

1 **Prediction and diagnosis of chronic kidney disease development and**
2 **progression using machine-learning: protocol for a systematic review and**
3 **meta-analysis of reporting standards and model performance**

4 **Short title:** Machine learning in chronic kidney disease

5 Fangyue Chen^{1,2,3}, Piyawat Kantagowit³, Tanawin Nopsophon^{3,4}, Arisa Chuklin³,

6 Krit Pongpirul^{3,5,6*}

7

8 ¹ Global Health Partnerships, Health Education England, United Kingdom

9 ² School of Public Health, Faculty of Medicine, Imperial College London, United Kingdom

10 ³ Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

11 ⁴ Harvard T.H. Chan School of Public Health, Boston, Massachusetts

12 ⁵ Department of International Health, Johns Hopkins Bloomberg School of Public Health,

13 Baltimore, MD, USA

14 ⁶ Clinical Research Center, Bumrungrad International Hospital, Bangkok, Thailand

15

16 *Corresponding author

17

18 **Abstract**

19 Chronic Kidney disease (CKD) is an important yet under-recognized contributor to morbidity
20 and mortality globally. Machine-learning (ML) based decision support tools have been developed
21 across many aspects of CKD care. Notably, algorithms developed in the prediction and diagnosis of
22 CKD development and progression may help to facilitate early disease prevention, assist with early
23 planning of renal replacement therapy, and offer potential clinical and economic benefits to patients
24 and health systems. Clinical implementation can be affected by the uncertainty surrounding the
25 methodological rigor and performance of ML-based models. This systematic review aims to evaluate
26 the application of prognostic and diagnostic ML tools in CKD development and progression.

27 The protocol has been prepared using the Preferred Items for Systematic Review and Meta-
28 analysis Protocols (PRISMA-P) guidelines. The systematic review protocol for CKD prediction and
29 diagnosis have been registered with the International Prospective Register of Systematic Reviews
30 (PROSPERO) (CRD42022356704, CRD42022372378). A systematic search will be undertaken of
31 PubMed, Embase, the Cochrane Central Register of Controlled Trials (CENTRAL), the Web of
32 Science, and the IEEE Xplore digital library. Studies in which ML has been applied to predict and
33 diagnose CKD development and progression will be included. The primary outcome will be the
34 comparison of the performance of ML-based models with non-ML-based models. Secondary analysis
35 will consist of model use cases, model construct, and model reporting quality.

36 This systematic review will offer valuable insight into the performance and reporting quality
37 of ML-based models in CKD diagnosis and prediction. This will inform clinicians and technical
38 specialists of the current development of ML in CKD care, as well as direct future model development
39 and standardization.

40

41 **Introduction**

42 Chronic kidney disease (CKD) is an important non-communicable disease that contributes to
43 significant morbidity and mortality on a global scale, directly or through cardiovascular
44 diseases attributable to impaired kidney function. Its estimated prevalence ranges from 9.1 to
45 15.1%. It has increased by 29.3% since 1990 due to the increase in the major chronic diseases
46 that contribute to its development, notably diabetes mellitus and hypertension (1,2).
47 Nevertheless, CKD is underrecognized by patients, clinicians, and health authorities. The
48 disease often progresses insidiously to end-stage kidney disease (ESKD) with late
49 presentation of symptoms and signs (3,4). The life-sustaining treatment for ESKD, renal
50 replacement therapy, poses a significant economic burden on patients and health systems,
51 meaning that currently, an estimated 47 to 73% of individuals are unable to receive it, leaving
52 around 2.3 million individuals dying prematurely (5). Strategies to prevent or delay CKD
53 onset and progression can potentially lower overall morbidity and mortality while minimizing
54 cost.

55 Machine Learning (ML), a subset of artificial intelligence (AI), has seen exponential
56 growth across healthcare (6,7). ML utilizes a specific dataset to generate an algorithm that
57 employs unknown or varied combinations of complex features and weights to predict the
58 outcome of future inputs (8). ML-based decision support tools have been developed across
59 many aspects of CKD care across disease prevention, diagnosis, and treatment (7), fuelled by
60 the growth in volume and variety of big data in nephrology and healthcare in general (9,10).
61 Notably, algorithms developed in the prediction and diagnosis of CKD development and
62 progression to ESKD may help to facilitate early disease prevention, assist with early care
63 planning, and allocate resources for the most significant clinical benefit (11-15).

64 Despite the growing promise of ML, several factors can hinder its clinical uptake.
65 These include uncertainty surrounding the performance of ML and the methodological rigor

66 behind its development. Non-ML-based prediction tools for CKD progression and prognosis
67 have been developed and validated, such as the Kidney Failure Risk Equation, which has
68 been used clinically to guide referrals to multidisciplinary CKD clinics (16-18). Comparisons
69 have been made between ML and non-ML-based prediction tools in general, specifically to
70 chronic diseases and prediction of acute kidney injury, which found similar performance
71 between prediction models developed with ML and conventional logistic regression (LR)
72 techniques (19-21). In addition, previous studies have questioned the reporting quality and
73 methodology of CKD prediction models (22,23), as well as other AI-based models in
74 imaging (24), oncology (25), and COVID-19 (26).

75 This systematic review aims to provide a comprehensive, in-depth summary and
76 evaluation of ML-based diagnostic and prognostic tools for CKD development and
77 progression, which will help to better direct future research strategy and methodology in
78 developing ML algorithms in CKD care.

79 The proposed systematic review aims to answer the following questions:

- 80 1. How do ML-based prediction tools in CKD development and progression perform
81 compared with tools developed using conventional techniques?
- 82 2. What are the use cases and constructs of these prediction tools?
- 83 3. How are the methodological characteristics and reporting quality of the ML-based tools?

84

85 **Materials and Methods**

86 The systematic review protocol was registered with the International Prospective Register of
87 Systematic Reviews (PROSPERO) on 26/09/22 for CKD prediction (CRD42022356704) and
88 CKD diagnosis (CRD42022372378). The protocol followed the Preferred Reporting Items

89 for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement (27). The
90 Checklist for critical Appraisal and data extraction for systematic Reviews of prediction
91 Modelling Studies (CHARMS) has been used to formulate review questions and data
92 extraction (28).

93 **Study eligibility criteria**

94 *Study designs*

95 Any peer-reviewed primary studies which assessed a prediction algorithm that utilizes ML
96 techniques applied to clinical problems in the prediction and diagnosis of chronic kidney
97 disease development and progression, including those for CKD screening, CKD prevention,
98 profiling of biomarkers contributing towards CKD, profiling of risk factors leading to CKD,
99 estimation of glomerular filtration rate (GFR) and creatinine levels, prediction of occurrence
100 of CKD, prediction of CKD stages, CKD diagnosis, CKD prognostication, prediction of CKD
101 progression to ESKD and/or requirement for renal replacement therapy, ESKD diagnosis will
102 be included.

103 Exclusion criteria are 1) Studies that utilize only image-based inputs as the different
104 model development processes require alternative extraction and appraisal tools; 2) studies
105 assessing prediction models of CKD complications other than its progression, including non-
106 exhaustively anemia, electrolyte disturbances, bone disorders, and cardiovascular events; 3)
107 prediction model of RRT including non-exhaustively the choice of RRT modalities which are
108 hemodialysis, peritoneal dialysis, and renal transplantation, and 4) studies reporting only
109 treatment-related outcomes of CKD such as adverse events, rate of complications, the
110 management of complications; 5) informal publication types such as case studies,
111 commentaries, letters to the editor, editorials, meeting abstracts, proceeding papers,

112 conference abstracts, protocols, guidelines, and recommendations; 6) review articles such as
113 narrative review, overview, systematic review, meta-analysis; 7) animal studies.

114

115 *Study participants*

116 Adult humans whose age was equal to or more than 18 years old.

117 *Types of interventions*

118 The studies will present prediction models utilizing ML techniques, including non-
119 exhaustively various regression techniques, decision trees, random forests, support vector
120 machines, K-nearest neighbor, and neural networks, as defined by individual studies. The
121 models will be for the prediction and diagnosis of chronic kidney disease development and
122 progression with or without mention of ESKD.

123 *Comparators*

124 We will include studies that compare the performance of ML-based prediction models with
125 those that utilize conventional techniques, including non-exhaustively those that use logistic
126 regression (including penalized LR), cox regression, Poisson regression, least squares linear
127 separation, generalized additive models, discriminant analysis, generalized estimation
128 equations, risk scores, and expert views. Studies that utilize only ML-based tools will also be
129 included.

130 **Study outcomes**

131 *Primary outcome*

- 132 • Performance comparison of ML-based and non-ML-based prediction tools in CKD
133 development and progression

134 *Secondary outcomes*

- 135 • ML-based model use case

- 136 • Performance of ML-based prediction tool in CKD development and progression
- 137 • Stages of model development (internal or external validation or clinical
- 138 implementation)
- 139 • Model development team specialty and the involvement of model end-user such as
- 140 clinicians during model development
- 141 • Evidence of model reporting quality description
- 142 • Characteristics of the dataset (size of training, validation and testing datasets, source
- 143 of dataset, population group, data period, length of follow-up)
- 144 • Prediction model construct including ML-based and non-ML based techniques
- 145 • Predictor characteristics and selection
- 146 • Model outcome characteristics and selection
- 147 • Model performance measures used

148 **Information sources and search strategy**

149 We will search through five databases: PubMed, Embase, the Cochrane Central Register of
150 Controlled Trials (CENTRAL), Web of Science, and the IEEE Xplore digital library. The
151 search strategy is constructed by two health information specialists with systematic review
152 experiences, combining search terms and subject headings (MeSH) related to "machine
153 learning," "artificial intelligence," "chronic kidney disease," and "End-stage kidney disease"
154 **(See Additional File)**. PubMed, Embase (OVID interface, 1947 onwards), Web of Science,
155 and CENTRAL were chosen for their broad coverage across biomedical, nursing, allied
156 health, and general scientific literature, while IEEE Xplore was included for coverage of
157 more technical literature in data science. Additional articles will be retrieved by manually
158 scrutinizing the reference lists of relevant publications.

159 **Study records**

160 *Data management*

161 Following database searching, studies will be populated into Covidence systematic review
162 software (29), which will manage study selection and data extraction.

163 *Selection process*

164 We will carry out two stages of screening. After study de-duplication through Covidence, two
165 reviewers will screen the titles and abstracts of potential studies independently. We will
166 eliminate abstracts in the initial screen if they do not report ML-based prediction models in
167 CKD.

168 Included studies will undergo full-text review against the full eligibility criteria. Reasons for
169 exclusion will be recorded for each study. Disagreement between two reviewers at each
170 article screening and selection stage will be resolved by consensus and a third person if
171 necessary. The PRISMA 2020 flow diagram will be generated to describe the workflow and
172 identification of included studies for the systematic review (30).

173 *Data collection and management*

174 The data extraction form will be designed prospectively before data collection and will be
175 pilot tested and refined. To ensure consistency across reviewers, explicit instructions will be
176 given, and calibration will take place before the extraction process. Two independent
177 reviewers will extract the following data items based on items included in the CHARMS
178 checklist: 1) source of data; 2) participant information; 3) outcome(s) to be predicted; 4)
179 candidate predictors; 5) sample size; 6) Missing data; 7) Model development; 8) Model
180 performance; 9) Model evaluation; 10) Results (final model presented) including model
181 performance; 11) interpretation and discussion. In addition, information will be extracted on
182 the study information (authors, year of publication, study design, journal, contact
183 information, study period, geographical location (area and country), and funding), the

184 assessment of reporting standards using an objective measure if mentioned in the study, and
185 any other relevant information. All relevant text, tables, and figures will be examined for data
186 extraction. Disagreements between the two independent reviewers will be resolved by
187 consensus. We will contact the study authors to request incompletely reported data in
188 included studies. We will conduct analyses using available data if no response is received
189 within 14 days.

190 **Reporting quality and risk of bias**

191 *Reporting quality assessment*

192 We will assess the reporting quality of studies against the TRIPOD (Transparent reporting of
193 a multivariable prediction model for individual prognosis or diagnosis) statement, which aims
194 to improve the transparent reporting of prediction modeling studies in all medical settings
195 (31).

196 The TRIPOD statement provides recommendations for reporting studies on
197 developing, validating, or updating a prediction model. To assess the completeness of
198 reporting amongst each publication, we will utilize the published "TRIPOD Adherence
199 Extraction Form," which evaluates 22 main items deemed essential in evaluating the
200 transparency of prediction model studies. Each article will only be assessed for items and
201 sub-items that it applies to (development, external validation, or incremental value reporting
202 of prediction models) based on guidance from the adherence form. Each TRIPOD item is
203 given adherence elements to help evaluate an item. The presence or lack of an adherence
204 element within the article will be marked down either as a "Yes" or a "No." For a TRIPOD
205 item to receive a score, all adherence elements must be present. Overall article's TRIPOD
206 score can be calculated by summing up adhered TRIPOD items and dividing by the total

207 number of applicable TRIPOD items for that article. Findings on reporting quality from the
208 TRIPOD adherence extraction will be summarised and graphically presented.

209 *Risk of bias assessment*

210 We will assess the risk of bias (ROB) of studies by applying PROBAST (prediction model
211 risk of bias assessment tool) (32). PROBAST was developed to assess ROB and applicability
212 concerns of a study that evaluates (e.g., develops or validates) a multivariable diagnostic or
213 prognostic prediction model. Reviewers will assess each study based on the published
214 "PROBAST Assessment Form." PROBAST is organized into four domains: participants,
215 predictors, outcomes, and analysis. A ROB rating (high, low, or unclear) will be assigned to
216 each domain based on answering signaling questions provided by the PROBAST assessment
217 form. Signaling questions can be answered as yes, probably yes, no, probably no, or no
218 information. Based on ROB results from the four domains, an overall ROB rating and
219 prediction model applicability rating will be given to the prediction model following the
220 recommendations in the PROBAST assessment form. A tabular presentation on PROBAST
221 results for each study will be available. Results will be summarized and graphically presented
222 for each domain.

223 *Comments on Methodology*

224 Two independent reviewers will assess the study quality and risk of bias. To improve review
225 consistency, reviewers will be onboarded on general decision rules and given review practice
226 for TRIPOD and PROBAST before conducting an independent review. Disagreements
227 between two reviewers will be resolved by consensus.

228 We will contact the author if not enough information is available for assessment. We
229 will utilize the available data if the authors do not respond for 14 days. We will present
230 reporting quality and risk of bias assessment in the respective tables.

231 **Data synthesis**

232 *Qualitative synthesis*

233 We will provide a qualitative analysis of the studies and their results following standard 4.2
234 and conduct a qualitative synthesis, chapter 4 of Finding What Works in Health Care:
235 Standards for Systematic Review (33). In addition, we will analyze studies by themes
236 following the primary and secondary outcomes. We will report details of prediction model
237 performance, comparing ML-based and non-ML-based models, the prediction model use
238 case, choice of predictors and outcomes, ML model construct, prediction model reporting
239 standard, and risk of bias.

240 *Quantitative synthesis*

241 The studies will likely show significant clinical, methodological and statistical heterogeneity.
242 We will therefore synthesize data quantitatively in appropriate subgroups (prediction of CKD
243 development, CKD diagnosis, and prediction of CKD progression to ESKD), guided by the
244 Cochrane Handbook of Systematic Reviews of Diagnostic Tests Accuracy, Chapter 10:
245 Understanding meta-analysis. Draft version (13 May 2022) (34).

246 *Measure of effect size*

247 Performance measures of the ML algorithms will be recorded, including 2x2 confusion
248 matrix, sensitivity, specificity, and AUROC or any other measures utilized by individual
249 studies. The most common measure will be selected for further meta-analysis if feasible.

250 *Assessment of heterogeneity*

251 We will assess the clinical heterogeneity of studies based on the ML algorithm use case, the
252 participant characteristics, the predictor choice, and selection. We will assess the
253 methodological heterogeneity based on the ML algorithm construct regarding the data size,

254 ML technique, and performance measures. We will assess statistical heterogeneity using the
255 χ^2 test and the I^2 statistic. We will consider an I^2 value greater than 50% indicative of
256 substantial heterogeneity.

257 *Quantitative Data Synthesis*

258 Suppose sensitivity, specificity, 2x2 confusion matrix results are available for the majority of
259 studies, we will use either a bivariate model or Rutter and Gatsonis HSROC models
260 depending on the consistency of study thresholds to calculate to define a summary point or
261 SROC curve of the algorithm performance estimate combining sensitivity and specificity.

262 We will undergo exploratory analysis comparing the performance of ML and non-ML
263 techniques. Depending on the number of studies that evaluated both methods, we will either
264 restrict the analysis to these studies or combine all studies that have utilized one of the
265 techniques. We will again use the bivariate model or the Rutter and Gatsonis HSROC model
266 to compare the relative accuracy of both tests.

267 If the majority of sensitivity, specificity, and 2x2 confusion data are not available, and
268 if AUROC is the most common performance measure, we will utilize methods described in
269 Christodoulou et al. (35) by analyzing pairwise differences in logit AUROCs between ML-
270 based and non-ML based techniques by random effects modeling by DerSimonian and Laird
271 method, either pooled or within subgroups stratified by the risk of bias, study outcomes, ML
272 techniques. Points and the 95% CI will describe the pairwise differences in logit AUROCs.

273 The meta-analysis will be performed using Review Manager version 5.4.1 (The
274 Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark) (36) and
275 Stata (37).

276 *Additional analysis*

277 Further subgroup analyses will be performed to explore possible sources of heterogeneity
278 based on the following: study quality (risk of bias and reporting quality assessment), ML
279 techniques, stages of validation, and dataset (stratified as determined by resulted studies).

280 We will conduct sensitivity analyses based on study quality, study publication years
281 (stratified by year), study populations, the ML model construct, the ML user case, or any
282 other relevant strata.

283 **Publication bias**

284 A funnel plot will be constructed to assess the risk of publication bias.

285 **Confidence in cumulative evidence**

286 Overall evidence quality will be assessed using The Grade of Recommendations of
287 Assessment, Development, and Evaluation (GRADE) guidance for assessing certainty
288 evidence for test accuracy incorporating domains including the risk of bias, indirectness,
289 applicability, inconsistency, imprecision, and publication bias (38). The overall level of
290 evidence: high, moderate, low, very low (one level deducted for each domain with serious
291 concerns/high risk of bias). No factors increased the level.

292 **Systematic review reporting**

293 As ML in CKD is a rapidly developing field, if the number of studies that meet the eligibility
294 criteria exceeds the capacity to report the study outcomes in one systematic review, the
295 research team will report diagnostic and prognostic tools as separate systematic reviews to
296 ensure a clear and focused reporting and appraisal of study outcomes.

297 **Discussion**

298 Artificial intelligence and machine learning have significant potential in modern healthcare.
299 Specifically, models developed for the prediction and diagnosis of CKD development and

300 progression can allow for early disease recognition and intervention, which may help to
301 facilitate early disease prevention and diagnosis, assist with early care planning, and allocate
302 resources for the most significant clinical and economic benefit.

303 As the number of algorithms grows exponentially, the focus should direct toward
304 addressing the barriers to clinical implementation. This review aims to assess the
305 methodological rigor of model development and compare ML-based algorithms' performance
306 with conventional methods. This will inform clinicians and technical specialists of the current
307 development of ML in CKD care, as well as direct future model development and
308 standardization.

309 **Acknowledgments**

310 None

311 **Authors' contributions**

312 Conceptualization: FC, TN, PK, AC, KP

313 Methodology: FC, TN, PK, AC, KP

314 Writing – Original Draft Preparation: FC, AC

315 Writing – Review & Editing: FC, PK, TN, KP

316 Supervision and guarantor: KP

317 **Financial support:** None

318

319 **References**

- 320 1. Hill, N. R., Fatoba, S. T., Oke, J. L., Hirst, J. A., O'Callaghan, C. A., Lasserson, D. S.,
321 & Hobbs, F. D. (2016). Global Prevalence of Chronic Kidney Disease - A Systematic
322 Review and Meta-Analysis. *PloS one*, *11*(7), e0158765.
- 323 2. GBD Chronic Kidney Disease Collaboration (2020). Global, regional, and national
324 burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global
325 Burden of Disease Study 2017. *Lancet (London, England)*, *395*(10225), 709–733.
- 326 3. Chen, T. K., Knicely, D. H., & Grams, M. E. (2019). Chronic Kidney Disease
327 Diagnosis and Management: A Review. *JAMA*, *322*(13), 1294–1304.
- 328 4. Levin, A., Tonelli, M., Bonventre, J., Coresh, J., Donner, J. A., Fogo, A. B., Fox, C.
329 S., Gansevoort, R. T., Heerspink, H., Jardine, M., Kasiske, B., Köttgen, A., Kretzler,
330 M., Levey, A. S., Luyckx, V. A., Mehta, R., Moe, O., Obrador, G., Pannu, N., Parikh,
331 C. R., ... ISN Global Kidney Health Summit participants (2017). Global kidney
332 health 2017 and beyond: a roadmap for closing gaps in care, research, and
333 policy. *Lancet (London, England)*, *390*(10105), 1888–1917.
- 334 5. Liyanage, T., Ninomiya, T., Jha, V., Neal, B., Patrice, H. M., Okpechi, I., Zhao, M. H.,
335 Lv, J., Garg, A. X., Knight, J., Rodgers, A., Gallagher, M., Kotwal, S., Cass, A., &
336 Perkovic, V. (2015). Worldwide access to treatment for end-stage kidney disease: a
337 systematic review. *Lancet (London, England)*, *385*(9981), 1975–1982.
- 338 6. Panch T, Duralde E, Mattie H, Kotecha G, Celi LA, Wright M, et al. (2022) A
339 distributed approach to the regulation of clinical AI. *PLOS Digit Health* *1*(5):
340 e0000040.
- 341 7. Busnatu, □., Niculescu, A. G., Bolocan, A., Petrescu, G., Păduraru, D. N., Năstasă, I.,
342 Lupu□oru, M., Geantă, M., Andronic, O., Grumezescu, A. M., & Martins, H. (2022).

- 343 Clinical Applications of Artificial Intelligence-An Updated Overview. *Journal of*
344 *clinical medicine*, 11(8), 2265.
- 345 8. Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P.
346 (2020). Introduction to Machine Learning, Neural Networks, and Deep
347 Learning. *Translational vision science & technology*, 9(2), 14.
- 348 9. Dash, S., Shakyawar, S.K., Sharma, M. *et al.* Big data in healthcare: management,
349 analysis and future prospects. *J Big Data* 6, 54 (2019).
- 350 10. Kaur, N., Bhattacharya, S., & Butte, A. J. (2021). Big Data in Nephrology. *Nature*
351 *reviews. Nephrology*, 17(10), 676–687.
- 352 11. Zhang, K., Liu, X., Xu, J., Yuan, J., Cai, W., Chen, T., Wang, K., Gao, Y., Nie, S., Xu,
353 X., Qin, X., Su, Y., Xu, W., Olvera, A., Xue, K., Li, Z., Zhang, M., Zeng, X., Zhang,
354 C. L., Li, O., ... Wang, G. (2021). Deep-learning models for the detection and
355 incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus
356 images. *Nature biomedical engineering*, 5(6), 533–545.
357 <https://doi.org/10.1038/s41551-021-00745-6>
- 358 12. Kao, H. Y., Chang, C. C., Chang, C. F., Chen, Y. C., Cheewakriangkrai, C., & Tu, Y.
359 L. (2022). Associations between Sex and Risk Factors for Predicting Chronic Kidney
360 Disease. *International journal of environmental research and public health*, 19(3),
361 1219.
- 362 13. Allen, A., Iqbal, Z., Green-Saxena, A., Hurtado, M., Hoffman, J., Mao, Q., & Das, R.
363 (2022). Prediction of diabetic kidney disease with machine learning algorithms, upon
364 the initial diagnosis of type 2 diabetes mellitus. *BMJ open diabetes research &*
365 *care*, 10(1), e002560.
- 366 14. Chuah, A., Walters, G., Christiadi, D., Karpe, K., Kennard, A., Singer, R., Talaulikar,
367 G., Ge, W., Suominen, H., Andrews, T. D., & Jiang, S. (2022). Machine Learning

- 368 Improves Upon Clinicians' Prediction of End Stage Kidney Disease. *Frontiers in*
369 *medicine*, 9, 837232.
- 370 15. Lee, K. H., Chu, Y. C., Tsai, M. T., Tseng, W. C., Lin, Y. P., Ou, S. M., & Tarng, D.
371 C. (2022). Artificial Intelligence for Risk Prediction of End-Stage Renal Disease in
372 Sepsis Survivors with Chronic Kidney Disease. *Biomedicines*, 10(3), 546.
- 373 16. Tangri, N., Stevens, L. A., Griffith, J., Tighiouart, H., Djurdjev, O., Naimark, D.,
374 Levin, A., & Levey, A. S. (2011). A predictive model for progression of chronic
375 kidney disease to kidney failure. *JAMA*, 305(15), 1553–1559.
- 376 17. Tangri, N., Grams, M. E., Levey, A. S., Coresh, J., Appel, L. J., Astor, B. C., Chodick,
377 G., Collins, A. J., Djurdjev, O., Elley, C. R., Evans, M., Garg, A. X., Hallan, S. I.,
378 Inker, L. A., Ito, S., Jee, S. H., Kovesdy, C. P., Kronenberg, F., Heerspink, H. J.,
379 Marks, A., ... CKD Prognosis Consortium (2016). Multinational Assessment of
380 Accuracy of Equations for Predicting Risk of Kidney Failure: A Meta-
381 analysis. *JAMA*, 315(2), 164–174.
- 382 18. Chan, C. T., Blankestijn, P. J., Dember, L. M., Gallieni, M., Harris, D., Lok, C. E.,
383 Mehrotra, R., Stevens, P. E., Wang, A. Y., Cheung, M., Wheeler, D. C., Winkelmayr,
384 W. C., Pollock, C. A., & Conference Participants (2019). Dialysis initiation, modality
385 choice, access, and prescription: conclusions from a Kidney Disease: Improving
386 Global Outcomes (KDIGO) Controversies Conference. *Kidney international*, 96(1),
387 37–47.
- 388 19. Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van
389 Calster, B. (2019). A systematic review shows no performance benefit of machine
390 learning over logistic regression for clinical prediction models. *Journal of clinical*
391 *epidemiology*, 110, 12–22.

- 392 20. Nusinovici, S., Tham, Y. C., Chak Yan, M. Y., Wei Ting, D. S., Li, J., Sabanayagam,
393 C., Wong, T. Y., & Cheng, C. Y. (2020). Logistic regression was as good as machine
394 learning for predicting major chronic diseases. *Journal of clinical epidemiology*, *122*,
395 56–69.
- 396 21. Song, X., Liu, X., Liu, F., & Wang, C. (2021). Comparison of machine learning and
397 logistic regression models in predicting acute kidney injury: A systematic review and
398 meta-analysis. *International journal of medical informatics*, *151*, 104484.
- 399 22. Collins, G. S., Omar, O., Shanyinde, M., & Yu, L. M. (2013). A systematic review
400 finds prediction models for chronic kidney disease were poorly reported and often
401 developed using inappropriate methods. *Journal of clinical epidemiology*, *66*(3), 268–
402 277.
- 403 23. Ramspek, C. L., de Jong, Y., Dekker, F. W., & van Diepen, M. (2020). Towards the
404 best kidney failure prediction tool: a systematic review and selection aid. *Nephrology*,
405 *dialysis, transplantation : official publication of the European Dialysis and*
406 *Transplant Association - European Renal Association*, *35*(9), 1527–1538.
- 407 24. Nagendran, M., Chen, Y., Lovejoy, C. A., Gordon, A. C., Komorowski, M., Harvey,
408 H., Topol, E. J., Ioannidis, J., Collins, G. S., & Maruthappu, M. (2020). Artificial
409 intelligence versus clinicians: systematic review of design, reporting standards, and
410 claims of deep learning studies. *BMJ (Clinical research ed.)*, *368*, m689.
- 411 25. Dhiman, P., Ma, J., Navarro, C. A., Speich, B., Bullock, G., Damen, J. A., Kirtley, S.,
412 Hooft, L., Riley, R. D., Van Calster, B., Moons, K., & Collins, G. S. (2021).
413 Reporting of prognostic clinical prediction models based on machine learning
414 methods in oncology needs to be improved. *Journal of clinical epidemiology*, *138*,
415 60–72.

- 416 26. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and
417 prognosis of covid-19: systematic review and critical appraisal [published correction
418 appears in *BMJ*. 2020 Jun 3;369:m2204]. *BMJ*. 2020;369:m1328.
- 419 27. Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle,
420 P., Stewart, L. A., & PRISMA-P Group (2015). Preferred reporting items for
421 systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and
422 explanation. *BMJ (Clinical research ed.)*, 350, g7647.
- 423 28. Moons, K. G., de Groot, J. A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman,
424 D. G., Reitsma, J. B., & Collins, G. S. (2014). Critical appraisal and data extraction
425 for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS*
426 *medicine*, 11(10), e1001744.
- 427 29. Covidence systematic review software, Veritas Health Innovation, Melbourne,
428 Australia. Available at www.covidence.org.
- 429 30. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al.
430 The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.
431 *BMJ* 2021;372:n71.
- 432 31. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent
433 reporting of a multivariable prediction model for individual prognosis or diagnosis
434 (TRIPOD): the TRIPOD statement. *BMJ (Clinical research ed.)*, 350, g7594.
- 435 32. Wolff, R. F., Moons, K., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S.,
436 Reitsma, J. B., Kleijnen, J., Mallett, S., & PROBAST Group† (2019). PROBAST: A
437 Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals*
438 *of internal medicine*, 170(1), 51–58.
- 439 33. Standard 4.2 conduct a qualitative synthesis, chapter 4 of Finding What Works in
440 Health Care: Standards for Systematic Reviews

- 441 34. Cochrane Handbook of Systematic Reviews of Diagnostic Tests Accuracy, Chapter
442 10: Understanding meta-analysis. Draft version (13 May 2022)
- 443 35. Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van
444 Calster, B. (2019). A systematic review shows no performance benefit of machine
445 learning over logistic regression for clinical prediction models. *Journal of clinical
446 epidemiology*, *110*, 12–22.
- 447 36. Review Manager 5 (RevMan 5). 5.3 ed. Copenhagen: Nordic Cochrane Centre, The
448 Cochrane Collaboration; 2014.
- 449 37. StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX:
450 StataCorp LLC.
- 451 38. Yang B, Mustafa RA, Bossuyt PM, et al. GRADE Guidance: 31. Assessing the
452 certainty across a body of evidence for comparative test accuracy. *J Clin Epidemiol*.
453 2021;136:146-156.