Prediction of Mortality in the First Two Years of Hemodialysis: Results from a Validation Study

Stephan Thijssen Len Usvyat Peter Kotanko

Renal Research Institute, New York, N.Y., USA

Key Words

Hemodialysis • Mortality • Logistic regression model

Abstract

Background: Chronic hemodialysis (HD) patients experience high rates of mortality. Alerting medical staff of patients at increased risk of death may support clinical decision making. **Methods:** A large cohort of incident HD patients was used to develop logistic regression models to predict death in the subsequent 6 months ('derivation cohort'). Predictors were age, gender, race, ethnicity, vascular access type, diabetic status, pre-HD systolic and diastolic blood pressure, pre-HD weight, pre-HD temperature, relative interdialytic weight gain, serum albumin, hemoglobin, phosphorus, serum creatinine, serum sodium, urea reduction ratio, equilibrated normalized protein catabolic rate, and equilibrated dialytic and renal Kt/V. These logistic regression models were then applied to validation cohorts. Predictive performance of the models was described in terms of sensitivity, specificity, and area under receiver-operating characteristic curves (AUC-ROC). **Results:** A total of 6,838 incident HD patients were followed over 2 years. The derivation cohort initially comprised 4,512 patients. In the validation cohort of initially 2,326 patients, the logistic regression models were able to predict mortality in subsequent 6-month periods with a sensitivity between 58 and 69%, and a specificity of

74–77%; the respective AUC-ROC were 0.67–0.72 (all $p < 0.0001$). Pre-HD weight and serum albumin levels were consistently significant predictors of mortality in all models. **Conclusion:** The results indicate that logistic regression models are useful tools in estimating incident HD patients' probability of death in 6-month intervals for at least up to 2 years after beginning dialysis. Model predictions may be used to alert medical staff to patients at increased risk of death and facilitate timely diagnostic and therapeutic interventions to improve outcomes.

Copyright © 2012 S. Karger AG, Basel

Background

Despite significant advances in hemodialysis (HD) technology, the mortality risk of chronic HD patients remains well above that seen in the general population, where individuals younger than 30 years of age have on average four times longer age-adjusted life expectancy than HD patients of comparable age, and HD patients aged 65 years or older have a mortality risk six times higher than the general population. Cardiovascular disease and infectious diseases are among the leading causes of mortality, and the difference in mortality risk between HD patients and the general population is most pronounced for heart disease, with threefold higher death

Fig. 1. Study design. All patients who survived the first 6 months on dialysis (n = 6,838) were randomly split 2:1 into a 'derivation' cohort for model development and a 'validation' cohort for model validation. The 2-year follow-up period was divided into 6-month segments (S1: months 1–6; S2: months 7–12; S3: months 13–18; S4: months 19–24). Using the derivation cohort, logistic regression models were developed to predict survival in S2 based on predictors in S1 (S1/S2 model), survival in S3 based on predictors in S2 (S2/S3 model), and survival in S4 based on predictors

in S3 (S3/S4 model). ROC and maximal Youden index were used to determine optimal death probability cutoffs for each model using the derivation cohort. Model performance was then assessed by applying the regression models and optimal cutoffs to the validation cohort and determining, for each model, sensitivity, specificity, positive and negative likelihood ratios, AUC-ROC, post-test probabilities of death as a function of pretest likelihood, and the diagnostic information gain resulting from a positive or negative test result.

rates (180.8 vs. 49.8 deaths per 1,000 patient years) in individuals aged 45–64 years [1].

Current epidemiologic studies seeking to investigate the determinants of mortality risk in dialysis patients usually consider either cross-sectional baseline characteristics (e.g. mean systolic blood pressure in the first 3 months after start of dialysis; serum albumin levels after 6 months) or time-dependent analyses, most commonly time-dependent Cox regression models. Another approach includes determining at least one of the patient's clinical or biochemical parameters periodically while the patient is undergoing HD treatments, and identifying a patient as having an increased risk for death if the patient has a substantial change in the rate of decline or rate of increase of the clinical or biochemical parameter [2]. The aim of this study was to develop and validate logistic regression models for identifying dialysis patients at increased risk of death in 6-month intervals for a period of up to 2 years after beginning dialysis.

Methods

We conducted a retrospective analysis of all incident in-center HD patients in Renal Research Institute clinics starting HD between July 1, 2000 and August 31, 2009. Patients were randomized into one of two cohorts: two thirds of them were used for model development ('derivation cohort'), the remainder for model validation ('validation cohort'). Patients were followed for up to 24 months. These 24 months were split into four nonoverlapping time segments: S1 = months 1–6, S2 = months 7–12, S3 = months 13–18, and S4 = months 19–24. Model development and validation were conducted in (1) all patients who survived the first 6 months of HD and were followed-up to month 12 (S1/S2 cohort), (2) all patients who survived the first 12 months and were followed-up to month 18 (S2/S3 cohort), and (3) all patients who survived the first 18 months and were followed-up to month 24 (S3/S4 cohort). Logistic regression models were constructed in the S1/S2, S2/S3, and S3/S4 derivation cohorts, so that the S1/S2 model predicted survival in S2 based on S1 predictors, the S2/S3 model predicted survival in S3 based on S2 predictors, and the S3/S4 model predicted survival in S4 based on S3 predictors (fig. 1).

Table 1. Baseline characteristics

Variable	Derivation cohort (n = 4,512)	Validation cohort (n = 2,326)
Male, %	56.6	56.5
White/Black/other, %	49.1/43.0/7.9	50.0/42.0/8.0
Hispanic, %	13.1	14.0
AV fistula or AV graft, %	30.5	29.4
Age, years	61.3 (15.5)	61.7 (15.5)
Diabetic patients, %	53.9	54.8
Pre-HD SBP, mm Hg	149.8 (17.9)	149.8 (18.2)
Pre-HD DBP, mm Hg	78.2 (10.8)	78.1 (11.1)
Pre-HD weight, kg	79.5 (21.3)	79.4 (21.2)
Pre-HD temperature, °C	36.5 (0.3)	36.5 (0.3)
IDWG, % of post-HD weight	3.1 (1)	3.1 (1)
Serum albumin, g/dl	3.7 (0.4)	3.7 (0.4)
Urea reduction ratio, %	69.4 (7.1)	69.4 (6.8)
Hemoglobin, g/dl	11.8 (1.1)	11.7 (1.1)
Serum phosphorus, mg/dl	5.2 (1.2)	5.2 (1.2)
Serum creatinine, mg/dl	7.2 (2.7)	7.3 (2.7)
Serum sodium, mmol/l	138.4 (2.6)	138.5 (2.6)
eKdrt/V	1.3 (0.3)	1.4 (0.3)
enPCR, g/day/kg body weight	0.8 (0.2)	0.8 (0.2)

Data are percentages or means (SD) in months 1–6. AV = Arteriovenous; SBP = systolic blood pressure; DBP = diastolic blood pressure; IDWG = interdialytic weight gain; eKdrt/V = equilibrated dialytic and renal Kt/V; enPCR = equilibrated normalized protein catabolic rate.

The following predictors were used in each of the three logistic regression models: age, gender, race, ethnicity, vascular access type, diabetic status, pre-HD systolic blood pressure, pre-HD diastolic blood pressure, pre-HD weight, pre-HD temperature, relative interdialytic weight gain (% of post-HD weight), serum albumin, urea reduction ratio, hemoglobin, serum phosphorus, serum creatinine, serum sodium, equilibrated normalized protein catabolic rate, and equilibrated dialytic and renal Kt/V. For continuous variables, averages over the respective 6-month periods were used. Access type was determined once at the beginning of each respective 6-month interval. For the other predictors, baseline values at the first treatment in month 1 were used.

Estimates of the model coefficients were obtained in the derivation cohort, and the models were validated in the validation cohort. Optimal thresholds of death probability were derived in the derivation cohort by computation of the maximal Youden index (Youden index = sensitivity + specificity – 1) over a wide range of thresholds [3]. Sensitivity, specificity, positive and negative likelihood ratios (+LR, –LR), and area under receiver-operating characteristic curves (AUC-ROC) were computed in the validation cohort to assess predictive accuracy of the models [4]. Conceptually, model predictions can be considered a diagnostic test. Given a pretest likelihood of a diagnosis (in our case, death in a

prespecified time interval), $P(D+)$, it is possible to compute the posttest likelihood of this diagnosis contingent upon the outcome of the test (positive or negative), $P(D+ | T+)$ and $P(D+ | T-)$. In general, the pretest likelihood for the presence of a disease, $P(D+)$, and posttest likelihood of a diagnosis can be linked by Bayes' theorem. $P(D+ | T+)$, the probability of death under the condition of a positive test result, and $P(D+ | T-)$, the probability of death under the condition of a negative test, can be calculated as follows:

$$P(D+ | T+) = \frac{P(D+) \cdot \text{sensitivity}}{P(D+) \cdot \text{sensitivity} + (1 - P(D+)) \cdot (1 - \text{specificity})} \quad (\text{Eq. 1})$$

$$P(D+ | T-) = \frac{P(D+) \cdot (1 - \text{sensitivity})}{P(D+) \cdot (1 - \text{sensitivity}) + (1 - P(D+)) \cdot \text{specificity}} \quad (\text{Eq. 2})$$

The difference between $P(D+)$ and $P(D+ | T+)$ and the difference between $P(D+)$ and $P(D+ | T-)$ can be interpreted as the diagnostic gain over and above $P(D+)$ in the presence of a positive or negative test result, respectively. Computations of $P(D+ | T+)$, $P(D+ | T-)$, and diagnostic gain as a function of $P(D+)$ are presented.

SAS version 9.2 and MedCalc version 11.6.1.0 were used for statistical analysis.

Results

A total of 6,838 patients were studied. The initial S1/S2 derivation cohort consisted of 4,512 patients, and the validation cohort had 2,326 patients. The baseline characteristics of these two cohorts are shown in table 1. The cohorts did not differ from each other significantly with respect to these characteristics.

Logistic regression models were constructed in the derivation cohort to predict mortality in S2 (based on S1 predictors), S3 (based on S2 predictors), and S4 (based on S3 predictors; table 2).

Pre-HD weight and serum albumin levels were significant predictors of mortality in all three models, whereas age, hemoglobin levels, gender, ethnicity, and serum sodium levels appeared to be significant predictors only in certain time periods after HD initiation (table 2). None of the remaining predictors (race, diabetic status, pre-HD systolic blood pressure, pre-HD diastolic blood pressure, relative interdialytic weight gain, urea reduction ratio, serum phosphorus, equilibrated normalized protein catabolic rate, equilibrated dialytic and renal Kt/V, vascular access type, serum creatinine, and pre-HD body temperature) were significant at any time period.

ROC analysis and computation of the Youden index in the derivation cohort showed the optimal 'alert' cutoffs for the probability of death to be 6.8% (S1/S2 model; fig. 2), 6.6% (S2/S3 model), and 5.8% (S3/S4 model).

Table 2. Logistic regression parameter estimates of the different models

	Estimate	p	OR	95% CI for OR
<i>S1/S2 model</i>				
Age, years	0.032	<0.0001	1.032	1.018–1.046
Pre-HD weight, kg	–0.019	<0.001	0.981	0.972–0.991
Serum albumin, g/dl	–1.258	<0.0001	0.284	0.206–0.391
Hemoglobin, g/dl	–0.201	0.001	0.818	0.724–0.924
<i>S2/S3 model</i>				
Male	0.418	0.020	1.519	1.068–2.160
Age, years	0.028	<0.0001	1.028	1.012–1.044
Pre-HD weight, kg	–0.019	0.001	0.982	0.971–0.992
Serum albumin, g/dl	–1.248	<0.0001	0.287	0.190–0.435
<i>S3/S4 model</i>				
Hispanic	–1.018	0.010	0.361	0.166–0.786
Pre-HD weight, kg	–0.018	0.003	0.982	0.970–0.994
Serum albumin, g/dl	–1.841	<0.0001	0.159	0.095–0.264
Serum sodium, mmol/l	–0.070	0.046	0.933	0.871–0.999

CI = Confidence interval; OR = odds ratio.

Fig. 2. ROC analysis in the S1/S2 derivation cohort to compute the optimal ‘alert’ threshold based on the maximal Youden index. In this model, the optimal threshold was a probability of death of 6.8%.

Table 3. Evaluation of prediction models (S1/S2, S2/S3, S3/S4) in the validation cohort

Model	Sensitivity, % (95% CI)	Specificity, % (95% CI)	+LR	–LR	AUC-ROC (95% CI)
S1/S2, predicting months 7–12 (n = 2,326)	65 (56–73)	75 (73–77)	2.58	0.47	0.698 (0.679–0.717)
S2/S3, predicting months 13–18 (n = 1,866)	69 (60–78)	74 (72–76)	2.67	0.41	0.717 (0.696–0.737)
S3/S4, predicting months 19–24 (n = 1,524)	58 (46–69)	77 (74–79)	2.46	0.55	0.670 (0.646–0.694)

All AUC-ROC are >0.5 (all $p < 0.0001$).

The logistic regression models developed in the derivation cohort (table 2) in conjunction with the aforementioned alert thresholds were then applied to the validation cohort. The results showed the significant predictive power of these models as indicated by AUC-ROC significantly greater than 0.5 (table 3; all $p < 0.0001$).

Knowing the predictive characteristics of the logistic regression models in S2, S3, and S4 in the validation cohort, the respective posttest likelihoods $P(D+ | T+)$ and $P(D+ | T-)$ and the diagnostic gains resulting from a positive or negative ‘test result’ were calculated. The result of this Bayesian analysis is shown in figure 3 for the prediction of death between months 7 and 12; similar analyses

for months 13–18 and months 19–24 showed materially the same results. As the pretest likelihood of death (x-axis) changes, the respective posttest likelihoods $P(D+ | T+)$ and $P(D+ | T-)$ change according to equations 1 and 2 (see Methods). For example, given a pretest likelihood of death of 0.20, a positive prediction of death by the logistic regression model increased the likelihood of death to 0.39 (increase by 0.19), whereas a negative prediction reduced the likelihood of death to 0.11 (decrease by 0.09). The diagnostic gain was greatest for pretest likelihoods around 0.4 for positive model results and around 0.6 for negative test results, respectively.

Fig. 3. Bayesian analysis of the S1/S2 model, predicting the probability of death in months 7–12 (S2 period) based on predictors in months 1–6 (S1 period). The posttest probability of death is plotted on the y-axis as a function of the pretest likelihood (x-axis). The pretest likelihood of death could be derived simply from population-based mortality data or in a more individualized fashion, e.g. via estimates by the attending physician. An example of the informational gain provided by the model is illustrated: if a patient had a pretest likelihood of death of 0.2 (or 20%; vertical dotted line), obtaining a positive test result (i.e. the model predicting this patient's death in the upcoming 6 months) would almost double that probability to 39%. The difference between this post-test likelihood and the pretest likelihood, 19 percentage points, represents the informational gain provided by the model. Similarly, at the same pretest likelihood of 20%, a negative test result would decrease the probability of death to 11%, almost half of the pretest likelihood. A pretest likelihood of 20% was chosen simply for illustrational purposes.

Discussion

Despite significant technological advances, mortality in chronic HD patients remains appallingly high, particularly in the first year of HD [5]. One way to potentially reduce mortality is to alert medical staff to patients at increased risk of death in order to facilitate timely diagnostic and therapeutic interventions. In the present study, logistic regression models were developed to predict mortality in 6-month intervals over a 2-year period in incident HD patients. In these models, readily available demographic and laboratory variables were used. Simple logistic regression models were developed in a large derivation cohort and rigorously evaluated in a separate validation cohort. This approach guards against overly optimistic estimates of model performance which

may arise from validating the models in the derivation cohort. In this study, both derivation and validation cohorts were drawn from dialysis units spread across the USA to assure a high degree of generalizability. Indeed, the demographics of the study population are comparable to the general US dialysis population, confirming a high degree of generalizability.

The prediction models presented here use untransformed variables readily available in the Renal Research Institute database. Additionally, other variables can be included in the logistic regression models. Analysis of variables not routinely measured in US clinics but measured in other parts of the world may demonstrate that some parameters carry valuable predictive power. Given the fact that clinical and laboratory indicators change over time and show a specific dynamic before death [2], descriptors of temporal change will be included in a future model iteration. Logistic regression models are one way to predict outcomes. Other models such as artificial neural networks, regression trees, and dynamic models (e.g. state space models) can be adapted to improve prediction of mortality. In addition, future model developments may also attempt to integrate the predictions made by the medical staff.

The value of subjective assessments has been shown in studies evaluating the predictive value of the 'surprise question' [6, 7]. In these studies, nurse practitioners who were primarily involved in the patients' long-term care in the dialysis unit were asked to answer the question 'Would you be surprised if this patient died in the next 6 months?' with either 'No, I would not be surprised' or 'Yes, I would be surprised'. Prediction models incorporating the answer to this question were validated in 514 patients and resulted in an AUC-ROC for the resulting prognostic model predictions of 6-month mortality of 0.80 (95% CI: 0.73–0.88), with the answer to the 'surprise question' being a significant predictor [7]. Other variables included were serum albumin levels, age, the presence or absence of peripheral vascular disease, and the presence or absence of dementia. These results indicate that prognostic assessment of the patient by medical staff can contribute significantly to the prediction of outcomes.

The approach we presented is not limited to a binary yes/no answer to the 'surprise question', but allows the input of prior probability $P(D+)$ given by physicians, thereby moving from estimates based on population data to an individualized pretest likelihood. The predictions provided by the logistic regression models can be combined with $P(D+)$ to calculate the posttest probability of death; this value, which can also be translated into a

score, could then be communicated to the attending physician. Whether or not the assessment by the medical staff should be an integral part of a prediction model is unclear. On one hand, under most circumstances, the predictions would be more individualized and likely more accurate. On the other hand, however, a subjectively defined P(D+) may heavily influence or even become the dominant component in the prediction process and outweigh the objective predictions made on the basis of, for example, laboratory data. It might, in fact, be argued that such a prediction model should not reflect any subjective assessment, but rather be reflective of the average risk of death of a cohort with similar characteristics as the patient in question. Such a prediction, by highlighting potential discrepancies between the average prognosis and the physician's subjective assessment of the individual patient, may direct the physician's attention to precisely those patients who need it most. Having the physician's subjective opinion heavily influence the model's prediction may take away the critical stimulus that such a discrepancy may provide.

The critical question is if the use of prediction models can eventually improve outcomes. Currently, there is no

indication whether or not knowing the prognosis (i.e. likelihood of death in the next 6 months) will improve outcomes. We hypothesize that medical staff will implement targeted diagnostic and therapeutic interventions when alerted to a patient's dismal prognosis. Nevertheless, the clinical utility of prediction models has to be studied rigorously in prospective randomized trials. Trial designs may involve randomizing entire dialysis units to either applying or not applying alert systems, and then following and comparing outcomes over a prolonged time period.

In conclusion, we developed simple logistic regression models to predict mortality in 6-month intervals over a period of up to 2 years in incident HD patients. Validation of these models in an independent patient cohort demonstrated a significant diagnostic information gain from applying the models. Future model iterations may include variables capturing the dynamics of clinical and laboratory indicators over time. Eventually, prediction models may evolve into 'alert systems' to facilitate timely diagnostic and therapeutic interventions in patients at increased risk of death. The clinical utility of such alert systems has to be evaluated rigorously in prospective trials.

References

- 1 United States Renal Data System: Mortality and causes of death. Annual Data Report 2007.
- 2 Kotanko P, Thijssen S, Usvyat L, Tashman A, Kruse A, Huber C, Levin NW: Temporal evolution of clinical parameters before death in dialysis patients: a new concept. *Blood Purif* 2009;27:38–47.
- 3 Akobeng AK: Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr* 2007;96:644–647.
- 4 Tripepi G, Jager KJ, Dekker FW, Zoccali C: Diagnostic methods 2: receiver operating characteristic (ROC) curves. *Kidney Int* 2009;76:252–256.
- 5 Himmelfarb J, Ikizler TA: Hemodialysis. *N Engl J Med* 2010;363:1833–1845.
- 6 Moss AH, Ganjoo J, Sharma S, Gansor J, Senft S, Weaner B, Dalton C, MacKay K, Pellegrino B, Anantharaman P, Schmidt R: Utility of the 'surprise' question to identify dialysis patients with high mortality. *Clin J Am Soc Nephrol* 2008;3:1379–1384.
- 7 Cohen LM, Ruthazer R, Moss AH, Germain MJ: Predicting six-month mortality for patients who are on maintenance hemodialysis. *Clin J Am Soc Nephrol* 2010;5:72–79.