Prediction of seven-year survival by artificial neural network and logistic regression: A comparison of results from medical and social data among 70-year-olds in Göteborg, Sweden.

Authors: Bo G Eriksson, Department of Sociology, University of Gothenburg
Valter Sundh, Department of Community Medicine and Public Health, University of Gothenburg

Abstract

Problem: Are there, for practical uses, any benefits of Artificial Neural Network analyses (ANN) compared with logistic regression analyses?

Data: A random sample of 2,294 70-year-old persons from Göteborg, Sweden, was investigated through interviews and medical examinations.

Methods: Seven-year mortality was studied by neural network analysis using the SPSS module Clementine 9.0 and by standard logistic regression analysis. The guiding problem was as follows: How do examples of ANN analyses perform compared with logistic regression models in the analyses of one set of social variables and one set of biological and health variables from the same individuals?

Result: The ANN generally did not perform better than logistic regression analyses, but in the data set with biological and health predictors, some ANN analyses produced much better results than logistic regression models when odds ratios were compared.

Discussion and Conclusion: The ANN can be used as a heuristic method to evaluate if there are hidden structures in data that are not revealed by regression methods and thus call for further analyses, either by more adequate regression models or by other methods. The ANN models could be used as predictors of outcomes of multifactor genesis, which has not been well understood using other investigation methods.

Keywords: Survival, logistic regression, artificial neural networks, random sample

Introduction

It is well known that longevity can be predicted in regression models and by using medical, social and lifestyle data [1]. It has also been discussed whether the predictions are equally good if either medical or social and lifestyle variables are used [2]. The commonly used analyses of survival are event history models and logistic regression models. Event history models use target survival as a continuous variable of lived time, while logistic models use a fixed survival length. The target is thus a binary variable—survived or not.

There are several other techniques for analysing survival, including artificial neural networks (ANN). A review article on the applicability of expert systems that included ANN concluded that these methods are problem oriented and useful in the social sciences [3]. ANN analyses have been used to study political attitudes related to educational and social status [4], social mobility and social networks [5], travel behaviour [6], and social capital [7]. ANN analyses are well suited to multidisciplinary studies, where data from different disciplines are assumed to have interaction effects on an outcome variable.

Comparisons of results from ANN analyses and more traditional regression analyses have provided divergent results. ANN has provided better results concerning cancer treatment [8-10] and survival [11, 12]. Similarly, ANN has presented equally better or poorer results in other studies on survival [13] and risk for occupational back injuries [14]. Reviews comparing ANN to regression models show the same mixed results [15, 16].

This brief sketch of a growing field of investigations is sufficient reason to conduct further analyses. This is a report on a larger investigation in which we hope to show the use of ANN analyses in sociological and multidisciplinary fields. The aim of this paper is to make two comparisons and discuss the prediction of survival. A comparison is first made between ANN and logistic regression. This is followed by a comparison of results concerning medical and social data. Our research question was as follows: Are there, for practical purposes, any benefits of ANN analyses compared with logistic regression analyses?

Five concerns in regression models

When using traditional statistical inference (e.g., regression models) in evaluating the ability of explanatory variables to predict a certain outcome, an uncertainty over the outcomes or findings always remains. There is always a possibility that if we use a more realistic, (i.e., more complicated) model, we could reach a better understanding of the association between the studied outcome and the available predictors. However, there are five good reasons why we almost always try to find a regression model that is as simple as possible: interpretability, non-linearity, interaction, multicollinearity, and overfitting.

Interpretability

To interpret complicated associations, we need to find the simplest model that gives an adequate fit of the observed data: the principle of parsimony. A model with few predictor variables and simple patterns of association is much easier to understand and report than a model with many predictors and/or with complicated association patterns. There are two main ways in which an association pattern can be complicated: non-linear association and interaction. These can combine to create non-linear interaction patterns.

Non-linearity

Non-linear association occurs when the effect of increasing exposure varies with different levels of exposure [17]. Examples are given in Figs. 1–3.

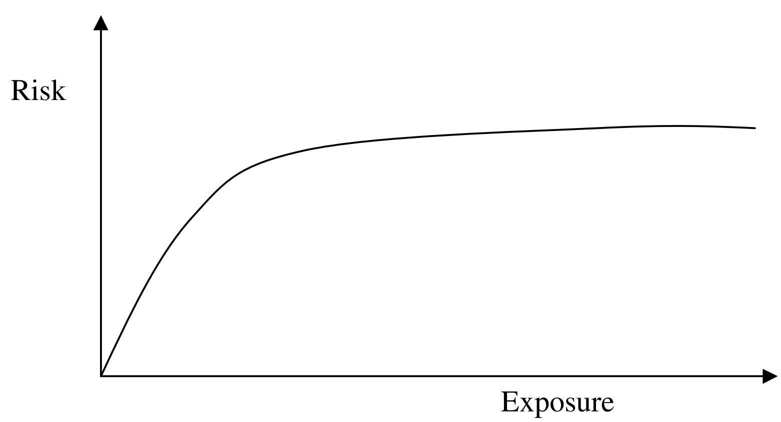

Figure 1. An example of non-linear association between exposure and risk

In figure 1 the change in risk when increasing the level of exposure by one unit on the scale is much larger at the low end of the scale than at the high end. It is likely that higher values of exposure above a certain exposure level will have no noticeable effect on the observed risk. Therefore, it is not realistic to describe this relation with a single parameter that would be assumed valid for all levels of exposure. We would thus be compelled to use a more complicated model to give an adequate description of the relationships that we study.

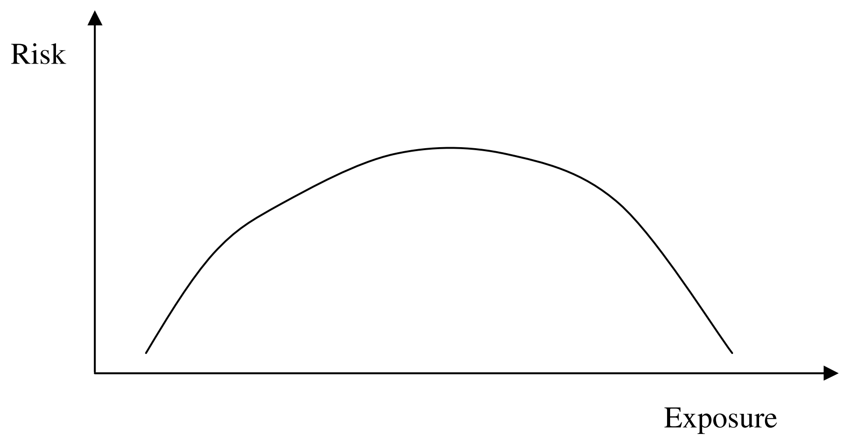

Figure 2. An example where both low and high levels of exposure are related to low risk

The easiest way to handle this situation of non-linear association is to add a quadratic term of the exposure to the model. To handle an association that is non-symmetric around the centre of the curve, a cubic model may be fitted to the data.

In slightly more complicated situations, but also more flexible, as in figure 3, modelling non-linear association can be done in spline models, where different models are applied to different intervals of the exposure variable.

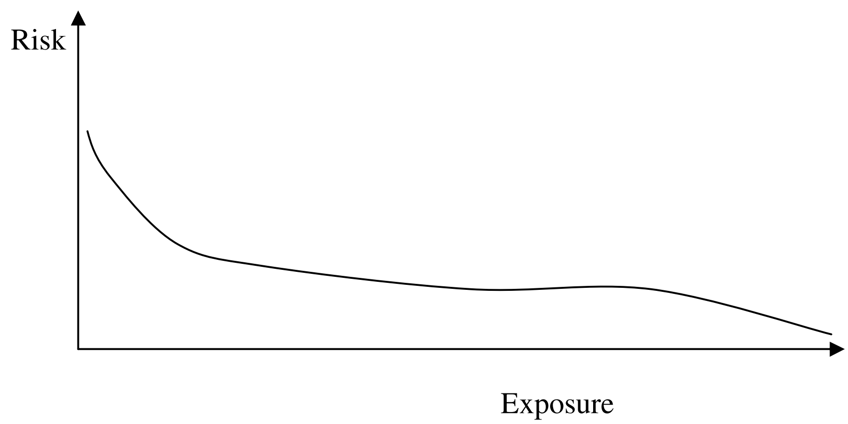

Figure 3. An example of a non-symmetric association between exposure and risk.

Interaction

Interaction means, in the simplest case, that the association between two variables is different on different levels of a third variable [17]. An example is given in Table 1.

In Table 1, we can see a significant difference in mortality between persons with low and high income for unmarried status, but a smaller difference in the opposite direction for married persons. When analysing this data it would be appropriate to include an interaction

Table 1. Proportion dead within seven years

Marital status	*Income> median*	*Income< median*	*Total*
Married	0.19643	0.15852	0.17370
Not married	0.14357	0.25597	0.18049
Total	0.16911	0.18375	0.17634

term for marital status and income level in any prediction model. More examples of interaction are shown in Figs. 4 and 5.

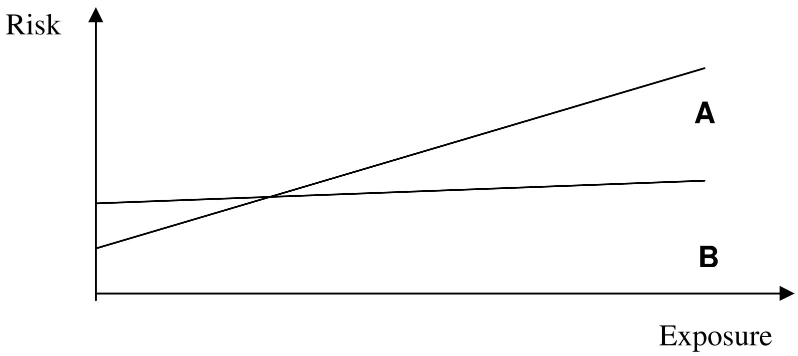

Figure 4. An example of linear interaction between exposure and risk for two groups A and B

In figure 4, the exposure is strongly associated with the outcome for group A, but weak for group B.

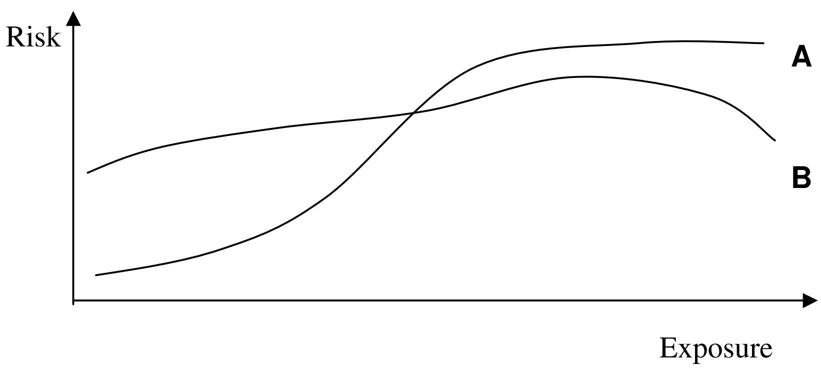

Figure 5. An example of non-linear interaction between risk and exposure for two groups A and B

In Fig. 5, the difference between the association between exposure and outcome cannot be described by one linear interaction term. In this case, it would probably be best to analyse group A and B separately and not try to reach a single conclusion that applies to all. More complicated interaction patterns involve more than three factors, and may be difficult to visualize concretely.

Multicollinearity

Multicollinearity [17] describes linear associations between the predictor variables in a model, with the consequence that the separate effect of each predictor on the outcome will be difficult or impossible to estimate with certainty. If the aim of the analysis is to evaluate the relative importance of a selected set of predictors, the presence of multicollinearity in the sample is a great disadvantage. However, if we only want to estimate the combined effect of the selected predictors, multicollinearity is not relevant. In population studies—our area of investigation—the aim of statistical analysis is to learn about how (and if) different predictors affect the studied outcome. Thus, in most cases it is necessary to guard against the effect of multicollinearity on the analysis.

Overfitting

Overfitting [18] is an effect of how regression models work—they find the best fit to the observed sample given the limitations of the models. The association patterns in the observed data, if we assume a perfectly random sample, depend on the association patterns in the population, and on random sample variation. Explaining the associations in our sample at an overly detailed level (i.e., adding too many parameters to the model) mainly results in explaining the random variation of the sample. There is no objective way to decide the level at which the results of an analysis will depend more on random sample variation than on true associations in the population. Because the level of detail in a model is dependent on the number of parameters (predictor variables), keeping the number of predictors as low as possible is advantageous.

Another aspect of overfitting is relevant when using regression models to predict the outcome for individual observations. To a certain extent, the parameters in regression models are influenced by random sample variation, where there will be an unrealistically high degree of agreement between the observed and model-predicted outcome when the model is applied to the same observations that are used to estimate the model. Therefore, it is necessary to protect against overfitting to obtain an unbiased estimate of the predictive power of a model. The established procedure is to use cross-validation (i.e., to split the observations into different subsets and calculate the model in one subset), apply the model to the rest of the observation, and then do the reverse.

Consequences of the five aspects of regression model building

Subscribing to the principle of parsimony, simple regression models are adapted because of interpretability, multicollinearity and overfitting—although we are well aware that it is almost always possible to find better fitting but also more complicated and unstable models. Often, there remains a nagging uncertainty that we may have missed potentially interesting findings in our data, which we may have found if we had had the time to delve deeper into the model searching process, (i.e., by testing additional or alternative non-linear and interaction effects). It would therefore be exciting if a possibility was given to examine thoroughly several standard regression models and compare them with methods that automatically perform a more exhaustive search for deeper structures in the data. This hopefully will confirm that in general, we can assume that when employing basic and relatively simple methods for statistical inference, we will at most times succeed in making essential discoveries about how things are connected in the real world. A valuable result would also be achieved if our results directed us as to how and when it might be worthwhile to consider adding additional terms to our regression models, to obtain a more complicated but also more realistic regression model.

Problem

The problem we address in this report is: How do examples of ANN analyses perform compared with logistic regression models in the analyses of one set of social variables and one set of biological and health variables from the same individuals?

Methods

We will compare the results of regression models with the results of a few different ANN models available in the Data Mining software package Clementine (Version 9.0) provided by SPSS. We will not try to describe or understand the internal workings of the ANN models, but will consider them as black boxes into which we pour data and then observe the output. The results of these investigations are discussed to provide an answer to the initial question regarding possible benefits of the ANN analyses.

Sample

H70–The Gerontological and Geriatric Population Studies in Göteborg, Sweden, began in 1971 with a random sample of 1/3 of all 70-year-olds in Göteborg and continued with new samples in 1976 and 1981 [19, 20]. For these three cohorts of 70-year-olds in Göteborg, we have two sets of data: set 1, on social, activity and situation from 2,291 probands; and set 2, on biological and health data from 2,194 probands. They are the same individuals and the difference is because of missing internal data. The social data and the medical data were used in two lines of analyses to compare the prediction of seven-year mortality. For both sets of predictor variables, the seven-year mortality was 17.6%. The variables of the two data sets are presented in Tables 2 and 3.

Table 2. Prediction data set 1: social, activity and situation variables

Variable	*Mean*	*St.dev.*	*Type*	*Values*
Gender (0=Female/1=Male)	0.460	0.499	Flag	0, 1
Income below median	0.494	0.500	Flag	0,1
Smoker at age 70	0.724	0.838	Ordered	0,1,2
Education above primary	0.186	0.389	Flag	0,1
No/low contact with children	1.559	1.484	Ordered	0,1,2,3,4,5,6.7
No/low helping children	4.162	1.657	Ordered	1,2,3,4,5,6
No/low participation in volunteer associations	5.351	2.004	Ordered	1,2,3,4,5,6,7
Low social activity level	3.732	1.190	Ordered	1,2,3,4,5,6
Low socioeconomic group	0.714	0.452	Flag	0,1
Physically inactive	0.177	0.382	Flag	0,1
Not married or cohabiting	0.389	0.488	Flag	0,1

Table 3. Prediction data set 2: biological and health variables

Variable	*Mean*	*St.dev.*	*Type*	*Values*
Gender (0=Female/1=Male)	0.461	0.499	Flag	0,1
Low BMI 10-15-75%	2.650	0.654	Ordered	1,2,3
High BMI 75-15-10%	1.350	0.654	Ordered	1,2,3
Cancer anamnestic	0.065	0.246	Flag	0,1
Drug treatment hypertension	0.241	0.428	Flag	0,1
Bronchitis	0.105	0.306	Flag	0,1
Congestive heart failure	0.144	0.352	Flag	0,1
Ischemic heart disease	0.254	0.435	Flag	0,1
Diabetes anamnestic	0.057	0.231	Flag	0,1
Peak Flow Meter standardized (lung function)	0.000	1.000	Continuous	–2.89 to 3.44
High total cholesterol (75-15-10%)	1.350	0.654	Ordered	1,2,3
Blood sedimentation rate	13.103	11.096	Continuous	1 to 138

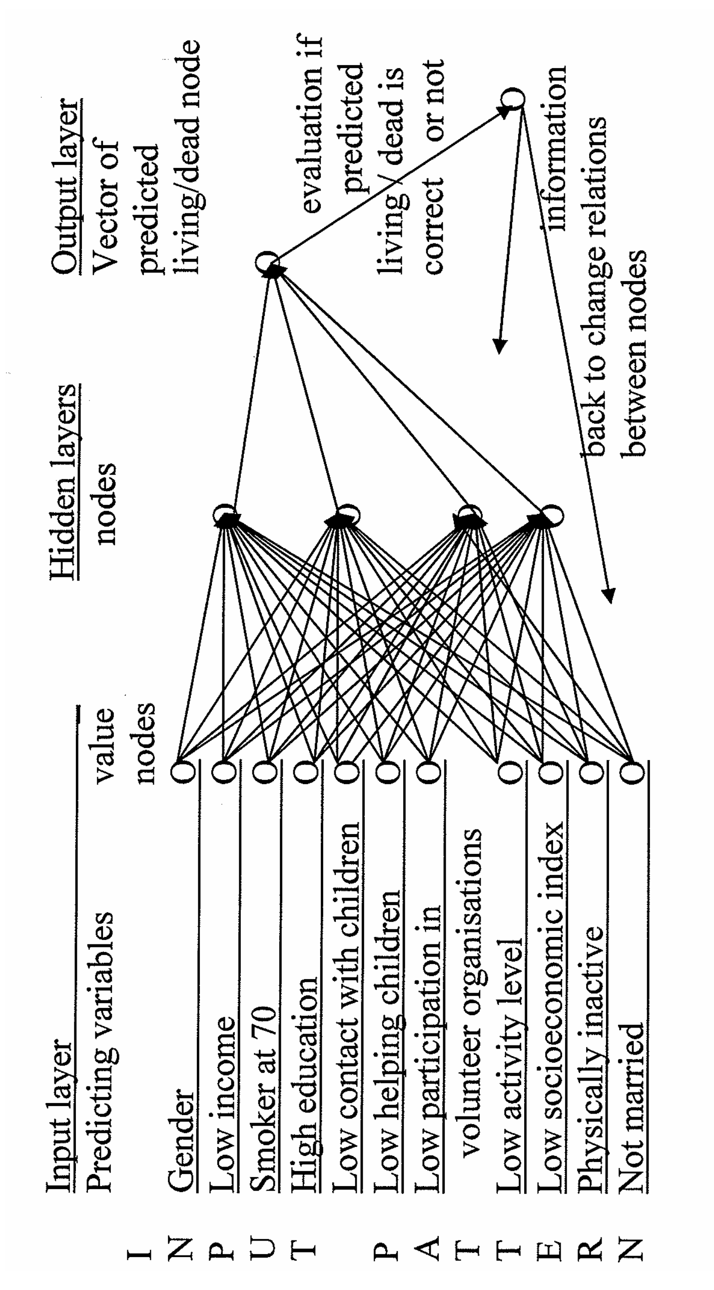

Figure 6. A sketch of ANN analyses process of 7-year survival

This description follows closely the manuals from the Clementine 9 documentation. A neural network, sometimes called a multilayer perceptron, is basically a simplified model of the way the human brain processes information. It works by simulating a large number of interconnected simple processing units or nodes, which resemble abstract versions of neurons that are connected.

The connections between nodes in Fig. 6 are symbolized by arrows. The weights of these connections are randomly distributed when the analyses start. The input layer for each proband comprises the assessed values of the variables. The nodes in the hidden layers are constructed by the programme. The value in the output layer for each individual is the ANN-predicted value for whether they survived the seven-year observation time or not. There predicted value is evaluated as correct or not compared to the observed value. This information is used by the programme to adjust the weights between the nodes: when the prediction is correct, the weights of the connections are increased; and when the prediction is wrong, the weights of the connections are decreased. This process continues until all cases are processed, and is called training the network. During the process, the weights are adjusted so that the output grows more accurate. This process invites overfitting problems, so to master overfitting, a common technique is to train the network on 50% of the cases and then evaluate its performance by applying the network to the second half of the cases. There are other ways to provide stopping criteria for the training process. There are several types of ANN architectures. The type used in this study is called a multilayer feedforward network, which propagates input signals forwards and error signals backwards.

The output vector from the ANN models used in this paper contains two variables for each case: predicted survival 0/1; and a number between 0 and 1 giving an estimate of the accuracy of the predicted value. Thus, in one case the prediction may be alive (1), with an accuracy of 0.8; in another case the prediction may be the same but the accuracy may be 0.3. We interpret this last figure, the accuracy, as a measure of the reliability of a given estimate.

Neural networks are powerful general function estimators. They usually perform prediction tasks at least as well as other techniques, if not significantly better. They also require minimal statistical or mathematical knowledge to train or apply. A major weakness is their interpretability—how the results are achieved is not well understood. The network takes advantage of structures in the data set that we do not understand. We shall discuss strengths and weaknesses with reference to the results of our analyses.

The Clementine program has many different ways of performing the ANN:

- Quick. This method uses rules of thumb and characteristics of the data to choose an appropriate shape (topology) for the network.
- Dynamic. This method creates an initial topology but modifies the topology by adding and/or removing hidden units as training progresses.
- Multiple. This method creates several networks of different topologies (the exact number depends on the training data). These networks are then trained in a pseudo-parallel fashion. At the end of training, the model with the lowest RMS error is presented as the final model.
- Prune. This method starts with a large network and removes (prunes) the weakest units in the hidden and input layers as training proceeds. This method is usually slow, but it often yields better results than other methods.
- RBFN. The radial basis function network (RBFN) uses a technique similar to k-means clustering to partition the data on the basis of values of the target field.
- Exhaustive prune. This method is related to the prune method. It starts with a large network and prunes the weakest units in the hidden and input layers as training proceeds. With exhaustive prune, network training parameters are chosen to ensure a very thorough search of the space of possible models to find the best one. This method is usually the slowest but often yields the best results.

In the analyses, we tried to optimize these methods and report examples of results from the different methods.

The data sets we used are taken from studies based on representative random population samples, and therefore contain a varying number of scattered missing values in almost all variables. In regression models, there are difficulties in handling cases with missing data. In ANN models, it is possible to include missing data in the analysis. To simplify the comparison, we have calculated all models on cases with complete data in all included variables.

How do we compare results from regression and ANN models?

To simplify the comparison, all examples cited in this paper use data where the outcome variable is dichotomous and where the basic question may be expressed as the problem of classifying the observations into one of two groups, e.g., survivors and non-survivors. The type of regression we will use in this study is therefore binary logistic regression, and the outcome variables for neural network models are defined in Clementine to be of type 'flag'.

Still, it is not obvious how we can compare the outcome from the two different kinds of models, so we choose a well-defined criterion for judging the prediction power of different models. Logistic regressions leave a result consisting of the overall goodness of fit for the model and estimates of the effect, with standard error, on the outcome of each predictor and estimated probability of each observation for the studied outcome (seven-year mortality in our examples).

The ANN result is the most probable outcome (dead or alive) for each case, and a quantitative estimate of how reliable this estimate is, with an accuracy value between 0 and 1. This means that there is no directly comparable parameter, either on the sample level or the individual level, so that a common comparable quantitative characteristic has to be constructed.

We decided to base this comparison on case ranks. For the logistic model, the ranking of cases is done on the predicted probabilities calculated from the regression model:

Probability of dying within seven years = $1-(e^{-LS})/(1+e^{-LS})$, where LS is the linear part of the regression equation: $LS = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$.

For ANN, we use the accuracy value. Values near 1 indicate that the classification is relatively certain, and ranking is based on this value for cases classified into the mortality group, and -1 * the values for those classified into the non-mortality group, meaning that the group with highest positive accuracy is the group with the highest estimated probability of dying.

The accuracy (or reliability) measure (i.e. how certain an estimation is) is not necessarily directly comparable to a level estimate of the probability of a certain outcome. Accuracy is more like a variance estimate than a level estimate. This may be a problem, as in principle the probability of an outcome (level of risk) and its reliability (or variance) are not necessarily strongly correlated. However, in practice, a combination of predictor values that result in a high proportion of deaths within seven years must mean that both the model estimate of mortality risk and accuracy or reliability of classification must be considered high, and vice versa. This is in accordance with the relation between expected value and variance for a binomial outcome. For a binomial variable, there is a direct relation between the parameter and its variance, so that a parameter value closer to 0 or 1 has smaller variance than a parameter value closer to 0.5.

For all practical purposes, we think it is realistic to assume that the two rankings defined by estimated probability (in the logistic model) and accuracy of classification (in ANN) are comparable.

Now we have two rankings of cases in terms of expected mortality, the next step is to decide on a criterion for judging which ranking gives the better prediction, and there are many possible alternatives.
Some are as follows:

1. Turn the rankings into two classes, predicted to die and predicted to survive, and compare the agreement between this classification and the actual outcome.
2. Transform the rankings to quantiles, for example, quintiles, and compare the correlation between quantiles and outcome. By applying a logistic regression model to this comparison, we can also calculate an overall odds ratio (OR) estimate that may be compared.
3. Calculate the receiver operating characteristic (ROC) curves, compare them visually or calculate the area under the curve and compare [21].
4. Compare separation of average rank in the observed outcome categories.

Because we are looking at a well-defined outcome measure in this example—dead or not dead—we think it is most natural to compare the ability of each model to classify cases correctly. This also gives the most concrete measure for comparison, giving us a result that is the easiest to report and interpret, which is an important advantage when there are many different models to consider.

A problem with the dichotomization of the model estimates is that unlike ANN, logistic regression models do not give any direct classification into the outcome categories. Thus, we have to decide on a cut-off level of the predicted probability from the model, unless we want to compare all possible cut-off levels simultaneously, as in a ROC analysis.

When choosing a cut-off level, two principal alternatives are to use an absolute level or to use a relative (percentile) value. It seems most natural for us to use a cut-off level that corresponds to the observed proportion of deaths in the sample, since this will give an unbiased prediction model, in the sense that the estimated proportion of death will be the same as the observed proportion. Further, it makes sense to limit the search for models to the subset of (approximately) unbiased models.

In our example data, the sample contains 404 persons (17.6% of the total sample) who have died within seven years. Because there are ties in the model estimates, the predicted proportion of non-survivors from the two models will not always be exactly the same, but will be chosen to be as close as possible to the observed proportion.

We present the results of logistic regression as main linear regression results and the results of regression with interaction terms. We explored the outcomes of OR for a large number of interaction possibilities. When we show results from regression with interaction, we always choose the models with the highest possible OR.

Results

A training run of the ANN provided a model accuracy estimation, which was less than 1. This figure increased, and after some time it stabilized around a particular value. We made a large number of analyses, and in our experience with 2,191 observations and 13 variables, the model accuracy value of training of the ANN stabilized after 2 min execution time. With the most promising ANN method, we allowed the execution to continue for 12 h without any noticeable increase in model accuracy. Thus, we conclude that for data of moderate size (no more than several thousand cases), there is no need to allow the training to proceed for more than 5 min (results achieved on a PC running Windows XP with 3 GHz Pentium processor and 1 Gb of working memory), and in fact results obtained after 2 min are very stable.

Social variables as predictors

For data set 1 with social variables as predictors, the best results of the analyses from the best performing ANN methods and from linear logistic regression are shown in Table 4.

Table 4. The odds ratios and approximated 95% confidence interval for predicting seven-year mortality by social variables from different methods of analyses

Method of analyses	*Odds ratio*	*Approx. 95% conf. interval*
ANN quick	1.929	1.50–2.48
ANN dynamic	1.773	1.37–2.29
ANN multiple	1.570	1.21–2.04
ANN RBFN	2.881	2.26–3.60
ANN exhaustive prune	1.773	1.37–2,48
Logistic main linear	2.927	2.29–3.74
Logistic main linear and interaction terms	3.466	2.72–4.41

In the analyses with the social predictor variables, the best ANN and RBFN models performed as well as or nearly as well as the logistic regression models judged on the approximated 95% confidence intervals.

Biological and health variables as predictors

For data set 2 with biological and health variables, the result is more complicated. Thus, we show results from several runs of the most promising analyses in Table 5.

Table 5. The odds ratios and approximated 95% confidence interval for predicting seven-year mortality by biological and health variables from different runs and methods of analyses

Method of analyses	*Run*	*Odds ratio*	*Approx. 95% conf. interval*
ANN Multiple	1	7.194	5.573–9.288
ANN Multiple	2	10.149	7.831–13.154
ANN Multiple	3	10.149	7.831–13.154
ANN Multiple	4	7.318	5.668–9.448
ANN Multiple	5	8.987	6.946–11.627
ANN Multiple	6	9.974	7.698–12.922
ANN Multiple	7	13.021	10.001–16.952
ANN Multiple	8	17.201	13.124–22.546
ANN Exhaustive prune	1	20.717	15.25–28.15
ANN Exhaustive prune	2	35.889	25.87–49.79
ANN Exhaustive prune	3	24.405	17.52–33.99
ANN Exhaustive prune	4	38.888	27.03–55.95
ANN Exhaustive prune	5	31.473	22.53–43.96
ANN Exhaustive prune	6	24.965	18.53–33.64
ANN Exhaustive prune	7	27.965	19.89–39.32
ANN Exhaustive prune	8	29.305	21.46–40.03
Logistic with interaction terms	1	3.981	3.08–5.14
Logistic with interaction terms	2	3.718	2.88–4.81
Logistic with interaction terms	3	4.050	3.14–5.23
Logistic with interaction terms	4	3.593	2.78–4.65
Logistic with interaction terms	5	4.050	3.14–5.23
Logistic with interaction terms	6	3.727	2.89–4.81

The first 16 lines in table 5 present the Clementine OR from 16 trainings on the same biological and health data using the multiple and exhaustive prune methods, which gave the best results for these data. We will keep those two methods when we calculate unbiased estimates (by using the observed mortality proportion as cut-off) of model-predicted classification. The models based on logistic regression with interaction terms that we judged as best motivated to enter into the models—described in the last six lines on the table above—have much smaller agreement with the observed outcome than the best ANN models.

Comparison between predictions

This last result is much more interesting than that of data set 1, where both the logistic linear main effects models and interaction terms were approximately as good as the ANN models. For the current data, both the linear and non-linear main-effects-only regression models gave much lower OR than the best ANN models. This is an indication that it ought to be possible to improve substantially the analyses of the regression models and attempt to find better models. After a systematic investigation of possible interaction terms, the best OR we managed to obtain was 4.409 with the approximated 95% confidence interval 3.42–5.69. When compared with the results in Table 5, the OR did not improve much, even with 22 interaction terms. Judged on the comparisons of the OR, the best ANN models performed better on data set 2 with biological and health variables.

When we examine the different runs of analyses in Table 5 and the best-found values obtained by the logistic models, it is clear that the OR are fairly stable between 3.593 and 4.050. The results from the ANN multiple methods varied between 7.194 and 17.201, and for the ANN exhaustive prune method varied between 20.717 and 38.888. It is possible that the high variation of the results of the ANN methods depends on the initial random assignment of weights of the relations between nodes. The initial values may lead the programme to use different paths through data, and an unfortunate start may make it more difficult for the training to yield a good result. If this is true, it is nevertheless the case that even such unfortunate starts can yield better results (judged by OR) than logistic regression methods. It might also be the case that the higher values obtained by an ANN method could be the result of overtraining, which for these models is analogues to overfitting in regression analyses.

Discussion

We have not found convincing proof that the use of ANN models in general would increase the quality of the statistical studies that use traditional tools, i.e., logistic regression models. However, that does not exclude the possibility that in special cases, the use of ANN and similar models would be a useful tool for improving our understanding of the association pattern in the data we analyse. One weakness with ANN appears to be a lack of stability on the individual level, although this observed phenomenon may be the natural consequence of multicollinearity in the data.

There is at least one heuristic use of ANN analyses. If the ANN analyses perform much better than an alternative traditional method, there is then information hidden in the data that the traditional method does not account for. In this case, the ANN result is a reason to be dissatisfied with the traditional analyses and to try to obtain a better understanding of the data, perhaps by using other methods. On the other hand, if the traditional methods give good results similar to the ANN methods, then there is a reason to be satisfied, at least for the moment, with the result. The ANN method did not reveal any hidden structure in the social data. In this case, we could be satisfied with the logistic regression results from the analyses of the set of social variables as predictors. However, there are hidden structures in the data set of biological and health variables that need further analysis by models other than logistic regression, such as those used in this study.

In some cases, the ANN analyses produce much better predictions, as demonstrated in the set of biological and medical data. In such cases where the outcome is of a multifactor genesis, the ANN models can be used for predictions. Examples of this could be deciding on alternative treatments [22] or recruiting staff [23].

In evaluating the usefulness of different approaches to statistical analysis, we attempted to cover our subject from as many different aspects as we felt relevant. However, we are aware of the disadvantage in being prevented from examining many of the issues as thoroughly as they deserve.

Further investigation may include:

1. What is the optimal criterion for comparing prediction models of different types? We have mainly used OR to compare the agreement between model-predicted and observed outcomes, but there are many other possible ways to compare different models.
2. Are different types of ANN more appropriate for different types of data and certain types of analysis, and less appropriate in other situations? We have not considered this at all.
3. Can repeated trainings be combined to form a better prediction model than the best single model?
4. In one data set, we found that some of the ANN methods gave much better prediction than the regression models, but we were not able to explain what caused this difference. Nor could we explain if and how these data differed from the other data set we used. It would be of great worth to be able to determine in advance if a certain data set has properties that indicate whether it could be worthwhile to consider ANN as a possible alternative in statistical analysis.
5. Do ANN adjust for overfitting in a satisfactory manner? A simulation study would be able to answer that question, but that would require another study.
6. It would be of great interest to repeat the comparisons we have done on a larger number of real-world data sets to reach a better understanding of the variability of the results we have obtained.
7. We have limited this study to comparing different methods for predicting a dichotomous outcome for practical reasons, but it would also be interesting to examine other types of outcomes, e.g., categorized data, continuous variables or continuous survival time.

Conclusions

In the analyses of seven-year mortality, the ANN model did not generally perform better than the logistic regression analyses. In the set of data with the biological and health variables, the ANN models performed much better when the OR were compared. This indicates that there are hidden structures in this data set that are not accounted for by the logistic regression. On the other hand, our evaluation is that the logistic regression models are good enough for the actual social data.

Thus, there are answers to our research question:

1. We recommend performing ANN as a heuristic tool to evaluate the results of regression analyses. If the ANN performs better, then there is a reason to try to find a better understanding of data than that given by the regression analyses. On the other hand, if the regression analyses perform as well as the ANN, at least until new findings appear, there is a reason to be satisfied with the regression results.
2. Well-performing ANN models can be used for predictions when there is an outcome of multifactor genesis that is not well understood by other methods of analyses.

Acknowledgements

This study was supported by a grant from The Faculty of Social Sciences, University of Gothenburg. We used the Clementine program by the kind permission of SPSS.

References:

1. Ljungquist, B., *Aging and survival: Studies of social, biobehavioral and genetic correlates.* 1996, Göteborg: Göteborgs Universitet.
2. Stalbrand, I.S., et al., *Subjective health and illness, coping and life satisfaction in an 80-year-old Swedish population-implications for mortality.* International Journal of Behavioral Medicine, 2007. **14**: p. 173-180.
3. Liao, S.H., *Expert system methodologies and applications - a decade review from 1995 to 2004.* Expert Systems with Applications, 2005. **28**(1): p. 93-103.
4. Sares, T.A., *Sociopolitical viewpoints as narrated by family and educational background.* Journal of Social Psychology, 1998. **138**(5): p. 637-644.
5. de Lillo, A. and C. Meraviglia, *The role of social determinants on men's and women's mobility in Italy. A comparison of discriminant analysis and artificial neural networks.* Substance Use & Misuse, 1998. **33**(3): p. 751-764.
6. Shmueli, D., I. Salomon, and D. Shefer, *Neural network analysis of travel behavior: Evaluating tools for prediction.* Transportation Research Part C-Emerging Technologies, 1996. **4**(3): p. 151-166.
7. Meraviglia, C., *Fuzzy Logic and Artificial Neural Networks in Classification.* Sociologia e Ricerca Sociale, 2001. **22**(64): p. 61-83.
8. Burke, H.B., et al., *Artificial neural networks improve the accuracy of cancer survival prediction.* Cancer, 1997. **79**(4): p. 857-862.
9. Ragde, H., et al., *Ten-year disease free survival after transperineal sonography-guided iodine-125 brachytherapy with or without 45-gray external beam irradiation in the treatment of patients with clinically localized, low to high Gleason grade prostate carcinoma.* Cancer, 1998. **83**(5): p. 989-1001.
10. Santos-Garcia, G., et al., *Prediction of postoperative morbidity after lung resection using an artificial neural network ensemble.* Artificial Intelligence in Medicine, 2004. **30**(1): p. 61-69.
11. Snow, P.B., et al., *Neural network and regression predictions of 5-year survival after colon carcinoma treatment.* Cancer, 2001. **91**(8): p. 1673-1678.

12. Macknight, C., et al., *Assessment of individual risk of death using self-report data: an artificial neural network compared with a frailty index.* Journal of the American Geriatrics Society, 2004. **52**(7): p. 1180-1184.
13. Freeman, R.V., et al., *Comparison of artificial neural networks with logistic regression in prediction of in-hospital death after percutaneous transluminal coronary angioplasty.* American Heart Journal, 2000. **140**(3): p. 511-520.
14. Landsittel, D.P., L.I. Gardner, and V.C. Arena, *Identifying populations at high risk for occupational back injury with neutral networks.* Human and Ecological Risk Assessment, 1998. **4**(6): p. 1337-1352.
15. Sargent, D.J., *Comparison of artificial neural networks with other statistical approaches - Results from medical data sets.* Cancer, 2001. **91**(8): p. 1636-1642.
16. Lisboa, P.J. and A.F.G. Taktak, *The use of artificial neural networks in decision support in cancer: A systematic review.* Neural Networks, 2006. **19**(4): p. 408-415.
17. Fox, J., *Regression diagnostics*, in *International handbooks of quantitative applications in the social sciences*, M.S. Lewis-Beck, Editor. 1993, Sage: London. p. 245-334.
18. Faber, N.M. and R. Rajkó, *How to avoid over-fitting in multivariate calibration--The conventional validation approach and an alternative.* Analytica Chimica Acta, 2007. **595**(1-2): p. 98-106.
19. Rinder, L., et al., *Seventy-year-old people in Gothenburg. A population study in an.* Acta Med Scand, 1975. **198**(5): p. 397-407.
20. Steen, B. and H. Djurfeldt, *The Gerontological and Geriatric Population Studies in Gothenburg, Sweden.* Zeitschrift Fur Gerontologie, 1993. **26**(3): p. 163-169.
21. Zweig, M., *Receiver-Operatin Characteristic (ROC) plots – A fundamental evaluation tool in clinical medicine.* Clinical Chemistry, 1993. **39**(4): p. 561-577.
22. Green, M., *Improving diagnosis of acute coronary syndromes in an emergency setting: a machine learning approach.* 2008, Lund :: Department of Theoretical Physics, Lund University.
23. Hausler, J. and M. Sommer, *Neural networks: Nonlinear methods of judgment formation in personnel selection.* Zeitschrift Fur Personalpsychologie, 2006. **5**(1): p. 4-15.