an episode.

To study clinically meaningful impact of the patients' age, age was divided into four groups: young (18–49), middle-aged (50–64), old (65–84), and very old (85+) based on theoretical [26] and the data frequency distribution.

LOS for each episode was a derived variable in the database. It was determined by summing the number of hours for all the nurse shifts in the episode. LOS was then partitioned into three groups: short (up to 48 hours), medium (48–119 hours), and long stay (120+ hours). Currently, patient stay of less than 48 hours is considered as observational visit. Average LOS is usually around 120 hours, which we consider a medium stay in our study. Stays longer than average are considered as long stays. [27].

Average nurse experience was a derived variable too. It was calculated by accumulating the experience of nurses, in years that provided care to the patient in an episode. The total number of years in nurse experience was then divided by the total number of unique nurses in that episode. In our study, we considered nurse with over two years of experience as an experienced nurse. The nurse with less than two years of experience was considered as an inexperienced nurse. These categories were also based on professional criteria [28].

II. Data modeling

The objective for conducting these experiments was to build models that can predict the comfortable death outcome for the EOL patients. Another aim was to evaluate the practicability of constructing the predictive models using data from the nursing EHR system.

After we had extracted and pre-processed the data from the previous step, we built multiple models on the refined data by means of different prediction algorithms and compared their performances. For the initial experiment, the models were based on Decision Trees [29], k-nearest neighbors (k-NN) [30], support vector machines (SVM) [31] and Logistic Regression [30]. We did not include Naïve-Bayes model [32] in the episode-level analysis as our definition of “NOC met” is directly related to Expected NOC outcome rating, and so, the Naïve-Bayes conditional independence assumption would be violated [33]. Naïve-Bayes models are included in our POC level analysis.

III. Experimental results

A. Coarse-Grained Episode-level models

The first experiment was to develop predictor models for predicting whether the NOC: Comfortable Death was met based on various nurse and patient characteristics and various nursing diagnoses and nursing interventions. The target variable was a binary variable (met/not met). In the first experiment, the final NOC rating was excluded as a predictor variable as the target variable is a deterministic function of the expected and final NOC outcome ratings.

We used 70% data for training whereas the rest 30% of the data was used for validation. The partition was done randomly. The models were evaluated using accuracy, f-measure, specificity and sensitivity. In this experiment, the positive class was NOC met.

1) Predicting Expected NOC outcome rating—We first present the results for episode level models that predict whether the NOC: Comfortable Death is met or not met. The results are given in the Table 2.

The table above shows that accuracies for all the models are generally high. Decision tree (Figure 1) has the highest accuracy, sensitivity, specificity and f-measure among all the models. SVM also has an accuracy of around 79%. K-NN models accuracies vary between 76–78%. Logistic Regression also has an accuracy of around 77%. Other measures of evaluation including sensitivity, specificity and f-measure are also generally high for all the

other models with the exception of specificity for k-NN ($k = 5$) model as the specificity for that particular model is less than 69%.

2) Feature Selection—Feature selection for this experiment was done using decision tree. As seen in Figure 1, the most important attribute was Expected NOC rating, followed by LOS when expected NOC rating was less than or equal to 4, and by First NOC rating when Expected NOC rating was 5. LOS was followed by Activity/Rest NANDA domain when patient had a medium LOS. Activity/Rest NANDA domain was followed by Perception/Cognition NANDA domain. As seen from the figure, only around 35% of the patients met their outcome when their expected score was set to be 5. This probably indicates that nurses are setting somewhat ambitious expectations from their patients, especially when the first NOC rating is less than or equal to 3, which most of them do not reach.

Feature selection was also performed using Least Absolute Shrinkage and Selection Operation method [34]. This method is a selection method for linear regression. Using this method, the results were slightly different. Using this method, Expected NOC rating was the most important attribute with the highest weight assigned, which was followed by age, and then the LOS. Activity/Rest NANDA domain was assigned the 4th highest weight, followed by Family NIC domain, which was given the 5th highest weight.

3) Predicting Final NOC Outcome Rating—We performed another experiment at episode-level in which we tried to predict the final NOC outcome rating using the same data. For this experiment, our target variable was Final NOC rating (values between 1 and 5). “NOC met” attribute was excluded for this experiment. As done previously, models were trained with 70% data and tested with the rest 30%. Because the target variable was polynomial, we did not use all of the performance matrices. The results are shown in Table 3.

Decision tree had the highest accuracy again for this experiment with over 75.5% accuracy. SVM model did not perform as well as it had in the previous experiment. Logistic Regression also had a lower accuracy. K-NN models had accuracy ranging from 68.7% to 73.3%, depending on value of k . These results once again showed that predictive capabilities exist in the data from the HANDS database.

B. Fine-grained Analysis of Plan of Cares (POC)

After the completion of episode-level analysis, we drilled down at the care of plan level. For this experiment, our objective was to predict the current NOC rating at the end of each nurse shift within each episode. The predictor variables were the rating from the previous nurse shift, age, experience of nurse in the current shift and NANDA and NIC domains in the shift. LOS was not included as the analysis was now on shift level data. Naïve-Bayes models were also used for this experiment. The Table 4 shows the results.

As seen from the table, SVM and LR again performed the worst among all the models. Decision tree and Naïve-Bayes models had almost the same accuracy at more than 78.8%. Both the K-NN models had an accuracy around 72%. Though the k-NN models were almost similar, their performance was better than SVM and LR models.

In all the experiments conducted above, Decision tree models have consistently performed better. Although in the fine-grained analysis of plan of cares, Naïve-Bayes had a better result, but the result was almost similar. Also, Naïve-Bayes models cannot be used in the coarse-grained episode level experiments. Therefore, for our data, we purpose decision tree algorithm for our experiments.

IV. Lessons Learned and Conclusion

Big data from EHRs provide opportunities for new discoveries, better care for the patients, and more efficient healthcare. However, predictive modeling on high dimensional and sparse dataset such as EHRs is a tedious task. Non-conventional methods for reducing dimensions, such as association mining, need to be used because traditional methods usually only consider the most frequent variables, which may not be of importance for a particular domain. On the other hand, non-traditional methods provide the context of the variables. Association mining can help in finding strong correlations in the data, which helps in eliminating those dimensions or variables that have limited impact on the results of the predictive models.

There are different algorithms for predictive modeling, however, not all models can be used with the available data. Each dataset has different characteristics, and hence, each dataset may require a different algorithm to get the best results. Therefore, selecting the correct technique to match the problem is important.

Mining the big data cannot be done properly without the help of domain experts. Before using the big data, there is a need to understand the domain to which the data belongs, with the help of these experts. Mining this data can help in revealing anomalies or unknown patterns, but they may be irrelevant with respect to the domain to which the data belongs. Therefore, results from EHR data need to be checked for clinical importance before implementation of these results.

In this study, we first tried to determine if any association exists between patient or nursing characteristics and NOC outcome being met or not. We also tried to reduce dimensions of our dataset during association mining step. We observed that older patients had a better chance of meeting the comfortable death outcome as compared with the younger patients. We also observed that patients with a longer stay at the hospital met their expected comfortable death outcome as compared to those patients who had a shorter LOS. Both of these findings were statistically significant ($p < 0.05$) using the chi-squared test.

Later, we developed prediction models with the objective of predicting comfortable death outcome, an important NOC for patients at EOL. Our secondary objective was to examine the practicability of these models using data from a current nursing EHR system. For our experiments, we used different models, including decision tree, SVM, LR, k-NN and Naïve-Bayes models, and achieved high accuracy results for our experiments at episode-level. Other performance measures were also high in general. In our second experiment at the episode level, we tried to predict the final NOC outcome rating. Once again, generally high accuracies were achieved except for SVM and LR models. In our fine-grained analysis of

plans of care, we also succeeded in getting high accuracies for decision tree and Naïve-Bayes. K-NN models also performed well. Decision tree models performed better than other models in episode-level analysis. In our plans of care analysis, Naïve-Bayes and decision tree model's performance was almost the same.

Our results from these experiments on big data from a nursing EHR system has shown that there is a robust capability in such systems to be used as a base in decision support framework for nursing. Most of the models we built had high accuracies for all the different analyses. Association mining also revealed interesting results. It is possible that through these methods, we can fully realize the potential of these EHR systems.

Using these different data mining approaches, that include association mining and predictive modeling, can help the hospital management in improving care for the general community. Consistent patient and nursing data can be used for devising effective nursing care for the patients. Integrating the EHRs with data modeling can complement the traditional research methods and improve care of general population [35]. These models can also help in reducing the healthcare costs along with improving nurse care. Despite the potential limitations in the data, standardization of different treatments is one important leap towards personalized care.

Acknowledgments

This research was made possible by Grant Number 1R01 NR012949 from the National Institutes of Health, National Institute for Nursing Research (NINR). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NINR.

The HANDS software that was used in this study is now owned and distributed by HealthTeam IQ, LLC.

References

1. Feliu J, et al. Development and Validation of a prognostic nomogram for terminally ill cancer patients. *Journal of National Cancer Institute*. 2011; 103(21):1613–1620.
2. Gagnon B, et al. Montreal prognostic score: estimating survival of patients with non-small cell lung cancer using clinical biomarkers. *British Journal of Cancer*. 2013; 109(8):2066–2071. [PubMed: 24064979]
3. [cited 2014 September 27] PubMed. Available from: <http://www.ncbi.nlm.nih.gov/pubmed>
4. Kavilanz PB. Health care's big money wasters. 2009 Aug 10. [cited 2014 June 29] 2009 Available from: http://www.money.cnn.com/2009/08/10/news/economy/healthcare_money_wasters/.
5. Cuckler, G. National Health Expenditures Projections 2012–2022. C.f.M.a.M.S. 2014. , editor. 2014.
6. Yeager D. Mining Untapped Data. *For the Record*. 2013; 25(5):10.
7. Riccardo, Bellazi; Zupan, B. Predictive data mining in clinical medicine: Current issues and guidelines. *International Journal of Medical Informatics*. 2008; 77(2):81–97. [PubMed: 17188928]
8. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: toward better research applications and clinical care. *Nature Reviews Genetics*. 2012; 13:395–405.
9. Vickers AJ. Prediction models in cancer care. *A Cancer Journal for Clinicians*. 2011; 61(5):315–326.
10. NANDA International, in *Nursing Diagnosis and definitions: 2007–2008*. Oxford, United Kingdom: John Wiley & Sons Ltd.; 2007.
11. Abdul-Khalek AM. Death Anxiety in clinical and nonclinical groups. *Death Studies*. 2005; 29(3): 251–259. [PubMed: 15818874]

12. Depaola SJG, M, Young JR, Neimeyer R. Death anxiety and attitudes towards the elderly among older adults: The role of gender and ethnicity. *Death Studies*. 2003; 27(4):335–354. [PubMed: 12749378]
13. Keller JW, Sherry D, Piotrowski C. Perspectives on Death: A developmental study. *The Journal of Psychology*. 1984; 116(1):137–142. [PubMed: 6699791]
14. Kisvetrova H, Klugar M, Kabelka L. Spiritual support interventions in nursing care for patients suffering death anxiety in the final phase of life. *International Journal of Palliative Nursing*. 2013; 19(12):599–605. [PubMed: 24356504]
15. Lehto RH, Stein KF. Death anxiety: An analysis of an evolving concept. *Research and Theory for Nursing Practice*. 2009; 23(1):23–41. [PubMed: 19418886]
16. Bengston VL, Cuellar JB, Ragan PK. Stratum contrasts and similarities in attitudes towards death. *Journal of Gerontology*. 1977; 32(1):76–88. [PubMed: 830739]
17. Feifel H, Branscomb AB. Who's afraid of death? *Journal of Abnormal Psychology*. 1973; 81(3): 282–288. [PubMed: 4710052]
18. Fortner BV, Neimeyer RA. Death Anxiety in older adults: A quantitative review. *Death Studies*. 1999; 23(5):387–411. [PubMed: 10558505]
19. Lo C, et al. Measuring death-related anxiety in advanced cancer; Preliminary psychometrics of the death and dying distress scale. *Journal of Pediatric Hematology Oncology*. 2011; 33(Supplement 2):S140–S145.
20. Levy C, Morris M, Kramer A. Improving End-of-life Outcomes in Nursing Homes by Targeting Residents at High Risk of Mortality for Palliative Care: Program Description and Evaluation. *Journal of Palliative Medicine*. 2008; 11(2)
21. Porock D, et al. Predicting Death in the Nursing Home: Development and Validation of the 6-Month Minimum Data Set Mortality Risk Index. *Journal of Gerontology*. 2005; 60A(4):491–498.
22. Zhang B, et al. Health care costs in the last week of life: Associations with end-of-lofe conversations. *Archives of internal medicine*. 2009; 169(5):480–488. [PubMed: 19273778]
23. Keenan G, et al. Maintaining a consistent big picture: Meaningful use of a Web-based POC EHR system. *International Journal of Nursing Knowledge*. 2012; 23(3):119–133. [PubMed: 23043651]
24. Moorhead, S.; Johnson, M.; Maas, M. Iowa Outcomes Project, Nursing outcomes classification (NOC). Mosby: St. Louis, MO; 2004.
25. Dochterman, J.; Bulechek, G. Nursing interventions classification (NIC). Mosby, , editor. St. Louis, MO; 2004.
26. Gronbach KW. The Age Curve: How to Profit from the Coming Demographic Storm. 2008
27. Hospital utilization (in non-federal short-stay hospitals). Centers for Disease Control and Prevention; 2014.
28. Benner P. From novice to expert. *American Journal of Nursing*. 1982; 82(3):402–407. [PubMed: 6917683]
29. Quinlan, J. C4.5: Programs for machine learning. Kaufmann, M., editor. San Francisco, CA: 2003.
30. Aha D, Kibler D, Albert M. Instance-based learning algorithms. *Machine Learning*. 1991; 6(1):37–66.
31. Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995; 20(3):273.
32. Pearl, J. The handbook of brain theory and neural networks. MIT Press; 1998. Bayesian networks.
33. Lewis, D. Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval; Proceedings of 10th European Conference on Machine Learning; 1998. p. 4-15.
34. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of Royal Statistical Society*. 1996; 58(1):267–288.
35. Desikan P, et al. Predictive Modeling in Healthcare: Challenges and Opportunities. [cited 2014 September 27] Available from: <http://lifesciences.ieee.org/publications/newsletter/november-2013/439-predictive-modeling-in-healthcare-challenges-and-opportunities>.

Figure 1.
Decision Tree for NOC: Comfortable Death NOC met/not met

Table 1

Example of Database

Episode ID	Shift ID	NANDA	NOC	NIC	Patient Age	Nurse Experience	Current Rating	Expected Rating
1	1	Death Anxiety	Comfortable Death	Dying Care	67	1	3	4
1	2	Death Anxiety	Comfortable Death	Life Span Care	67	5	5	4
2	1	Death Anxiety	Comfortable Death	Dying Care	80	2	4	5
2	2	Death Anxiety	Comfortable Death	Dying Care	80	2	2	5

Table 2

Model Performance of Experiment 1: Predicting NOC met/not met

Model	Accuracy	Sensitivity	Specificity	F-Measure
SVM	79.39	82.86	75.41	0.81
Logistic Regression	77.10	81.43	72.13	0.79
k-NN (k = 2)	76.52	75.71	77.42	0.77
k-NN (k = 3)	77.10	78.57	75.41	0.79
k-NN (k = 4)	77.86	78.57	77.05	0.79
k-NN (k = 5)	76.92	84.06	68.85	0.79
Decision Tree	86.26	90.00	81.97	0.88

Table 3

Model Performance for Experiment 2: Predicting the Final NOC Rating

Model	Accuracy
Decision Tree	75.57
k-NN (k = 2)	68.70
k-NN (k = 3)	73.28
k-NN (k = 4)	71.75
k-NN (k = 5)	70.99
SVM	68.92
Logistic Regression	68.86

Table 4

Model Performance for Experiment 3: Predicting NOC rating at end of current shift

Model	Accuracy
Decision Tree	78.81
k-NN (k = 5)	72.11
K-NN (k = 2)	71.75
SVM	66.31
LR	65.16
Naïve-Bayes	78.85