Comparison of Logistic Regression and Neural Network Analysis
Applied to Predicting Living Setting after Hip Fracture

KENNETH J. OTTENBACHER, PhD, RICHARD T. LINN, PhD, PAMELA M. SMITH, DNS, RN,
SANDRA B. ILLIG, MS, RN, MELODEE MANCUSO, PhD, AND CARL V. GRANGER, MD

PURPOSE: Describe and compare the characteristics of artificial neural networks and logistic regression to develop prediction models in epidemiological research.

METHODS: The sample included 3708 persons with hip fracture from 46 different states included in the Uniform Data System for Medical Rehabilitation. Mean age was 75.5 years (sd = 14.2), 73.7% of patients were female, and 82% were non-Hispanic white. Average length of stay was 17.0 days (sd = 10.6). The primary outcome measure was living setting (at home vs. not at home) at 80 to 180 days after discharge.

RESULTS: Statistically significant variables (p < .05) in the logistic model included follow-up therapy, sphincter control, self-care ability, marital status, age, and length of stay. Areas under the receiver operating characteristic curves were 0.67 for logistic regression and 0.73 for neural network analysis. Calibration curves indicated a slightly better fit for the neural network model.

CONCLUSIONS: Follow-up therapy and independent bowel and/or bladder function were strong predictors of living at home up to 6 months after hospitalization for hip fracture. No practical differences were found between the predictive ability of logistic regression and neural network analysis in this sample.

Ann Epidemiol 2004;14:551–559. © 2004 Elsevier Inc. All rights reserved.

KEY WORDS: Rehabilitation Outcomes, Joint Replacement.

INTRODUCTION

Recently, artificial neural networks (ANN) have been proposed as an alternative to modeling complex non-linear relationships in health care research (1–3). Several articles have appeared comparing the predictive ability of neural networks to multivariable logistic regression (4–6). These studies have not indicated the clear superiority of either approach. The benefits of a particular method appear to depend on the research question, the nature and number of variables involved, and several other complex factors (7). Additional research is needed to determine when a specific approach is best for a given question. In this article, we present introductory information on neural networks and compare the ability of neural networks and logistic regression to address prediction questions in epidemiological research. An illustration is provided comparing neural networks and logistic regression to predict living setting (home vs. not home) following hip fracture.

Logistic Regression

Readers of the Journal will be familiar with multivariable logistic regression and we present only a brief overview. The logistic regression model (8, 9) relates y to x in assuming that

$$Pr(y = 1 \mid x) = \Lambda(\beta_0 + \sum \beta_i x_i)$$  \hspace{1cm} (1)$$

where $$\Lambda(\mu) = 1/1 + \exp(-\mu)$$ denotes the logistic function. The unknown regression coefficients $$\beta_i$$, which have to be estimated from the data, are directly interpretable as log-odds ratios, or in terms of $$\exp(\beta_i)$$, as odds ratios. The assumption is the predictor variables are related in a linear manner to the log odds [$\log(p/(1 – p))]$ for the outcome of interest. Variables are often selected for inclusion in logistic regression models using some form of the backward or forward stepwise regression technique (9). Widely accepted criteria such as the Hosmer-Lemeshow statistic have been developed for assessing the goodness-of-fit for logistic regression.
models (8). Logistic regression is a widely used statistical method in epidemiological research (10–14).

**Artificial Neural Networks**

Neural networks are computer-based, self-adaptive models and have been used in the medical literature to analyze complex non-linear relationships (1, 15). Neural networks consist of a set of processing units (nodes) that simulate neurons. These nodes are interconnected to other nodes by a set of weights that are analogous to synaptic connections. The connections allow signals to travel through the network in parallel as well as serially. The synaptic weight is interpreted as the strength of the connection across nodes. The nodes are simple computing elements based on the model of the neuron that generates an action potential when a certain stimulus level is achieved. This has been modeled mathematically as a weighted sum of all incoming signals to a node, which is compared with a threshold. When the threshold stimulus is exceeded, the node fires; otherwise, it remains quiet. The relationship is presented in Fig. 1.

The brain learns by adapting the strength of the synaptic connections. Likewise, the synaptic weights in neural networks, similar to coefficients in regression models, are adjusted to solve problems presented to the network (15). Learning or training is the term used to describe the process of finding the values of these weights. Many different training algorithms have been developed for establishing neural network weights. Backpropagation is the most commonly used training method in developing prediction models (1, 15).

Training a neural network is achieved by using a training data set consisting of predictor variables and the known outcomes. Networks are programmed to adjust their internal weights based on the mathematical relationships identified between the inputs and outputs in a data set (1, 15). Once a network has been trained, it can be used for prediction or classification tasks in a separate test (or validation) data set.

Neural networks are constructed with layers of nodes. A feedforward network is one where the nodes in one layer are connected only to nodes in the next layer and not to nodes in a preceding layer or units in the same layer. Figure 2 shows a multi layered feedforward network. The first layer of the network consists of the input nodes. These nodes correspond to the independent variables in regression models. The last layer contains the output node, or dependent variable in a regression model. All other nodes in the neural network are called “hidden” nodes. This hidden layer has no correspondent in regression models, although it may be considered roughly as a layer that represents interactions among independent variables (these interactions are not clearly interpretable, however). This extra layer of hidden nodes adds flexibility to neural network models and allows them to model complex mathematical functions. As opposed to the logistic regression model, a neural network may have several output units. An equivalent model would be a set of logistic regression models, one for each dependent variable. Most applications reported in health care research, however, include a single output node (2).

Each input node is connected to each hidden node by a connection weight. Figure 2 includes illustrative nodes and weights. At each hidden node, a weighted linear combination of the inputs is summed (including the bias weight that is analogous to the intercept in regression) to determine the net input to that node. This result is then passed through an activation function, most commonly the logistic or sigmoid function. A logistic transformation of the weighted inputs to the output node is applied to determine the overall output of the network. The output of the network (prediction) will range from 0 to 1, if the logistic transformation is used.

The weights are determined iteratively, with the goal of minimizing error for all cases. The connection weights in an untrained neural network are initially random values. Estimating the optimal values of these connection weights is the purpose behind training the neural network. The network

![Figure 1](image1.png)  **FIGURE 1.** Example of neuron and neural network connections. Adapted from Ohno-Machano and Rowland (31).

![Figure 2](image2.png)  **FIGURE 2.** Example of multilayer neural network with input, hidden and output layers (nodes).
Multilayered neural networks with several hidden units may be able to reach 100% accuracy, by over training or "memorizing" all the cases. The challenge is to train a network to recognize patterns but not overfit them. There are numerous ways to prevent overfitting, but some come at the cost of "losing" information. Cross validation is one method and is described in the example presented below. Linear and logistic regression models have less potential for overfitting because the range of functions that they can model is more limited.

An illustration of the concepts described above is provided in the next section. In this example, we compared logistic regression and neural networks using living setting (home vs. not home) following rehabilitation for hip fracture as the outcome variable. We selected living setting for the following reasons: 1) living at home vs. not living at home represents a dichotomous variable with clear cost and social implications (18,19); 2) living setting following hospital discharge is a complex phenomenon influenced by multiple variables that are poorly understood (20); and 3) we had access to a large representative national database with information on a unique combination of sociodemographic and functional performance variables for patients receiving medical rehabilitation services (21).

METHODS

Source of Data

Data were analyzed from 186 hospitals in 46 states contributing to the Uniform Data System for Medical Rehabilitation (UDSMR™) and the National Follow-up Services (NFS). The UDSMR is a data collection and analysis service associated with the State University of New York at Buffalo (21). At the time of the study, more than 850 rehabilitation centers used the UDSMR data service, representing approximately 80% of all comprehensive medical rehabilitation facilities in the United States. Detailed information on the UDSMR is available at: www.udsmr.org.

The UDSMR collects demographic variables, diagnoses (ICD-9 codes), facility characteristics, patient pre-hospital living arrangements, marital status, pre-disability employment status, length of stay (LOS), source of payment, hospital charges, follow-up therapy, and information on discharge setting (21, 22). Performance on a standardized measure of basic daily living skills, the Functional Independence Measure (FIM instrument™) is also included in the database (21). The FIM instrument is an 18-item assessment measuring: self-care, sphincter control, transfers, mobility, communication, and social cognition in adults. Extensive research has demonstrated the validity, reliability, and responsiveness of the FIM instrument in assessing motor and cognitive

training algorithm is used to gradually adjust the weights in the network to minimize the difference between the predicted output of the network and the known value of the outcome variable. This difference is referred to as the error of a neural network and is similar to the concept of minimizing the residuals in regression. The backpropagation algorithm is a popular method to find weights for multilayer feedforward networks and to minimize error. The development of the backpropagation algorithm is primarily responsible for the surge of interest in neural networks (3). The exact mathematical equations for the backpropagation algorithm are complex and available elsewhere (3, 16, 17). The basic goal is to optimize the connection weights. The most common performance metric used in neural networks is the sum of squared errors defined as:

\[ E = \frac{1}{2} \sum_{p=1}^{n} \sum_{k=1}^{0} (y_{pk} - \hat{y}_{pk})^2 \]  

where \( p \) refers to the patterns (observations) within a total of \( n \) patterns, the subscript \( k \) to the output unit with a total of \( 0 \) output units, \( y \) is the observed response, and \( \hat{y} \) is the model (predicted) response. This is a sum of the squared difference between the predicted response and the observed response averaged over all output and observations (patterns). In the case of predicting a single outcome, \( k = 1 \), the equation reduces to:

\[ E = \frac{1}{2} \sum_{p=1}^{n} (y_{pk} - \hat{y}_{pk})^2 \]  

“Training” a neural network is equivalent to fitting a regression model. It is performed by iteratively modifying weights such that outputs produced by the network will have a minimal amount of error. Typically, initial small random weights are updated gradually. The training rule is simple: Whenever the network’s output is not close enough to the desired output, a change in weights occurs in the direction that minimizes the error. The change is proportional to the difference between the network’s output and the desired output, that is, the error is a function of the difference between the result that the network produced and the result that it should have produced. Although the same principle applies to regression models, in the latter case another “training” algorithm is used. In the case of a simple logistic regression model, there is only one logistic function being used, and only parameters referring to this single function have to be estimated. In the neural network case, there may be several logistic functions at intermediate and output nodes, and the parameters for all of them are being estimated. This makes “training” in neural networks more difficult.
functions associated with basic daily living skills (23–26). Using the UDSMR protocol, the FIM instrument is administered to persons receiving inpatient medical rehabilitation within 72 hours of both admission and discharge. In this investigation, we used FIM instrument ratings obtained at discharge (see additional description below). Follow-up data was collected by telephone interview 80 to 180 days after discharge by the National Follow-up Services and aggregated with the complete UDSMR patient record. Detailed information regarding psychometric properties and validity of the UDSMR data has been reported by several independent researchers (25–28) including the Centers for Medicare and Medicaid Services (CMS) (formerly Health Care Financing Administration or HCFA) (27).

Study Population

Complete admission, discharge, and follow-up information was available for 4122 patients with hip fracture who received inpatient rehabilitation services from 1996 through 1998. Patient records with ICD-9 codes for hip fracture were eligible for analysis. We excluded records with missing or out of range data values (n = 221) and cases with nonspecific impairment codes (n = 91). Non-specific impairment codes occurred when a facility selected a UDSMR major impairment group code (based on ICD-9 codes) that contained insufficient information to classify the patient’s condition. In addition, we excluded patients who were younger than 55 years of age (n = 157). The remaining 3708 patients comprised the study sample, and represented 89% of the usable patient records from the original sample.

Study Variables

Living setting. The UDSMR definition of home is a private, community-based dwelling (house, apartment, or mobile home) that is the home of the subject, his/her family, or friends (21). The remaining settings are: board and care, transitional living, intermediate care, skilled nursing, acute hospital unit, chronic hospital, rehabilitation hospital, subacute setting, assisted living residence, and other alternate levels of care. Living setting at follow-up was dichotomized into: home vs. not at home. This decision was based on estimated cost and use of resources. We assumed that costs and use of other resources would be higher for persons not living in their home setting.

Length of stay (LOS). LOS was calculated as the total number of medical rehabilitation days.

Ethnicity. Ethnicity/race was recorded from the medical record as: non-Hispanic white, African American (black), Hispanic, Asian, Native American, and other. This information was based on patient self-report at admission.

Marital status. Marital status was recorded as married, single (never married), widowed, separated, or divorced.

Information on marital status was based on self-report at admission.

Primary payer source. The UDSMR data set includes 14 categories for primary payer source (21). For this investigation the categories were collapsed to four: Medicare, Medicaid, commercial insurance, or health maintenance organization (HMO).

Follow-up therapy. At follow-up interview, information was collected on whether the patients received follow-up therapy after discharge. The categories of response included: outpatient therapy, home-based paid professional therapy, and both outpatient and home-based therapy. Information on intensity or duration of follow-up therapy is not collected. We dichotomized this variable into those receiving follow-up therapy vs. not receiving follow-up therapy.

FIM Instrument. The FIM item is 18 items were divided into 6 subcales: self-care, sphincter control, mobility, locomotion, communication, and social cognition (21). Scoring decisions are based on whether or not help is needed for an activity to be completed and if so, how much help is required. A score must be recorded for each item. The lowest possible score is 1 (most dependent) and the highest possible score is 7 (most independent). Total FIM scores range from 18 to 126. Reliability of the FIM instrument has been established by independent investigators (23–25).

Statistical Analysis

Logistic regression analysis was used to predict follow-up living setting. The dependent (criterion) variable was the dichotomized living setting obtained at follow-up (home vs. not home). The independent variables entered into the logistic regression equation were age, gender, ethnicity, marital status, LOS, primary payer source, follow-up therapy, and FIM instrument ratings for the six subcales. Variables that were not continuous were dummy coded. Race was coded as white vs. non-white. Marital status was coded as married vs. not married (single, widowed, separated, or divorced).

We specified a full factorial model that included all two-way interaction terms (13). The analysis generated Wald-statistics, regression coefficients, standard errors, confidence intervals, Nagelkerke $R^2$, Hosmer-Lemeshow goodness-of-fit chi-square, and predicted group membership. The Nagelkerke $R^2$ attempts to quantify the proportion of explained variance in the logistic regression model, similar to the $R^2$ in linear regression, although the variation in a logistic regression model must be defined differently (29). Logistic regression models were also computed to examine the independent association of demographic and FIM instrument scores on living setting. In these models most independent variables were dichotomized and odds ratios with 95% confidence intervals reported.
The neural network analysis was conducted using Neural Connection software (Version 2.1) (30). We built a three-layered feedforward neural network with 13 input nodes, 4 hidden nodes, and 1 binary output node indicating living at home vs. not at home. The variables included in the input layer were the same variables entered in the logistic regression model. The number of nodes in the hidden layer was determined as a best fit by the software. Training and testing were done using randomly selected subsets of the data (see description below). During the training, connection weights were adjusted by use of the Langevin extension of the back propagation updating algorithm (15). The learning rate \( \eta \) had a start value of 0.5. Eta was decreased geometrically every epoch (training cycle) using the following equation: \( \eta = k(\eta) \), with \( k = 0.998 \). The momentum \( (\alpha) \) was set to 0.7. Updating occurred after every 100th pattern. The network weights were initiated with random numbers between −0.025 and 0.025. A stopping criterion was established to determine when to terminate the training process to achieve optimum performance and to avoid overtraining. This criterion was calculated by use of a threefold cross-validation procedure described below.

The data set \( (n = 3708) \) was randomly divided into three equal parts. One part was used as a test set, and training was performed on the remaining two parts. This procedure was repeated three times so that each part was used once in a test set. Each time all variables in the training set were presented to the network, the performance was evaluated with respect to the error obtained in the training and test sets. This evaluation did not alter the connection weights. The error in the training set decreased with an increased number of training cycles, whereas the error in the test set reached a minimum, after which it increased despite the further decrease in training error. The error in the training set, which corresponds to the minimum error in the test set, was assessed. The mean of the three training errors calculated in the threefold cross validation procedure was defined as the stopping criterion in the training procedure (1, 30). Following the suggestion of Ohno-Machado and Rowland (31), the mode for missing information was calculated using the 1996–1998 data. Missing data were then assigned the appropriate mode for each variable.

Discrimination and calibration (goodness-of-fit) were measured. The discriminatory power of the logistic regression and neural network (ability to classify persons living at home vs. those not living at home) was analyzed using the area under the receiver operating characteristic (ROC) curves. Calibration of the models (how accurately the models predicted over the entire range) was measured both by construction of calibration curves and computation of the Hosmer-Lemeshow goodness-of-fit chi-square statistic. Calibration curves were constructed using two methods. First, calibration curves were generated using observed and expected numbers of persons not living at home. The actual and expected number not living at home was plotted on a linear axis. Standard linear regression was then performed and correlation coefficients calculated. Second, the output data from the neural network and logistic regression were rank ordered, then divided into deciles, and the average “not living at home” rate for each decile was calculated. The actual and calculated average not living at home rate was plotted on a linear axis. The calibration curves based on the average for each decile are easier to interpret visually since they have a smaller number of data points, but the data points (average for each decile) are not independent.

The Hosmer-Lemeshow goodness-of-fit chi-square statistic was calculated as originally described (13). The predicted number of persons not living at home was again calculated for each decile (8 degrees of freedom). This number was then subtracted from the total number of patients in that decile giving the expected number of persons living at home. Chi-squared values were calculated as the squared differences between observed and expected values divided by the expected number for each decile. These were then summed for each decile giving the \( \chi^2 \) value (8 degrees of freedom).

### RESULTS

Descriptive and demographic information for the sample appear in Table 1. Average age and length of stay are represented as mean ± SD. There were approximately 48,204 data points for the neural network analysis (13 variables with 3708 cases) and approximately 384 missing data points (4%). The percent of missing data points across the 13 variables range from 0 to 6 percent. The percentage of patients not living at home was 18.7%. There was, therefore, significantly more living at home test and covariant patterns than not living at home patterns in the neural network analysis.

Logistic regression results. None of the interaction terms in the first logistic regression model were statistically significant. Based on the lack of significant interaction we ran a follow-up logistic regression model using a forward stepwise procedure. We used the likelihood-ratio (LR) test to enter variables into the model. This involved estimating the model with each variable eliminated in turn and looking at the change in \(-2 \log\text{-likelihood}\) when each variable is deleted. Seven variables were entered into the training model before the forward stepwise procedure was terminated. Variables identified as significantly associated with living at home vs. not living at home in the logistic regression model were: follow-up therapy, sphincter control, self-care, marital status, age, and length of stay. The Nagelkerke \( R^2 \) for the logistic regression model was 0.19. The Hosmer-Lemeshow Chi-square was 15.0 \( (df = 8, p = .19) \). The logistic regression equation for the statistically significant
TABLE 1. Demographic characteristics and descriptive statistics for entire sample of hip fracture patients (N = 3,708)

<table>
<thead>
<tr>
<th>Variable</th>
<th>All patients</th>
<th>Home</th>
<th>Not home</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>3,708</td>
<td>3,015 (81.3%)</td>
<td>693 (18.7%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>976</td>
<td>794 (81.4%)</td>
<td>182 (18.6%)</td>
</tr>
<tr>
<td>Female</td>
<td>2,732</td>
<td>2,257 (82.6%)</td>
<td>475 (17.4%)</td>
</tr>
<tr>
<td>Age (sd)</td>
<td>75.52 (sd = 14.2)</td>
<td>74.73 (sd = 14.4)</td>
<td>81.32 (sd = 11.6)</td>
</tr>
<tr>
<td>Ethnicity*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>2,945</td>
<td>2,385 (81%)</td>
<td>560 (19.1%)</td>
</tr>
<tr>
<td>Non-white</td>
<td>647</td>
<td>550 (85%)</td>
<td>97 (15%)</td>
</tr>
<tr>
<td>Marital status**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1,379</td>
<td>1,172 (83%)</td>
<td>207 (17%)</td>
</tr>
<tr>
<td>Not married</td>
<td>2,157</td>
<td>1,726 (80%)</td>
<td>431 (20%)</td>
</tr>
<tr>
<td>LOS</td>
<td>17.0 (sd = 10.6)</td>
<td>16.5 (sd = 10.4)</td>
<td>20.4 (sd = 11.6)</td>
</tr>
<tr>
<td>FIM instrument</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-care</td>
<td>33.4 (sd = 6.3)</td>
<td>34.1 (sd = 5.7)</td>
<td>27.8 (sd = 7.5)</td>
</tr>
<tr>
<td>Sphincter control</td>
<td>11.7 (sd = 2.7)</td>
<td>12.0 (sd = 2.5)</td>
<td>9.6 (sd = 3.6)</td>
</tr>
<tr>
<td>Mobility (9–21)</td>
<td>7.3 (sd = 2.9)</td>
<td>7.5 (sd = 4.9)</td>
<td>5.5 (sd = 2.8)</td>
</tr>
<tr>
<td>Locomotion (5–14)</td>
<td>14.6 (sd = 5.6)</td>
<td>15.0 (sd = 3.3)</td>
<td>11.7 (sd = 4.0)</td>
</tr>
<tr>
<td>Communication</td>
<td>12.7 (sd = 2.2)</td>
<td>13.0 (sd = 1.9)</td>
<td>11.0 (sd = 3.1)</td>
</tr>
<tr>
<td>Social cognition</td>
<td>18.0 (sd = 5.9)</td>
<td>18.7 (sd = 3.5)</td>
<td>14.4 (sd = 5.1)</td>
</tr>
</tbody>
</table>

*Ethnicity/race missing for 116 patients.
**Marital status missing for 172 patients.

The association between sociodemographic variables, FIM scores, and living setting was examined in three logistic regression models. These models are presented in Table 2 and include odds ratios for variables added in each model. A ROC curve was also computed using the predicted probabilities for group membership from the logistic regression model. The area under the ROC curve based on the logistic regression model is 0.67 (SE = 0.11) (see Fig. 3).

Neural network results. The neural network was stable and training was uncomplicated [root mean square (RMS) < 0.01]. RMS < 0.01 indicates the neural network successfully completed all training and validation runs. The analysis produced a total of 89 weights in building the model. The following weights were associated with the connection of four hidden nodes to the output node: −0.731, 1.465, −0.982, and 0.799 (Bias = 0.625). The input variables (nodes), connection weights for the hidden nodes, and model. The area under the ROC curve based on the logistic regression model is 0.67 (SE = 0.11) (see Fig. 3).
output node are presented in Fig. 4. The Hosmer-Lemeshow chi-square for the neural network analysis was 10.1 (df = 8, p = .21). As noted above, two sets of calibration curves were computed—one based on observed and expected “not living at home” counts and one based on deciles. The linear regression values from the data calculated using observed and expected “not living at home” counts were similar for the logistic model (R = 0.71) and neural network model (R = 0.82). The calibration curves based on the deciles are presented in Fig. 5.

The neural network output is a single continuous variable with range 0 to 1. A threshold in this interval was used above which all values were regarded as consistent with not living at home. By varying this threshold, a ROC curve was obtained. The area under the ROC curve for the neural network analysis was 0.73 (SE = 0.11) (see Fig. 3). The performance of the neural network model and logistic regression models were not statistically different (p = 0.12) using the method proposed by Hanley and McNeil (32).

DISCUSSION

The purpose of this study was to present basic descriptive information on the use of neural networks and to compare logistic regression and neural network models in predicting living setting for patients with hip fracture 3 to 6 months after discharge. Seven statistically significant predictor variables were identified in the logistic regression analysis. The logistic regression models suggest that the odds of not living at home are greater for older patients, for patients who did not receive follow-up therapy, and for patients with impaired bowel or bladder function, or deficits in self-care (e.g., dressing, feeding). Other variables associated with not living at home at follow-up included marital status, ethnicity, and LOS.

Both logistic regression and the neural network models did an excellent job of predicting persons who were living at home. This is inherent in the distribution of covariant patterns, i.e., there were more persons living at home than not living at home in the follow-up period. If a model predicted living at home for every test cohort in our data set it would be correct approximately 81% of the time. The distribution of cohorts is also important in the development of models. Because there are many more covariant patterns for persons living at home, the neural network model will train very well in detecting persons living at home. The challenge is developing a model to correctly classify persons who will not be living at home at follow-up. In the calibration analysis, logistic regression produced a larger number of persons not living at home in its lowest decile, i.e., these

FIGURE 4. Neural network model showing input variables (nodes), hidden nodes, and connection weights with output node for data on living setting following hip fracture (N = 3,708).
cases were more serious misclassifications since the probability of not living at home was smallest in the lower deciles. The neural network also produced a set of these misclassified cases, but less than the logistic regression approach.

The comparison methodology used in this study constrained the neural network analysis by limiting the number of potential predictor variables to the same set of predictor variables used in the logistic regression analysis. One of the advantages of neural network analysis is that it allows the inclusion of a large number of variables (1, 15).

In developing future models to predict complex phenomenon, such as living status following hospitalization, the advantages and limitations of the approaches must be carefully considered. Neural networks are able to model complex non-linear relationships between independent and dependent variables—this is a significant advantage (2, 31). Another advantage of the neural network approach is that there are not many assumptions to be verified before the models can be constructed and, as noted above, a large number of variables can be included in the analysis (15).

The most popular training algorithm for neural networks is back propagation (1). Although there are numerous variations of this algorithm, they are all based on an iterative search for adequate weights using a pre-determined error function. This search is not exhaustive, so it is possible that the optimal solution for a given problem may never be considered. Advocates of neural networks argue that finding a “good enough” solution is acceptable and better than conventional statistical approaches (1, 15).

A limitation of the neural network approach is that standardized coefficients and odds ratios corresponding to each independent variable cannot be easily calculated and presented as they are in logistic regression models. Weights are generated in a neural network analysis, but their interpretation is difficult and the weights may be influenced by the program used to generate them (15). This lack of interpretability at the level of individual variables (predictors) is one of the most criticized features in neural network models (3, 14). Advocates of the method argue that the existing trade-off between being able to model complex non-linear functions and interpretation of weights favors neural networks for applications where the primary goal is to obtain a reliable prediction rather than to gain an understanding of the contribution of individual variables (3, 14). In those cases where the relationship among predictor and criterion variables is known to be linear, or the goal of the analysis is explanation rather than prediction, conventional regression techniques are preferred.

Several early applications of neural networks in medicine reported an excellent fit of the model to a given set of data (2). The impressive results usually were derived from overfitted models, where too many free parameters were allowed (2, 15). As noted previously, the challenge is to train a network to recognize patterns but not overfit them. Linear and logistic regression models have less potential for overfitting because the range of functions they can model is more limited. Our experience in modeling living setting in persons with hip fracture suggests that using both approaches may be useful. Logistic regression helps the investigator identify those variables that may be good predictors of a particular outcome and helps to narrow the number of parameters included in the neural network analysis. Logistic regression also allows the investigator to place confidence limits around model outputs and parameter estimates after the underlying structure of relationships among variables is identified. Clinical predictive ability may be enhanced through the use of neural network analysis that is able to examine non-linear interactions among variables. We agree with Tu (33) that logistic regression remains the clear choice when the primary goal of model development is to examine possible causal relationships among variables, but that some form of hybrid technique incorporating features of both logistic regression and neural networks might lead to the development of optimum prediction models. For example, Duh et al. (34) developed and cross-validated prediction models for newly diagnosed cases of liver disorders by using a combination of logistic regression and neural networks. Hajmeer and Basheer (35) recently reported a hybrid Bayesian-neural network approach to modeling bacterial (e-coli) growth.
Developing a better understanding of the “hidden” layers in neural networks is essential to creating future methods that will incorporate the best features of neural networks and logistic regression. This research is progressing using relevance and sensitivity analysis to identify important predictor variables in neural networks (36). Continued examination of the advantages and limitations of neural networks compared with traditional multivariate statistics will eventually produce better analytical tools for use in epidemiological and health care research.

REFERENCES