Journal of Clinical Epidemiology 145 (2022) 126–135

Journal of Clinical Epidemiology

REVIEW

Prediction models for living organ transplantation are poorly developed, reported, and validated: a systematic review

Maria C. Haller[a,b], Constantin Aschauer[c], Christine Wallisch[a], Karen Leffondré[d], Maarten van Smeden[e], Rainer Oberbauer[c], Georg Heinze[a,*]

[a] *Center for Medical Statistics, Informatics and Intelligent Systems (CeMSIIS), Section for Clinical Biometrics, Medical University of Vienna, Vienna, Austria*

[b] *Department for Internal Medicine III, Nephrology and Hypertension Diseases, Transplantation Medicine and Rheumatology, Ordensklinikum Linz, Linz, Austria*

[c] *Division of Nephrology and Dialysis, Department of Medicine III, Medical University of Vienna, Vienna, Austria*

[d] *University of Bordeaux, INSERM, Bordeaux Population Health Research Center, UMR1219, Bordeaux, France*

[e] *Julius Center for Health Science and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands*

Accepted 31 January 2022; Available online 4 February 2022

Abstract

Objective: To identify and critically appraise risk prediction models for living donor solid organ transplant counselling.

Study Design and Setting: We systematically reviewed articles describing the development or validation of prognostic risk prediction models about living donor solid organ (kidney and liver) transplantation indexed in Medline until April 4, 2021. Models were eligible if intended to predict, at transplant counselling, any outcome occurring after transplantation or donation in recipients or donors. Duplicate study selection, data extraction, assessment for risk of bias and quality of reporting was done using the CHARMS checklist, PRISMA recommendations, PROBAST tool, and TRIPOD Statement.

Results: We screened 4691 titles and included 49 studies describing 68 models (35 kidney, 33 liver transplantation). We identified 49 new risk prediction models and 19 external validations of existing models. Most models predicted recipients outcomes ($n = 38$, 75%), e.g., kidney graft loss (29%), or mortality of liver transplant recipients (55%). Many new models ($n = 46$, 94%) and external validations ($n = 17$, 89%) had a high risk of bias because of methodological weaknesses. The quality of reporting was generally poor.

Conclusion: We advise against applying poorly developed, reported, or validated prediction models. Future studies could validate or update the few identified methodologically appropriate models.

Keywords: Risk prediction models; Living donor; Kidney transplantation; Liver transplantation; Systematic review; Risk of bias; Quality of reporting

1. Introduction

Kidney transplantation is the preferred treatment for eligible end-stage renal disease patients and liver transplantation is a life-saving rescue treatment for end stage liver disease patients [1–6]. Due to the organ shortage from deceased donors, both transplants are nowadays routinely performed with organs from related or altruistic living donors imposing the risk of organ donation to these otherwise healthy individuals[7,8]. Although advances in transplantation medicine improved transplant outcomes over the last decades, long-term risks beyond the first decade after donation for the living donor are still under debate. Decision making about living donor transplants is further complicated by the fact that benefits for the recipient have to be weighed against risks for the healthy donor. Risk prediction models are a valuable tool to quantitatively estimate the individualized risk of future events and are therefore of great clinical interest to inform transplant counselling. However, the optimal combination of donor and recipient information to achieve accurate prognosis estimation of posttransplant events for the donor and recipient are currently unknown.

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

* Corresponding author. Tel.:+ 43-01-40400 66880; fax: + 43-01-40400 66870.

E-mail address: georg.heinze@meduniwien.ac.at (G. Heinze).

https://doi.org/10.1016/j.jclinepi.2022.01.025

What is new?

Key findings

- The majority of risk of prediction models developed for living donor kidney or liver transplantation are at high risk of bias and poorly reported.
- Few new developed prediction models had a low risk of bias. Those with low risk of bias lacked appropriate external validation.

What this adds to what is known

- This is the first review to systematically identify and critically appraise risk prediction models that were developed or validated for use in living donor kidney or liver transplantation.

What is the implication, and what should change now

- Transplant recipients, their potential living organ donors and clinicians should be aware that reliable, externally validated risk prediction models for post-transplant/post-donation prognosis estimation are not yet available and should be critical about their use until external validation studies at low risk of bias are available.

We therefore aimed to systematically review and critically appraise currently available prognostic models about living donor kidney and liver transplantation.

2. Material and methods

We conducted a systematic review to identify risk prediction models in living donor kidney or liver transplantation that were developed to aid transplant candidates, their potential living donors and health care providers in estimating the risk of future events occurring after transplantation for the donor or recipient. Reporting of our systematic review is in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) and the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) checklist. The study protocol including the search syntax is available from the Appendix A [9].

2.1. Search and study selection process

We developed a search strategy using Boolean combinations of Medical Subject Headings (MeSH) and text words for "kidney transplantation"; "liver transplantation," and "living donor" filtered for risk prediction models [10]. Searches included all articles indexed in MEDLINE via Ovid from inception until April 4, 2021 without language restriction. Additionally, articles were identified by cross-checking reference lists of included studies and asking experts in the field for relevant citations. Two reviewers (MCH, CA) independently screened title and abstracts and assessed full texts of potentially relevant studies for eligibility using predefined eligibility criteria. We included models that gave a prognostic estimation of any outcome occurring after transplantation/donation in the recipient or donor as a function of at least two predictor variables measured in the recipient or the donor. Risk prediction modelling studies that either developed a new model (derivation studies) or validated an existing model (validation studies) intended for use at the time of transplant counseling for individual prediction of posttransplant events regardless of age, prediction horizon, outcomes, or publication date were eligible. We excluded predictor finding studies, prediction model impact studies, prediction models that were not applicable to transplants from living donors, and diagnostic models.

2.2. Data extraction, assessment of risk of bias, and quality of reporting

Two reviewers (M.C.H., A.C.) independently extracted predefined items according to the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist [11], assessed risk of bias using the Prediction model Risk of Bias Assessment Tool (PROBAST) [12], and appraised the quality of reporting using the TRIPOD Statement [13]. As one of the included studies was published by authors among the review team (H.M., C.W., G.H., R.O.), two independent experts in risk prediction (MvS, K.L.) assessed the risk of bias of the respective models [14].

The CHARMS checklist provides guidance to design and conduct a systematic review of prediction models, in particular to frame a proper review question and to define relevant items for data extraction.

The PROBAST tool consists of 20 signaling questions to facilitate structured and transparent judgment of risk of bias defined to occur when shortcomings in study design, conduct, or analysis lead to systematically distorted estimates for individualized predictions. Following recommendations of PROBAST, a high risk of bias in at least one domain justified rating the overall risk of bias as high [15].

The TRIPOD Statement was developed to improve the transparency of reporting of risk prediction studies and consists of 31 items essential for good reporting [16].

We extracted each model's predictive performance including measures of discrimination and calibration sorted by the strength of the reported performance measures. External validations were considered the strongest possible form of validation, followed by internal validation (including bootstrap validation, cross validation, random splits,

temporal splits), and apparent performance (i.e., calculation of predictive performance in the development dataset).

Disagreements were resolved through discussion among the author team for all steps of the review conduct.

2.3. Analysis

Data of eligible studies was extracted and risk of bias and quality of reporting was assessed for each model within a study. We presented absolute and relative frequencies to describe model characteristics, risk of bias and reporting. A simple TRIPOD score was computed by awarding one point for each item on the checklist if reporting met requirements according to the TRIPOD Explanation and Elaboration Document [16], and adding the points for a model (range 0–31 points).

3. Results

The search and study selection process are summarized in Fig. 1. After screening 4,691 titles, we retained 143 titles for full text assessment, of which 49 studies describing 68 prediction models met eligibility criteria (reference list in Appendix B).

3.1. New developed risk prediction models

Characteristics of new risk prediction models are summarized in Table 1. Full data extraction tables of all models are available from Appendix C.

Overall, 49 different predictors were included across the 29 new kidney transplant models, and 34 different predictors across the 20 new liver transplant models (Fig. 2). Predictors that were included in at least a third of the models about kidney transplantation (≥10 times) were recipient age, recipient weight or height measure, donor age, predonation kidney function measure, donor gender, donor weight or height measure, and a measure of tissue compatibility. Predictors in at least a third of liver transplantation (≥6 times) were pretransplant recipient laboratory value, model of end stage liver disease (MELD) or paediatric end stage liver disease) score, a measure of graft weight, and peritransplant blood loss or number of transfusions required.

Based on the studies reporting performance measures, c-indices for models to predict recipient mortality ranged from 0.75 to 0.77 in kidney transplantation, and varied widely from 0.63 to 0.95 in liver transplantation. Studies to predict graft loss in kidney transplantation (including composite outcomes of graft loss and recipient death) reported c-indices with a wide range from 0.57 to 0.88. Calibration was assessed in 16 models only ($n = 12$ for kidney transplantation, $n = 4$ for liver transplantation), whereby a calibration plot was shown for 12 models ($n = 9$ for kidney transplantation, $n = 3$ for liver transplantation) and a Hosmer–Lemeshow test was reported for four models ($n = 3$ for kidney transplantation, $n = 1$ for liver transplantation).

Eleven models predicted donor outcomes, one of which was about living liver donors, and 10 were about living kidney donors. Of the 38 models predicting recipient out-

Table 1. Characteristics of included new developed risk prediction models ($n = 49$) for living donor kidney and liver transplantation

	N (%) of models or range
Type of data used	
Single center cohort	37 (76%)
National registry data	9 (18%)
Multicenter cohort	1 (2%)
Data source not reported	2 (4%)
Geographical region	
Asia	25 (51%)
North America	15 (31%)
Europe	7 (14%)
Afrika	2 (4%)
Study dates for data collection	
Kidney transplant counselling models	1963 to 2018
Liver transplant counselling models	1994 to 2018
Prediction horizon	
Kidney transplant counselling models	3 mo to 40 yr
Liver transplant counselling models	Peritransplant to 5 yr
Predicted outcomes	
Recipient outcome	38 (78%)
Donor outcome	11 (22%)
Outcomes in kidney transplant models (n = 2	
Mortality (recipient)	3 (10%)
Graft loss	8 (29%)
Renal graft function	7 (24%)
Uretheral obstruction of renal graft	1 (3%)
Kidney function (donor)	6 (21%)
Proteinuria (donor)	1 (3%)
End-stage renal disease (donor)	2 (7%)
Mortality (donor)	1 (3%)
Outcomes in liver transplant models (n = 20)	
Mortality (recipient)	11 (55%)
Acute kidney injury	3 (15%)
Hyperglycaemia	2 (10%)
Liver graft function	1 (5%)
Liver graft regeneration (recipient)	1 (5%)
Survival (unclear if graft or recipient)	1 (5%)
Liver regeneration (donor)	1 (5%)

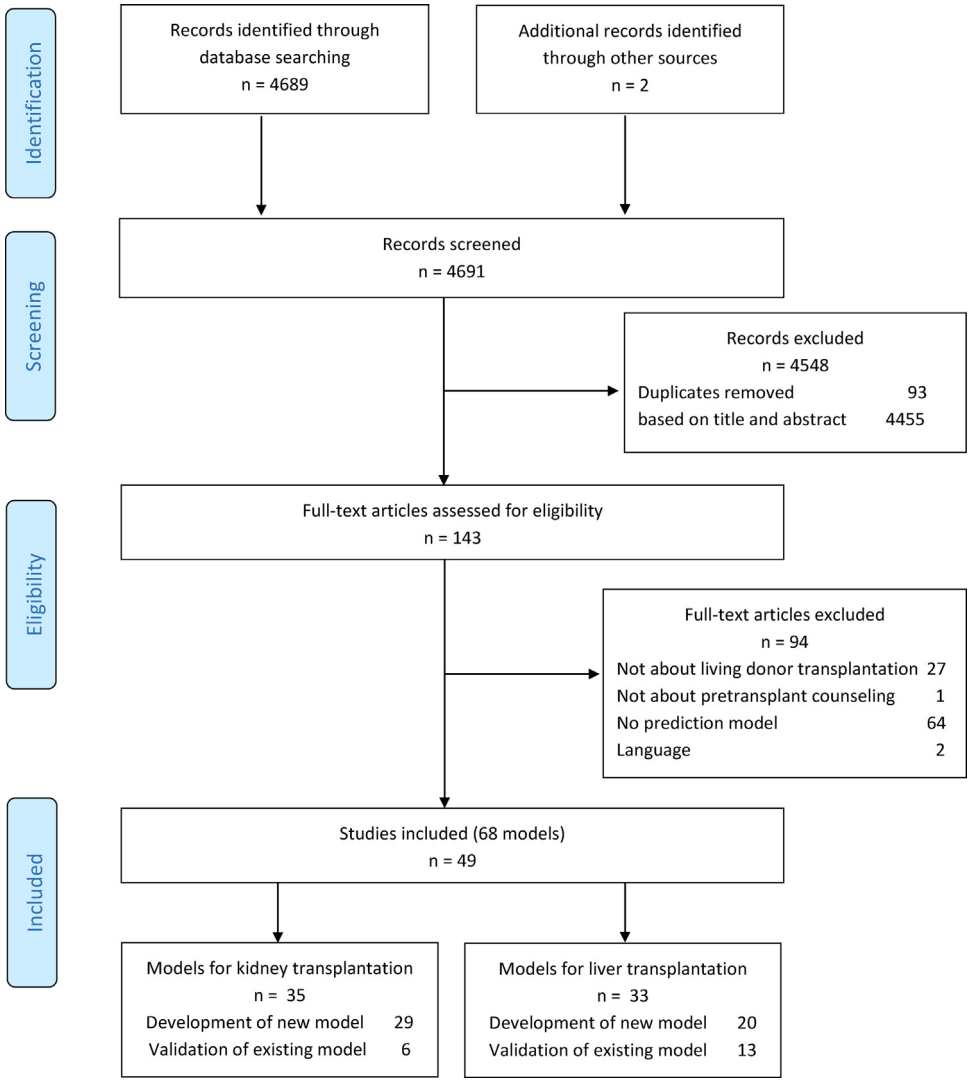

Fig. 1. PRISMA flow diagram of study selection process.

comes, 21 models (55%) included donor information. This was achieved by using donor information as single predictors of donor characteristics (e.g., donor age), a measure of tissue compatibility (e.g., HLA mismatch), a calculated ratio of donor and recipient information (e.g., donor to recipient weight ratio) or as linear predictor of overall donor prognosis. None of the models reported the consideration of interaction terms between donor and recipient variables.

3.2. External validation of existing risk prediction models

Overall there were 19 external validation studies of existing models ($n = 6$ in kidney transplantation, $n = 13$ in liver transplantation). In kidney transplantation, the model to predict postdonation eGFR for the kidney donor developed by Benoit 2017 was externally validated by two studies (Appendix C), and the graft failure score developed by Massie 2016 was externally validated in four studies (Appendix C). In liver transplantation, external validation studies concerned a large number of models that were originally developed for deceased donor liver transplantation, intensive care patients or patients with hepatocellular carcinoma. Two studies externally validated the D-MELD score, one study reported the external validation of a previously developed model by the same authors (Appendix C). Due to insufficient reporting of performance measures (e.g., confidence intervals), number of events, and prediction horizons in combination with heterogeneously defined endpoints and different types of c-indices calculated, a meta-analysis was unfeasible.

3.3. Risk of bias assessment

Assessment of risk of bias according to PROBAST is shown in Fig. 3. In kidney transplantation overall, new models ($n = 29$) were at low ($n = 3$, 10%) or high ($n = 26$, 90%) risk of bias, and external validations of existing models ($n = 6$) were at unclear ($n = 2$, 33%), and high ($n = 4$,

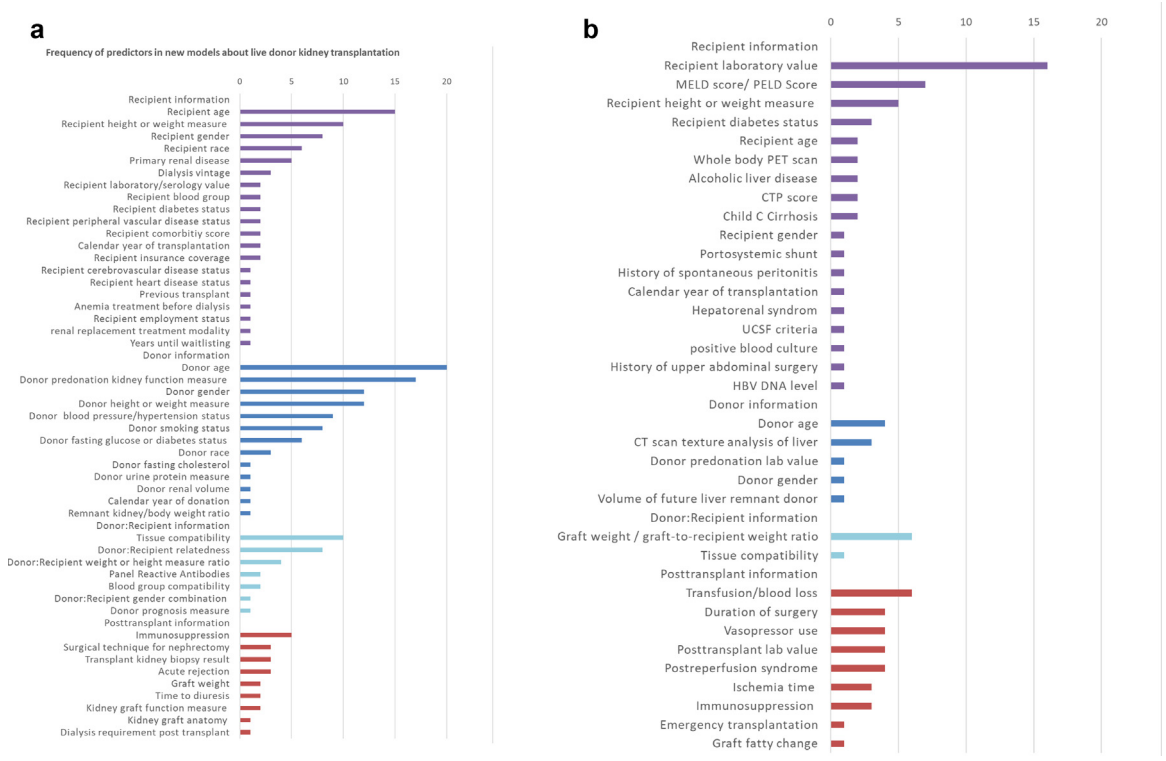

Fig. 2. Inclusion frequencies for each predictor are shown. Frequency of included predictors in all new prediction models for living donor kidney transplantation (A) and liver transplantation (B). Predictors are sorted in descending frequency, predictors measured in the recipient are colored purple, donor derived predictors are colored blue, predictors based on recipient and donor information are colored turquoise, and predictors unknown at the time of risk estimation are colored red. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article).

Fig. 3. Prediction model risk of bias assessment tool (PROBAST) risk of bias for included models ($n = 35$ for kidney transplant models, $n = 29$ for new kidney transplant models, $n = 6$ for external validations of kidney transplant models; $n = 33$ for liver transplant models, $n = 20$ for new liver transplant models, $n = 13$ for external validations of liver transplant models).

67%) risk of bias. The overall risk of bias in liver transplantation was high for all new models ($n = 20$), and for all external validations of existing models ($n = 13$).

The participant domain was judged at low risk of bias in about two-thirds of new models ($n = 21$, 69% for kidney transplantation, $n = 13$, 65% for liver transplantation), and at unclear risk due to insufficient reporting in six (21%) kidney transplant models and three (15%) liver transplant models. The risk of bias was high in three (10%) kidney transplant models and four (20%) liver transplant models because participants were selected based on subjective eligibility criteria ("suboptimal CT quality") or posttransplant information ("posttransplant pathology reading," "immediate graft function," "absence of posttransplant complications," "recurrence of hepatocellular carcinoma"). In external validation studies, the participant domain mostly had a low risk of bias ($n = 4$, 67% kidney transplantation, $n = 8$, 62% liver transplantation), was unclear for one (17%) kidney transplant model and five (38%) liver transplant models. One externally validated kidney transplant model was at high risk of bias, because in this study half of participants included in the validation set were already used in the training set, but the study was still declared as external validation.

The predictor domain had low risk of bias in 14 (48%) new kidney transplant models and five (25%) new liver transplant models, and an unclear risk in three (10%) kidney transplant models and three (15%) liver transplant models. A large number of models were rated high risk of bias for predictor domain ($n = 12$, 42% for kidney transplantation, $n = 12$, 60% for liver transplantation), because posttransplant information was used for model development that is unavailable at the time of model application. In validation studies, the predictor domain was rated low risk of bias for all kidney transplant models and three (23%) liver transplant models. External validations of ten (77%) liver transplant models were at high risk of bias for predictor domain because included predictors were measured posttransplant.

The risk of bias for the outcome domain was low in 20 (69%) new kidney transplant models, eight (40%) new liver transplant models, and unclear due to unclear definitions of endpoints in six (21%) kidney transplant models, and six (30%) liver transplant models. Reasons for high risk of bias in six (30%) liver transplant models and three (10%) kidney transplant models were arbitrary definitions of outcomes or a too short time frame between predictor assessment and outcome determination. External validations of four (67%) kidney and five (38%) liver transplant models had a low risk of bias for the outcome domain, an unclear risk in one (17%) kidney and one (8%) liver transplant model. One (17%) kidney and seven (54%) liver transplant models were at high risk of bias, because outcomes were not appropriately defined, and the time interval between predictor assessment and outcome determination was inadequate.

The analysis domain was at low risk of bias in a quarter of new kidney transplant models ($n = 7$, 24%), and none of the new liver transplant models. Insufficient transparency in reporting of analysis lead to an unclear risk of bias in three (15%) liver transplant models. Twenty two (76%) kidney and 17 (85%) liver transplant models were at high risk of bias, because of small sample sizes, mishandling of missing data, weak strategies for model building and model performance evaluation. Sample size ranged from 18 [17] to 89,009 [18] in kidney transplant models and from 77 [19] to 538 [20] in liver transplant models. Models predicting a binary or time to event endpoint ($n = 41$ overall, $n = 23$ kidney transplantation, $n = 18$ liver transplantation) had a nadir of observed events at 31 [21] in kidney and seven in liver transplantation [22]. The number of candidate predictors for predictor selection was clearly reported in only a third of the models ($n = 9$, 31% kidney transplantation, $n = 7$, 35% for liver transplant models). The median number of events per candidate predictors was 3.8. Consequently, the sample size was judged unreasonably small in 27 (55%) models overall. Another major methodological concern was handling of missing data, which was adequate in only 11 (22%) models overall, as most studies did not describe their methods of dealing with missing data or frequently excluded participants with missing data. Model building strategies included univariate screening in a large number of models ($n = 13$, 45% kidney transplant models, $n = 19$, 95% liver transplant models), and details on accounting for complexities in the data (e.g., censoring, competing risk), handling of predictors or correcting for optimism as recommended by PROBAST were infrequently provided. Many models were neither internally nor externally validated ($n = 11$, 38% kidney transplant, $n = 10$, 50% liver transplant), and four (14%) kidney and four (20%) liver transplant models even lacked reporting of apparent performance measures. In validation studies, two (33%) kidney transplant models had an unclear risk of bias for the analysis domain. Four out of six (67%) kidney and all liver transplant models were at high risk of bias, because the datasets contained less than the recommended minimum of 100 events for external validations [23].

3.4. Quality of reporting

The quality of reporting for selected items from the TRIPOD checklist is depicted in Fig. 4. The quality of reporting was generally poor according to TRIPOD (maximum score = 31) with a median of 21 (interquartile range [IQR] 16–24) for kidney and 20 (IQR 16–24) for liver transplant models. For most models reporting was adequate about the data source ($n = 30$, 85% in kidney transplantation, $n = 32$, 97% in liver transplantation), patient characteristics ($n = 22$, 65% in kidney transplantation, $n = 19$, 58% in liver transplantation), and source of funding ($n = 26$, 76% in kidney transplantation, $n = 20$, 61% in

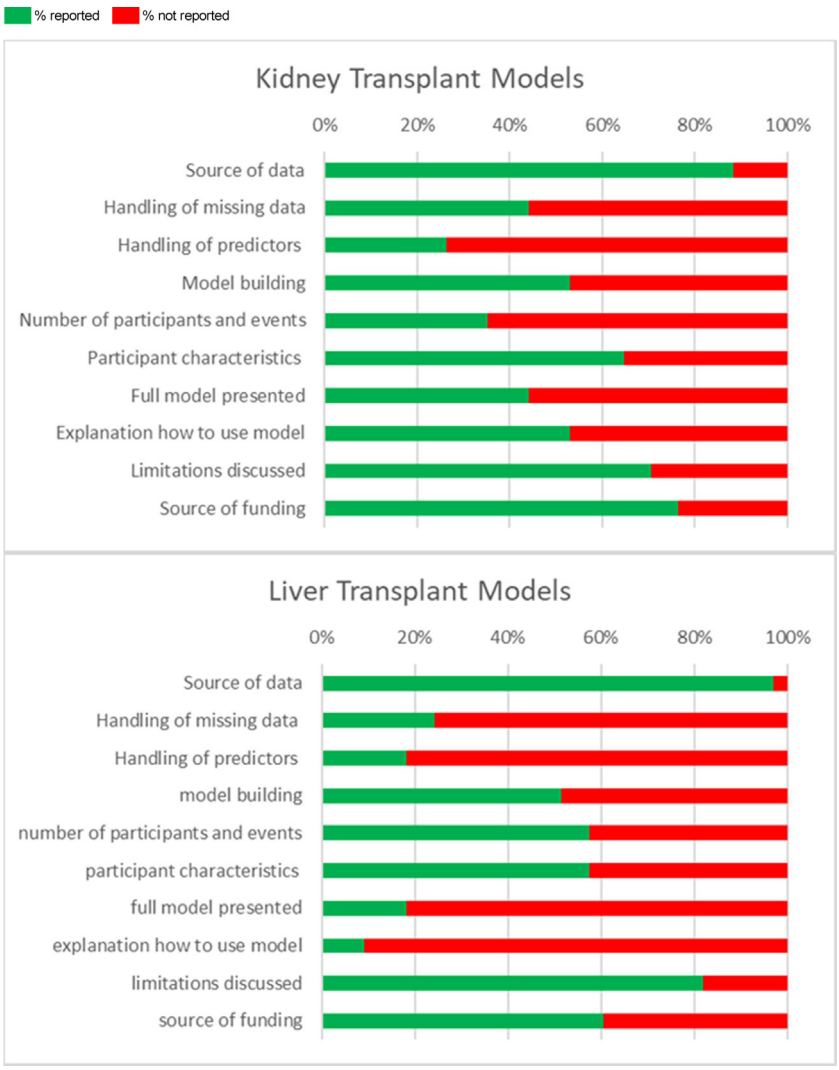

Fig. 4. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis checklist (TRIPOD) quality of reporting for selected items for included models ($n = 35$ kidney transplant models, $n = 33$ for liver transplant models).

liver transplantation). Reporting was particularly poor concerning items relevant for reproducibility and applicability of proposed models as the full model information including intercept or baseline survival was presented for only 16 kidney transplant models (46%) and six liver transplant models (18%). An explanation on how to use the model or a calculator was provided for 19 kidney transplant models (54%) and only three liver transplant model (9%).

4. Discussion

This systematic review identified and critically appraised 68 risk prediction models about living donor solid-organ (kidney and liver) transplant counselling reported in 49 studies. Eleven models were intended for use in donors to predict their prognosis after organ donation, and 38 models predicted recipient outcomes following transplantation. A superior number of models were new models, and only a minority of models were externally validated. We found most risk prediction models included in our review were poorly reported and at high risk of bias.

Our findings mirror those of previous systematic reviews that appraised risk prediction models for various medical conditions demonstrating a widespread use of poor methods to develop risk prediction models, incomplete reporting and a lack of external validation studies compromising reliability of predictions [24–30].

In contrast to prediction models in other medical areas, selection of participants and study designs used were appropriate in most models, likely because organ transplantations are performed in transplant centres only and often recorded in a single centre database or national registry [30]. On the downside however, more than half of the models violated basic principles by including predictors that are unknown at the time of risk estimation (i.e., before transplantation) but partly measured as late as 6 months

after transplantation. Interestingly, all models with time to event endpoints reported hazard ratios rather than directly predicting expected (restricted) survival time [31]. Living donor transplants are medical interventions of rather low frequency, limiting the number of participants and events available for modeling. Determinants of effective sample size (i.e., number of participants with outcome and number of candidate predictors) demonstrated that most models had a substantially lower event per candidate predictor ratio than recommended by PROBAST [15]. These limited sample sizes increased the risk of overfitting and likely yielded too optimistic prediction estimations for many models, but optimism corrected prediction estimates and consideration of shrinkage factors that have been proposed to improve accurate prediction in case of multicollinearity or high-dimensional data were rarely considered [32]. Furthermore, the risk of overfitting was rarely discussed as a potential limitation of proposed models. We also found weak approaches to model building, which were predominated by univariate prescreening to select predictors and categorization of continuous predictors rather than state-of-the-art approaches of functional form specification for such predictors[33,34]. Another common cause for concern was inadequate handling of missing data. Missing values were rarely handled using multiple imputation and participants with missing data were often simply excluded. Evaluation and reporting of predictive performance received little attention in general. For a number of models performance measures were not reported, and many models were only assessed for apparent performance, which overestimates true performance. Calibration and discrimination are equally important to judge a model's accurate predictive ability[35,36], and yet calibration was ignored in most models, although some type of discriminant performance, most frequently expressed as c-statistic, was reported for most models. If calibration was evaluated, calibration plots or Hosmer–Lemeshow tests were provided, although the Hosmer–Lemeshow test has a low power to detect miscalibration[37,38]. Calibration-in-the-large or the calibration slope were not reported [39]. Furthermore, there is dearth of externally validated models, despite performance evaluation in independent datasets being a requirement to confirm a model's generalizability for clinical use[40,41].

Transparent reporting is a prerequisite for reproducibility and critical assessment of research in general, and for external validation of proposed models and their bedside use specifically for prediction models. Most severely, only half of included models were presented in full with intercept or baseline risk or with a calculator. Consequently, many models are not even available in a format allowing their use for a single individual or external validation.

From a methodological perspective prediction models in transplantation of organs from living donors are unique in the sense that they should include variables characterizing donors and recipients. The spectrum of approaches to deal with candidate predictors from two individuals covered a wide range from purposefully excluding donor characteristics [42] to including a linear predictor of donor mortality risk as a summary estimate of donor-derived risk [14]. However, most models included well-known transplant-related predictors expressing some type of tissue compatibility, graft size measure, or calculated ratio of donor to recipient characteristics. Interactions between donor and recipient characteristics beyond tissue compatibility were not considered in any of the included models.

4.1. Limitations of this review

Although this systematic review follows current recommendations for systematic reviewing of risk prediction studies, some limitations need to be acknowledged. Despite using validated search strings to identify risk prediction models, we might have missed eligible studies due to inconsistent terminology of risk prediction models in title and abstract. We only evaluated risk prediction models in kidney or liver transplants from a living organ donor. Therefore, our findings do not generalize to risk prediction models in deceased donor kidney or liver transplants. In addition to applying PROBAST for risk of bias assessment, we used current reporting recommendations to evaluate whether reporting of included models met these standards. However, TRIPOD reporting guidelines were not developed as a tool to assess quality of reporting but rather to support researchers in writing transparent reports of their original work.

4.2. Future research

Future research should primarily focus on external validation and recalibration of existing models at low risk of bias. Second, if new models are needed prospectively planned collaborations to develop and validate models using individual patient data meta-analysis, and prospectively calculating the required sample size are recent initiatives to improve generalizability of proposed models and to reduce methodological challenges when facing limited sample size[43–45]. Future prediction model developers should pay special attention to transparent reporting to allow use of developed models for updating or external validation in other datasets as well as synthesis with other models[13,46]. Herein, journals play a pivotal role in demanding adequate reporting of risk prediction models to reduce the volume of misleading research published.

5. Conclusion

This systematic review demonstrates that, despite a fair number of published prediction models for living donor solid organ transplantation, most proposed models are poorly reported, inappropriately conducted with key analyses, in particular, adequate performance evaluation not carried out, resulting in an overall high risk of bias for

most models. Consequently, predictions for individual patients cannot be derived from these models or are likely less accurate than suggested by the primary study. Few models were rated at low risk of bias, but lack appropriate external validation [14]. Based on this systematic review we do not recommend using any of the identified models for living donor kidney or liver transplant counselling, because of their high risk of bias, unclear generalizability or lack of adequate extern validation.

Authors' contributions

Maria C. Haller: conceptualization, methodology, formal analysis, investigation, data curation, writing – original draft preparation, visualization. Constantin Aschauer: formal analysis, investigation, data curation. Christine Wallisch: formal analysis, data curation. Karen Leffondré: formal analysis, investigation, data curation. Maarten Van Smeden: formal analysis, investigation, data curation. Rainer Oberbauer: writing – review & editing, supervision. Georg Heinze: conceptualization, validation, writing – review & editing, supervision, Project administration. All authors interpreted the findings, critically revised the work for intellectual content and approved the final version submitted for publication.

Funding

This study received no funding.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jclinepi.2022.01.025.

References

[1] Wolfe RA, Ashby VB, Milford EL, Ojo AO, Ettenger RE, Agodoa LY, et al. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. New England J Med 1999;341(23):1725–30.

[2] Haller M, Gutjahr G, Kramar R, Harnoncourt F, Oberbauer R. Cost-effectiveness analysis of renal replacement therapy in Austria. Nephrol Dial Transplant 2011;26(9):2988–95 Epub 2011 Feb 10. doi:10.1093/ndt/gfq780.

[3] Wong G, Howard K, Chapman JR, Chadban S, Cross N, Tong A, et al. Comparative survival and economic benefits of deceased donor kidney transplantation and dialysis in people with varying ages and co-morbidities. PLoS One 2012;7(1) e29591Epub 2012 Jan 18. doi:10.1371/journal.pone.0029591.

[4] Tonelli M, Wiebe N, Knoll G, Bello A, Browne S, Jadhav D, et al. Systematic review: kidney transplantation compared with dialysis in clinically relevant outcomes. Am J Transplant 2011;11(10):2093–109.

[5] Haydon GH, Neuberger J. Liver transplantation of patients in end-stage cirrhosis. Baillieres Best Pract Res Clin Gastroenterol 2000;14(6):1049–73.

[6] Starzl TE, Demetris AJ, Van Thiel D. Liver transplantation (1). New England J Med 1989;321(15):1014–22.

[7] Lieber SR, Schiano TD, Rhodes R. Should living donor liver transplantation be an option when deceased donation is not? J Hepatol 2018;68(5):1076–82.

[8] Haugen AJ, Mjøen G. Living donation - any news on the donor? Curr Opin Organ Transplant 2021;26(1):85–90.

[9] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62(10):1006–12 doi: 10.16/j.jclinepi.2009.06.005. Epub Jul 23.

[10] Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One 2012;7(2):e32844 Epub 2012 Feb 29. doi:10.1371/journal.pone.0032844.

[11] Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014;11(10):e1001744 eCollection 2014 Oct. doi:10.1371/journal.pmed.

[12] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170(1):51–8.

[13] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350 g7594. doi:10.1136/bmj.g7594.

[14] Haller MC, Wallisch C, Mjoen G, Holdaas H, Dunkler D, Heinze G, et al. Predicting donor, recipient and graft survival in living donor kidney transplantation to inform pretransplant counselling: the donor and recipient linked iPREDICTLIVING tool - a retrospective study. Transplant Intl 2020;22:22.

[15] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, 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.

[16] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162(1):W1–73. doi:10.7326/M14-0698.

[17] Zapletal C, Lorenz MW, Woeste G, Wullstein C, Golling M, Bechstein WO. Predicting creatinine clearance by a simple formula following live-donor kidney transplantation. Transplant Intl 2004:490–4.

[18] Ashby VB, Leichtman AB, Rees MA, Song PX, Bray M, Wang W, et al. A kidney graft survival calculator that accounts for mismatches in age, sex, hla, and body size. Clin J Am Soc Nephrol 2017;12(7):1148–60.

[19] Ko YC, Tsai HI, Lee CW, Lin JR, Lee WC, Yu HP. A nomogram for prediction of early allograft dysfunction in living donor liver transplantation. Medicine (Baltimore) 2020;99(42) e22749.

[20] Park MH, Shim HS, Kim WH, Kim HJ, Kim DJ, Lee SH, et al. Clinical risk scoring models for prediction of acute kidney injury after living donor liver transplantation: a retrospective observational study. PLoS One 2015;10(8):e0136230 eCollection 2015. doi:10.1371/journal.pone.

[21] Rook M, Hofker HS, van Son WJ, Homan van der Heide JJ, Ploeg RJ, Navis GJ. Predictive capacity of pre-donation GFR and renal reserve capacity for donor renal function after living kidney donation. Am J Transplantation 2006:1653–9.

[22] Chung HS, Lee YJ, Jo YS. Proposal for a new predictive model of short-term mortality after living donor liver transplantation due to acute liver failure. Ann Transplant 2017;22:101–7.

[23] Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Statistics Med 2016;35(2):214–26.
[24] Kleinrouweler CE, Cheong-See FM, Collins GS, Kwee A, Thangaratinam S, Khan KS, et al. Prognostic models in obstetrics: available, but far from applicable. Am J Obstet Gynecol 2016;214(1):79–90 .e36Epub Jun 10. doi:10.1016/j.ajog.2015.06.013.
[25] Onland W, Debray TP, Laughon MM, Miedema M, Cools F, Askie LM, et al. Clinical prediction models for bronchopulmonary dysplasia: a systematic review and external validation study. BMC Pediatr 2013;13:207. doi:10.1186/471-2431-13-207.
[26] Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012;9(5):1–12 Epub 2012 May 22. doi:10.1371/journal.pmed.1001221.
[27] Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med 2011;9:103. doi:10.1186/741-7015-9-103.
[28] Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. BMC Med 2010;8:21. doi:10.1186/741-7015-8-21.
[29] Kaboré R, Haller M, Harambat J, Heinze G, Leffondré K. Risk prediction models for graft failure in kidney transplantation: a systematic review. NDT 2017. doi:10.1093/ndt/gfw405.
[30] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020;369:m1328.
[31] Tian L, Zhao L, Wei LJ. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. Biostatistics 2014;15(2):222–33.
[32] Dunkler D, Sauerbrei W, Heinze G. Global, parameterwise and joint shrinkage factor estimation. J Statistical Software 2016;69(8):1–19.
[33] Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Binder H, et al. State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. Diagn Progn Res 2020;4:3.
[34] Steyerberg EW. Clinical prediction models. A practical approach to development, validation, and updating. New York, USA: Springer; 2009.
[35] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21(1):128–38. doi:10.1097/EDE.0b013e3181c30fb2.
[36] Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19(4):453–73.
[37] Bertolini G, D'Amico R, Nardi D, Tinazzi A, Apolone G. One model, several results: the paradox of the Hosmer-Lemeshow goodness-of-fit test for the logistic regression model. J Epidemiol Biostat 2000;5(4):251–3.
[38] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 2014;35(29):1925–31.
[39] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17(1):230.
[40] Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338:b605 (doi):. doi:10.1136/bmj.b605.
[41] Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol 2003;56(9):826–32.
[42] van Walraven C, Austin PC, Knoll G. Predicting potential survival benefit of renal transplantation in patients with chronic kidney disease. CMAJ 2010;182(7):666–72 Epub 2010 Mar 29. doi:10.1503/cmaj.091661.
[43] Ahmed I, Debray TP, Moons KG, Riley RD. Developing and validating risk prediction models in an individual participant data meta–analysis. BMC Med Res Methodol 2014;14:3.
[44] Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441.
[45] Riley RD, Debray TPA, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Statistics Med 2021;40(19):4230–51.
[46] Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017;356:i6460. doi:10.1136/bmj.i6460.