Comparison of Multiple Prediction Models for Ambulation Following Spinal Cord Injury

Todd Rowland MD [1] Lucila Ohno-MachadoMD PhD [1] Aleksander Ohrn MSc [2]

[1] Harvard Decision Systems Group
Brigham and Women's Hospital
Harvard Medical School
Boston, MA trowland@dsg.harvard.edu

[2] Knowledge Systems Group
Dept. of Computer and Information Science
Norwegian University of Science and Technology
Trondheim, Norway aleks@idi.ntnu.no

Abstract: *Few studies have properly compared predictive performance of different models using the same medical data set. We developed and compared 3 models (logistic regression, neural networks, and rough sets) in the in prediction of ambulation at hospital discharge following spinal cord injury. We used the multi-center Spinal Cord Injury Model System database. All models performed well and had areas under the receiver operating characteristic curve in the 0.88-0.91 range. All models had sensitivity, specificity, and accuracy greater than 80% at ideal thresholds. The performance of neural network and logistic regression methods was not statistically different (p = 0.48). The rough sets classifier performed statistically worse than either the neural network or logistic regression models (p-values 0.002 and 0.015 respectively).*

INTRODUCTION

One of the most common questions asked by patients following acute spinal cord injury is: "Will I be able to walk?" One of the ways our society defines outcome following spinal cord injury is one's ability to walk. Numerous studies have evaluated this issue. Most commonly studies have segmented patients by type of injury, and American Spinal Injury Association[1] (ASIA) or Frankel impairment scores[2] prior to evaluation of ambulation potential. Predictors for ambulation include: ASIA score D or E at admission, age [3,4], ASIA motor scores for LEMS (Lower Extremity Motor Score, ASIA) greater than or equal to 10 by one month[5,6,7], greater than 2/5 MMT (Manual Motor Testing) quadriceps[8], pin sensation below level of injury[9,5], SEPs[10], Yale Score[4], Modified Barthel Index[11]. Patients with motor and sensory complete injuries at the time of admission rarely ambulate independently.

Information from these studies is useful for clinicians to estimate ambulation potential; however, no previous studies have attempted to build a predictive model which can be used for all types of spinal cord injury patients using information available at the time of admission. A number of studies have used neural networks to model medical and functional outcomes following catastrophic injury[12,13,14,15]. These studies did not compare model performance with other methods such as logistic regression. The purpose of this project was to compare various outcome models that predict ambulation at discharge from rehabilitation for all individuals with acute spinal cord injury.

Outcome events are often related to multiple factors that are difficult to adequately model using linear discriminant analysis or other linear models. Logistic regression is a commonly used modeling technique; however, there are alternative methods that may result in better classification performance or simpler implementation in clinical settings. It is unclear which techniques are most valuable for a given problem. Experimental comparison using rigorous evaluation methods is necessary. We created models using logistic regression, backpropagation neural networks, and rough sets to predict ambulation following spinal cord injury.

METHODS AND MATERIALS

Methods

The study population consisted of 17,861 patients who sustained a spinal cord injury between 1973 and 1997, who survived at least 24 hours after injury, and were admitted to one of 24 federally funded designated regional SCI care systems. The data was collected at the time of admission to these centers and was aggregated into the Spinal Cord Injury Model System Database (SCIMS) [16]. The database is carefully managed for quality control and has a well defined data dictionary.

Materials/Model Design

The SCIMS database contains many medical,

1091—8280/98/$5.00 © 1998 AMIA, Inc.

neurologic, functional, and demographic variables. In this study, items which would be available at the time of admission to the hospital were used to construct the outcome prediction models. Input variables included those shown to be predictive of functional outcome in prior studies and items chosen by spinal cord injury experts involved with the project. These items included system days (*days from date of admission minus start of database 1/1/73*) [1], injury to system admission, age, gender, racial/ethnic group, traumatic etiology, marital status, level of education, primary occupation, category of neurologic impairment (e.g. complete tetraplegia), level of preserved neurologic function, ASIA impairment scale score, presence of fractures, presence of hemo/pneumothorax, use of mechanical ventilation. A data set which excluded records from the original study population with missing or unknown data was created. This data set contained 5626 records and was used to construct the models. Before building models we randomly divided the data into two sets; one set (n= 3772) was used as a training set and the other data set (n=1854) was used as a test set. Table 1 shows that there were not significant differences between training and test sets. All models used identical randomized test and training sets.

Table 1	Training set	Test set
# patients	3772	1854
Mean age (yrs)	32.8	32.7
Males (%)	80.4	82.7
Race (% white)	66.0	66.3
Race (% black)	30.7	30.7
Marital (% single)	54.4	55.4
Marital (% married)	28.6	29.8
Education (% HS)	49.8	51.3
Education (% bachelor)	6.9	5.9
Occupation (% working)	54.7	55.2
Occupation (% retired)	5.5	5.6
Occupation (% student)	16.4	16.1
Causes (% auto accident)	33.1	31.3
Causes (% fall)	19.7	21.4
ASIA (% A)	49.3	50.4
ASIA (% D)	18.3	19.6
Neuro impair (% tetra, complete)	28.0	28.9
Neuro impair (% tetra, incomplete)	29.7	29.9
Neuro impair (% para, complete)	21.3	21.6
Neuro impair (% para, incomplete)	19.8	18.4
Neuro level (C1-C8) right/left	51.2/51.0	51.8/51.6
Neuro level (T1-T12) right/left	36.1/35.9	36.9/37.3
Neuro level (LS levels) right/left	12.7/13.0	11.2/11.1
Pneumo/hemothorax (%)	18.2	19.5
Ventilator (% dependent)	10.3	9.4
Ventilator (% brief)	9.3	10.3
Fracture (% 1 or more)	28.4	26.5
Average days injury-admit	11.2	10.3
Ambulates at discharge (%)	19.4	21.4

[1] System days was included to provide a measure of how recently a patient was admitted to the system, i.e. how "modern" was the care provided. This is important because there have been numerous improvements in critical and trauma care during the period 1973-1997.

Neural networks and logistic regression models require either continuous or binary data; therefore, variables which contained categorical data types were converted to dummy binary variables. For example, if a database field contained 3 different categories (e.g. the race database field might contain white, black, or asian categories) it was converted to 3 binary fields of the form 1=true and 0=false (e.g. black =1 means black is true.).

We built a 3-layered feedforward neural network with 60 input units, 20 hidden units, and 1 binary output unit which indicated ambulation at discharge from the hospital using NevProp3 neural network software[17]. The network was trained by backpropagation[18], with the goal of minimizing an error function that corresponded to the cross-entropy (or maximizing the log likelihood). We further divided the training set into 2, and used the first set for the actual training (n=1386) and the second (holdout set n=1386) for monitoring overfitting. Initial weights were randomly assigned in each of 10 runs of the same network. Areas under the Receiver Operator Characteristic (ROC) curve were measured and averaged for the 10 runs.

The logistic regression model was developed using SAS software with the same data sets. No automated variable selection method was chosen. A SAS macro was created to apply the logistic equation resulting from the training set to the test set.

The rough set model was implemented using the knowledge discovery tool ROSETTA. Details of this modeling tool can be found in Ohrn et al 1998[19]. Rough set theory is a soft-computing technique that uses Boolean reasoning to classify imprecise, uncertain or incomplete data. For each possible subset of variables the data set gives rise to an equivalence relation called an indiscernibility relation, where two objects are in this relation (i.e. are members of the same equivalence class) if and only if they cannot be discerned from each other on the basis of the attribute subset. This relation can be used to approximate sets. Subsets of interest to approximate in a supervised learning setting would typically be the sets of objects with the same values for the outcome variable (the decision classes). For a detailed discussion of rough set theory, see Pawlak 1991[20].

RESULTS

Each model had variable sensitivity, specificity, and accuracy over a range of thresholds from 0 to 1. Depending on the utilities assigned to (1) missing patients who will ambulate (false negatives) and (2) falsely predicting ambulation (false positives), a threshold to optimize model performance can be chosen. Table 2 summarizes the sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), and accuracy using a threshold which minimizes the sum of $(1\text{-sensitivity})^2$ and $(1\text{-specificity})^2$. This threshold determines the point in the receiver operating characteristic (ROC) curve that is closest to (0,1). The logistic regression, neural network, and rough set models had an area under the ROC curve in the 0.88-0.91 range, (see Table 3 and Figure 1). The performance of neural network and logistic regression methods was not statistically different, $p = 0.48$ (using the method proposed by Hanley & McNeil)[21]. The rough sets method performed statistically worse than the neural network and logistic regression models ($p = 0.002$ and 0.015 respectively).

Table 2: Model Performance at Optimal Threshold*

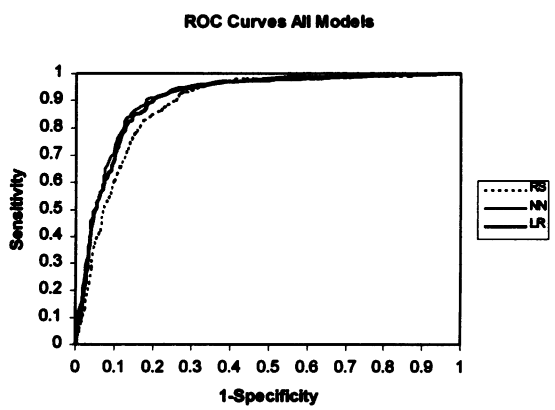
	Logistic Regression	Neural Networks	Rough Sets
Sensitivity	0.849	0.861	0.816
Specificity	0.849	0.852	0.829
NPV	0.954	0.958	0.943
PPV	0.606	0.614	0.566
Accuracy	0.849	0.854	0.827

* threshold that corresponds to point in ROC curve closest to (0,1)

Table 3:

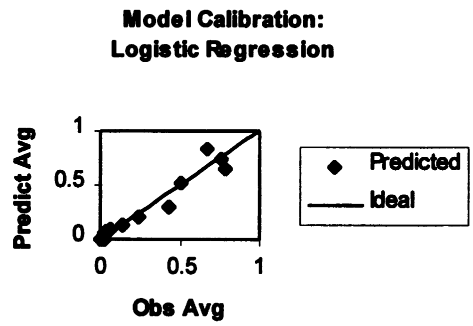
Model	ROC Curve Area	Standard Error
Logistic Regression	0.909	.008
Neural Network	0.909	.008
Rough Set	0.889	.008

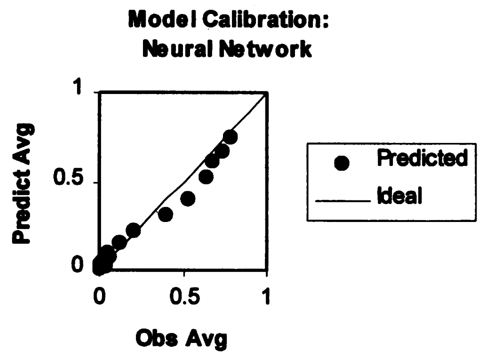
Figure 1

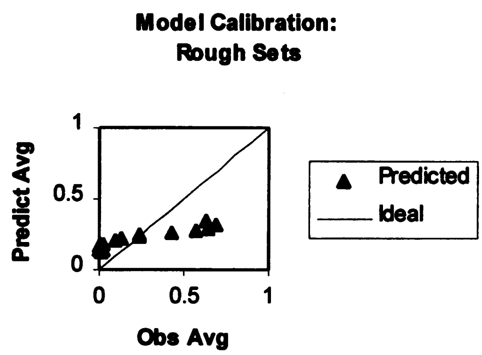


All of the models produce a prediction in the interval 0 and 1 for each case. Calibration curves were generated by comparing weighted average values of predicted versus actual outcome values, sorted by predicted values, and are shown in Figure 2. In a well calibrated model, the predicted value is equivalent to the probability of the outcome for each patient. The neural network and logistic regression models were relatively well calibrated, in contrast to the rough sets model.

Figure 2







which provide some explanation of the models prediction[22]. Neural networks do not provide methods to help explain the prediction.

DISCUSSION

All of the models performed well at predicting ambulation at discharge for spinal cord injured individuals (ROC curve areas ranging from 0.88-0.91). The logistic regression and neural network models performed statistically better than rough sets and provided well calibrated models. For well calibrated models, the raw predicted outcome is the "probability" of walking for a particular patient of interest. All models had sensitivity, specificity, and accuracy greater than 80% at ideal thresholds.

It is important, however, to note that this study has several limitations. ASIA scoring was modified slightly in 1992. Some patients injured prior to 1992 and categorized as an ASIA D (incomplete injury, motor strength useful) would now be categorized as ASIA C (incomplete injury, motor strength non-useful) using the revised ASIA system. Using patients injured after 1992 would eliminate this potential problem, but would also eliminate over 3000 patients from the models. We minimized this problem by including the *system days* variable, which adjusts for changes to the model systems over time. The SCIMS database evolved over several years and included more data fields in recent years. One variable of particular value for predicting ambulation is the ASIA motor index score[5,6,7] at admission. This variable was included starting in 1986 and is available for 5855 of 17861 patient records in the database. We excluded records with missing or unknown data. Other methods, such as substitution of the mean or mode values, could have been used to prepare data sets for modeling; however, we chose a more conservative data exclusion method, given the uncertainties involved in substitution methods. We acknowledge that this elimination may have introduced certain biases and that a prospective study should be performed to validate our results.

In addition to prediction of outcome, the logistic regression and rough sets models provide methods to help explain the prediction. Logistic regression yields a regression equation with coefficients for each significantly associated covariate. This regression equation allows one to make inferences regarding variable contribution to the model. Rough sets provides a set of conditional rules or statements which provide some explanation of the models prediction[22]. Neural networks do not provide methods to help explain the prediction.

Prior studies that predicted ambulation selected subsets of spinal cord injured individuals by neurologic categories, such as incomplete quadriplegia, incomplete paraplegia, or ASIA score. These studies included a smaller number of variables than our analysis and may have excluded important input variables. To our knowledge, this is the first study that included all neurologic categories of spinal cord injury for prediction of ambulation. We feel that this approach yields models that may be more useful to clinicians, since a single prediction model can be used for all spinal cord injured individuals, rather than multiple neurologic category subset models. Well calibrated models provide predictions which are analogous to probabilities.

CONCLUSIONS

The intent of this study was to demonstrate a comparison of model performance and not to provide a detailed analysis of each model. All models performed well at prediction of ambulation at hospital discharge for spinal cord injured individuals; however, the logistic regression and neural network models performed significantly better than the rough sets model. We concluded that the logistic regression model is the method of choice to predict ambulation in this data set, given its good classificatory performance, calibration, and potential insight into variable relevance. Its popularity among health care researchers is also a major advantage. This study is the first to provide and compare models which predict ambulation across all neurologic categories of spinal cord injury. For such models to be useful to clinicians, it is still necessary to minimize the number of input variables. We are currently investigating whether such models can be constructed.

Acknowledgements

We would like to thank William McKinley, M.D. and the Spinal Cord Injury Model System for their assistance with acquiring access to the multi-center database. This study was supported in part by grant no. H133N50015 from the National Institute on Disability and Rehabilitation Research, Department of Education, Washington, DC and grants no. R55LM/OD6538-01 and no.2 T15 LM07092 from the National Library of Medicine, Washington, DC.

References

1. American Spinal Injury Association: Standards for neurological and functional classification of spinal cord injury, revised 1992. 1992; Chicago: American Spinal Injury Association.

2. American Spinal Injury Association: Standards for neurological classification of spinal cord injury, revised 1982. 1982; Chicago: American Spinal Injury Association.

3. Burns SP, Ditunno JF Jr, Graziani V, Rolle WA Jr, and Golding DG. Recovery of ambulation in motor-incomplete tetraplegia. Arch Phys Med Rehabil 1997;78(11):1169-72.

4. Daverat P, Sibrac MC, Dartigues JF, Mazaux JM, Marit E, Debelleix X, Barat M. Early prognostic factors for walking in spinal cord injuries. Paraplegia 1988;26(4):255-61.

5. Waters RL, Adkins RH, Yakura JS, Sie I. Motor and sensory recovery following incomplete tetraplegia. Arch Phys Med Rehabil 1994;75(3):306-11.

6. Waters RL, Adkins R, Yakura J, Vigil D. Prediction of ambulatory performance based on motor scores derived from standards of the American Spinal Injury Association. Arch Phys Med Rehabil 1994;75(7):756-60.

7. Waters RL, Adkins RH, Yakura JS, Sie I. Motor and sensory recovery following incomplete paraplegia. Arch Phys Med Rehabil 1994;75(1):67-72.

8. Crozier KS, Cheng LL, Graziani V, Zorn G, Herbison G, Ditunno JF, Jr. Spinal cord injury: prognosis for ambulation based on quadriceps recovery. Paraplegia 1992;30(11):762-7.

9. Crozier KS, Graziani V, Ditunno JF, Jr., Herbison GJ. Spinal cord injury: prognosis for ambulation based on sensory examination in patients who are initially motor complete. Arch Phys Med Rehabil 1991;72(2):119-21.

10. Jacobs SR, Yeaney NK, Herbison GJ, Ditunno JF, Jr. Future ambulation prognosis as predicted by somatosensory evoked potentials in motor complete and incomplete quadriplegia [see comments]. Arch Phys Med Rehabil 1995;76(7):635-41.

11. Lazar RB, Yarkony GM, Ortolano D, Heinemann AW, Perlow E, Lovell L, Meyer PR. Prediction of functional outcome by motor capability after spinal cord injury. Arch Phys Med Rehabil 1989;70(12):819-22.

12. Ebell MH. Artificial neural networks for predicting failure to survive following in-hospital cardiopulmonary resuscitation. J Fam Pract 1993;36(3):297-303.

13. Grigsby J, Kooken R, Hershberger J. Simulated neural networks to predict outcomes, costs, and length of stay among orthopedic rehabilitation patients. Arch Phys Med Rehabil 1994;75(10):1077-81.

14. Oczkowski WJ, Barreca S. Neural network modeling accurately predicts the functional outcome of stroke survivors with moderate disabilities. Arch Phys Med Rehabil 1997;78(4):340-5.

15. McGonigal M, Cole J, and Schwab W. A new approach to probability of survival scoring for trauma quality assurance. The Journal of Trauma 1993;34(6):863-8.

16. Stover SL, Delisa JA, Whiteneck GG, editors.Spinal Cord Injury: Clinical Outcomes from the Model Systems. Gaithersburg, MD: Aspen Publishers; 1995

17. NevProp3 [computer program]. Goodman P. 3. 1997; ftp://ftp.scs.unr.edu/pub/cbmr/nevpropdir/.

18. Rumelhart DE, Hinton GE, Williams RJ. Rumelhart DE, McClelland JL, editors.Parallel Distributed Processing. Cambridge, MA: MIT Press; 1986;learning internal representation by error propagation.

19. Ohrn A, Komorowski J, Skowron A, et al. Skowron A, Polkowski L, editors.Rough Sets in Knowledge Discovery. Physica Verlag; 1998;The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets -- The ROSETTA System.

20. Pawlak Z. Rough Sets - Theoretical Aspects of Reasoning about Data. Dordrecht: Kluwer Academic Publishers; 1991.

21. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148(3):839-43.

22. Ohrn A, Ohno-Machado L, and Rowland T. Building Manageable Rough Classifiers. Submitted to AMIA 1998 Annual Symposium, Orlando, FL, Nov 7-11 1998