External validation and clinical utility of a prediction model for 6-month mortality in patients undergoing hemodialysis for end-stage kidney disease

Brian Forzley[1,2], Lee Er[3], Helen HL Chiu[3], Ognjenka Djurdjev[3], Dan Martinusen[4], Rachel C Carson[1,4], Gaylene Hargrove[1,4], Adeera Levin[1,3] and Mohamud Karim[1,5]

Abstract

Background: End-stage kidney disease is associated with poor prognosis. Health care professionals must be prepared to address end-of-life issues and identify those at high risk for dying. A 6-month mortality prediction model for patients on dialysis derived in the United States is used but has not been externally validated.
Aim: We aimed to assess the external validity and clinical utility in an independent cohort in Canada.
Design: We examined the performance of the published 6-month mortality prediction model, using discrimination, calibration, and decision curve analyses.
Setting/participants: Data were derived from a cohort of 374 prevalent dialysis patients in two regions of British Columbia, Canada, which included serum albumin, age, peripheral vascular disease, dementia, and answers to the "the surprise question" ("Would I be surprised if this patient died within the next year?").
Results: The observed mortality in the validation cohort was 11.5% at 6 months. The prediction model had reasonable discrimination (c-stat = 0.70) but poor calibration (calibration-in-the-large = −0.53 (95% confidence interval: −0.88, −0.18); calibration slope = 0.57 (95% confidence interval: 0.31, 0.83)) in our data. Decision curve analysis showed the model only has added value in guiding clinical decision in a small range of threshold probabilities: 8%–20%.
Conclusion: Despite reasonable discrimination, the prediction model has poor calibration in this external study cohort; thus, it may have limited clinical utility in settings outside of where it was derived. Decision curve analysis clarifies limitations in clinical utility not apparent by receiver operating characteristic curve analysis. This study highlights the importance of external validation of prediction models prior to routine use in clinical practice.

Keywords

End-stage kidney disease, chronic kidney disease, prediction model, prognosis, dialysis, end-of-life care, palliative care

---

What is already known about the topic?

- Patients with end-stage kidney disease (ESKD) experience high mortality and morbidity.
- Despite the importance of prognostication in patients with ESKD, there is limited understanding of existing prognostic tools that were developed to support the practice for guiding clinical management of individual patients.
- There is a 6-month mortality prediction model for patients on hemodialysis that reports particularly high discriminative ability and is available for use on mobile and online apps; however, its utility and external validity have not been assessed.

---

[1]Department of Medicine, Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada
[2]Penticton Regional Hospital Renal Program, Penticton, BC, Canada
[3]British Columbia Provincial Renal Agency, Vancouver, BC, Canada
[4]Vancouver Island Health Authority, Victoria, BC, Canada

[5]Fraser Health Authority, Surrey, BC, Canada

Corresponding author:
Brian Forzley, Penticton Regional Hospital Renal Program, 110-2504 Skaha Lake Road, Penticton, BC V2A 6G1, Canada.
Email: brian.forzley@interiorhealth.ca

What this paper adds?

- This study is the first external validation that assesses performance for an ESKD mortality prediction model in a population external to and independent from the validation cohort.
- This study is also the first ESKD mortality prediction model assessment that utilizes decision curve analysis to help clarify clinical utility.
- Using robust methodology, this study indicates that the 6-month mortality prediction model has limited clinical utility despite reasonable discrimination in the validation cohort.

Implications for practice, theory, or policy

- Future studies need to clarify appropriate uses for this 6-month mortality prediction model for guiding aspect of care that are lower risk, for example, coordinating uptake of advance care planning discussion, where it could add value in guiding clinical decision within a narrow range of threshold probabilities (8%–20%) as per the decision curve analysis.
- The findings support the importance of evaluating prediction model using local data for relevance and suitability of routine use.
- Limitations and applicability of prognostic tools should be recognized prior to uptake in clinical practice.

Introduction

Poor survivorship in end-stage kidney disease (ESKD) remains an important issue in the care of patients with chronic kidney disease (CKD) around the world. The International guidelines on CKD emphasize the high morbidity and mortality experienced by patients with ESKD.[1] The Canadian Organ Replacement Register indicates that in 2013, 5333 Canadians with ESKD began some form of renal replacement therapy, an almost three-quarters increase from two decades ago.[2] While patients aged ⩾75 years were the largest group to start dialysis, their 5-year survival was only 27%.[2] Among the elderly in facility care, studies have shown both poor survivorship and poor functional status associated with dialysis treatment for ESKD.[3] Health care professionals must therefore be prepared to address issues and recommend appropriate care related to end of life. The associated issues and the shared decision-making process are complex and include advance care planning (ACP), revised medication strategies that prioritize symptom management, and consideration of non-dialytic management of ESKD including planned conservative care as well as withdrawal of dialysis.[4–6]

While initiatives across countries support earlier ACP for patients with chronic diseases,[5,7] the actual conversations around life expectancy and optimal timing for these conversations may be challenging in the absence of a reliable prognostic tool.[7,8] Despite the distressing nature of these discussions, patients want to be informed of their prognosis.[9,10] Nonetheless, risk prediction models to guide ACP or withdrawal discussions are often not routinely utilized for clinical management of individual patients. Given the importance of prognostication in patients with ESKD, improved prognostication strategies and tools are needed.

Uptake of prediction models into clinical care requires that the internal validity, external validity, and clinical usefulness are established.[11] In recent years, existing prognostic models that predict mortality among patients with ESKD have been quite extensively studied with internal validations.[12–22] Studies from the United States, Canada, Taiwan, and European countries have reported models with variable but reasonably accurate predictive ability.[12,13,15–22] However, none of these models have been independently validated in a population different from the original study population.

The 6-month mortality prediction model for mortality in patients on hemodialysis (HD) by Cohen et al.[12] is of particular interest given its higher degree of discrimination (area under the curve (AUC), 0.87), with predictor variable content similar to comparable studies but has the highest c-statistics, a measure of predictive accuracy,[23] among them (details in Supplemental Material). The prognostic model is also well received in clinical practice given its availability on smartphones and Internet tools.[24] The objective of this study is to independently evaluate its performance and clinical utility in a Canadian cohort of ESKD patients using robust methodology.

Methods

Patients and data collection

Prevalent patients on dialysis in two geographic regions of British Columbia, Canada, were identified from quality improvement datasets at two distinct periods: one over a 4-month period in 2011 and the other over a 16-month period in 2013–2014. The primary nephrologists were asked to answer the "Surprise Question (SQ)," "Would I be surprised if this patient died within the next year?", pertaining to their respective patients. The date of

SQ was regarded as start date of the study (i.e. Time 0).In addition to the SQ, four other key variables in the prediction model derived by Cohen et al.[12] and demographic variables were also collected at the time of SQ: age, serum albumin, peripheral vascular disease (PVD), and dementia, as well as sex, race, and dialysis vintage. All serum albumin measures were normalized to bromcresol purple measures.[25] Albumin values were truncated at 3 and 4.5 g/dL as in Cohen et al.[12]

Exclusion criteria included patients who were on peritoneal dialysis, missing data in any of the five key variables, age less than 19 years old or greater than 92 years old, and serum albumin less than 1.7 g/dL or greater than 5 g/dL (these ranges were not valid in the model by Cohen et al.[12]).

Mortality outcomes were collected up to 24 months after recruitment. Deaths occurring within 6.5 months were regarded as the observed death events at 6 months, as per the same definition used in the development cohort. This study was approved by the Research Ethics Board of the Providence Health Care Research Ethics Board (ID: H14-03325).

6-month predicted mortality risk

The 6-month predicted mortality risk for each patient by Cohen et al.[12] was calculated as $1 - 0.58^{\exp(\text{prognostic index})}$, where prognostic index $= \log(2.71) \times SQ + \log(0.27) \times Albumin + \log(1.36) \times Age + \log(1.88) \times PVD + \log(2.24) \times Dementia$. SQ was coded as not surprised versus surprised, and age was expressed in a unit of 10 years.

Statistical analysis

Descriptive statistics for continuous variables were reported in mean with standard deviation or median with interquartile range as appropriate. Categorical variables were summarized in counts and percentages. Baseline characteristics of our cohort were compared to the development cohort, where feasible, using a two-sample *t*-test and Pearson's chi-square test.

Evaluation of model performance: discrimination, calibration, and clinical utility analyses. The performance of the 6-month prediction model was evaluated in three aspects: discrimination, calibration, and clinical utility:

1. Discrimination referring to the ability of the model to distinguish a patient with an event from a patient without an event was examined by the concordance "C" statistic and by discrimination plot which depicts the difference between mean predicted risks for those who survived and those who died.
2. Calibration referring to the agreement between observed events and predictions (i.e. accuracy of predicted risks) was displayed in a calibration plot with the observed proportions of death against groups of patients by deciles of predicted mortality risks, where the distributions of actual death and alive events were also displayed. A LOESS smoothing algorithm was used to estimate the observed probabilities of death in relation to the predicted probabilities. The estimated calibration-in-the-large and calibration slope were examined against the ideal values of 0 and 1, respectively.
3. Clinical utility of prediction was assessed using a novel decision curve analysis proposed by Vickers and Elkin[26] Net benefit (NB) is the difference between the proportion of true positives and weighted proportion of false positives for a given probability threshold (P_t). The weight calculation is $P_t/(1-P_t)$, reflecting the ratio of the harms of unnecessary treatment to the benefits of appropriate treatment at probability threshold P_t. The decision curve is a plot of net benefit against a range of increasing probability thresholds. The decision curve from the model was compared to the decision curves derived from two strategies: classifying all patients as predicted to die (i.e. the "treat all" curve) and classifying all patients as predicted to survive (i.e. the "treat none" curve). Threshold probabilities that achieve a net benefit greater than the "treat all" and "treat none" curves identify a segment of threshold probabilities where the model provides clinical utility. We supplemented conventional metrics such as sensitivity (Sens), specificity (Spec), positive predictive value (PPV), negative predictive value (NPV), and NB at a selected set of threshold probabilities for reference.

Investigation of suboptimal performance. Suboptimal performance of a prediction model can be affected by differences in case-mix (distribution of observed and/or unobserved variables) and/or differences in regression model coefficients (different predictive nature of the variables). The expected model performance was obtained from simulation assuming the model is valid under our validation cohort case-mix. We also recalibrated the model; the estimates reflected the performance of refitted models after accounted for case-mix and optimal regression coefficients in the validation sample (details in Supplemental Material).

All tests were two-sided with *p*-value <0.05 considered as statistically significant. All analyses were performed in SAS software version 9.3 (SAS Institute, Cary, NC) and R software version 2.15.2. The methods conform with the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement[27] and recent developments for improved prediction model assessment.[28]
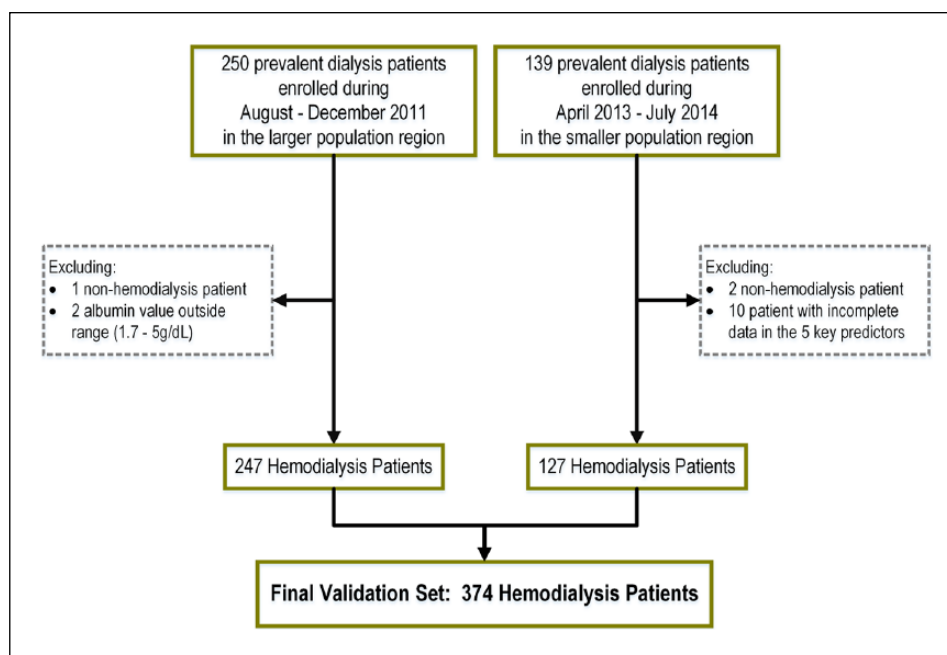
Figure 1. Cohort derivation.

Results

Study cohort

A total of 389 patients were identified, of which 15 patients were excluded: 3 non-HD patients, 10 patients with incomplete data for variables in the prediction model, and 2 patients with out-of-range serum albumin values, leaving 374 patients for analyses (Figure 1). Table 1 depicts the baseline characteristics of our study cohort compared to the derivation cohort for the prediction model by Cohen et al.[12] Our cohort was older (mean age: 68 vs 60; p-value < 0.001), greater proportion were Caucasian, and longer vintage on HD. We had a greater proportion of patients with a "not surprised" assignment (45% vs 16%; p-value < 0.001) and PVD (24% vs 3%; p-value < 0.001), slightly lower serum albumin (3.2 g/dL vs 3.8 g/dL; p-value < 0.001) and fewer with dementia (9% vs 20%; p-value < 0.001). By the end of 24-month follow-up, a total of 127 deaths were observed, with 43 deaths (11.5%) occurred during the first 6 months.

Evaluation of model performance

Discrimination analysis. Figure 2(a) left panel shows the receiver operating characteristic (ROC) curve with a c-statistic of 0.70 (95% confidence interval (CI): 0.62, 0.78). Figure 2(a) right panel displays the distribution of predicted risks among those alive and dead with a discrimination slope of 0.11 (95% CI: 0.05, 0.17; p-value = 0.001). These indicate that the model provided reasonable discriminative ability in our cohort.

Calibration analysis. The calibration plot is shown in Figure 2(b). The calibration-in-the-large was estimated as −0.53 (95% CI: −0.88, −0.18; p-value = 0.003), indicating an average of overestimation of observed mortality. The calibration slope was 0.57 (95% CI: 0.30, 0.83; p-value < 0.001), significantly deviated from the ideal slope of 1. The deviation was more overt in the higher predicted mortality risks, indicating the predicted risks were higher than the observed risks.

Clinical utility analysis. The model adds clinical utility only for threshold probabilities between 8% and 20%, as depicted in the decision curve in Figure 2(c). Comparing to a "treat-all" strategy, using the model to guide clinical decisions for the range of 8%–20% would lead to an equivalent of 12–42 fewer false positives per 100 patients for equal number of true positives. For probability thresholds greater than 20%, the net benefit for the model was negative compared to the "treat-none" strategy. This results from predicted risk being overestimated such that the increase of false-positive classification was larger than the loss of inappropriate withholding treatment. Table 2 depicts the comparison of actual mortality status against classification status for clinical decision at a few selected threshold probabilities for illustration and supplementary purposes.

Investigation of suboptimal performance

From Table 3, under the simulation scenario, the expected value for c-statistic was 0.80 assuming model is valid under our validation case-mix. It was much higher than the

Table 1. Baseline characteristics.

	Development cohort	BC validation cohort	p-value
Number of patients	449	374	
Age (years; mean ± SD)	60 ± 17	68 ± 15	<0.001
Male (n (%))	254 (57)	221 (59)	0.47
Race: Caucasian (n (%))	282 (65)	268 (72)	0.007
Time on hemodialysis (months)			
Median (IQR)	18 (4, 39)	26 (9, 48)	NA[a]
<3 months (n (%))	71/339 (19)	23 (6)	<0.001
Surprise question: not surprised (n (%))	71 (16)	168 (45)	<0.001
Serum albumin (g/dL)			
Mean ± SD	3.8 (0.4)	3.2 (0.4)	<0.001
<3.5 (n (%))	67 (15)	24 (71)	<0.001
Peripheral vascular disease (n (%))	15 (3)	91 (24)	<0.001
Dementia (n (%))	88 (20)	35 (9)	<0.001

IQR: interquartile range; SD: standard deviation.
The compositions of the development and validation cohorts were different.
[a]Unable to perform statistical testing from available summary statistics.

reported 0.70 from the validation sample, suggesting that case-mix could not be seen as a source for poor external performance. Under the partial recalibrated model, the calibration slope deviated from the ideal value of 1, and c-statistic (0.70 (95% CI: 0.62, 0.78)) was much lower compared with the expected value. The c-statistic remained unchanged (0.71 (95% CI: 0.63, 0.78)) under the fully recalibrated model despite optimal calibration. This indicates that the difference in the predictive ability of the five variables between our data and the development data may have contributed to the poor performance. As demonstrated in Table 3, such effects were weakened in serum albumin, dementia, age (not statistically significant), and PVD (not statistically significant) while the association between SQ and outcome remained equally strong.

Discussion

In this study, the model performance and clinical utility of a 6-month mortality prediction model developed by Cohen et al.[12] were assessed in a Canadian cohort of ESKD patients on HD, using multiple dimensions of performance. Overall, the results demonstrate some limitations related to model performance and clinical utility of the model, which should be known prior to applying the tool into routine clinical use outside the United States.

Model performance of the US-based prognostic model is suboptimal in external cohort

The findings demonstrate that the model has reasonable discrimination, when comparing the Canadian cohort to the derivation cohort in the United States.[12] The findings are similar to those of other prediction models that are widely applied, such as the CHADS2-vasc score for predicting

stroke in atrial fibrillation,[29] with respect to the values of discrimination. However, the discrimination in our cohort is not as strong as it was demonstrated when the original derivation cohort was compared against the validation cohorts. Several potential reasons may have contributed to the suboptimal discrimination. One possibility is that the prediction model may be over-fitted, as demonstrated by a shrinkage factor of 0.57 on the regression coefficients, leading to optimistic and biased estimates of performance in the development set. Another possibility may be attributed to the differences in how variables were defined within our study cohort and between the two studies. The associations between dementia and PVD and mortality outcome were found to be weaker in our cohort compared to the development set. The details of dementia classification are not explicitly reported by Cohen et al.[12] and thus differences in diagnostic criteria for this variable may account for some of the issues. Furthermore, practice patterns for dialysis withdrawal when dementia is diagnosed may also differ between populations and could explain the difference in association. There may be similar definition issues for PVD. Other possibilities include collinearity in some variables (e.g. SQ and age) or different factors impacting survival between the two HD populations. This issue merits further investigations.

The potential clinical applications of the model are broad, and thus, identifying appropriate and inappropriate roles for this tool in clinical decision-making is complex but important. The findings in our study suggest that the prognostic model might not be as useful in clinical decision-making that requires a high degree of accuracy around prognosis (i.e. certainty of death within 6 months) and where a discrepancy of prognosis implies harm to the patient. An example of this would include withdrawal of cardioprotective medications and considering and having
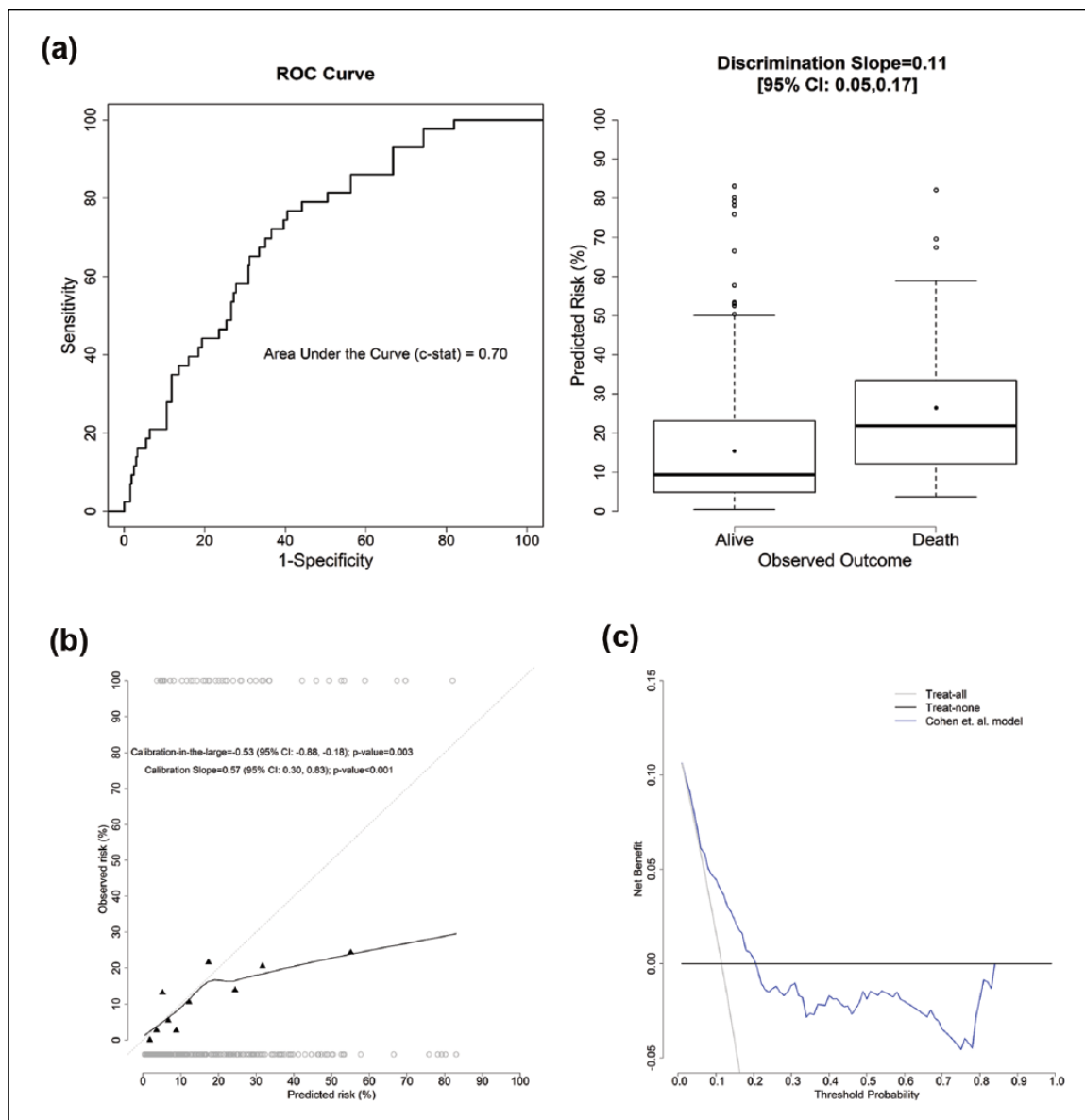
Figure 2. (a) ROC curve (left) and discrimination plot (right). The 6-month prediction model by Cohen et al. provided reasonable discrimination ability in the validation cohort. (b) Calibration plot, where the dark gray circles corresponded to the predicted risk for those who died (top) and alive (bottom) at 6 months. The 6-month prediction model by Cohen et al. was not well calibrated in the validation cohort. (c) Decision curve analysis. The 6-month prediction model by Cohen et al. only has added value in guiding clinical decision in a small range of threshold probabilities: 8%–20%, compared to both the "Treat-all" and "Treat-none" strategies.

conversations regarding withdrawal of dialysis: if death within 6 months is overestimated by the model, then these actions may not be warranted and may cause distress and harm to the patients and their family.

Robust evaluation of prognostic models is crucial prior to implementation

This study demonstrates the use of a robust approach—decision curve analysis—to evaluate clinical utility of a prediction model in nephrology. Decision curve analysis has been advocated as a means of bridging the mathematical features of a model with clinical applicability[26] and has been proposed as part of a standard means of assessing prediction models as per recommendations of the TRIPOD statement.[27] This type of analysis has been used in a variety of settings.[30,31] In this study, the model was shown to have clinical utility for probability threshold below 20% but net harm above 20%.

Further work may involve testing whether or not physicians would be willing to use this prognostic model for guiding aspect of care management that are low risk, for example,

Table 2. Sensitivity, specificity, and positive and negative predicted values at selected probability thresholds.

	Predict death (Probability > 8%)	Predict alive (Probability ≤ 8%)	Total
Death	35	8	43
Alive	186	145	331
Total	221	153	374
Sensitivity = 81.4%		Positive predictive value = 15.8%	
Specificity = 43.8%		Negative predictive value = 94.8%	

	Predict death (Probability > 50%)	Predict alive (Probability ≤ 50%)	Total
Death	6	37	43
Alive	13	318	331
Total	19	355	374
Sensitivity = 13.4%		Positive predictive value = 31.6%	
Specificity = 96.1%		Negative predictive value = 89.6%	

	Predict death (Probability > 20%)	Predict alive (Probability ≤ 20%)	Total
Death	24	19	43
Alive	92	239	331
Total	116	258	374
Sensitivity = 55.8%		Positive predictive value = 20.7%	
Specificity = 72.2%		Negative predictive value = 92.6%	

	Predict death (Probability > 75%)	Predict alive (Probability ≤ 75%)	Total
Death	1	42	43
Alive	6	325	331
Total	7	367	374
Sensitivity = 2.3%		Positive predictive value = 14.3%	
Specificity = 98.2%		Negative predictive value = 88.5%	

Table 3. Assessments for sources of suboptimal model performance.

Performance metrics	Development cohort	BC validation cohort	Reference values for performance		
			Simulation[a]	Partially recalibrated[b]	Fully recalibrated[c]
c-Statistics	0.87 (0.82, 0.92)	0.70 (0.62, 0.78)	0.80 (0.79, 0.80)	0.70 (0.62, 0.78)	0.71 (0.63, 0.78)
Calibration-in-the-large	NA	−0.53 (−0.88, −0.18)	−0.004 (−0.013, 0.006)	0.06 (−0.28, 0.41)	0.01 (−0.32, 0.34)
Calibration slope	NA	0.57 (0.30, 0.83)	1.00 (0.99, 1.01)	0.61 (0.33, 0.89)	0.89 (0.48, 1.30)
Net benefit	NA	8%–20%	6%–80%	7%–21%	4%–30%
Hazard ratio estimates					
Surprise question (not surprise vs surprise)	2.71 (1.75, 4.17)	Based on estimates from development	Based on estimates from development	Based on estimates from development	2.98 (1.96, 4.54)
Serum albumin (per 1 g/dL)	0.27 (0.15, 0.50)				0.45 (0.21, 0.94)
Age (per 10 years)	1.36 (1.17, 1.57)				1.11 (0.96, 1.28)
Peripheral vascular disease (Yes vs No)	1.88 (1.24, 2.84)				1.40 (0.96, 2.03)
Dementia (Yes vs No)	2.24 (1.11, 4.48)				1.64 (1.03, 2.63)
Baseline survival at 6 months	0.58			0.72	0.71

NA: not applicable.

The suboptimal external performance of the 6-month prediction model by Cohen et al. may be explained by the difference in the predictive ability of the five variables but not case-mix.

95% confidence interval of the true value is noted in the parentheses.

[a]Randomly assigning outcome to the underlying case-mix distribution in the BC validation cohort.

[b]New estimate for baseline survival function at 6 months with original hazard ratio estimates from the development cohort.

[c]New estimate for baseline survival function at 6 months and new estimates for hazard ratios based on the BC validation cohort.

coordinating uptake of ACP discussions, versus aspects of care that may require more accuracy and precision such as withdrawing from life-sustaining dialysis treatment. It is a limitation that the model can identify a patient group at higher risk of death, yet does not perform well at the individual level; thus, clinicians must take into account the limitation of the prognostic information yielded from the model when engaging patients in discussions about prognosis.

Strengths and limitations

To our knowledge, this is the first study to undertake a robust external assessment of the model derived by Cohen et al.[12] using a fully independent validation cohort, that is, different HD population and different investigators. Our results showing reduced discrimination and different effect of the variables are an important reminder that these models should be independently validated in several populations before clinicians should place great confidence in their routine use.[32]

Note that the model tested was derived in a different time period than the validation cohort, and the impact of changes in practice prior to dialysis initiation (CKD care), and selection criteria for dialysis, ethnicity of the cohort, and age, may have all impacted the results. Another relative limitation of this study includes the lack of a more well-defined variable classification for PVD and dementia that may have impacted the differences in the predictive effect of PVD and dementia

on mortality in our study cohort. Finally, this study did not assess a broader range of predictor variables beyond the Cohen model, in part due to limitations in our own dataset.

Conclusion

This study indicates that the mortality prediction model by Cohen et al.[12] has limited clinical utility despite reasonable discrimination in a Canadian cohort of HD patients in the current era. The findings of this study do not support its use in guiding specific clinical decisions for individuals that require a high degree of accuracy since the model overpredicts mortality. Our results highlight the need for robust independent validation of prediction models prior to clinical uptake by a broader spectrum of practice, especially given the increased availability of prediction tools.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Kidney Disease Improving Global Outcomes (KDIGO). Chapter 5: referral to specialists and models of care. *Kidney Int* 2013; 3(1): 112–119.
2. Canadian Organ Replacement Register (CORR) and Canadian Institute for Health Information. 2015 CORR annual report: treatment of end-stage organ failure in Canada 2004–2013, https://secure.cihi.ca/estore/product-Family.htm?locale=en&pf=PFC2864
3. Kurella Tamura M, Covinsky KE, Chertow GM, et al. Functional status of elderly adults before and after initiation of dialysis. *N Engl J Med* 2009; 361(16): 1539–1547.
4. Davison R and Sheerin NS. Prognosis and management of chronic kidney disease (CKD) at the end of life. *Postgrad Med J* 2014; 90(1060): 98–105.
5. Davison SN. Integrating palliative care for patients with advanced chronic kidney disease: recent advances, remaining challenges. *J Palliat Care* 2011; 27(1): 53–61.
6. Association RP. *Shared decision-making in the appropriate initiation of and withdrawal from dialysis*. Rockville, MD: Renal Physicians Association, 2010.
7. Galla JH. Clinical practice guideline on shared decision-making in the appropriate initiation of and withdrawal from dialysis. The Renal Physicians Association and the American Society of Nephrology. *J Am Soc Nephrol* 2000; 11(7): 1340–1342.
8. Billings JA and Bernacki R. Strategic targeting of advance care planning interventions: the Goldilocks phenomenon. *J Am Med Assoc Intern Med* 2014; 174(4): 620–624.
9. Wittenberg SM and Cohen LM. Estimating prognosis in end-stage renal disease. *Prog Palliat Care* 2009; 17(4): 165–169.
10. Davison SN. End-of-life care preferences and needs perceptions of patients with chronic kidney disease. *Clin J Am Soc Nephrol* 2010; 5(2): 195–204.
11. Brar R and Tangri N. Predicting death without dialysis in elderly patients with CKD. *Clin J Am Soc Nephrol* 2015; 10(3): 341–343.
12. Cohen LM, Ruthazer R, Moss AH, et al. Predicting six-month mortality for patients who are on maintenance hemodialysis. *Clin J Am Soc Nephrol* 2010; 5(1): 72–79.
13. Couchoud C, Labeeuw M, Moranne O, et al. A clinical score to predict 6-month prognosis in elderly patients starting dialysis for end-stage renal disease. *Nephrol Dial Transplant* 2009; 24(5): 1553–1561.
14. Moss AH, Ganjoo J, Sharma S, et al. Utility of the "surprise" question to identify dialysis patients with high mortality. *Clin J Am Soc Nephrol* 2008; 3(5): 1379–1384.
15. Thamer M, Kaufman JS, Zhang Y, et al. Predicting early death among elderly dialysis patients: development and validation of a risk score to assist shared decision making for dialysis initiation. *Am J Kidney Dis* 2015; 66(6): 1024–1032.
16. Floege J, Gillespie IA, Kronenberg F, et al. Development and validation of a predictive mortality risk score from a European hemodialysis cohort. *Kidney Int* 2015; 87(5): 996–1008.
17. Wagner M, Ansell D, Kent DM, et al. Predicting mortality in incident dialysis patients: an analysis of the United Kingdom Renal Registry. *Am J Kidney Dis* 2011; 57(6): 894–902.
18. Mauri JM, Cleries M and Vela E. Design and validation of a model to predict early mortality in haemodialysis patients. *Nephrol Dial Transplant* 2008; 23(5): 1690–1696.
19. Chen JY, Tsai SH, Chuang PH, et al. A comorbidity index for mortality prediction in Chinese patients with ESRD receiving hemodialysis. *Clin J Am Soc Nephrol* 2014; 9(3): 513–519.
20. Holme I, Fellstrom BC, Jardin AG, et al. Prognostic model for total mortality in patients with haemodialysis from the Assessments of Survival and Cardiovascular Events (AURORA) study. *J Intern Med* 2012; 271(5): 463–471.
21. Chung SH, Noh H, Jeon JS, et al. Impact of incremental risk factors on peritoneal dialysis patient survival: proposal of a simplified clinical mortality risk score. *Blood Purif* 2009; 27(2): 165–171.
22. Quinn RR, Laupacis A, Hux JE, et al. Predicting the risk of 1-year mortality in incident dialysis patients: accounting for case-mix severity in studies using administrative data. *Med Care* 2011; 49(3): 257–266.
23. Pencina MJ and D'Agostino RB Sr. Evaluating discrimination of risk prediction models: the C statistic. *JAMA* 2015; 314(10): 1063–1064.
24. QxMD Software Inc. Calculate by QxMD, www.QxMD.com (accessed 16 December 2017).
25. Clase CM, St Pierre MW and Churchill DN. Conversion between bromcresol green- and bromcresol purple-measured albumin in renal disease. *Nephrol Dial Transplant* 2001; 16(9): 1925–1929.
26. Vickers AJ and Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; 26(6): 565–574.
27. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015; 13: 1.
28. Steyerberg EW and Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014; 35(29): 1925–1931.
29. van den Ham HA, Klungel OH, Singer DE, et al. Comparative performance of ATRIA, CHADS2, and CHA2DS2-VASc risk scores predicting stroke in patients with atrial fibrillation: results from a National Primary Care Database. *J Am Coll Cardiol* 2015; 66(17): 1851–1859.
30. Siddiqui MM, Rais-Bahrami S, Turkbey B, et al. Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer. *JAMA* 2015; 313(4): 390–397.
31. Bouzas-Mosquera A, Peteiro J, Broullon FJ, et al. Incremental value of exercise echocardiography over exercise electrocardiography in a chest pain unit. *Eur J Intern Med* 2015; 26(9): 720–725.
32. Kadatz MJ, Lee ES and Levin A. Predicting progression in CKD: perspectives and precautions. *Am J Kidney Dis* 2016; 67: 779–786.