



## Original Investigation | Nephrology

# Mortality Risk Prediction Models for People With Kidney Failure A Systematic Review

Faisal Jarrar, MD; Meghann Pasternak, MD; Tyrone G. Harrison, MD, PhD; Matthew T. James, MD, PhD; Robert R. Quinn, MD, PhD; Ngan N. Lam, MD, MSc; Maoliosa Donald, PhD, MSc; Meghan Elliott, MD, MSc; Diane L. Lorenzetti, PhD; Giovanni Strippoli, MD; Ping Liu, PhD; Simon Sawhney, MBChB, PhD; Thomas Alexander Gerdts, PhD; Pietro Ravani, MD, PhD

## Abstract

**IMPORTANCE** People with kidney failure have a high risk of death and poor quality of life. Mortality risk prediction models may help them decide which form of treatment they prefer.

**OBJECTIVE** To systematically review the quality of existing mortality prediction models for people with kidney failure and assess whether they can be applied in clinical practice.

**EVIDENCE REVIEW** MEDLINE, Embase, and the Cochrane Library were searched for studies published between January 1, 2004, and September 30, 2024. Studies were included if they created or evaluated mortality prediction models for people who developed kidney failure, whether treated or not treated with kidney replacement with hemodialysis or peritoneal dialysis. Studies including exclusively kidney transplant recipients were excluded. Two reviewers independently extracted data and graded each study at low, high, or unclear risk of bias and applicability using recommended checklists and tools. Reviewers used the Prediction Model Risk of Bias Assessment Tool and followed prespecified questions about study design, prediction framework, modeling algorithm, performance evaluation, and model deployment. Analyses were completed between January and October 2024.

**FINDINGS** A total of 7184 unique abstracts were screened for eligibility. Of these, 77 were selected for full-text review, and 50 studies that created all-cause mortality prediction models were included, with 2 963 157 total participants, who had a median (range) age of 64 (52-81) years. Studies had a median (range) proportion of women of 42% (2%-54%). Included studies were at high risk of bias due to inadequate selection of study population (27 studies [54%]), shortcomings in methods of measurement of predictors (15 [30%]) and outcome (12 [24%]), and flaws in the analysis strategy (50 [100%]). Concerns for applicability were also high, as study participants (31 [62%]), predictors (17 [34%]), and outcome (5 [10%]) did not fit the intended target clinical setting. One study (2%) reported decision curve analysis, and 15 (30%) included a tool to enhance model usability.

**CONCLUSIONS AND RELEVANCE** According to this systematic review of 50 studies, published mortality prediction models were at high risk of bias and had applicability concerns for clinical practice. New mortality prediction models are needed to inform treatment decisions in people with kidney failure.

## Key Points

**Question** What are the quality and clinical applicability of existing mortality prediction models for people with kidney failure?

**Findings** This systematic review identified and evaluated 50 studies with more than 2.9 million participants reporting mortality prediction models for people with kidney failure. Models were found to be at high risk of bias and to have applicability concerns for clinical practice, and none demonstrated promise in terms of clinical usability or incorporation into guidelines.

**Meaning** These findings suggest that new mortality prediction models are needed to inform treatment decisions in people with kidney failure.

## + Supplemental content

Author affiliations and article information are listed at the end of this article.

JAMA Network Open. 2025;8(1):e2453190. doi:10.1001/jamanetworkopen.2024.53190

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

JAMA Network Open. 2025;8(1):e2453190. doi:10.1001/jamanetworkopen.2024.53190

January 3, 2025 1/15

## Introduction

Risk prediction models are increasingly endorsed to help patients understand their treatment preferences and promote personalized care.<sup>1,2</sup> These models are of particular value for people with kidney failure.<sup>3</sup> People with kidney failure have higher morbidity and mortality than those with nonmetastatic cancer<sup>4,5</sup> and often face difficult treatment decisions.<sup>6</sup> Those who cannot receive a kidney transplant need to consider the trade-off between starting or continuing dialysis therapy (peritoneal dialysis or hemodialysis) and choosing conservative management without dialysis. Personalized risk predictions can inform the management of kidney failure for an individual and support decisions that best reflect their unique goals, preferences, and values.<sup>5-7</sup> Yet, current guidelines do not include any recommendation to consult a mortality prediction model.<sup>8</sup>

One possible barrier to sharing mortality information in clinical practice may be a lack of mortality risk prediction models suitably designed to assist with decision-making.<sup>9-11</sup> Alternatively, promising models may exist that provide potentially relevant and useful information but need further evaluation to determine their clinical applicability. Incorporating a risk prediction model into clinical practice requires evidence that a model is relevant to its intended use, outperforms alternative strategies when challenged with data that it has not been trained on, could feasibly be used, and would help make better informed decisions. For example, consider a model that predicts 1-year mortality risk for people newly diagnosed with kidney failure. Such a model could be used to decide whether a person wants to receive dialysis. The model would need to be designed, created, and evaluated with data that represent the target population of model users—in this case, people newly diagnosed with kidney failure not receiving dialysis.

This systematic review assessed the quality and applicability of mortality prediction models for people with kidney failure. We evaluated whether any of them could be implemented or be adapted for implementation or whether new models are required. We followed critical appraisal guidelines<sup>12,13</sup> and used recommended tools for assessment of their risk of bias and applicability to the intended population and settings.<sup>14</sup>

## Methods

### Protocol

The protocol of this systematic review (eMethods and eAppendix in [Supplement 1](#)) was registered with PROSPERO ([CRD42023486220](#)). This systematic review was conducted according to the guidelines set out in the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS)<sup>12</sup> and the Prediction Model Risk of Bias Assessment Tool (PROBAST)<sup>14</sup> and was reported according to the Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis reporting guideline for Systematic Reviews of Multivariable Prediction Models (TRIPOD-SRMA).<sup>15</sup>

### Data Sources and Searches

An information specialist and medical librarian (D.L.L.) searched Ovid MEDLINE, Ovid Embase, and the Cochrane Library from 2004, when the Kidney Disease: Improving Global Outcomes was originally established to develop and implement guidelines for the care of people with kidney disease,<sup>16</sup> to September 30, 2024. Searches combined terms from 3 concepts: (1) chronic kidney failure (eg, *CKD*, *renal insufficiency*); (2) mortality (eg, *mortality*, *death*), and (3) prediction modeling (eg, *calibration*, *measures of discrimination*) and recommended filters for prediction models. Terms were searched as keywords and subject headings (eg, MEDLINE MeSH).<sup>17</sup> No language restrictions were applied to the search strategy. A complete description of the search strategy is provided in eTable 1 in [Supplement 1](#). The reference lists of all included articles were also searched for any additional, relevant articles.

## Study Selection

This review included studies that created mortality prediction models (all-cause mortality or mortality from specific causes) for people with kidney failure treated with long-term dialysis (ie, hemodialysis or peritoneal dialysis) or people with sustained estimated glomerular filtration rate (eGFR) below 15 mL/min/1.73 m<sup>2</sup> with a prediction horizon of at least 3 months.<sup>8</sup> We also considered model evaluation studies to search for the study where the model was created. We excluded studies that were not restricted to people with kidney failure, studies that exclusively included kidney transplant recipients, studies that were limited to patients with acute kidney injury in hospital, and studies that reported associations rather than predictions. We also excluded letters, editorials, narrative reviews, commentaries, and case reports, but their reference lists were used to identify potential primary studies.

## Data Extraction and Quality Assessment

Two reviewers (F.J. and M.P.) independently screened all abstracts of studies in duplicate based on titles and abstracts, and then reviewed full texts to determine eligibility based on inclusion and exclusion criteria. Reviewers followed the CHARMS<sup>12</sup> and PROBAST<sup>14</sup> recommendations to extract elements of the prediction framework and analysis strategies as explained later. Reviewers also considered the extent to which each study adhered to the TRIPOD+AI reporting standard for all prediction models, irrespective of whether regression or machine learning methods were used.<sup>13</sup> Discrepancies between the 2 reviewers in study selection for inclusion and data extraction were resolved by discussing with 2 arbitrators (P.R. and P.L.). Analyses were completed between January and October 2024.

## Elements for Critical Appraisal

Reviewers used the CHARMS checklist (eTable 2 in [Supplement 1](#))<sup>12</sup> and the PROBAST signaling questions (eTable 3 in [Supplement 1](#))<sup>14</sup> on the domains of participants, predictors, outcome, and analysis. PROBAST provides a structured approach to rate the risk of bias and applicability concerns as low, high, or unclear.

A medical risk prediction model is developed in a suitable framework that defines the target population, the prediction time origin (when the model is applied to the patient data), the prediction time horizon, the predictors, the outcome, and competing events (if any). The framework implicitly defines who can use the model as well as how and when it can be used.<sup>18</sup> Risk of bias and applicability concerns are high if study participants, predictors, and outcome do not represent the settings in which the model will be used.

For the risk of bias, PROBAST also includes signaling questions for statistical analysis (analysis domain). Here the focus is on data quality, the statistical methods used to analyze the data, the modeling algorithm that produces the prediction model, and the methods used to evaluate the prediction performance, utility, and usability.

Reviewers assessed the modeling algorithms that the primary studies used to make the prediction model (eMethods in [Supplement 1](#)). Reviewers considered modeling algorithms inappropriate if they based the selection of the predictors on univariate analyses and then used expert-driven model building and goodness-of-fit testing to obtain the prediction model.<sup>14,17</sup> Reviewers also considered backward variable selection as inappropriate unless the modeling algorithm was evaluated using cross-validation in which all steps of the modeling algorithm, including the backward selection, were repeated in training sets and evaluated in independent test sets.<sup>17,19</sup> Similar considerations apply to the selection of hyperparameters or tuning of machine learning algorithms.<sup>18</sup>

Reviewers assessed whether the primary study used a single split of the data or repeated splits of the data for cross-validation (evaluation using the learning data or internal testing). A single random split is not recommended because the results will typically depend on the random seed (Monte-Carlo error), it is prone to manipulation, and it conceals part of the learning data.<sup>18</sup> Finally,

reviewers assessed whether the final model was evaluated using temporally or geographically distinct data (ie, external testing; eMethods in Supplement 1).

Reviewers extracted the criteria for model evaluation and comparison of rival models as reported by the authors (eMethods in Supplement 1), including calibration plots and the time-dependent area under the receiver operating characteristic curve (AUC), a measure of discrimination, and the time-dependent Brier score or prediction error, a measure of both calibration and discrimination.<sup>18</sup> Reviewers reported whether the primary study used improper performance measures, including the C index (Harrell concordance index)<sup>20</sup> and measures of reclassification.<sup>21,22</sup> These measures are not proper because they may erroneously show that a misspecified model systematically outperforms the data-generating model.<sup>18</sup>

Reviewers assessed whether study authors applied decision curve analysis<sup>23</sup> or whether they linked specific clinical decisions with categories defined by predicted risks.<sup>24</sup> Finally, reviewers considered whether a model was tested in a clinical trial, as this is the ultimate test of model utility.<sup>25</sup> For usability, reviewers noted whether authors provided a calculator, nomogram, or an alternative tool that would ease clinician access to patient predicted risks. Data were collected in Excel version 16.91 for Mac (Microsoft Corp) tables.

Results

Study Characteristics

Of the 7184 titles and abstracts screened, 77 studies were selected for full-text review, and 50 articles<sup>26-75</sup> met the inclusion criteria (Figure 1). These studies reported 50 unique models including a total of 2 963 157 participants (range, 173-1 150 195 participants), with a median (range) age of 64 (52-81) years and median (range) proportion of women of 42% (2%-54%). All models were designed to predict all-cause mortality with time horizons ranging from 3 months to 10 years. Study characteristics are shown in Table 1.

Adherence to Recommended Reporting Guidelines

None of the studies met all criteria of the TRIPOD+AI statement for methods. Sample size and power considerations were not reported in any of the studies, and information on whether there were missing data and how they were handled was not reported in 26 studies (52%).<sup>26, 28, 29, 33, 43, 45, 46, 49, 51, 53, 54, 57-63, 66-69, 71-74</sup> As for the TRIPOD+AI results checklist, none of the studies met all criteria.

Figure 1. Study Selection

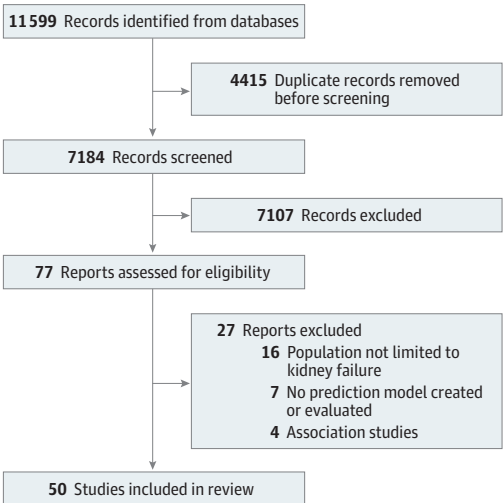


Table 1. Characteristics of the 50 Studies Included in the Systematic Review<sup>a</sup>

Source	Design	Enrolment period	Study setting	Study region	Participant No. <sup>b</sup>	Age at entry <sup>c</sup>	Participants, %		
							Women	HD	Incident
Chen et al, <sup>26</sup> 2014	Retrospective cohort	2005-2010	Multicenter	Taiwan	30 303	64.3 (13.3)	48.4	100	100
Obi et al, <sup>27</sup> 2018	Retrospective cohort	2007-2015	Multicenter	US	35 878	69.4 (11.1)	2	NA	100
Inaguma et al, <sup>28</sup> 2019	Prospective cohort	2011-2013	Multicenter	Japan	1520	70 (60.8)	32.4	NA	100
Santos et al, <sup>29</sup> 2020	Retrospective cohort	2009-2016	Single center	Portugal	421	75.5 (70.8)	46.3	97.6	100
Pladys et al, <sup>30</sup> 2020	Retrospective cohort	2015	Multicenter	France	9052	68.4 (15.1)	35.6	90.7	100
Hemke et al, <sup>31</sup> 2013	Prospective cohort	1997-2007	Multicenter	Netherlands	13 868	59.7 (15.1)	38.7	64.5	100
Chen et al, <sup>32</sup> 2017	Retrospective cohort	2006-2009	Database	US	159 362	NA	46	95.21	100
Chua et al, <sup>33</sup> 2014	Retrospective cohort	2005-2010	Single center	Singapore	983	60 (13)	47.7	72.4	100
van Diepen et al, <sup>34</sup> 2014	Prospective cohort	1997-2007	Multicenter	Netherlands	394	65.3 (54.5-72.4)	45	69	100
Floege et al, <sup>35</sup> 2015	Prospective cohort	2007-2009	Fresenius medical care	Multinational	9722	64.4 (14.7)	40.2	100	100
Dusseaux et al, <sup>36</sup> 2015	Retrospective cohort	2002-2006	French national registry	France	8955	78 (74-82)	40	NA	100
Doi et al, <sup>37</sup> 2015	Retrospective cohort	2006-2011	Multicenter	Japan	688	69 (59-77)	33.4	100	0
Thamer et al, <sup>38</sup> 2015	Retrospective cohort	2009-2010	Multicenter	US	52 796	76.9 (6.5)	46.2	95.8	100
Couchoud et al, <sup>39</sup> 2015	Retrospective cohort	2005-2012	Multicenter	France	24 348	81.1	39.6	NA	100
Hemke et al, <sup>40</sup> 2015	Retrospective cohort	1997-2007	Multicenter	Netherlands	1835	59.7 (15.1)	38.7	64.5	100
Patzner et al, <sup>41</sup> 2016	Retrospective cohort	2005-2011	Multicenter	US	663 860	64.0 (14.9)	44.2	NA	100
Haapio et al, <sup>42</sup> 2017	Retrospective cohort	2000-2008	Multicenter	Finland	4335	62.3 (21.2)	36.5	76.1	100
Lin et al, <sup>43</sup> 2019	Retrospective cohort	2000-2011	Multicenter	Taiwan	48 153	75 (69.5-79.0)	54	NA	NA
Akbulic et al, <sup>44</sup> 2019	Retrospective cohort	2007-2014	Multicenter	US	27 615	68.7 (11.2)	1.9	94	100
Cho et al, <sup>45</sup> 2017	Retrospective cohort	2005-2008	Multicenter	South Korea	7606	54.5 (13.8)	44	0	100
Wick et al, <sup>46</sup> 2017	Retrospective cohort	2003-2012	Multicenter	Canada	2199	75.2 (6.5)	39.2	85.5	100
Ivory et al, <sup>47</sup> 2017	Retrospective cohort	2000-2009	Multicenter	Australia, UK, New Zealand	23 658	NA	40	NA	100
Geddes et al, <sup>48</sup> 2006	Combined data <sup>d</sup>	1997-2002	Multicenter	Multinational	2310	63.4 (13.5)	42.3	NA	100
Mauri et al, <sup>49</sup> 2008	Prospective cohort	1997-2003	Multicenter	Spain	5738	64.6 (14.4)	37.8	100	100
Couchoud et al, <sup>50</sup> 2009	Retrospective cohort	2002-2006	Multicenter	France	4991	80.9 (4.1)	39.4	NA	100
Liu et al, <sup>51</sup> 2010	Retrospective cohort	1999-2005	Multicenter	US	33 077	65 (15)	48.1	NA	100
Jacob et al, <sup>52</sup> 2010	Prospective cohort	1990-2007	Multicenter	US	242 576	NA	46.68	89.23	100
Marinovich et al, <sup>53</sup> 2010	Retrospective cohort	2004-2005	Multicenter	Argentina	5360	58.9 (15.5)	43.8	100	100
Quinn et al, <sup>54</sup> 2011	Retrospective cohort	1998-2005	Multicenter	Canada	16 205	62.81 (15.7)	41.64	75.75	100
Wu et al, <sup>55</sup> 2022	Retrospective cohort	2006-2015	Multicenter	Taiwan	210 174	NA	42.6	NA	100
Noh et al, <sup>56</sup> 2020	Retrospective cohort	2008-2014	Multicenter	Korea	1730	52.7 (12.6)	42.7	0	NA
Siddiqi et al, <sup>57</sup> 2021	Retrospective cohort	2013-2017	Multicenter	Pakistan	758	NA	NA	100	NA
McAdams-DeMarco et al, <sup>58</sup> 2018	Retrospective cohort	2009-2016	Multicenter	US	1975	53.7 (13.5)	40.5	67.7	NA
Gao et al, <sup>59</sup> 2022	Retrospective cohort	2011-2019	Multicenter	China	200	52.26 (13.3)	45	60.5	NA
Chaudhuri et al, <sup>60</sup> 2023	Retrospective cohort	2000-2019	Multicenter	Global	76 113	61.7	42.6	100	NA
Thijssen et al, <sup>61</sup> 2012	Retrospective cohort	2000-2009	Multicenter	US	4512	61.3 (15.5)	43.4	100	100
Wagner et al, <sup>62</sup> 2011	Retrospective cohort	2002-2004	Multicenter	UK	3631	64 (49.73)	37.8	70.1	100
Zhu et al, <sup>63</sup> 2021	Retrospective cohort	2010-2016	Single center	China	173	58	30	100	100
Cohen et al, <sup>64</sup> 2010	Prospective cohort	2006-2008	Multicenter	US	512	61 (17)	44	100	NA
Siga et al, <sup>65</sup> 2020	Prospective cohort	2010-2014	Multicenter	France	4915	NA	NA	100	100
Jung et al, <sup>66</sup> 2018	Prospective cohort	2008-2011	Multicenter	Korea	3309	61.7 (13.6)	41.2	71.2	NA
Wang et al, <sup>67</sup> 2021	Retrospective cohort	2007-2016	Multicenter	China	1200	55 (NA)	36.5	100	NA
Tapak et al, <sup>68</sup> 2020	Retrospective cohort	2007-2017	Multicenter	Iran	785	NA	45.1	100	100
Holme et al, <sup>69</sup> 2012	Prospective cohort (secondary analysis)	2003-2004	Multicenter	International	1868	64 (8.6)	38	100	0
Rankin et al, <sup>70</sup> 2022	Retrospective cohort	2008-2017	Multicenter	US	1 150 195	63 (15)	42	NA	100

(continued)

Table 1. Characteristics of the 50 Studies Included in the Systematic Review<sup>a</sup> (continued)

Source	Design	Enrolment period	Study setting	Study region	Participant No. <sup>b</sup>	Age at entry <sup>c</sup>	Participants, %		
							Women	HD	Incident
Fernandez Lucas et al, <sup>71</sup> 2007	Retrospective cohort	2000-2004	No information	Spain	304	64 (16-84)	37	80	100
Goldstein et al, <sup>72</sup> 2024	Retrospective cohort	2003-2013	Single center	US	42 351	63.4 (52.5-73.6)	43	100	100
Noppakun et al, <sup>73</sup> 2024	Retrospective cohort	2005-2016	Thai registry	Thailand	17 354	76.9 (5.1)	53.5	100	100
Yang et al, <sup>74</sup> 2023	Retrospective cohort	2007-2020	Single center	China	551	61 (51-70)	47.2	100	100
Okada et al, <sup>75</sup> 2024	Retrospective cohort	2006-2007	Japanese registry	Japan	2739	80 (77,83)	42.7	100	100

Abbreviations: HD, hemodialysis; NA, not available.

<sup>a</sup> No study included people with kidney failure not treated with kidney replacement. All studies were written in English language.

<sup>b</sup> Participant number refers to the number of incident and/or prevalent patients receiving dialysis. As opposed to incident patients for whom the study entry date corresponded to the dialysis start date, prevalent patients had been receiving dialysis for a variable amount of time when they entered the study. One study (Patzner et al,<sup>41</sup> 2016) included transplant recipients (whose data were not included in this review).

<sup>c</sup> Age in years summarized as mean (SD) or median (range).

<sup>d</sup> Design combined retrospective and prospective data.

Criteria for model specification was not met in 28 studies (56%)<sup>26, 28, 30, 32, 36, 43-46, 50-57, 63-69, 71, 73, 75</sup> and participants in 34 studies (68%).<sup>26, 28, 32-34, 37, 39, 43-47, 49, 51-54, 56-61, 63-69, 71-74</sup> None of the studies met all criteria for usability of the model in the context of current care. Data related to TRIPOD+AI recommendations are shown in eTable 4 in Supplement 1.

Prediction Framework

All studies included people who had already made a treatment choice for kidney failure, ie, dialysis therapy. None included people who had chosen conservative care without dialysis or people who still had to decide their preferred treatment. Most studies (39 [78%])<sup>26-36, 38-42, 44-55, 61-63, 65, 68, 70-75</sup> included incident patients who entered the study on the dialysis start date. Two studies (4%)<sup>37, 69</sup> included prevalent patients who had been receiving dialysis for a variable time before the enrolment date. The remaining 9 studies (18%)<sup>43, 56-60, 64, 66, 67</sup> did not specify whether patients were prevalent patients already receiving dialysis or incident patients who started dialysis at study entry. Information on the type of dialysis was not reported in 12 studies (24%),<sup>27, 28, 36, 39, 41, 43, 47, 48, 50, 51, 55, 70</sup> 18 studies (36%)<sup>26, 35, 37, 49, 53, 57, 60, 61, 63-65, 67-69, 72-75</sup> included only patients receiving hemodialysis, 2 studies (4%)<sup>45, 56</sup> only patients receiving peritoneal dialysis patients, and the remaining (18 studies [36%])<sup>29-34, 38, 40, 42, 44, 46, 52, 54, 58, 59, 62, 66, 71, 76</sup> a variable proportion of the 2 dialysis modalities. All studies used a prediction framework with a single time origin (cohort entry). In 16 studies (32%),<sup>34, 43, 51, 56-60, 62-69</sup> the prediction time origin was not described, and 3 studies (6%)<sup>51, 58, 72</sup> did not specify a prediction time horizon. Sixteen studies (32%)<sup>26, 29, 51, 52, 56, 58-61, 63, 65-69, 71</sup> did not describe the intended use of their model. Four studies (8%)<sup>31, 35, 40, 45</sup> used predictors measured after the prediction time origin, while 11 studies (22%)<sup>35, 43, 56, 57, 59, 63-68</sup> did not report when predictors were measured. Prediction framework data are summarized in Table 2 and detailed in eTable 5 in Supplement 1.

Modeling Algorithms and Evaluation of Prediction Performance

Primary studies used traditional and/or machine learning methods for prediction model creation (eTable 6 in Supplement 1), including logistic regression (19 studies [38%])<sup>29, 31-34, 36-39, 41, 42, 46, 47, 49, 50, 54, 55, 61, 75</sup>, Cox regression (21 studies [42%])<sup>26-28, 30, 35, 40, 45, 48, 51, 53, 57-59, 62-64, 66, 69, 71, 73, 74</sup>, and random forests (1 study [2%])<sup>68</sup> for survival outcomes. Six (12%) used machine learning algorithms for binary outcomes: random forests (1 study [2%])<sup>44</sup>, random forests and neural networks (1 study [2%])<sup>43</sup>, extreme gradient boosting machines (2 studies [4%])<sup>76, 9</sup>, logistic regression and Bayesian networks (1 study [2%])<sup>65</sup>, and neural networks (1 study [2%])<sup>67</sup>. Two studies (4%)<sup>56, 72</sup> used regression methods, random forests, extreme gradient boosting, and neural networks for binary and



survival outcomes. One study (2%)<sup>52</sup> used linear regression and neural networks for continuous outcomes (time). None of the studies that used methods for binary outcomes provided information on how censoring was handled. Most predictor variable selection strategies were based on univariate analyses. Of the 21 studies (42%)<sup>26, 27, 31, 33, 34, 37, 38, 40-42, 46, 50, 52, 54, 55, 59, 60, 64, 69, 72, 73</sup> that applied automatic variable selection methods, 7 studies (33%)<sup>27, 34, 37, 40, 46, 54, 73</sup> accounted for optimism, but none used cross-validation to evaluate the prediction performance of the selected prediction model. Results of internal evaluation of the prediction performance were reported in 34 studies (68%), of which 7 studies (21%)<sup>26, 43, 44, 46, 55, 65, 67</sup> used *k*-fold cross-validation, 6 studies (18%)<sup>28, 29, 34, 37, 54, 68</sup> used a not otherwise explained bootstrapping procedure, and 21 studies (62%)<sup>27, 30-32, 39-41, 49, 50, 52, 56, 57, 60-62, 69, 70, 72-75</sup> used a single random split of the data. Results of an external evaluation of the prediction performance were reported in 13 studies (26%). One of these studies (8%)<sup>47</sup> tested the

**Table 2. Studies Meeting Criteria for Optimal Prediction Framework Design, Model Training and Testing Strategies, and Measures of Prediction Performance and Usefulness**

Criterion	Criterion definition	Studies addressing each item
<b>Prediction framework</b>		
Target population	Describes whether the study addressed or specified for whom the prediction model is intended to be used. Study outlines which inclusion and exclusion criteria should be applied when using the model, eg, if making predictions in people starting dialysis (target population), the prediction model should be trained and tested in incident patients.	Kidney replacement: 38 studies (76%) <sup>26, 29-35, 37, 38, 40, 42, 44-46, 49, 52-54, 56-69, 71-75</sup> ; kidney failure not treated with kidney replacement: 0 studies; unclear: 12 studies (24%) <sup>27, 28, 36, 39, 41, 43, 47, 48, 50, 51, 55, 70</sup>
Target of the analysis	Describes the parameter the study aimed to estimate (eg, individualized risk of death). The analysis target is the mortality risk prediction for an individual.	All clearly defined the event of death; none identified the target of the analysis as the predicted risk for an individual
Time origin for prediction	Reflects when time zero of survival analysis was in the study. Does the time zero in the study reflect the time when predictions are made? Is time zero a common time point in participants' disease trajectory?	First dialysis: 30 studies (60%) <sup>26-30, 32, 33, 36-39, 41, 42, 44, 46-50, 52-55, 61, 70-75</sup> ; after dialysis start: 4 studies (8%) <sup>31, 35, 40, 45</sup> ; unclear: 16 studies (32%) <sup>34, 43, 51, 56-60, 62-69</sup>
Prediction time horizons	Describes what time horizons were chosen in the study, ie, how far in time from the time origin the prediction is projected (eg, 10-y survival probability). Do the time horizons reflect a range that covers age-dependent patient needs (early for older people and more distant for younger people)?	Not reported: 3 studies (6%) <sup>51, 58, 72</sup> ; reported: 47 studies (94%) <sup>26-50, 52-57, 59-71, 73-75</sup>
Static vs dynamic risk prediction	Describes whether the models are meant to be used once at baseline or repeatedly over time (ie, static vs dynamic). If repeatedly, was immortal time addressed by landmarking analysis?	Static: 50 studies (100%) <sup>26-75</sup> ; dynamic: 0 studies
Outcome of interest	Describes the event of interest (eg, all-cause death, disease-specific death). Does the definition and the method to capture death parallel real-life situation?	All-cause death: 50 studies (100%) <sup>26-75</sup>
Competing risks	Describes whether competing risks were considered, if applicable.	Not applicable to all-cause mortality
Intended use	Describes who is going to use the model and how. Did the authors specifically address who the intended model users are?	Described: 34 studies (68%) <sup>27, 28, 30-50, 53-55, 57, 62, 64, 70, 72-75</sup> ; unclear: 16 studies (32%) <sup>26, 29, 51, 52, 56, 58-61, 63, 65-69, 71</sup>
<b>Model testing</b>		
Internal testing	Describes whether internal testing techniques were used by the authors. Only cross-validation is appropriate. Measured by statistics of model performance, such as discrimination and calibration.	Not performed: 16 studies (32%) <sup>33, 35, 36, 38, 42, 45, 47, 48, 51, 53, 58, 59, 63, 64, 66, 71</sup> ; performed: 34 studies (68%) <sup>26-32, 34, 37, 39-41, 43, 44, 46, 49, 50, 52, 54-57, 60-62, 65, 67-70, 72-75</sup> with 21 (62%) <sup>27, 30-32, 39-41, 49, 50, 52, 56, 57, 60-62, 69, 70, 72-75</sup> single split, 7 (21%) <sup>26, 43, 44, 46, 55, 65, 67</sup> cross-validation, and 6 (18%) <sup>28, 29, 34, 37, 54, 68</sup> bootstrapping
External testing	Describes whether the authors tested the performance of the model using data not seen during model training.	Performed: 13 studies (26%) <sup>27, 35, 36, 38, 42, 45, 47, 48, 51, 55, 57, 64, 74</sup> with 1 study (8%) <sup>47</sup> using temporally and geographically distinct data; 8 studies (62%) <sup>36, 38, 42, 48, 51, 55, 64, 75</sup> using temporally distinct data only; and 4 studies (31%) <sup>27, 35, 45, 57</sup> using data from different data sources
<b>Model performance</b>		
Performance measures	Common performance measures, including the time-dependent AUC and Brier score (prediction error, a measure of both calibration and discrimination), were considered.	C index: 43 studies (86%) <sup>26, 27, 29-34, 39-47, 49-51, 53-70, 73, 75</sup> ; AUC graph: 22 studies (44%) <sup>28-30, 34, 35, 44-46, 48, 53, 59-61, 63, 65-67, 70-72, 74, 75</sup> ; Brier score: 2 studies (4%) <sup>47, 63</sup> . calibration plot: 25 studies (50%) <sup>26, 27, 30, 31, 34-38, 40-42, 46, 47, 50, 54, 55, 62, 63, 65, 66, 70, 72, 74</sup>
<b>Clinical usefulness</b>		
Utility	Do the authors discuss a threshold risk or risk category above which a diagnostic or therapeutic decision is made. Is there a decision curve analysis showing net benefit of using the model?	Met criteria: 1 study (2%) <sup>74</sup>
Usability	There is an easy-to-use risk calculator or nomogram that facilitates risk prediction at the bedside.	Met criteria: 15 studies (30%) <sup>27, 29, 30, 34, 35, 37, 39, 41, 46, 47, 55, 59, 60, 63, 74</sup>

Abbreviation: AUC, area under the receiver operating characteristic curve.

model using temporally distinct data and geographically distinct data, 8 (62%)<sup>36,38,42,48,51,55,64,75</sup> used temporally distinct data, and 4 (31%)<sup>27,35,45,57</sup> used data from different data sources (eTable 5 in Supplement 1).

Performance Measures

The most used measures of prediction performance were measures of discrimination ability, including the C index and the time-independent AUC (41 studies [82%]<sup>26-42,45-51,53-55,57-59,61-64,66,68,69,71</sup>). A calibration plot was reported in 24 studies (48%).<sup>26,27,30,31,34-38,40-42,46,47,50,54,55,62,63,65,66,70,72,74</sup> Two studies (4%)<sup>47,63</sup> reported Brier scores (eTable 6 in Supplement 1).

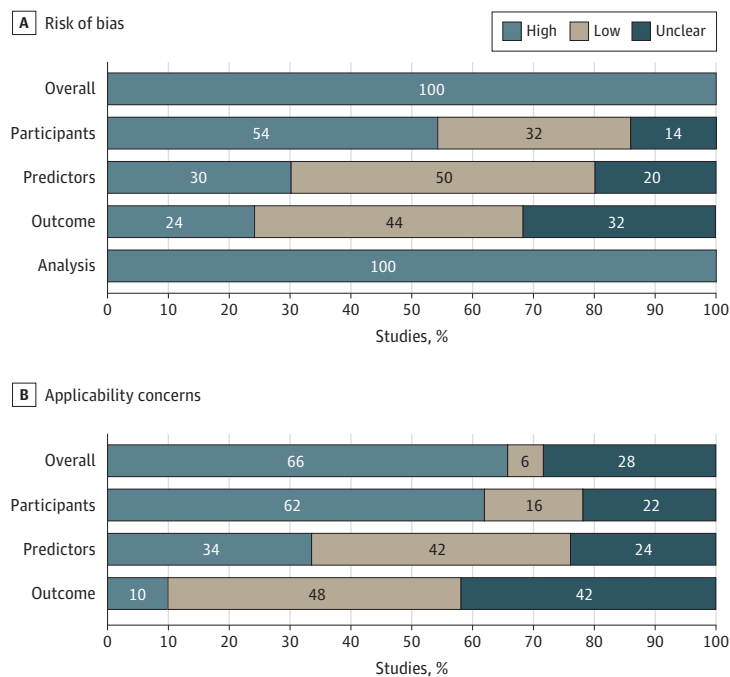
Clinical Usefulness

Criteria for usability (publication of a risk calculator or nomogram) were met in 15 studies (30%)<sup>27,29,30,34,35,37,39,41,46,47,55,59,60,63,74</sup> (eTable 5 in Supplement 1). One study (2%)<sup>74</sup> reported decision curve analysis.

Risk of Bias and Applicability Concerns

Risk of bias and applicability assessment are summarized in Figure 2 and detailed in eTable 6 in Supplement 1. All included studies were at a high risk of bias due to inadequate selection of study population (27 studies [54%]<sup>26-28,33,37,41,43-45,47,48,51-53,55,56,59,60,62,64-69,71,75</sup>), shortcomings in methods of measurement of predictors (15 studies [30%]<sup>26,33,37,48,51,52,57-59,61-64,67,72</sup>) and outcome (12 studies [24%]<sup>28,33,38,45,48,53-55,58,59,65,67</sup>), and flaws in the analysis strategy (all studies [100%]<sup>26-75</sup>). Concerns for applicability were also high, as study participants (31 studies [62%]<sup>26-28,33,36,37,41,43-45,47,48,51,52,55-60,62-71,75</sup>), predictors (17 studies [34%]<sup>26,33,40,48,51,52,56-59,62-64,67,69,70,72</sup>), and outcome (5 studies [10%]<sup>28,48,58,65,67</sup>) did not fit the intended target clinical setting.

Figure 2. Risk of Bias and Applicability Concerns



Risk of bias and applicability concerns were rated using the Prediction Model Risk of Bias Assessment Tool assessment tool as low, high, and unclear considering 4 domains for bias (A) and 3 for applicability (B). For bias and applicability, an overall rating is presented above the domain-specific ratings. See eTable 5 in Supplement 1 for details.



## Discussion

This systematic review identified 50 mortality prediction models for people with kidney failure published in the last 2 decades. All models were created using data from people with kidney failure receiving hemodialysis or peritoneal dialysis. No study included people with kidney failure before they had made a treatment decision. While some models were created with a well-designed prediction framework, thus far none demonstrated promise in terms of clinical usability or incorporation into guidelines. None of the prediction models were ready to be updated<sup>77</sup> or retrained.<sup>24</sup> New research is needed to create clinical prediction models for people with kidney failure before and after a treatment decision has been made.

A recent meta-analysis<sup>78</sup> completed a quantitative comparison of mortality prediction performance measures obtained from studies reporting mortality prediction models developed for patients starting dialysis. However, model performance measures do not have a direct clinical use, and the meta-analysis did not propose a new prediction model. Hence the value of meta-analytic summaries of prediction performance is limited.<sup>18</sup> Our review focused on the quality and appropriateness of existing prediction models with emphasis on risk of bias and applicability concerns as defined by the CHARMS and PROBAST tools.<sup>12,14,17</sup> According to these reporting guidelines, existing prediction models that this review identified should not be used in clinical practice without further evaluation.

Most studies included in this review focused on measures of discrimination for the evaluation of their prediction models. The most common measures of discrimination for survival analyses are the time-dependent AUC and the C index.<sup>20</sup> Both measures assess the ability of the prediction model conditional on the outcome, and hence, they have limited clinical value because the outcome is unknown at the prediction time origin when treatment decisions are made. Calibration, on the other hand, has a direct clinical interpretation. For example, if a well-calibrated model predicts that the 5-year risk of death for a patient is 27%, then we can expect that 27 of 100 persons similar to that patient will die within 5 years.<sup>18</sup> Note also that a model can systematically predict too high (or too low) risks, and the discrimination measures may not detect this bias (miscalibration).

Mortality prediction models for people with kidney failure can support patient-clinician discussions at a crucial point where decisions are made about treatment options, including long-term kidney replacement therapy, conservative kidney management, and/or end-of-life planning.<sup>7</sup> A predicted risk can guide people with kidney failure when they engage in informed discussions on prognosis, decision-making, and life planning according to their personal views and values. Existing guidelines and the American Society of Nephrology Choosing Wisely campaign emphasize that individualized prognostic information should be included in the decision to initiate dialysis.<sup>79</sup> However, current CKD guidelines do not recommend the use of any specific prediction model for mortality.<sup>8</sup> Our review identified an abundance of mortality prediction models for people with kidney failure, suggesting that limitations in study design rather than number of studies is the primary problem. None of the existing studies involved patients, caregivers, and clinicians in prediction model design, which may help to promote acceptability and enhance usability and uptake in clinical practice. End-user engagement is key to ensure that the needs and personal preferences are addressed when a prediction model is created starting with the target population, including people who are making treatment decisions about dialysis, conservative kidney management, and/or dialysis withdrawal.

Our study has implications for future research. New mortality prediction models should be created for people with kidney failure. To inform the choice of conservative kidney management vs dialysis, the study cohort should include people with kidney failure before the treatment has been decided. The prediction time origin should be set at the time when treatment decisions are made. Predictors should be measured up to the prediction time origin and not after. If updated information on predictors or treatments is available,<sup>80</sup> eg, after 6 months from the initial prediction origin, an updated predicted risk can be provided for people who have survived the first 6 months. Key

elements of the analysis plan for the creation of a medical prediction model include the preparation of a learning dataset that reflects the prediction framework, the choice of an adequate scoring rule to compare rival modeling strategies, and some form of cross-validation to evaluate the prediction performance of the final model. External testing could be performed using temporally or geographically distinct data considering that heterogeneity in model performance across time and geography is expected due to differences in patient characteristics, clinical practice, health policy, and measurement procedures across regions and over time within the same region.<sup>76</sup> Temporal testing works by splitting the study cohort according to a calendar date into training and testing data. This is advised to mimic the performance of the model in future patients and to identify any deterioration in model performance due to population or health care practice changes over time. The model can subsequently be updated using more recent data and its reproducibility verified as new temporally distinct data become available. Geographical testing for transportability is useful if the target populations and settings do not deviate too much from those for which the model was originally designed. Finally, the model should be evaluated in real-life conditions, eg, by a cluster randomized study that compares treatment uptake and outcomes between people who use standard care and people who use a care strategy informed by the model.<sup>25</sup>

## Strengths and Limitations

Our study is strengthened by a comprehensive and sensitive search strategy that covered several databases without language restrictions. We used the CHARMS and PROBAST tools, which were specifically created for the critical appraisal of prediction models designed to inform treatment decisions.

Our study also has limitations, largely related to the limitations of the included studies. First, included prediction models were likely designed to address a range of different research needs, as opposed to inform clinical practice. Second, none involved people with kidney failure who still had to make treatment decisions or who chose conservative care without dialysis. Additionally, many studies included in this review were published before recommended reporting standards and quality assessment tools for risk prediction modeling were published.

## Conclusions

According to this systematic review of 50 studies, published mortality prediction models for people with kidney failure were not found suitable to inform clinical decision-making. We advocate for the use of existing guidelines and checklists to design, conduct, and report prediction modeling studies and the involvement of stakeholders in the study design to enhance model usability and clinical uptake.

## ARTICLE INFORMATION

**Accepted for Publication:** November 1, 2024.

**Published:** January 3, 2025. doi:10.1001/jamanetworkopen.2024.53190

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](#). © 2025 Jarrar F et al. *JAMA Network Open*.

**Corresponding Author:** Pietro Ravani, MD, PhD, Departments of Medicine and Community Health Sciences, Cumming School of Medicine, University of Calgary, 3230 Hospital Dr NW, Calgary, AB T2N 4Z6, Canada ([pravani@ucalgary.ca](mailto:pravani@ucalgary.ca)).

**Author Affiliations:** Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada (Jarrar, Pasternak, Harrison, James, Quinn, Lam, Donald, Elliott, Liu, Ravani); Libraries and Cultural Resources, University of Calgary, Calgary, Alberta, Canada (Lorenzetti); Department of Precision and Regenerative Medicine and Jonian Area, University of Bari, Bari, Italy (Strippoli); School of Public Health, University of Sydney, Sydney, New South Wales, Australia (Strippoli); Aberdeen Centre for Health Data Science, University of Aberdeen,

Aberdeen, Scotland, United Kingdom (Sawhney); Department of Public Health, University of Copenhagen, Copenhagen, Denmark (Gerds).

**Author Contributions:** Drs Jarrar and Ravani had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

**Concept and design:** Jarrar, Pasternak, Donald, Lorenzetti, Strippoli, Liu, Sawhney, Ravani.

**Acquisition, analysis, or interpretation of data:** Jarrar, Pasternak, Harrison, James, Quinn, Lam, Elliott, Lorenzetti, Liu, Sawhney, Gerds, Ravani.

**Drafting of the manuscript:** Jarrar, Sawhney, Ravani.

**Critical review of the manuscript for important intellectual content:** All authors.

**Statistical analysis:** Jarrar, Sawhney, Gerds, Ravani.

**Obtained funding:** Lam, Liu, Ravani.

**Administrative, technical, or material support:** Harrison, James, Quinn, Liu, Ravani.

**Supervision:** Harrison, Strippoli, Ravani.

**Conflict of Interest Disclosures:** Dr Quinn reported receiving speaker fees, attending advisory boards, and receiving research support from Baxter Corp and co-owning a patent for the Dialysis Measurement Analysis and Reporting Systems. No other disclosures were reported.

**Funding/Support:** We disclose the following financial support for the research, authorship, or publication of this article: Dr Liu received the Kidney Research Scientist Core Education and National Training (KRESCENT) New Investigator Award, cosponsored by the Kidney Foundation of Canada and Canadian Institutes of Health Research (FRN 2023KNIA-1058404) and held a grant (LevMax) from Alberta Innovates. Dr Harrison is supported by a KRESCENT New Investigator Award and is supported as a new investigator by the Roy and Vi Baay Chair for Kidney Research and the Kidney Health and Wellness Institute at the University of Calgary. Dr Ravani held Canadian Institutes for Health Research funding (FRN 173359) to support studies in chronic kidney disease and was supported by the Baay Chair in Kidney Research at the University of Calgary.

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Data Sharing Statement:** See [Supplement 2](#).

## REFERENCES

1. Deardorff WJ, Barnes DE, Jeon SY, et al. Development and external validation of a mortality prediction model for community-dwelling older adults with dementia. *JAMA Intern Med*. 2022;182(11):1161-1170. doi:10.1001/jamainternmed.2022.4326
2. Chapman BP, Lin F, Roy S, Benedict RHB, Lyness JM. Health risk prediction models incorporating personality data: motivation, challenges, and illustration. *Personal Disord*. 2019;10(1):46-58. doi:10.1037/per0000300
3. Levin A. predicting outcomes in nephrology: lots of tools, limited uptake: how do we move forward? *J Am Soc Nephrol*. 2024;35(3):361-363. doi:10.1681/ASN.0000000000000288
4. Tonelli M, Lloyd A, Cheung WY, et al. Mortality and resource use among individuals with chronic kidney disease or cancer in Alberta, Canada, 2004-2015. *JAMA Netw Open*. 2022;5(1):e2144713. doi:10.1001/jamanetworkopen.2021.44713
5. Davison SN, Levin A, Moss AH, et al; Kidney Disease: Improving Global Outcomes. Executive summary of the KDIGO Controversies Conference on Supportive Care in Chronic Kidney Disease: developing a roadmap to improving quality care. *Kidney Int*. 2015;88(3):447-459. doi:10.1038/ki.2015.110
6. Kurella Tamura M, Meier DE. Five policies to promote palliative care for patients with ESRD. *Clin J Am Soc Nephrol*. 2013;8(10):1783-1790. doi:10.2215/CJN.02180213
7. 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. doi:10.2215/CJN.05960809
8. Stevens PE, Ahmed SB, Carrero JJ, et al; Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int*. 2024;105(4S):S117-S314. doi:10.1016/j.kint.2023.10.018
9. Axelsson L, Alvariza A, Lindberg J, et al. Unmet palliative care needs among patients with end-stage kidney disease: a national registry study about the last week of life. *J Pain Symptom Manage*. 2018;55(2):236-244. doi:10.1016/j.jpainsymman.2017.09.015

10. Schell JO, Patel UD, Steinhauser KE, Ammarell N, Tulskey JA. Discussions of the kidney disease trajectory by elderly patients and nephrologists: a qualitative study. *Am J Kidney Dis*. 2012;59(4):495-503. doi:10.1053/j.ajkd.2011.11.023
11. Wong SP, Kreuter W, O'Hare AM. Treatment intensity at the end of life in older adults receiving long-term dialysis. *Arch Intern Med*. 2012;172(8):661-663. doi:10.1001/archinternmed.2012.268
12. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10):e1001744. doi:10.1371/journal.pmed.1001744
13. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378
14. Wolff RF, Moons KGM, Riley RD, et al; PROBAST Group†. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51-58. doi:10.7326/M18-1376
15. Snell KIE, Levis B, Damen JAA, et al. Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis: Checklist for Systematic Reviews and Meta-Analyses (TRIPOD-SRMA). *BMJ*. 2023;381:e073538. doi:10.1136/bmj-2022-073538
16. Eknayan G, Lameire N, Barsoum R, et al. The burden of kidney disease: improving global outcomes. *Kidney Int*. 2004;66(4):1310-1314. doi:10.1111/j.1523-1755.2004.00894.x
17. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med*. 2019;170(1):W1-W33. doi:10.7326/M18-1377
18. Gerds TA, Kattan MW. *Medical Risk Prediction Models: With Ties to Machine Learning*. Chapman and Hall; 2021. doi:10.1201/9781138384484
19. Austin PC, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *J Clin Epidemiol*. 2004;57(11):1138-1146. doi:10.1016/j.jclinepi.2004.04.003
20. Blanche P, Kattan MW, Gerds TA. The C-index is not proper for the evaluation of  $t$ -year predicted risks. *Biostatistics*. 2019;20(2):347-357. doi:10.1093/biostatistics/kxy006
21. Pepe MS, Fan J, Feng Z, Gerds T, Hilden J. The net reclassification index (NRI): a misleading measure of prediction improvement even with independent test data sets. *Stat Biosci*. 2015;7(2):282-295. doi:10.1007/s12561-014-9118-0
22. Hilden J, Gerds TA. A note on the evaluation of novel biomarkers: do not rely on integrated discrimination improvement and net reclassification index. *Stat Med*. 2014;33(19):3405-3414. doi:10.1002/sim.5804
23. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574. doi:10.1177/0272989X06295361
24. Liu P, Sawhney S, Heide-Jørgensen U, et al Predicting the risks of kidney failure and death in adults with moderate to severe chronic kidney disease: multinational, longitudinal, population based, cohort study. *BMJ*. 2024;385:e078063. doi:10.1136/bmj-2023-078063
25. Simon R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Per Med*. 2010;7(1):33-47. doi:10.2217/pme.09.49
26. 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. doi:10.2215/CJN.03100313
27. Obi Y, Nguyen DV, Zhou H, et al. Development and validation of prediction scores for early mortality at transition to dialysis. *Mayo Clin Proc*. 2018;93(9):1224-1235. doi:10.1016/j.mayocp.2018.04.017
28. Inaguma D, Morii D, Kabata D, et al. Prediction model for cardiovascular events or all-cause mortality in incident dialysis patients. *PLoS One*. 2019;14(8):e0221352. doi:10.1371/journal.pone.0221352
29. Santos J, Oliveira P, Malheiro J, et al. Predicting 6-month mortality in incident elderly dialysis patients: a simple prognostic score. *Kidney Blood Press Res*. 2020;45(1):38-50. doi:10.1159/000504136
30. Pladys A, Vigneau C, Raffray M, et al. Contribution of medico-administrative data to the development of a comorbidity score to predict mortality in end-stage renal disease patients. *Sci Rep*. 2020;10(1):8582. doi:10.1038/s41598-020-65612-x
31. Hemke AC, Heemskerk MB, van Diepen M, Weimar W, Dekker FW, Hoitsma AJ. Survival prognosis after the start of a renal replacement therapy in the Netherlands: a retrospective cohort study. *BMC Nephrol*. 2013;14(1):258. doi:10.1186/1471-2369-14-258

32. Chen LX, Josephson MA, Hedeker D, Campbell KH, Stankus N, Saunders MR. A clinical prediction score to guide referral of elderly dialysis patients for kidney transplant evaluation. *Kidney Int Rep*. 2017;2(4):645-653. doi:10.1016/j.ekir.2017.02.014
33. Chua HR, Lau T, Luo N, et al. Predicting first-year mortality in incident dialysis patients with end-stage renal disease—the UREA5 study. *Blood Purif*. 2014;37(2):85-92. doi:10.1159/000357640
34. van Diepen M, Schroijen MA, Dekkers OM, et al. Predicting mortality in patients with diabetes starting dialysis. *PLoS One*. 2014;9(3):e89744. doi:10.1371/journal.pone.0089744
35. 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. doi:10.1038/ki.2014.419
36. Dusseux E, Albano L, Fafin C, et al. A simple clinical tool to inform the decision-making process to refer elderly incident dialysis patients for kidney transplant evaluation. *Kidney Int*. 2015;88(1):121-129. doi:10.1038/ki.2015.25
37. Doi T, Yamamoto S, Morinaga T, Sada KE, Kurita N, Onishi Y. Risk score to predict 1-year mortality after haemodialysis initiation in patients with stage 5 chronic kidney disease under predialysis nephrology care. *PLoS One*. 2015;10(6):e0129180. doi:10.1371/journal.pone.0129180
38. Thamer M, Kaufman JS, Zhang Y, Zhang Q, Cotter DJ, Bang H. 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. doi:10.1053/j.ajkd.2015.05.014
39. Couchoud CG, Beuscart JBR, Aldigier JC, Brunet PJ, Moranne OP; REIN registry. Development of a risk stratification algorithm to improve patient-centered care and decision making for incident elderly patients with end-stage renal disease. *Kidney Int*. 2015;88(5):1178-1186. doi:10.1038/ki.2015.245
40. Hemke AC, Heemskerk MBA, van Diepen M, Dekker FW, Hoitsma AJ. Improved mortality prediction in dialysis patients using specific clinical and laboratory data. *Am J Nephrol*. 2015;42(2):158-167. doi:10.1159/000439181
41. Patzer RE, Basu M, Larsen CP, et al. iChoose Kidney: a clinical decision aid for kidney transplantation versus dialysis treatment. *Transplantation*. 2016;100(3):630-639. doi:10.1097/TP.0000000000001019
42. Haapio M, Helve J, Grönhagen-Riska C, Finne P. One- and 2-year mortality prediction for patients starting chronic dialysis. *Kidney Int Rep*. 2017;2(6):1176-1185. doi:10.1016/j.ekir.2017.06.019
43. Lin SY, Hsieh MH, Lin CL, et al. Artificial intelligence prediction model for the cost and mortality of renal replacement therapy in aged and super-aged populations in Taiwan. *J Clin Med*. 2019;8(7):995. doi:10.3390/jcm8070995
44. Akbilgic O, Obi Y, Potukuchi PK, et al. Machine learning to identify dialysis patients at high death risk. *Kidney Int Rep*. 2019;4(9):1219-1229. doi:10.1016/j.ekir.2019.06.009
45. Cho H, Kim MH, Kim HJ, et al. Development and validation of the modified Charlson Comorbidity Index in incident peritoneal dialysis patients: a national population-based approach. *Perit Dial Int*. 2017;37(1):94-102. doi:10.3747/pdi.2015.00201
46. Wick JP, Turin TC, Faris PD, et al. A clinical risk prediction tool for 6-month mortality after dialysis initiation among older adults. *Am J Kidney Dis*. 2017;69(5):568-575. doi:10.1053/j.ajkd.2016.08.035
47. Ivory SE, Polkinghorne KR, Khandakar Y, et al. Predicting 6-month mortality risk of patients commencing dialysis treatment for end-stage kidney disease. *Nephrol Dial Transplant*. 2017;32(9):1558-1565. doi:10.1093/ndt/gfw383
48. Geddes CC, van Dijk PCW, McArthur S, et al. The ERA-EDTA cohort study—comparison of methods to predict survival on renal replacement therapy. *Nephrol Dial Transplant*. 2006;21(4):945-956. doi:10.1093/ndt/gfi326
49. Mauri JM, Cléries M, Vela E; Catalan Renal Registry. Design and validation of a model to predict early mortality in haemodialysis patients. *Nephrol Dial Transplant*. 2008;23(5):1690-1696. doi:10.1093/ndt/gfm728
50. Couchoud C, Labeuw M, Moranne O, et al; French Renal Epidemiology and Information Network (REIN) registry. 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. doi:10.1093/ndt/gfn698
51. Liu J, Huang Z, Gilbertson DT, Foley RN, Collins AJ. An improved comorbidity index for outcome analyses among dialysis patients. *Kidney Int*. 2010;77(2):141-151. doi:10.1038/ki.2009.413
52. Jacob AN, Khuder S, Malhotra N, et al. Neural network analysis to predict mortality in end-stage renal disease: application to United States Renal Data System. *Nephron Clin Pract*. 2010;116(2):c148-c158. doi:10.1159/000315884
53. Marinovich S, Lavorato C, Moriñigo C, et al. A new prognostic index for one-year survival in incident hemodialysis patients. *Int J Artif Organs*. 2010;33(10):689-699. doi:10.1177/039139881003301001

54. Quinn RR, Laupacis A, Hux JE, Oliver MJ, Austin PC. 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. doi:[10.1097/MLR.0b013e318202aa0b](https://doi.org/10.1097/MLR.0b013e318202aa0b)
55. Wu MY, Hu PJ, Chen YW, et al. Predicting 3-month and 1-year mortality for patients initiating dialysis: a population-based cohort study. *J Nephrol*. 2022;35(3):1005-1013. doi:[10.1007/s40620-021-01185-w](https://doi.org/10.1007/s40620-021-01185-w)
56. Noh J, Yoo KD, Bae W, et al. Prediction of the mortality risk in peritoneal dialysis patients using machine learning models: a nation-wide prospective cohort in Korea. *Sci Rep*. 2020;10(1):7470. doi:[10.1038/s41598-020-64184-0](https://doi.org/10.1038/s41598-020-64184-0)
57. Siddiqi M, Kimber AC, Shabbir J. Multivariable prognostic model for dialysis patients with end stage renal disease: an observational cohort study of Pakistan by external validation. *Saudi Med J*. 2021;42(7):714-720. doi:[10.15537/smj.2021.42.7.20210082](https://doi.org/10.15537/smj.2021.42.7.20210082)
58. McAdams-DeMarco MA, Ying H, Thomas AG, et al. Frailty, inflammatory markers, and waitlist mortality among patients with end-stage renal disease in a prospective cohort study. *Transplantation*. 2018;102(10):1740-1746. doi:[10.1097/TP.0000000000002213](https://doi.org/10.1097/TP.0000000000002213)
59. Gao X, Wang J, Huang H, et al. Nomogram model based on clinical risk factors and heart rate variability for predicting all-cause mortality in stage 5 CKD patients. *Front Genet*. 2022;13:872920. doi:[10.3389/fgene.2022.872920](https://doi.org/10.3389/fgene.2022.872920)
60. Chaudhuri S, Larkin J, Guedes M, et al. Predicting mortality risk in dialysis: Assessment of risk factors using traditional and advanced modeling techniques within the Monitoring Dialysis Outcomes initiative. *Hemodial Int*. 2023;27(1):62-73. doi:[10.1111/hdi.13053](https://doi.org/10.1111/hdi.13053)
61. Thijssen S, Usvyat L, Kotanko P. Prediction of mortality in the first two years of hemodialysis: results from a validation study. *Blood Purif*. 2012;33(1-3):165-170. doi:[10.1159/000334138](https://doi.org/10.1159/000334138)
62. 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. doi:[10.1053/j.ajkd.2010.12.023](https://doi.org/10.1053/j.ajkd.2010.12.023)
63. Zhu J, Tang C, Ouyang H, Shen H, You T, Hu J. Prediction of all-cause mortality using an echocardiography-based risk score in hemodialysis patients. *Cardiorenal Med*. 2021;11(1):33-43. doi:[10.1159/000507727](https://doi.org/10.1159/000507727)
64. 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(1):72-79. doi:[10.2215/CJN.03860609](https://doi.org/10.2215/CJN.03860609)
65. Siga MM, Ducher M, Florens N, et al. Prediction of all-cause mortality in haemodialysis patients using a Bayesian network. *Nephrol Dial Transplant*. 2020;35(8):1420-1425. doi:[10.1093/ndt/gfz295](https://doi.org/10.1093/ndt/gfz295)
66. Jung HY, Kim SH, Jang HM, et al. Individualized prediction of mortality using multiple inflammatory markers in patients on dialysis. *PLoS One*. 2018;13(3):e0193511. doi:[10.1371/journal.pone.0193511](https://doi.org/10.1371/journal.pone.0193511)
67. Wang Y, Zhu Y, Lou G, Zhang P, Chen J, Li J. A maintenance hemodialysis mortality prediction model based on anomaly detection using longitudinal hemodialysis data. *J Biomed Inform*. 2021;123:103930. doi:[10.1016/j.jbi.2021.103930](https://doi.org/10.1016/j.jbi.2021.103930)
68. Tapak L, Sheikh V, Jenabi E, Khazaei S. Predictors of mortality among hemodialysis patients in Hamadan province using random survival forests. *J Prev Med Hyg*. 2020;61(3):E482-E488. doi:[10.15167/2421-4248/JPMH2020.61.3.1421](https://doi.org/10.15167/2421-4248/JPMH2020.61.3.1421)
69. Holme I, Fellström BC, Jardin AG, Schmieder RE, Zannad F, Holdaas H. 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. doi:[10.1111/j.1365-2796.2011.02435.x](https://doi.org/10.1111/j.1365-2796.2011.02435.x)
70. Rankin S, Han L, Scherzer R, et al. A machine learning model for predicting mortality within 90 days of dialysis initiation. *Kidney360*. 2022;3(9):1556-1565. doi:[10.34067/KID.0007012021](https://doi.org/10.34067/KID.0007012021)
71. Fernandez Lucas M, Teruel JL, Zamora J, Lopez Mateos M, Rivera M, Ortuno J. A Mediterranean age-comorbidity prognostic index for survival in dialysis populations. *J Nephrol*. 2007;20(6):696-702.
72. Goldstein BA, Xu C, Wilson J, et al. Designing an implementable clinical prediction model for near-term mortality and long-term survival in patients on maintenance hemodialysis. *Am J Kidney Dis*. 2024;84(1):73-82. doi:[10.1053/j.ajkd.2023.12.013](https://doi.org/10.1053/j.ajkd.2023.12.013)
73. Noppakun K, Nochaiwong S, Tantraworasin A, et al; Nephrology Society of Thailand. Mortality rates and a clinical predictive model for the elderly on maintenance hemodialysis: a large observational cohort study of 17,354 Asian patients. *Am J Nephrol*. 2024;55(2):136-145. doi:[10.1159/000535669](https://doi.org/10.1159/000535669)
74. Yang M, Yang Y, Xu Y, et al. Development and validation of prediction models for all-cause mortality and cardiovascular mortality in patients on hemodialysis: a retrospective cohort study in China. *Clin Interv Aging*. 2023;18:1175-1190. doi:[10.2147/CIA.5416421](https://doi.org/10.2147/CIA.5416421)



75. Okada H, Ono A, Tomori K, et al. Development of a prognostic risk score to predict early mortality in incident elderly Japanese hemodialysis patients. *PLoS One*. 2024;19(4):e0302101. doi:[10.1371/journal.pone.0302101](https://doi.org/10.1371/journal.pone.0302101)
76. la Roi-Teeuw HM, van Royen FS, de Hond A, et al. Don't be misled: 3 misconceptions about external validation of clinical prediction models. *J Clin Epidemiol*. 2024;172:111387. doi:[10.1016/j.jclinepi.2024.111387](https://doi.org/10.1016/j.jclinepi.2024.111387)
77. Booth S, Riley RD, Ensor J, Lambert PC, Rutherford MJ. Temporal recalibration for improving prognostic model development and risk predictions in settings where survival is improving over time. *Int J Epidemiol*. 2020;49(4):1316-1325. doi:[10.1093/ije/dyaa030](https://doi.org/10.1093/ije/dyaa030)
78. Anderson RT, Cleek H, Pajouhi AS, et al. Prediction of risk of death for patients starting dialysis: a systematic review and meta-analysis. *Clin J Am Soc Nephrol*. 2019;14(8):1213-1227. doi:[10.2215/CJN.00050119](https://doi.org/10.2215/CJN.00050119)
79. Williams AW, Dwyer AC, Eddy AA, et al; American Society of Nephrology Quality, and Patient Safety Task Force. Critical and honest conversations: the evidence behind the "Choosing Wisely" campaign recommendations by the American Society of Nephrology. *Clin J Am Soc Nephrol*. 2012;7(10):1664-1672. doi:[10.2215/CJN.04970512](https://doi.org/10.2215/CJN.04970512)
80. van Geloven N, Swanson SA, Ramspek CL, et al. Prediction meets causal inference: the role of treatment in clinical prediction models. *Eur J Epidemiol*. 2020;35(7):619-630. doi:[10.1007/s10654-020-00636-1](https://doi.org/10.1007/s10654-020-00636-1)

#### SUPPLEMENT 1.

**eMethods.** Detailed Methods

**eTable 1.** MEDLINE, Embase, and Cochrane Search Strategies

**eTable 2.** CHARMS Checklist

**eTable 3.** PROBAST Signaling Questions

**eTable 4.** TRIPOD+AI Checklist

**eTable 5.** Prediction Framework, Model Training and Testing, and Usefulness for the Included Studies

**eTable 6.** Characteristics of the Studies Included in the Systematic Review and Critical Appraisal for Risk of Bias and Applicability According to PROBAST

**eAppendix.** Study Protocol

**eReferences.**

#### SUPPLEMENT 2.

**Data Sharing Statement**