NDT
Nephrology Dialysis Transplantation

Original Article

Predicting technique survival in peritoneal dialysis patients: comparing artificial neural networks and logistic regression

Navdeep Tangri[1], David Ansell[2] and David Naimark[3]

[1]Department of Internal Medicine, McGill University, Montreal, QC, Canada, [2]United Kingdom Renal Registry, Bristol, UK and
[3]Division of Nephrology, Sunnybrook Health Sciences Centre and the Departments of Medicine and of Health Policy Management
and Evaluation, University of Toronto, Toronto, ON, Canada

Abstract

Background. Early technique failure has been a major limitation on the wider adoption of peritoneal dialysis (PD). The objectives of this study were to use data from a large, multi-centre, prospective database, the United Kingdom Renal Registry (UKRR), in order to determine the ability of an artificial neural network (ANN) model to predict early PD technique failure and to compare its performance with a logistic regression (LR)-based approach.

Methods. The analysis included all incident PD patients enrolled in the UKRR from 1999 to 2004. The event of interest was technique failure. For both the ANN and LR analyses a bootstrap approach was used: the data were divided into 20 random training (75%) and validation (25%) sets. Models were derived on the latter and then used to make predictions on the former. Predictive accuracy was assessed by area under the ROC curve (AUROC). The 20 AUROC values and their standard errors were then averaged.

Results. There were 3269 patients included in the analysis with a mean age of 59.9 years and a mean observation time of 430 days. Of the patients, 38.3% were female and 90.8% were Caucasian. 1458 patients (44.6%) suffered technique failure. The AUROC for the ANN model was 0.760 ± 0.0167 and the LR model was 0.709 and 0.0208. ($P = 0.0164$)

Conclusions. Using UKRR data, both ANN and LR models predicted early PD technique failure with moderate accuracy. In this study, an ANN outperformed an LR-based approach. As the scope and the completeness of the UKRR increases, the question of whether more sophisticated ANN models will perform even better remains for further study.

Keywords: artificial neural networks; early technique failure; logistic regression; peritoneal dialysis; technique survival

Introduction

End-stage renal disease (ESRD) prevalence continues to rise worldwide [1]. Peritoneal dialysis (PD) is a clinically and economically attractive mode of therapy for ESRD [2]. The survival of patients treated with PD is equivalent to those who receive haemodialysis and PD has been associated with a greater self-reported quality of life [3,4]. Most patients have no medical contraindications to either haemo- or peritoneal dialysis and are free to choose a modality on the basis of social and logistical considerations [5]. Despite the advantages of PD, the proportion of the ESRD population who have adopted this modality is declining in both Europe and North America [1,6]. The decrease in the prevalence of PD patients may be due to the emergence of factors that limit its adoption. For example, as the median age and degree of co-morbid illness increase among incident ESRD patients, the fraction of patients with significant social and/or logistical barriers to PD adoption also increases [7]. Early technique failure is another major constraint on the growth of PD as a treatment option. Technique failure necessitates a switch to haemodialysis that increases costs and decreases patient self-reliance and social flexibility.

Factors affecting technique survival in PD have been studied using single centre and registry data [8–12]. These studies have used regression methods to determine the relative importance of a variety of factors on the risk of early technique failure in groups of patients. Models that combine these factors in order to predict early PD technique failure for individual patients are lacking in the literature. An accurate prediction model would be a potentially useful way of identifying patients at particularly high risk of early technique failure so that increased clinical scrutiny and timely intervention could be brought to bear. Yet, predicting early technique failure is difficult due to the myriad medical and social factors that may influence the outcome. These factors may also have a non-linear relationship with early technique failure and may be subject to complex variable interactions.

Artificial neural networks (ANNs) are a relatively new class of statistical prediction tools that are particularly
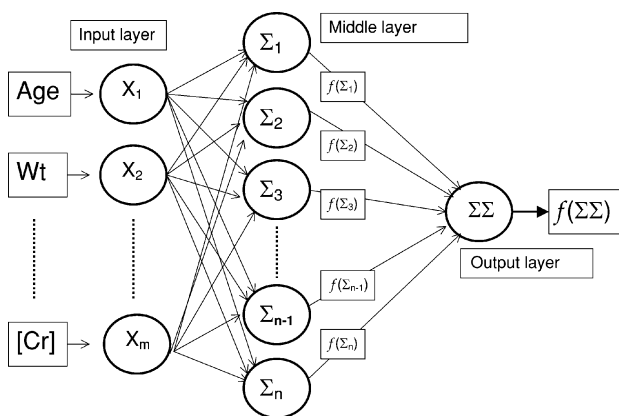
Fig. 1. Artificial neural network (ANN) architecture. ANNs consist of artificial neurons. Each artificial neuron has a processing node ('body') represented by circles in the figure as well as connections from ('dendrites') and connections to ('axons') other neurons which are represented as arrows in the figure. In a commonly used ANN architecture, the multilayer perceptron, the neurons are arranged in layers. An ordered set (a vector) of predictor variables is presented to the input layer. Each neuron of the input layer distributes its value to all of the neurons in the middle layer. Along each connection between input and middle neurons there is a connection weight so that the middle neuron receives the product of the value from the input neuron and the connection weight. Each neuron in the middle layer takes the sum of its weighted inputs and then applies a non-linear (usually logistic) function to the sum. The result of the function then becomes the output from that particular middle neuron. Each middle neuron is connected to the output neuron. Along each connection between a middle neuron and the output neuron there is a connection weight. In the final step, the output neuron takes the weighted sum of its inputs and applies the non-linear function to the weighted sum. The result of this function becomes the output for the entire ANN. More details are provided in the appendix.

suited to complex pattern recognition tasks (Figure 1 and the appendix) [13]. ANNs have the advantage of automatically detecting and modelling complex non-linear relationships between 'inputs' to the network (i.e. patient demographic, clinical and laboratory data) and the 'output' (i.e. early technique failure) and can consider all possible interactions between the input variables. In contrast, conventional, regression-based, methods require non-linear relationships between input and output variables to be specified *a priori*. Interactions must be pre-specified in regression analyses and relatively few of these can be accommodated [14,15]. ANNs have been used successfully as a prediction tool in a variety of medical and non-medical situations [13]. In nephrology, ANNs have been used successfully to screen for glomerulopathy using urine biomarkers, to predict erythropoietin responsiveness, to stratify PD membrane characteristics and to predict delayed renal allograft dysfunction [16–25].

The United Kingdom Renal Registry (UKRR) is a large, comprehensive, validated and prospective data source that includes information from the majority of the ESRD patients in the United Kingdom [6]. In the current analysis, we used data from the UKRR in order to determine the predictive performance of ANN models and to compare the ability of the ANN approach with a traditional logistic regression (LR) model to predict early PD technique failure.

Methods

Description of the data source

The UKRR is operated under the auspices of the UK Renal Association and provides independent audit and analysis of renal care in the UK. The UKRR data collection methods have been described in detail elsewhere [6]. In brief, renal units using registry-compatible information systems are required to electronically export data to the UKRR on a quarterly basis. Local software extraction routines identify all patients on dialysis or with a renal transplant and gather a predefined dataset which includes socio-demographic data, ESRD diagnosis, any modality changes during the current quarter, date of death, transfers to other centres and 3-monthly recordings of weight, blood pressure and laboratory parameters. Data arriving at the UKRR are subject to algorithms that identify incongruent values that are then verified with the renal units and corrected if required. Completeness of returns approaches 100% for primary diagnosis-related information and >60% overall for other data supplied by the renal units [6].

Data from 60 renal units servicing a population base of 53.4 million were included in the current analysis.

Subjects

The present analysis included all incident dialysis patients older than the age of 18 in the UKRR who started PD from 1 January 1999 until 31 December 2004. Patients were considered to have selected PD as their initial dialysis modality if they were receiving PD at 90 days after starting renal replacement therapy.

Data abstraction

Data abstracted for each eligible subject included the dates of dialysis initiation, transition to haemodialysis, transplantation, loss to follow-up and/or death. The primary outcome of interest was PD technique failure. Patients who died or were lost to follow-up while on PD, received a renal transplant or who remained on PD until 31 December 2004 were defined as technique survivors. Technique failure was defined as a change in dialysis modality to haemodialysis for a period exceeding 1 month. For patients who changed dialysis modality more than once, the starting date of the first period of haemodialysis that lasted greater than 1 month was considered to be the technique failure date. For each subject, an end date was defined as the date of the earliest event among the possible outcomes of remaining on PD until 31 December 2004, technique failure, transplantation, loss to follow-up or death. The observation time for each subject was then calculated as the number of days between dialysis initiation and the end date. The outcome variable, 'PD technique failure', was coded as '1' if the end date corresponded to technique failure and a '0' for all other outcome events. A Kaplan–Meier cumulative probability curve for PD technique failure was plotted with SPSS version 15.0, Chicago, IL, USA.

The subset of potential predictor variables abstracted from the UKRR included demographic, clinical and

Table 1. Summary of the predictor variables included in the artificial neural network and logistic regression analyses for subjects who did not and who did suffer peritoneal dialysis technique failure.

Variable	Technique survival Mean ± SD or % (N)	Technique failure Mean ± SD or % (*N*)	*P*-value
Observation time (days)	617 ± 494 (1811)	299 ± 344 (1458)	<0.001*
Demographics			
Age (years)	60.9 ± 15.8 (1811)	55.8 ± 16 (1458)	<0.001*
Female sex	38.5 (1811)	38 (1458)	0.775
Caucasian race	91.1 (1383)	90.6 (1242)	0.641
Cause of ESRD			
Diabetes	22.1 (1811)	18.2 (1458)	0.005
Glomerulopathy	12.1 (1811)	15.9 (1458)	0.002
Renovascular disease	12.9 (1811)	11.8 (1458)	0.355
PCKD	5.6 (1811)	7.6 (1458)	0.021
Pyelonephritis	6.5 (1811)	8 (1458)	0.088
Other	12.2 (1811)	12.1 (1458)	0.956
Unknown	28.7 (1811)	26.3 (1458)	0.139
Comorbid conditions			
Sympt. CVD	11.5 (615)	8.3 (504)	0.072
Angina	24.1 (615)	15.5 (502)	<0.001*
Past MI	14.5 (615)	11.3 (504)	0.114
Past CABG	7.8 (614)	5.4 (504)	0.096
Angioplasty	4.9 (613)	2.8 (498)	0.069
PVD	3.1 (613)	2.2 (501)	0.346
Leg ulcer	4.9 (610)	3.6 (501)	0.272
Claudication	15.7 (613)	9.6 (500)	0.002
Smoking	20.5 (599)	18.8 (490)	0.467
COPD	7 (613)	5.1 (505)	0.190
Diabetes	10.3 (610)	10.6 (502)	0.901
Malignancy	9.5 (612)	9.5 (504)	0.979
Liver disease	1.6 (615)	2.2 (502)	0.495
Physical examination			
Weight (kg)	69.8 ± 14.3 (497)	71.4 ± 14.6 (438)	0.091
Height (cm)	168 ± 10 (431)	169 ± 10 (386)	0.207
SysBP (mmHg)	141 ± 26 (819)	144 ± 25 (701)	0.051
DiasBP (mmHg)	79 ± 14.5 (819)	82.1 ± 14.1 (701)	<0.001*
Laboratory data			
Calcc (mmol/L)	2.43 ± 0.21 (1685)	2.43 ± 0.21 (1380)	0.827
Phos (mmol/L)	1.59 ± 0.49 (1667)	1.61 ± 0.46 (1357)	0.159
Alb (g/L)	32 ± 6 (1701)	33.5 ± 5.1 (1376)	<0.001*
IPTH (pmol/L)	26 ± 28.3 (909)	25.6 ± 29.4 (789)	0.756
Creat (µmol/L)	608 ± 211 (1737)	664 ± 226 (1405)	<0.001*
Urea (mmol/L)	18.7 ± 7.2 (1727)	18.9 ± 6.4 (1397)	0.459
Haem (g/dL)	11.1 ± 1.7 (1715)	11 ± 1.7 (1377)	0.425
Ferritin (µg/L)	323 ± 497 (1387)	292 ± 377 (1182)	0.082
Cholesterol (mmol/L)	5.22 ± 1.28 (787)	5.3 ± 1.45 (629)	0.330
Bicarbonate (mmol/L)	26.3 ± 4.2 (1534)	26.1 ± 3.8 (1212)	0.112
HbA1c (%)	7.48 ± 1.86 (176)	7.51 ± 1.91 (164)	0.893
Aluminium (µmol/L)	0.318 ± 0.634 (305)	0.307 ± 0.47 (307)	0.804

Continuous variables are shown as means ± standard deviations (SD) while categorical variables are shown as percentages (%). The number of subjects on which a mean value or percentage was calculated is given in the parentheses. Statistical significance was computed with independent *t*-tests for continuous variables and chi-square tests for categorical variables. The nominal significance level of 0.05 was Bonferroni-adjusted for the number of tests such that a *P*-value of 0.00125 or less (*) was considered to be significant. For each subject, blood pressure values, weight and laboratory values represent measurements closest to the date of peritoneal dialysis initiation while height values are the average of all available measurements.
ESRD = end-stage renal disease, CVD = cardiovascular disease, MI = myocardial infarction, CABG = coronary artery bypass grafting, PVD = peripheral vascular disease, COPD = chronic obstructive pulmonary disease, SysBP = systolic blood pressure, DiasBP = diastolic blood pressure, Calcc = calcium concentration adjusted for albumin, Phos = phosphate concentration, Alb = albumin concentration, iPTH = intact parathyroid hormone concentration, Creat = creatinine concentration, Haem = haemoglobin concentration and HbA1c = haemoglobin A1c percentage.

laboratory variables (Table 1). The demographic variables included the date of birth, and binary indicators of gender and Caucasian race. The age at dialysis initiation was calculated as the number of days between the dates of birth and dialysis initiation divided by 365.25.

The clinical variables included the aetiology of ESRD that was encoded as a set of mutually exclusive binary indicators for diabetic nephropathy, glomerulonephritis, reno-

vascular disease, polycystic kidney disease, pyelonephritis, other and unknown causes. The following binary indicators of co-morbid illnesses were included in the analysis: diabetes mellitus (excluding subjects already categorized as having diabetic nephropathy), symptomatic cardiovascular disease, angina pectoris, past myocardial infarction, a history of coronary artery bypass surgery, a history of angioplasty, peripheral vascular disease and/or non-traumatic

lower limb amputation, lower limb ulceration, claudication, past or present smoking, chronic obstructive pulmonary disease, known malignancy and liver disease. Dialysis centre was indicated by a set of mutually exclusive binary indicators for each renal unit with at least 20 prevalent PD patients on 31 December 2004. Subjects belonging to centres with fewer than 20 PD patients on that date were assigned a generic binary centre indicator. Patients were assigned to the dialysis unit where their PD was initiated regardless of subsequent migration to another renal unit. A centre-size variable was assigned to each subject that was equal to the number of patients served by their assigned renal unit on 31 December 2004. For patients who were assigned the generic centre code, the sum of the patients served by the units included in the generic code was employed as their centre-size value. The measurements of systolic and diastolic blood pressure and weight closest to the date of dialysis initiation were chosen for inclusion in the analysis while the value for height was the average of all available measurements for each subject.

The quarterly laboratory data closest to the date of dialysis initiation were included in the analysis. The laboratory variables that were abstracted included the concentrations of creatinine, urea, calcium, phosphate, intact parathyroid hormone, bicarbonate, albumin, total cholesterol, ferritin and haemoglobin (Table 1). Each calcium value was corrected for albumin using the formula CorrCa = Ca + [(40−Alb)×0.025]. Clinical and laboratory data were compared between patients with technique survival and failure using t-tests and chi-squared tests for continuous and categorical variables, respectively, using SPSS version 15.0, Chicago, IL, USA.

Each laboratory variable was evaluated for normality by visual inspection of histograms and by normal plots. Skewed variables were transformed and re-evaluated for normality: the ferritin level was transformed by the log10 function and the intact PTH and aluminium values were transformed by the natural logarithm function. For binary input variables, missing values were imputed by replacing the value with the proportion of positive cases across all subjects in whom the value of the binary variable was not missing. A given missing continuous input variable was imputed with a multiple regression model such that the missing variable was considered as the dependent and the rest of the continuous variables were considered as the independent variables. For the purpose of missing data imputation, a set of such regression models was created for each of the continuous input variables.

ANN bootstrap procedure

Multilayer perceptron ANNs with 40-80-1 nodal architectures were constructed and trained using the back propagation approach with Neuroshell 2 version 3.0. (Ward Systems Group, Frederick, MD, USA). In order to enhance ANN training, by eliminating inputs with the value '0', all input factors were transformed to values between 1 and 2 using the equation $x' = [(x - \min(x))/(\max(x) - \min(x))] + 1$ where $\min(x)$ and $\max(x)$ are the minimum and maximum of the input variable 'x' across all subjects.

The predictive performance of the application of the ANN approach to the analysis of PD technique failure was determined with a bootstrap approach [26]. For each of 20 bootstrap iterations, 75% of the data (∼2450 cases) were randomly selected and used to train a network. The training of the ANN was stopped when the average difference between the known outcome of the training cases (2 for event and 1 for no event) and the predicted outcomes from the ANN (numbers between 1 and 2) converged to a pre-set minimum (see the appendix). The trained ANN was then used to make predictions on a validation set consisting of the remaining 25% of cases in the dataset. Twenty random training and validation sets and ANNs were created in this way.

The accuracy of the 20 sets of predictions was each assessed by the area under the receiver operating characteristic curve (AUROC). An AUROC of 1.0 implies perfect discrimination between cases and controls in the validation set while a value of 0.5 indicates no predictive ability. The AUROC was computed using CLABROC software version 1.9.1 [27,28]. The 20 AUROC values and their standard errors were then averaged.

Using the 'slope' parameters provided by the CLABROC software for each of the 20 validation sets, the optimum thresholds for discriminating between patients with PD technique success and failure were calculated (see the appendix) [29,30]. Using these thresholds, the resulting sensitivity, specificity, positive predictive value and negative predictive values were calculated for each bootstrap sample. In addition, the classification accuracy was calculated for each sample as sum of the number of true positives and true negatives divided by the total number of patients in the validation set. For a given bootstrap sample, the improvement in accuracy beyond that which would be expected by chance was computed as the ratio of the observed classification accuracy to the accuracy expected by chance.

Logistic regression bootstrap procedure

For the LR analyses, the outcome of interest was the development of technique failure within 1 year of starting PD. Patients who were observed for a period shorter than 1 year and who were censored (functioning PD on 31 December 2004, death or loss to follow-up with functioning PD or transplant) were excluded from the LR analyses (704/3269). The data transformation that converted inputs and outputs to a range between 1 and 2 for the ANN training and validation was not undertaken for the LR analyses.

Otherwise, a similar strategy was employed for the LR bootstrap. Twenty random samples consisting of 75% of the cases were each used to derive a LR model. Each of the 20 models incorporated all the potential predictor variables that were used to train the ANNs without any interaction terms. For each of the 20 regression equations, the intercept parameter and model coefficients were then used to make predictions on the remaining 25% of cases in the dataset. Twenty random 'training' and validation sets and LR models were created in this way. The average AUROC statistic and its standard error were computed as described above. Likewise, the sensitivity, specificity, positive
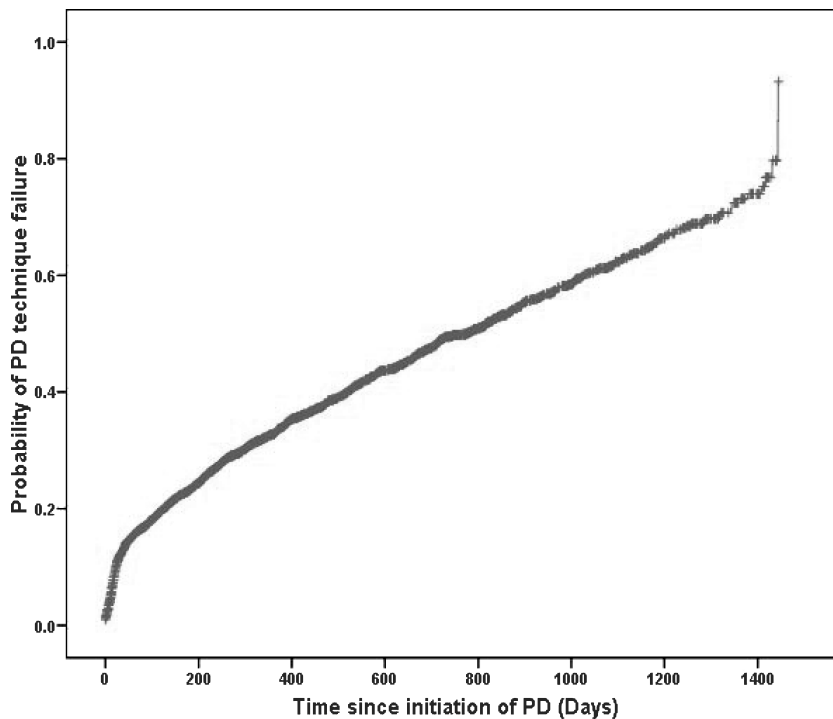
Fig. 2. The Kaplan–Meier curve for probability of peritoneal dialysis (PD) technique failure after the initiation of PD. Survival until 31 December 2004, death, transplantation or loss to follow-up with functioning PD were considered to be censored observations.

predictive value and negative predictive values, and the classification accuracy were calculated as described above.

Comparison of the ANN and logistic models

The 20 ROC curves from the ANN bootstrap were compared with the 20 from the logistic bootstrap that yielded 400 paired comparisons. For each comparison, the ratio of the difference in the areas of the ANN and logistic ROC curves to the standard error of the difference yielded a normally distributed z-statistic and a two-sided P value [31]. The overall significance of the difference in AUROC for the ANN and logistic bootstrap samples was taken as the average of the P values for the 400 pairs.

Results

Patient characteristics

A Kaplan–Meier plot of the cumulative probability of PD technique failure as a function of time since the initiation of PD is shown in Figure 2. Baseline demographic and laboratory characteristics of the patients are presented in Table 1. The mean age of the patients was 59.9 years. Technique survivors were, on average, 5 years older than patients who suffered technique failure ($P < 0.001$). The majority of the patients were Caucasian and 38% were female. The mean observation time was 430 days. Forty-five percent of the patients suffered from technique failure during the observation period. Patients who failed PD had higher values for diastolic blood pressure, serum creatinine and albu-

min (all P values < 0.001). Subjects who suffered from PD technique failure were less likely to have had a previous myocardial infarction ($P < 0.001$). After Bonferroni correction for multiple testing, there were no other significant differences between the predictor variables in the two groups of patients.

ANN bootstrap results

The results for the bootstrap iterations are shown in Table 2. The average AUROC and standard error of the AUROC were 0.760 and 0.0167, respectively. Each AUROC calculation using the CLABROC software yielded two parameters, which, when averaged over the 20 samples, allowed for the construction of an average receiver operating characteristic curve as shown in Figure 3. One of the 20 validation sets was chosen at random in order to construct a histogram comparing the ANN outputs for patients who suffered from technique failure versus those who did not (Figure 4). The average of the optimal thresholds was 1.46 that yielded average sensitivity, specificity, positive predictive value and negative predictive value of 70, 68, 64 and 74%, respectively. Using the optimum threshold the average classification accuracy in the validation set was 69% whereas the expected accuracy by chance was 51%. This represents a 37% improvement in classification accuracy beyond chance ($P < 0.0001$).

Logistic regression bootstrap results

The results for the logistic bootstrap iterations are shown in Table 2. The average AUROC and standard error of the

Table 2. Mean predictive performance for 20 artificial neural network (ANN) models and logistic regression models created using a bootstrap approach.

Model	Artificial neural network	Logistic regression
AUROC	0.7602	0.7090*
Standard error	0.0167	0.0208
Optimal threshold	1.4627	0.4016
Sensitivity	0.7043	0.6021
Specificity	0.6818	0.6856
Positive predictive value	0.6392	0.5469
Negative predictive value	0.7421	0.7320
IABC	1.3655	1.2411

*$P = 0.016$.

AUroc = area under the receiver operating characteristic curve (a value of 1.0 implies perfect discrimination between PD technique failure and success whereas 0.5 implies no discrimination); Optimal threshold. = the optimal threshold ANN output value that maximizes sensitivity and specificity; IABC = improvement in accuracy beyond chance (the ratio of the observed number of true positive plus true negative cases at the optimal threshold to the number expected by chance).
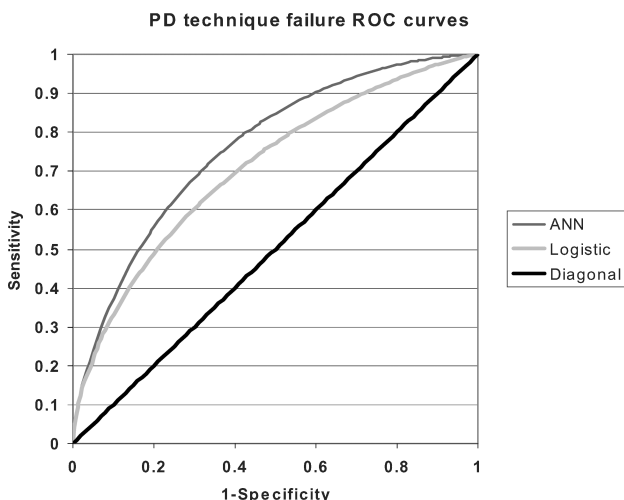

Fig. 4. Histogram of the artificial neural network (ANN) model output when applied to the validation set for subjects who did and did not suffer PD technique failure. For each of 20 bootstrap samples, the data were randomly divided into a training set from which an ANN model was derived, and a validation set on which the ANN was validated. The histogram data represent one of the 20 sets of validation set predictions selected at random.

Fig. 3. Receiver-operating characteristic (ROC) curves for the artificial neural network (ANN) and logistic regression bootstrap analyses (see the text for a description of the bootstrap procedure). The curves represent the average curves for the 20 ANN and 20 logistic models. The area under the ROC curve (AUROC) is an index of predictive performance: an AU-ROC of 1.0 represents perfect discrimination while a value of 0.5 indicates no discrimination between subjects with and without PD technique failure. The average AUROC values for the ANN and logistic regression models were 0.760 and 0.709, respectively ($P = 0.0164$).

Fig. 5. Histogram of the logistic regression model output when applied to the validation set for subjects who did and did not suffer PD technique failure. For each of 20 bootstrap samples, the data were randomly divided into a 'training set' from which a regression model was derived, and a validation set on which the regression model was validated. The histogram data represent one of the 20 sets of validation set predictions selected at random.

AUROC were 0.709 and 0.0208, respectively. As described above, the average receiver operating characteristic curve is shown in Figure 3. A histogram of the distribution of the predictions for a randomly selected LR model for subjects who actually did and did not suffer from technique failure in the first year of PD is shown in Figure 5. For the logistic models, the average of the optimal thresholds was 0.40 that yielded average sensitivity, specificity, positive predictive value and negative predictive value of 60, 68, 55 and 74%, respectively. Using the optimum threshold the average classification accuracy in the validation set was 65% whereas the expected accuracy by chance was 53%. This
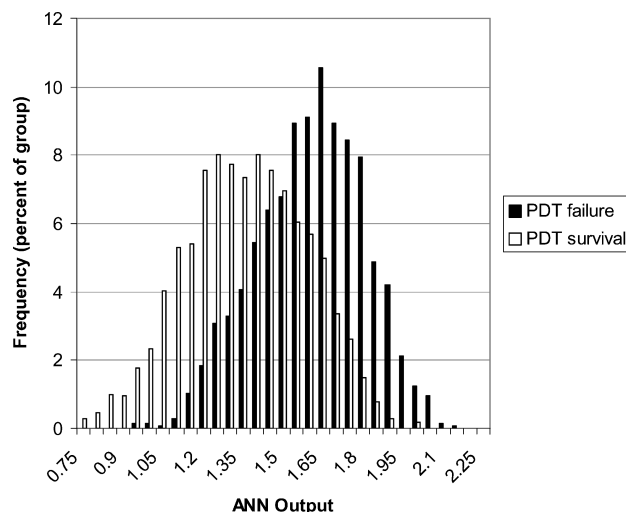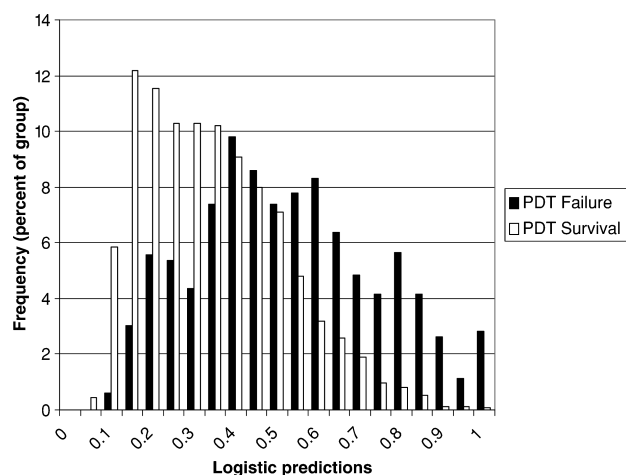
represents a 24% improvement in classification accuracy beyond chance ($P < 0.0001$).

Comparison of the ANN and logistic regression models

Overall, the ANN models performed better than the logistic ones. The average difference in the AUROC values for the ANN and LR models was 0.0512 (Figure 3, Table 2). In order to put this gain into perspective, consider that the maximum possible improvement in predictive performance for the average ANN model would be $1 - 0.709 = 0.291$. Thus the observed gain in performance represents 17.6%

of the theoretical maximum. The *P* value averaged over all 400 possible comparisons between ANN and logistic ROC curves was 0.0164.

Discussion

In the United Kingdom, 24% of patients starting renal replacement therapy choose PD as their initial treatment modality [6], with PD being twice as common in patients who are under the age of 65 compared with those who are older. Early technique failure with PD remains a significant problem in the United Kingdom ESRD population. In the cohort of patients who were included in the present study, 45% developed technique failure over a mean observation period of 430 days. This high rate is consistent with data from other large renal registries [10,11]. Thus, early technique failure is a major impediment to the growth of PD as a treatment option globally. The premise underlying the current study was that the accurate identification of patients at particularly high risk of early technique failure at the initiation of PD would allow for greater clinical scrutiny and timely intervention in order to forestall the outcome.

Previous investigators have used regression methods in order to identify factors that influence PD technique survival in groups of patients. For example, McDonald *et al.* found a significant relationship between body-mass index and early technique failure using data from the ANZDATA registry in Australia and New Zealand [10]. Likewise, Tonelli found that aboriginal ethnicity had a significant, independent effect on PD mortality and technique survival in Canada [9]. Huisman *et al.* studied the impact of centre effect using data from the Dutch renal registry and found that the number of PD patients treated in a renal unit was inversely related to the probability of early technique failure [11]. Although these investigations have provided very important contributions to our understanding of the nature of early PD technique failure, they were not designed with the goal of predicting the likelihood of this outcome for individual patients.

The UKRR presents a unique opportunity for the development of predictive models. The registry is a large repository of data that is subject to stringent quality control [6]. The automated and electronic submission from the participating renal units ensures that information regarding all patients receiving renal replacement therapy is captured prospectively. We hypothesized that the combination of the high-quality data contained in the UKRR combined with a sophisticated prediction method, the ANN (Figure 1 and the appendix), would be able to predict early PD technique failure accurately. Furthermore, a secondary hypothesis was that the ANN method would perform better than a traditional, LR-based prediction model.

We found that application of an ANN to the UKRR dataset predicted PD technique survival with moderate accuracy (AUROC 0.760). This can be understood intuitively to mean that, given two patients, one who ultimately will suffer PD technique failure and one who will not, our average ANN model will produce a higher score for the former patient 76% of the time [32,33]. If one were to use the optimal threshold, a clinically and statistically signifi-

cant improvement in classification accuracy beyond that expected by chance would be observed. The average AUROC value observed in the current study compares favourably with previous ANN-based prediction models in medical applications such as predicting psychosis outcomes, predicting response to chemotherapy and classifying tumours (AUROCs 0.70–0.91) [13,34,35]. In nephrologic applications, such as screening for glomerulopathy using urine biomarkers, predicting erythropoeitin responsiveness, stratifying PD membrane characteristics and predicting delayed renal allograft dysfunction, AUROC values ranged from 0.65 to 0.95 and sensitivities and specificities ranged from 64 to 92% and 65 to 92%, respectively [16–25].

ANNs have distinct advantages compared to the more familiar LR models. Logistic models assume linear behaviour which means that as the value of a given predictor variable increases, the predicted risk of the outcome increases. However, non-linear, '*U*-shaped', relationships between predictor variables and outcome risk have been noted in other areas of nephrology such as the effect of serum biochemical markers–urea, potassium, bicarbonate, phosphate, and cholesterol–on the mortality risk of haemodialysis patients [36–38]. Logistic models can accommodate non-linear behaviour by first transforming the variable using a logarithmic or polynomial function, but the analyst must know *a priori* that the non-linearity exists and also which transforming function to apply.

In general, ANN-based prediction models have outperformed LR-based ones in medical applications [13]. For example, Green *et al.* found that ANNs were superior to LR in the prediction of acute myocardial infarction (AUROC values: ANN = 0.811, LR = 0.764, *P* = 0.03) [34]. Studies that have assessed the performance of logistic models in other areas of nephrology—such as predicting the progression to ESRD among patents with chronic kidney disease—have yielded disappointing results. For example, Hemmelgran *et al.* applied a regression model to a cohort of 10 184 elderly patients to predict rapid progression of CKD and found an AUROC of 0.59 in their validation set [39]. The results of the present study are consistent with this theme: we found the ANN approach to be superior to an LR approach for the outcome of PD technique survival (AUROC 0.760 versus 0.709, *P* = 0.0146). Comparing the distributions of model outputs (Figures 4 and 5), the logistic predictions did not discriminate between subjects who did and did not suffer technique failure as cleanly as the ANN predictions did.

The current study has limitations. There was an inherent bias in the selection of the study cohort since the subjects had already chosen PD as a modality. This may limit the ability to apply the prediction models to pre-dialysis patients who may be considering all forms of renal replacement therapy. The UKRR is a superb data source; however, some values were missing for the predictor variables used in the study. For example, only about one-third of the subjects had information regarding co-morbid illnesses. The ANN and logistic models lacked information on residual renal function and peritoneal membrane fluid and solute clearance characteristics that may have improved their predictive ability [8]. Information regarding the aetiology of technique failure was also not available. The predictive performance

may have been improved by the use of a more refined outcome, that is knowledge of not only when PD technique failure occurred but why. However, the fact that the ANN models were able to achieve a respectable performance despite these data limitations provides a basis for optimism that, when such data become available in the UKRR, the performance of future models will improve substantially.

Whether or not the performance of the ANN models improves, there are some practical issues regarding their implementation in a PD clinic. The ANN models included observation time as an input variable. This would seem to preclude the use of the ANN approach in the clinic since the observation time for a given incident PD patient cannot be known *a priori*. However, this is not really an issue because a fixed time could be selected (such as 1, 2 or 5 years) and entered as an input to the trained ANN in order to produce predictions for that time horizon. Another potential concern is that, unlike LR-based prediction models, the output from an ANN is not a probability *per se*, but, rather, a risk score. To make the output of an ANN model comprehensible to health care providers and patients, it would have to be recalibrated as a probability value. However, even in its raw form, the output of an ANN could be used, along with a threshold value, to help clinicians to make a dichotomous decision regarding whether a given patient should receive extra clinical scrutiny or not. LR models, in theory, can be used to calculate the probability of an outcome with a handheld calculator while ANN prediction models must be implemented on a computer. Given that the provision of modern PD care is computationally advanced, with computerized modelling of dialysis prescription for example, the addition of another computer-based tool should not be overly burdensome.

In conclusion, an ANN-based model performed reasonably well in predicting early technique failure among incident PD patients. The ANN performed significantly better than a traditional, LR-based prediction model. As the UKRR repository grows, in terms of the number of patients captured and the detail of the data, whether even more sophisticated ANN technology will provide better predictive performance will remain as an area of active investigation.

Appendix

ANN fundamentals

ANNs are implemented as software programs that simulate the information processing architecture of a network of biological neurons. Each artificial neuron consists of an information processing node ('body'), its connections from other neurons ('dendrites') and its connection to other neurons ('axons'). A typical ANN consists of layers of artificial neurons. In the most common arrangement, the multilayer perceptron (MLP) (Figure 1), a set of input neurons each receives one of the values of an ordered set (a vector) of predictor variables. Information from the predictor variables is passed through the layers of the ANN such that, between layers, a set of weight factors modifies the information. The neurons within a layer each sum the weighted inputs from their 'dendrites' and then apply a non-linear function

(usually the logistic) to the sum that is sent out as an output along their 'axons'. Ultimately, the modified information reaches the output neuron that performs a final summation and non-linear function application. The result of this function becomes the output for the entire ANN (Figure 1).

In order for an ANN to be useful, it must be trained. Training involves presenting a set of cases that each have values for the predictor variables as well as a known outcome [e.g. PD technique failure (outcome = 1) versus no failure (outcome = 0)]. Initially, the weights inside the ANN are set to random values so that its output is meaningless. However, with each case that is presented to the ANN, an error value, which is the difference between its output and the actual outcome (1 or 0), is used to adjust the weights within the ANN so as to minimize the error on subsequent presentations. The procedure for adjusting the weight values is known as the generalized delta rule [40]. The error signals are propagated backwards layer-by-layer through the ANN and, hence, this training approach is known as back-propagation. Each neuron in the middle layer receives the error value from the output neuron multiplied by the weight connecting the neurons. The modified error values for the neurons in the middle layer are then used to compute the error terms for the neurons in the input layer. Each input neuron takes a weighted sum of the error values of the neurons to which it connects in the middle layer. The weights used in this calculation are the same connection weights between the two layers that were used to generate the output.

After the error value has backpropagated, the weights are adjusted using the following formula: $\Delta w_{ij} = \alpha^* e_i^* [o^*(1-o)]$ where 'Δw_{ij}' is the weight change for the connection between the ith neuron in the input layer and the jth neuron in the middle layer, 'α' is the learning rate coefficient (which determines the fraction of a weight change that is produced by a given error value) and 'o' is the current value of the ANN output. Likewise, the weights connecting the middle layer and the output neuron are adjusted. After each set of 'n' input vectors and known outputs is presented to the ANN, an overall error measure is calculated such as the mean square error, $MSE = (1/n)\Sigma(t_k - o_k)^2$ where 't_k' is the actual value for associated with the kth input vector and 'o_k' is the ANN output for that vector. Eventually, after many presentations of the set of training vectors, the MSE value converges to a minimum. At this point the ANN has been trained and is ready to make predictions on a new set of cases. In order to validate the performance of the ANN, it is tested against new cases with known outcomes (the validation set) and a performance statistic such as the AUROC is computed.

Optimal threshold values for ROC curves

Building on the work of previous authors [30,41–43], it is possible to generate a closed form equation for the optimal threshold of an ROC curve [29]. Let x represent the possible values of the output from a prediction model (ANN or LR) when applied to a validation set. Assume that x is distributed as two Gaussian distributions: $x_D \sim N(\mu_D, \sigma_D)$ and $x_N \sim N(\mu_N, \sigma_N)$ for individuals with and without PD technique failure, respectively. The optimal threshold, x_t,

is the value for x which maximizes $y = \text{sens}(x) + \text{spec}(x)$ where the sensitivity and specificity at x_t are $\text{sens}(x_t) = \text{Prob}(x \geq x_t \, ; \mu_D, \sigma_D) = 1 - \Phi[(x_t - \mu_D)/\sigma_D]$ and $\text{spec}(x_t) = \text{Prob}(x \leq x_t \mid \mu_N, \sigma_N) = \Phi[(x_t - \mu_N)/\sigma_N]$, respectively and where $\Phi[g]$ is the standard normal probability mass function at 'g'. Setting the first derivative of y with respect to x_t to 0 and solving for x_t yields $x_t = (b\mu_D + \mu_N)/(1+b)$ where $b = (\sigma_N/\sigma_D)$. An estimate for 'b' is provided by CLABROC [28] while μ_D and μ_N can be estimated from the average outputs of the subjects with and without PD technique failure, respectively, in the validation set.

Conflict of interest statement. None declared.

References

1. Grassmann A, Gioberge S, Moeller S. ESRD patients in 2004: global overview of patient numbers, treatment modalities and associated trends. *Nephrol Dial Transplant* 2005; 20: 2587–2593

2. Sennfalt K, Magnusson M, Carlsson P. Comparison of hemodialysis and peritoneal dialysis—a cost-utility analysis. *Perit Dial Int* 2002; 22: 39–47

3. Liem YS, Wong JB, Hunink MG *et al*. Comparison of hemodialysis and peritoneal dialysis survival in The Netherlands. *Kidney Int* 2007; 71: 153–158

4. Rubin HR, Fink NE, Plantinga LC *et al*. Patient ratings of dialysis care with peritoneal dialysis versus hemodialysis [see comment]. *JAMA* 2004; 291: 697–703

5. Little J, Irwin A, Marshall T *et al*. Predicting a patient's choice of dialysis modality: experience in a United Kingdom renal department. *Am J Kidney Dis* 2001; 37: 981–986

6. Ansell D, Feest TG, Tomson C *et al. UK Renal Registry Report 2006*. Bristol, UK: UK Renal Registry, 2007

7. Jager KJ, Korevaar JC, Dekker FW *et al*. Netherlands Cooperative Study on the Adequacy of Dialysis (NECOSAD) Study Group. The effect of contraindications and patient preference on dialysis modality selection in ESRD patients in The Netherlands [see comment]. *Am J Kidney Dis* 2004; 43: 891–899

8. Rumpsfeld M, McDonald SP, Johnson DW. Higher peritoneal transport status is associated with higher mortality and technique failure in the Australian and New Zealand peritoneal dialysis patient populations. *J Am Soc Nephrol* 2006; 17: 271–278

9. Tonelli M, Hemmelgarn B, Manns B *et al*. Use and outcomes of peritoneal dialysis among Aboriginal people in Canada. *J Am Soc Nephrol* 2005; 16: 482–488

10. McDonald SP, Collins JF, Johnson DW. Obesity is associated with worse peritoneal dialysis outcomes in the Australia and New Zealand patient populations. *J Am Soc Nephrol* 2003; 14: 2894–2901

11. Huisman RM, Nieuwenhuizen MGM, Th de Charro F. Patient-related and centre-related factors influencing technique survival of peritoneal dialysis in The Netherlands. *Nephrol Dial Transplant* 2002; 17: 1655–1660

12. Churchill DN, Thorpe KE, Nolph KD *et al*. The Canada-USA (CANUSA) Peritoneal Dialysis Study Group. Increased peritoneal membrane transport is associated with decreased patient and technique survival for continuous peritoneal dialysis patients. *J Am Soc Nephrol* 1998; 9: 1285–1292

13. Penny W, Frost D. Neural networks in clinical medicine [Review] [91 refs]. *Med Decis Making* 1996; 16: 386–398

14. Itchhaporia D, Snow PB, Almassy RJ *et al*. Artificial neural networks: current status in cardiovascular medicine [Review] [39 refs]. *J Am Col Cardiol* 1996; 28: 515–521

15. Lisboa PJ. A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Netw* 2002; 15: 11–39

16. Varghese SA, Powell TB, Budisavljevic MN *et al*. Urine biomarkers predict the cause of glomerular disease. *J Am Soc Nephrol* 2007; 18: 913–922

17. Wang YF, Hu TM, Wu CC *et al*. Prediction of target range of intact parathyroid hormone in hemodialysis patients with artificial neural network. *Computer Methods Programs Biomed* 2006; 83: 111–119

18. Gabutti L, Ferrari N, Mombelli G *et al*. Does cystatin C improve the precision of Cockcroft and Gault's creatinine clearance estimation? *J Nephrol* 2004; 17: 673–678

19. Gabutti L, Lotscher N, Bianda J *et al*. Would artificial neural networks implemented in clinical wards help nephrologists in predicting epoetin responsiveness? *BMC Nephrol* 2006; 7: 13

20. Chen CA, Lin SH, Hsu YJ *et al*. Neural network modeling to stratify peritoneal membrane transporter in predialytic patients. *Intern Med* 2006; 45: 663–664

21. Parekattil SJ, Kumar U, Hegarty NJ *et al*. External validation of outcome prediction model for ureteral/renal calculi. *J Urol* 2006; 175: 575–579

22. Oates JC, Varghese S, Bland AM *et al*. Prediction of urinary protein markers in lupus nephritis [see comment]. *Kidney Int* 2005; 68: 2588–2592

23. Nielsen M, Granerus G, Ohlsson M *et al*. Interpretation of captopril renography using artificial neural networks. *Clin Physiol Funct Imaging* 2005; 25: 293–296

24. Dimitrov BD, Ruggenenti P, Stefanov R *et al*. Chronic nephropathies: individual risk for progression to end-stage renal failure as predicted by an integrated probabilistic model. *Nephron* 2003; 95: c47–c59

25. Brier ME, Ray PC, Klein JB. Prediction of delayed renal allograft function using an artificial neural network. *Nephrol Dial Transplant* 2003; 18: 2655–2659

26. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993

27. Metz CE, Jiang Y, MacMahon H *et al*. ROCKIT: maximum likelihood estimation to fit a binormal ROC curve to continuously-distributed data and/or ordinal category data, 2001. http://www-radiology.uchicago.edu/krl/KRL_ROC/software_index6.htm

28. Metz CE, CLABROC. Statistical analysis of ROC data in evaluating diagnostic performance (1.9.1) Computer Software. 1998

29. Naimark DMJ. A comparison of regression methods with artificial neural networks for the prediction of the outcome of patients on chronic dialysis. 1-1-2000, masters thesis, University of Toronto

30. Hanley JA. The use of the 'binormal' model for parametric ROC analysis of quantitative diagnostic tests. *Stat Med* 1996; 15: 1575–1585

31. Hanley JA, McNeil B. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148: 839–843

32. Hanley JA, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29–36

33. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988; 240: 1285–1293

34. Green M, Bjork J, Forberg J *et al*. Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. *Artif Intell Med* 2006; 38: 305–318

35. Peng SY, Wu KC, Wang JJ *et al*. Predicting postoperative nausea and vomiting with the application of an artificial neural network. *Br J Anaesth* 2007; 98: 60–65

36. Lowrie EG, Huang WH, Lew NL. Death risk predictors among peritoneal dialysis and hemodialysis patients: a preliminary comparison. *Am J Kidney Dis* 1995; 26: 220–228

37. Lowrie EG, Lew NL, Huang WH. Race and diabetes as death risk predictors in hemodialysis patients. *Kidney Int Suppl* 1992; 38: S22–S31

38. Lowrie EG, Lew NL. Death risk in hemodialysis patients: the predictive value of commonly measured variables and an evaluation of

death rate differences between facilities. *Am J Kidney Dis* 1990; 15: 458–482

39. Hemmelgarn BR, Culleton BF, Ghali WA. Derivation and validation of a clinical index for prediction of rapid progression of kidney dysfunction. *Q J Med* 2007; 100: 87–92

40. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by Error propagation. In Rumelhart DE, McLelland JL (eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press Cambridge, MA, 1994, pp 318–364

41. Dorfman DE, Alf E. Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating-method data. *J Math Psychol* 1969; 6: 487–496

42. Peng F, Hall WJ. Bayesian analysis of ROC curves using Markov-chain Monte Carlo methods. *Med Decis Making* 1996; 16: 404–411

43. Van Der Schouw YT, Straatman H, Verbeek AL. ROC curves and the areas under them for dichotomized tests: empirical findings for logistically and normally distributed diagnostic test results. *Med Decis Making* 1994; 14: 374–381