

Artificial neural network for risk assessment in preterm neonates

B Zernikow, K Holtmannspoetter, E Michel, W Pielemeier, F Hornschuh, A Westermann, K H Hennecke

Vestische Kinderklinik
Witten/Herdecke
University
Lloydstr. 5
D-45711 Datteln
Germany
B Zernikow
K Holtmannspoetter
W Pielemeier
F Hornschuh
A Westermann
K H Hennecke

Medizinischer Dienst
der
Krankenversicherung
Westfalen-Lippe,
Muenster,
Germany
E Michel

Correspondence to:
Dr Boris Zernikow.
Email: Boris.Zernikow@
t-online.de

Accepted 9 March 1998

Abstract

Aim—To predict the individual neonatal mortality risk of preterm infants using an artificial neural network “trained” on admission data.

Methods—A total of 890 preterm neonates (<32 weeks gestational age and/or <1500 g birthweight) were enrolled in our retrospective study. The neural network trained on infants born between 1990 and 1993. The predictive value was tested on infants born in the successive three years.

Results—The artificial neural network performed significantly better than a logistic regression model (area under the receiver operator curve 0.95 vs 0.92). Survival was associated with high morbidity if the predicted mortality risk was greater than 0.50. There were no preterm infants with a predicted mortality risk of greater than 0.80. The mortality risks of two non-survivors with birthweights >2000 g and severe congenital disease had largely been underestimated.

Conclusion—An artificial neural network trained on admission data can accurately predict the mortality risk for most preterm infants. However, the significant number of prediction failures renders it unsuitable for individual treatment decisions.

(Arch Dis Child Fetal Neonatal Ed 1998;79:F129–F134)

Keywords: artificial neural network; mortality; prediction

Mortality has traditionally been used to compare the therapeutic performances of neonatal intensive care units.^{1–3} As individual mortality risk depends on numerous confounding factors, crude mortality can be adjusted for prematurity,^{4,5} birthweight,^{1,6} initial illness severity,^{7–9} presence of congenital malformations, or a combination of these factors, for the purposes of quality assessment.^{2,3,7}

The Score for Neonatal Acute Physiology (SNAP), Perinatal Extension (SNAP-PE),⁷ Clinical Risk Index for Babies (CRIB) score^{7–9} and alternative models¹⁰ are excellent predictors of overall inpatient mortality of preterm neonates if accuracy is judged by the area under the receiver operating characteristic (ROC) curve.^{7–9,11} Despite such excellent results a significant number of individual prediction failures still occur. Case analysis of non-survivors with a low predicted individual mortality risk, or of survivors with a very high predicted individual mortality risk might yield further insight into the pitfalls of scoring systems.

As the quality of early therapeutic interventions might significantly influence the degree of illness severity,¹² data on initial illness severity are compromised if collected later than a few hours after birth.^{7–9} Thus several variables contributing to CRIB or SNAP scores are regarded as input as well as output data with respect to intensive care unit performance.^{7–9,11} Shortening the data collection period is desirable.^{7–9}

Artificial neural networks (ANN) are software tools with the capacity to learn. An ANN behaves like a child learning to differentiate between cats and dogs, by means of examples, under the supervision of his/her parents. An ANN can learn the relation between input variables (size, fleece, voice, behaviour patterns) and the output variable (cat or dog) by presenting multiple input–output pairs to it (supervised learning). After this learning or “training” period the ANN can “predict” the output (cat or dog) on inputs of further unknown examples.^{13,14} This capability is called generalisation. ANNs have shown excellent predictive accuracy in medicine even on inaccurate or incomplete input data. Recent

Table 1 Comparison of training and test set and validation set patients (binary scaled input and output variables)

Variable	Training set	Validation set	p Value
Time period	04/04/90 –05/11/93	06/11/93 –24/10/96	
Output			
28 day survivors	397	419	NS
28 day non-survivors	38	36	NS
Input			
Prenatal			
White race	342	354	NS
Singleton birth	328	334	NS
Perinatal			
Apgar, 1 minute < 3	34	26	NS
Apgar, 1 minute < 5	129	71	0.001
Apgar, 1 minute < 7	254	219	0.002
Apgar, 5 minutes < 3	4	0	0.040
Apgar, 5 minutes < 5	14	9	NS
Apgar, 5 minutes < 7	70	67	NS
Apgar, 10 minutes < 3	2	0	NS
Apgar, 10 minutes < 5	7	5	NS
Apgar, 10 minutes < 7	21	21	NS
Condition stable	191	209	NS
Condition fair/unstable	215	206	NS
Condition vitally endangered	29	40	NS
Congenital malformation	11	11	
Emergency delivery	287	304	NS
Inborn	276	352	0.001
pH < 7.10	17	28	NS
pH < 7.15	26	43	NS
pH < 7.20	52	93	0.001
pH < 7.25	109	175	0.001
PROM (24 h)	147	161	NS
Caesarean section	329	334	NS
Female gender	212	224	NS
SGA	80	84	NS
Body temperature < 35.0°C	29	21	NS
Body temperature < 35.5°C	63	41	0.011
Body temperature < 36.0°C	126	134	NS
Ventilated on transport	254	174	0.001

PROM: premature rupture of membranes at least 24 hours before birth; pH value obtained from the first capillary blood analysis on admission.

Table 2 Comparison of training and test set and validation set patients (continuous input variables) (5%, 95% centiles in parentheses)

Variable	Training set	Validation set	p Value
Input			
Prenatal			
Mother's age (years)	28.0 (19.5/37.6)	29.2 (20.7/38.4)	0.001
Perinatal			
Apgar score, 1 minute	6 (2/9)	7 (2/9)	0.001
Apgar score, 5 minutes	8 (5/10)	8 (5/10)	0.047
Apgar score, 10 minutes	8 (7/10)	8 (7/10)	0.048
Birthweight (g)	1240 (634/2030)	1330 (634/2100)	NS
Date of birth (years)	2.1 (0.6/3.5)	5.1 (3.7/6.4)	
Gestational age (weeks)	29 (25/33)	29 (24/33)	NS
pCO ₂ (mm Hg)	46 (30/67)	53 (33/76)	0.001
pH	7.31 (7.13/7.47)	7.27 (7.08/7.43)	0.001
Body temperature (°C)	36.5 (34.8/37.5)	36.4 (35.2/37.8)	NS
Transport duration (min)	5 (2/45)	5 (1/40)	0.001

Date of birth: number of years elapsed since start of the study; pH and PCO₂: from capillary blood gas, analysis on admission.

Table 3 Matched pairs analysis of survivors with high (>0.5) or low (<0.5) predicted individual mortality risk*

Variable	Predicted individual mortality risk ≥ 0.50 (n=25)	Predicted individual mortality risk < 0.50 (n=25)	p Value
Male gender	13	11	NS
Gestational age (completed weeks)	26.3 (2.3)	26.3 (2.3)	NS
Birthweight (g)	878 (318)	876 (325)	NS
Days intubated	26.2 (17.7)	24.1 (17.9)	NS
Duration of antibiotic treatment (days)	23.1 (10.9)	21.4 (12.1)	NS
Number of PRC transfusions	8.2 (5.1)	7.4 (4.9)	NS
Bronchopulmonary dysplasia	16	12	NS
IVH III/IV or PVL	14	6	0.02
Diseases requiring surgical intervention	16	8	0.02

PRC: packed red cells; IVH: intraventricular haemorrhage; PVL: periventricular leucomalacia
*For normally distributed variables SD in parentheses.

reviews describe the many clinical applications of ANN.¹⁵⁻²⁰

Using an ANN for admission data on preterm neonates we aimed to: predict the individual neonatal mortality risk of preterm neonates from their admission data; compare the net's performance with that of a logistic regression model; and characterise survivors despite a high predicted individual mortality risk, and non-survivors despite a low predicted individual mortality risk.

Methods

Our investigation was based on a pre-existing data set of 890 preterm neonates (gestational age < 32 completed weeks gestation and/or birthweight < 1500 g) born between 4 April 1990 to 24 October 1996. Admission and outcome data were available on computer files (tables 1 and 2). Outcome was defined with respect to neonatal death (death within the first 28 days of life).

The patient sample was split into a training and test (n=435) and a validation set (n=455) according to date of birth. The cutoff date was arbitrarily set to 5 November 1993. The training and test set data were subjected to a forward stepwise logistic regression analysis

Table 4 Logistic regression model for six predictor variables

Variable	Odds ratio (95% confidence interval)	p Value
Congenital malformations	18.73 (2.13 to 164.8)	0.008
Condition vitally endangered	16.32 (5.59 to 47.6)	0.001
pH < 7.10	6.43 (1.44 to 28.9)	0.014
Apgar score, 1 minute < 5	3.97 (1.52 to 10.3)	0.005
Gestational age (weeks)	0.68 (0.55 to 0.81)	0.001
Condition stable	0.22 (0.05 to 1.11)	0.066

(cutoff p value to include a variable Pin=0.05, cutoff p value to exclude a variable Pout=0.1) (SPSS/PC release 6.1.2, SPSS Inc., Chicago, IL, USA), and were used to train the ANN (Predict, NeuralWare Inc., Pittsburgh, USA). Predict helped to develop a feed-forward fully connected three layer perceptron. The descent gradient learning rule was applied for error propagation.^{21, 22} The ANN input variable selection mode implemented in Predict is based on genetic algorithms and helps to find the optimal combination of inputs.²¹ The training process was stopped as soon as the prediction error on a subset of the training data set, namely the test set, stopped declining further. After the training process the ANN had settled to 13 input and three hidden processing elements. There was one output processing element.

For validation the inputs were derived from the independent validation data set (6 November 1993 to 24 October 1996). The logistic regression model and the ANN delivered the individual mortality risk of each single preterm neonate. Prediction and actual outcome were compared to assess performance.

For the purpose of performance assessment and method comparison ROC (receiver operating characteristic) curves were used. The outcome predictions as delivered by both models were continuous variables (individual mortality risk). The cutoff point used to transform them into dichotomous mortality predictions was varied to obtain ROC curves.^{23, 24} The maximum likelihood estimation of the binormal ROC curves, and the area under the fitted ROC curves were calculated using LABROC1 and CLABROC software (IBM-PC version 1.2.1; Metz CE *et al.* 1993, Department of Radiology, Chicago Medical Center, Chicago, IL, USA). CLABROC was used to test corresponding ROC curves for any significant difference (univariate Z-score test of the difference between the areas under the two paired ROC curves, and univariate Z-score test of the difference between sensitivity values on the two paired ROC curves at a selected specificity level²⁵).

Surviving patients with an ANN based individual mortality risk of more than 0.50 were matched with infants identical in gestational age and birthweight (within ±5%), but with a predicted individual mortality risk of less than 0.50. Morbidity of the two groups was compared with respect to the variables listed in table 3. Non-survivors with very low predicted individual mortality risk (<0.10) were analysed individually.

To test for associations between the variables in the training or validation set univariate tests were performed (χ^2 test for categorical variables; *t* test, or Mann-Whitney U test for continuous variables). A p value of < 0.05 was considered significant.

Results

The training data set overall neonatal mortality was 8.7%, while in the validation set it was 7.9%. The difference is not significant (table 1). The validation dataset differed in several respects from the training set (table 1 and 2)

Table 5 ANN inputs and area under ROC curves of ANNs with reduced input variable set*; comparison to ANN with complete input set

Skipped variable	Area under the ROC curve (SD)	p Value
None (complete input set)	0.954 (0.015)	
Gestational age	0.924 (0.015)	0.004
Birthweight	0.928 (0.017)	0.033
Condition on admission vitally endangered and condition on admission fair	0.929 (0.015)	0.040
Condition on admission stable and condition on admission fair	0.930 (0.015)	0.042
Inborn	0.936 (0.016)	NS
Capillary pH < 7.10	0.943 (0.015)	NS
Temperature on admission < 35.5°C	0.945 (0.014)	NS
Condition on admission stable and condition on admission unstable	0.945 (0.014)	NS
Emergency delivery	0.946 (0.013)	NS
10 minute Apgar score < 5	0.947 (0.013)	NS
5 minute Apgar score < 3	0.947 (0.013)	NS
Congenital malformation	0.949 (0.012)	NS
1 minute Apgar score	0.950 (0.013)	NS
Gestational age and birthweight	0.881 (0.029)	0.001
All except gestational age and birthweight	0.890 (0.021)	0.001

*Table is ranked according to the estimated relative predictive impact of the omitted variables in descending order.

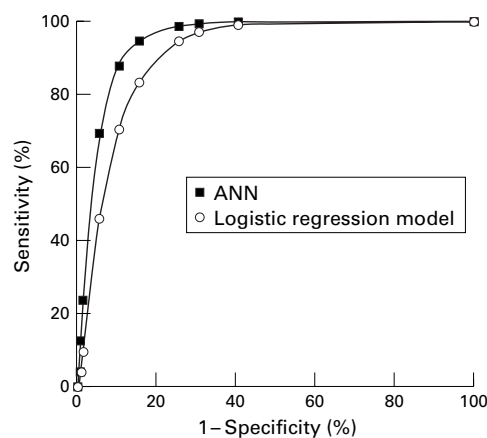


Figure 1 ROC plots: ANN compared with the logistic regression model.

mainly because of the higher rate of inborn infants in the validation time period.

Stepwise logistic regression revealed the items low gestational age, one minute Apgar score < 5, pH of the first blood gas analysis < 7.10, a life threatening condition on admission, and the presence of congenital malformation, as significant risk factors for neonatal mortality. The factor stable condition on admission was included in the model by the forward stepwise algorithm of SPSS. This factor did not reach significance when the odds ratio was calculated (p = 0.066) (table 4).

The optimal ANN used 13 different input variables (table 5). To investigate the specific impact of individual ANN input variables on neonatal mortality prediction, 14 ANNs were trained, each omitting one or two input variables. An additional ANN was trained with just two inputs—namely, birthweight and gestational age. The resulting areas under the

ROC curves reflecting each ANN’s performances are given in table 5.

When adjusted for a specificity level of 80, 85, 90, or 95%, in terms of sensitivity, the predictions of the ANN were significantly better than the logistic regression model (table 6). The ANN area under the ROC curve was 0.954 (SD 0.011), which is larger than the 0.917 (SD 0.017) of the logistic regression model area (p = 0.002, area test) (fig 1).

No preterm neonate had an ANN predicted mortality risk of >0.80. Thirteen (68.4%) of the 19 preterm neonates with an individual predicted mortality risk of >0.70 died in the neonatal period. Of the 67 preterm babies with a predicted mortality risk of >0.5, 36 (53.7%) died before 29 days of life and three thereafter (fig 2).

When morbidity was graded according to grossly abnormal brain ultrasound scans, or to the number of postnatally acquired diseases requiring one or more surgical interventions, survivors with an ANN based mortality risk >0.50 had a significant higher morbidity than matched survivors with an ANN based mortality risk < 0.50 (p=0.02)(table 3).

The clinical characteristics, cause of death, and the individual ANN derived mortality risk of the three low risk non-survivors (predicted mortality risk < 0.10) are summarised in table 7.

Case 1, a preterm infant of 27 weeks gestation, in his first hours of life, developed respiratory distress syndrome that was unresponsive to surfactant rescue treatment. Cultures on blood and tracheal fluid grew *Escherichia coli*. The infant died on his third day of life, presenting with septic shock syndrome.

Case 2 had a positive history with respect to neonatal death in three maternal siblings. Polyhydramnios had been present in each case, and all the mother’s siblings had died of respiratory insufficiency. The mother turned out to have myotonic dystrophy. Her infant was in need of aggressive stabilisation immediately after birth, including adrenalin and sodium bicarbonate. After stabilisation the infant was transported to neonatal intensive care. On admission her capillary pH was within an acceptable range (table 7).

Case 3 had complex cardiac malformation (double outlet right ventricle, hypoplastic pulmonary artery, pulmonary valve stenosis)

Table 6 Comparison of prediction accuracy of ANN and logistic regression models

Specificity (%)	Sensitivity (95% confidence interval) (%)		p Value
	ANN	Logistic regression model	
75	98.8 (93.1 to 99.9)	94.7 (85.1 to 98.6)	0.061
80	97.5 (89.8 to 99.6)	90.6 (78.8 to 96.7)	0.029
85	94.6 (84.2 to 98.7)	83.5 (69.5 to 92.5)	0.010
90	87.8 (74.2 to 95.4)	70.6 (54.4 to 83.5)	0.002
95	69.6 (51.6 to 83.7)	46.1 (28.9 to 64.0)	0.001

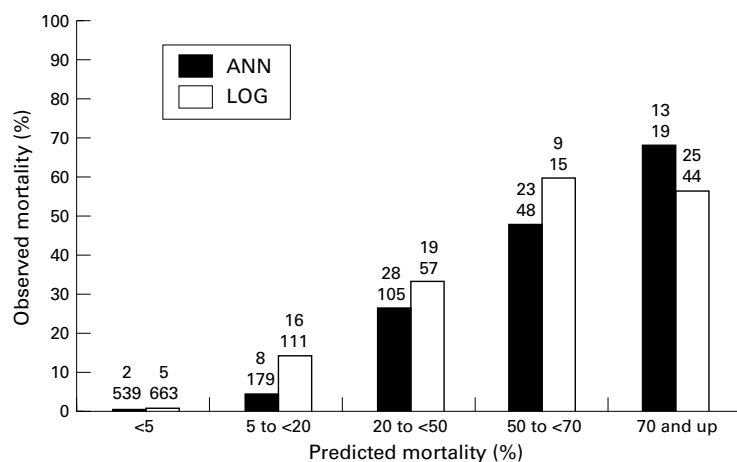


Figure 2 Predicted and observed neonatal mortality: solid column represents ANN; open column represents logistic regression analysis. Absolute numbers of fatal (numerator) and total cases (denominator) of the corresponding patient group are shown.

and respiratory distress syndrome grade IV. Blood pressure and blood pH could be stabilised for only a short period.

In the latter two cases the individual predicted mortality risks increased when predictions were obtained from an ANN trained without information on gestational age and birthweight (table 7).

Discussion

Artificial neural networks have been widely used for outcome prediction.¹⁶ Mortality has been predicted with great accuracy for patients on intensive care units, and for patients who received cardiopulmonary resuscitation in hospital. (Doig GS, et al. Proceedings of annual symposium on computers applied to medical care, 1993: 361-65).^{26 27}

Solely based on routine data for the first minutes of life, our ANN was capable of accurately predicting individual neonatal mortality risk in preterm neonates despite changes in patient characteristics between the training and validation time period. The ANN performed better than a logistic regression model.

While the logistic regression model comprised six items, the ANN used 13 items (table 5). The input variable selection of the ANN seems reasonable as most selected variables were reportedly associated with mortality^{1-4 7-9 28-32} and as all predictive variables isolated by the stepwise logistic regression analysis were also included in the ANN input variable set. Hence

subtle and/or non-linear associations between input variables and neonatal mortality not detected by logistic regression analysis might account for the superior performance of the ANN.

Gestational age, birthweight, and condition on admission are the only items that must not be omitted during the training process without severely compromising performance (table 5). The special impact of the first two of those items on neonatal mortality risk prediction is illustrated by the fact that the area under the ROC curve of the ANN trained with just those two items (table 5) is larger than comparable reports suggest.⁷⁻¹⁰ We cannot exclude that our neonatologists on duty underestimated the gestational age of some of the neonates, including more mature and "healthier" infants (who are at almost no risk for neonatal mortality and where mortality risk is easier to predict) into our study population. Any comparable studies focused on prediction of inpatient mortality (table 8). Thus neonatal mortality may be easier to predict than inpatient mortality.

The method to find the optimal ANN input variable set still awaits standardisation despite recent advances in the field.^{33 34} It seems reasonable to use pruning or genetic algorithms for that purpose. We relied on the intrinsic variable selection mode of Predict,²¹ but—a feature inherent in genetic algorithms—we cannot be absolutely sure that we have found the variable set with the maximum predictive power.³⁵

There is no easy way to assess the relative impact of the individual ANN input variables. We favoured a pragmatic approach by training several ANNs, subsequently omitting one input of the optimal input variable set, and comparing the different performances.

Twenty five survivors with a high (> 0.50) ANN derived individual mortality risk underwent a matched pairs analysis with counterparts of lower risk (< 0.50) (table 3). All 50 preterm neonates were very immature (median gestational age 26.3 weeks, median birthweight about 880 g), and most of them survived but at the expense of major sequelae. The ANN identified a subgroup of preterm neonates with significantly higher morbidity when they survived. Individuals in this subgroup presenting with a predicted individual neonatal mortality risk of > 0.50 had a higher rate of cerebral com-

Table 7 Case analysis of non-survivors with ANN predicted individual mortality risk < 0.10

Variable	Case 1	Case 2	Case 3
Apgar 1/5/10 minutes	09/10/10	04/05/05	04/07/08
Birthweight (g)	1085	2005	2025
Capillary blood pH on admission	7.35	7.32	7.17
Condition on admission	Stable	Vitally endangered	Vitally endangered
Visible congenital malformation on admission	None	None	None
Emergency delivery	Yes	Yes	Yes
Gestational age (completed weeks)	27	31	31
Inborn	Yes	No	Yes
Body temperature on admission (°C)	36.5	36.2	36.2
Cause of death	<i>E coli</i> sepsis	Congenital myotonic dystrophia, lung hypoplasia, RDS IV° and pneumonia on postmortem examination	Complex congenital heart disease
ANN derived mortality risk	0.022	0.044	0.098
ANN-2i derived mortality risk	0.006	0.378	0.624

*ANN trained omitting the items gestational age and birth weight

Table 8 Comparison of different mortality prediction models

Patient group	Institutions	Patient group size - validation set included	Study period	Data collection period	End point	Sensitivity at 0.74 specificity	Area under ROC curve for validation sample	Name of score (reference)
Inborn 501-1500 g	7	3603	1987-9	Pre- and perinatal	Death	*0.77	0.82	Horbar 1993
< 1500 g and ventilated; patients who died in the first 12 h or had severe congenital malformations were excluded	2	262	1983-9	Pre-, peri-, and neonatal (12 h)	Death > 12 h	0.81		Tarnow 1990
Newborns; late readmitted and moribund patients excluded	3	1621	1989-90	Neonatal (24 h)	Death	*0.88	0.90 (0.82†)	SNAP (Richardson 1993)
Newborns; late readmitted and moribund patients excluded	3	1621	1989-90	Peri- and neonatal (24 h)	Death	*0.91	0.93 (0.79†)	SNAP-PE (Richards 1993)
< 1500 g or < 31 weeks gestation; lethal congenital malformations excluded	4	1201	1988-90	Pre-, peri-, and neonatal (12 h)	Death	*0.87	0.90	CRIB (INN 1993)
< 1501 g or < 32 weeks gestation	1	890	1990-6	Pre- and perinatal	Neonatal death	0.99	0.95	This article

#: about; † according to Rantonen *et al* (1994)³⁹ Prenatal: input data collection from conception until labour
Perinatal: input data collection from labour until admission (admission data included)
Neonatal: input data collection after admission (time span as indicated)

plications, and developed more conditions that eventually required surgical intervention. Those findings agree with those of recent studies on illness severity scores (SNAP, CRIB) where a positive association between high scoring and the presence of a major cerebral abnormality on ultrasound brain scan was found.^{7 11 36}

Although overall prediction of neonatal mortality is accurate enough for the purpose of intrahospital quality management and risk stratification, for the purpose of individual non-treatment decisions, scoring systems as well as the ANN still fail to differentiate between high risk preterm neonates who will eventually die and those who will survive for at least 28 days. Important factors contributing to preterm mortality risk were probably not included in the development of established scoring systems and in the training data of our ANN. And perhaps those contributing factors (for example, antioxidant status, etc) have not been measured yet.

One probable reason for the prediction failures in cases 2 and 3 of low risk non-survivors is the tendency of the ANN to underestimate the mortality risk the heavier and more mature the babies are. Our database lacked important information on physiological measures such as cord blood pH value, oxygen requirement in the first hours of life, or the presence of signs of infection. Hence the ANN overestimated the impact of immaturity on mortality. In case 1 sepsis started during day 1 of life and the baby died on day 3. The data collecting period required to detect physiological derangements would have been 12 to 24 hours—the same as in the CRIB or SNAP scores. Our admission data did not adequately reflect individual mortality risk in this case.

Several models that can predict mortality in preterm neonates have been developed.^{1-3 7-9 37} A performance comparison of these models proved difficult as the studies differ in respect to authors' intention, study population, number of institutions enrolled, data collecting period, items used to characterise the neonate, primary endpoint, and general applicability (table 8).

Although high impact items were missing in our database, and data collecting was restricted to the prenatal and immediate perinatal periods,

the ANN's performance is equivalent to established scoring systems. In contrast to most other mortality prediction models, we have opted for neonatal mortality instead of inpatient mortality as the endpoint. The reason for this is that individual diseases contributing substantially to mortality after 28 days (late necrotising enterocolitis, central catheter sepsis, respiratory syncytial virus infection, etc.) are unlikely to be associated with admission data. It should be mentioned that in contrast to published reports,^{2 3} surfactant treatment and high frequency oscillation, both of which can influence mortality,³⁸⁻⁴⁰ had already been introduced into routine care during the study period.

We relied on pre-existing admission data. We feel that information on important topics was missing in our database—for example, illness severity in the first hours of life. One of the items used was a very subjective measure of illness severity—the assessment of the overall condition on admission, as performed by the physician on duty. Stevens *et al*¹ and Richardson *et al*⁸ reported that the physician's and nurse's estimation of neonatal mortality risk was highly correlated with the actual mortality risk, or with the SNAP score. The assessment of the overall condition on admission had not been intended to serve as an estimate of mortality risk at the time our data were recorded. Notably, an ANN trained with those subjective assessments made by 30 different physicians delivered reliable results when fed with independent data of the subsequent validation time period, when overall condition on admission was assessed by another 30 physicians. The use of this kind of subjective input variable is a major limitation of the present study. Efforts are under way to substitute those variables with physiological measures, and to broaden the database. The aim is to train an ANN exclusively on objectively measurable physiological data, to render it broadly applicable to other neonatal intensive care units.

A feature intrinsic to an ANN is its retraining potential in case of advances in medical care, or simply in case of additional data becoming available. In contrast to multiple regression analysis, expanding the database allows the ANN to improve its knowledge using non-

linear relations between input and output data, whereas multiple regression at some point reaches its intrinsic limitations.²⁰ With multiple regression, the limit of accuracy is determined by two factors—namely, “noise” (missing, inaccurate, or false data), and the degree of non-linearity between the input and outcome variables. With the ANN approach it is noise alone that determines the limit of accuracy. As Richardson and Tarnow-Mordi have pointed out,¹¹ “as sophisticated systems for automatic acquisition of large quantities of routine clinical information become available, emphasis may shift in favour of more complex clinical scoring systems to maximise predictive accuracy.” Clearly, the ability of an ANN to use a very large number of variables from different sources (parameters, settings, and monitoring data from mechanical ventilators, ECG signals or clinical chemistry data) is a distinct advantage. Neural networks are able to include both quantitative and qualitative data into the same model. There are no limitations with respect to ordinal scaled data as there are in logistic regression analysis.

With the recent advances in hardware and software any clinician is able to develop ANNs to depersonalise his/her experience, and to make it accessible to junior colleagues. This can even happen under circumstances where rules are difficult to formulate and situations are too complex to be analysed by classic statistics.¹⁶

Trained ANNs predicting morbidity and mortality might help to assure high quality of care through comparison of predicted and actual outcome; to assess the time course of intrahospital advances in treatment; and to warn physicians and nurses, when individual infants are at high risk and deserve intensified attention. But prediction failures tell us that neonatal intensive care medicine is too complex to base an individual no-treatment policy on a prediction that might turn out to be very inaccurate for individual patients.

The help of Miss Ulla Ollesch and the staff of the neonatology department at the Vestische Kinderklinik for data entry is gratefully acknowledged. The expert linguistic advice of Heidi Walsh and Neil McIntosh is also very much appreciated. Special thanks to Charles E Metz who kindly provided the software for the ROC curve analyses.

Part of the data in this study are an integral part of Antje Westermann's thesis.

- 1 Horbar JD, McAuliffe TL, Adler SM, Albersheim S, Cassady G. Variability in 28-day outcomes for very low birthweight infants: an analysis of 11 neonatal intensive care units. *Pediatrics* 1988;**82**:554-9.
- 2 Horbar JD, Onstad L, Wright E. Predicting mortality risk for infants weighing 501-1500 grams at birth: A National Institute of Health Neonatal Research Network report. *Crit Care Med* 1993;**21**:12-18.
- 3 Tarnow-Mordi W, Ogston S, Wilkinson AR, et al. Predicting death from initial disease severity in very low birth weight infants: a method for comparing the performance of neonatal units. *BMJ* 1990;**300**:1611-4.
- 4 Verloove-Vanhorick SP, Verwey RA, Brand R, Gravenhorst JB, Keirse MJNC, Ruys JH. Neonatal mortality risk in relation to gestational age and birthweight. *Lancet* 1986;**i**:55-7.
- 5 Cooke RWI. Trends in survival and incidence of cerebral haemorrhage 1980-9. *Arch Dis Child* 1991;**66**:403-7.
- 6 Anonymous. The fate of the baby under 1501 g at birth. [Editorial]. *Lancet* 1980;**i**:461-3.
- 7 International Neonatal Network. The CRIB (critical risk index for babies) score: a tool for assessing initial neonatal risk and comparing the performance of neonatal intensive care units. *Lancet* 1993;**342**:193-8.
- 8 Richardson DK, Gray JE, McCormick MC, Workman-Daniels K, Goldmann DA. Score for neonatal acute physiology: a physiology severity index for neonatal intensive care. *Pediatrics* 1993;**91**:617-23.
- 9 Richardson DK, Phibbs CS, Gray J, McCormick MC, Workman-Daniels K, Goldmann DA. Birth weight and illness severity: independent predictors of neonatal mortality. *Pediatrics* 1993;**91**:969-75.
- 10 Maier RF, Rey M, Metz BC, Obladen M. Comparison of mortality risk: a score for low birthweight infants. *Arch Dis Child* 1997;**76**:F146-F51.
- 11 Richardson DK, Tarnow-Mordi WO. Measuring illness severity in newborn intensive care. *J Intensive Care Med* 1994;**9**:20-3.
- 12 Boyd O, Grounds RM. Physiological scoring systems and audit. *Lancet* 1993;**341**:1573-4.
- 13 Hertz J, Krogh A, Palmer R. *Introduction to the theory of neural computing*. Massachusetts: Addison Wesley, 1991.
- 14 Rumelhart DE, Hinton GE, Williams RJ. Learning internal representation by error propagation. In: Rumelhart DE, McClelland JL, eds. *Parallel distributed processing: Exploration in the microstructure of cognition*. Massachusetts: MIT Press, 1986:318-64.
- 15 Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet* 1995;**346**:1135-8.
- 16 Baxt WG, Skora J. Prospective validation of artificial neural network trained to identify acute myocardial infarction. *Lancet* 1996;**347**:12-15.
- 17 Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995;**346**:1075-9.
- 18 Dybowski R, Gant V. Artificial neural networks in pathology and medical laboratories. *Lancet* 1995;**346**:1203-7.
- 19 Forsström JJ, Dalton KJ. Artificial neural networks for decision support in clinical medicine. *Ann Med* 1995;**27**:509-17.
- 20 Smith M. *Neural networks for statistical modeling*. New York: Van Nostrand Reinhold, 1993.
- 21 Anonymus. *Predict Manual*. Pittsburgh: Neural Ware Inc., 1995.
- 22 Jacobs RA. Increased rates of convergence through learning rate adaption. *Neural Networks* 1988;**1**:295-308.
- 23 Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;**8**:283-6.
- 24 Hanley JA, McNeil BJ. The meaning and use of the area under the receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29-36.
- 25 Metz CE, Wang PL, Kronman HB. A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck F, ed. *Information processing in medical imaging*. The Hague: Martinus Nijhoff, 1984:432-45.
- 26 Ebell MH. Artificial neural networks for predicting failure to survive following in-hospital cardiopulmonary resuscitation. *J Fam Pract* 1993;**36**:297-303.
- 27 Dybowski R, Weller P, Chang R, Gant V. Predicting outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *Lancet* 1996;**347**:1146-50.
- 28 Phelbs DL, Brown DR, Tung B, et al. 28-day survival rate of 6676 neonates with birth weights of 1250 grams or less. *Pediatrics* 1991;**87**:7-17.
- 29 Rantonen J, Mäkelä A, Boyd H, Apajala M, Pohjavuori M. CRIB and SNAP: assessing the risk of death for preterm neonates. *Lancet* 1994;**343**:1272-3.
- 30 Perlman M, Claris O, Hao Y, et al. Secular changes in the outcome to eighteen to twenty four months of age of extremely low birth weight infants, with adjustment for change in risk factors and severity of illness. *J Pediatr* 1995;**126**:75-87.
- 31 Stevens SM, Richardson DK, Gray JE, Goldmann DA, McCormick MC. Estimating neonatal mortality risk: an analysis of clinicians' judgments. *Pediatrics* 1994;**93**:945-50.
- 32 Schmidt B, Kirpalani H, Rosenbaum P, Cadman D. Strengths and limitations of the Apgar score: a critical appraisal. *J Clin Epidemiol* 1988;**41**:843-50.
- 33 Tetko IV, Villa AEP, Livingstone DJ. Neural network studies. 2. Variable selection. *J Chem Inf Comput Sci* 1993;**36**:794-803.
- 34 Reed R. Pruning algorithms - a survey. *IEEE Trans Neural Networks* 1993;**4**:740-7.
- 35 Mitchie D, Spiegelhalter DJ, Taylor C, eds. *Machine learning, neural nets and statistical classification*. Chichester: Ellis Horwood, 1994:213-27.
- 36 Gray JE, Richardson DK, McCormick MC, Workman K, Goldmann DA. Score for Neonatal Acute Physiology (SNAP) and risk of intraventricular hemorrhage (IVH). *Pediatr Res* 1992;**31**:249A.
- 37 Patterson CC, Halliday HL. Prediction of outcome after delivery for the very low birth weight (<1500g) infant. *Pediatr Perinat Epidemiol* 1988;**2**:221-8.
- 38 Corbet A, Gerdes J, Long W, et al and American Exosurf Neonatal Study Groups I and II. Double-blind, randomized trial of one versus three prophylactic doses of synthetic surfactant in 826 neonates weighing 700 to 1100 grams: effect on mortality rate. *J Pediatr* 1995;**126**:969-78.
- 39 Egberts J, Brand R, Walth H, Bevilacqua G, Breat G, Gardini F. Mortality, severe respiratory distress syndrome, and chronic lung disease of the newborn are reduced more after prophylactic than after therapeutic administration of the surfactant Curosurf. *Pediatrics* 1997;**100**. Pages URL: <http://www.pediatrics.org/cgi/content/full/100/1/e4>.
- 40 Gerstmann DR, Minton SD, Stoddard RA, et al. The provo multicenter early high-frequency oscillatory ventilation trial: improved pulmonary and clinical outcome in respiratory distress syndrome. *Pediatrics* 1996;**98**:1044-57.