Table S1. TRIPOD Checklist for Prediction Model Validation.

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	4
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	4
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	6
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	6
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	6-8
	5b	Describe eligibility criteria for participants.	7
	5c	Give details of treatments received, if relevant.	
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	7-8
	6b	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	7
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	Explain how the study size was arrived at.	7
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	7-8
Statistical analysis methods	10c	For validation, describe how the predictions were calculated.	7-8
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	8
	10e	Describe any model updating (e.g., recalibration) arising from the validation, if done.	8
Risk groups	11	Provide details on how risk groups were created, if done.	8
Development vs. validation	12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	Supplementary tables
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	10
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	8, Table 1
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	
Model performance	16	Report performance measures (with CIs) for the prediction model.	Fig.2
Model-updating	17	If done, report the results from any model updating (i.e., model specification, model performance).	Fig.2
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	13-14
Interpretation	19a	For validation, discuss the results with reference to performance in the development data, and any other validation data.	12-13
	19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	13

Implications	20	Discuss the potential clinical use of the model and implications for future research.	13-14
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	S Tables
Funding	22	Give the source of funding and the role of the funders for the present study.	14

Table S2: PICOTS items framing the review aim, search strategy, and study inclusion and exclusion criteria for the systematic review

Item	Description
Population	People with type 2 diabetes or applicable to people with type 2 diabetes by including it as predictor
Intervention or Model	All prognostic models from discovery studies to predict risk of nephropathy
Comparator	Not applicable
Outcome(s)	Albuminuria, diabetic kidney disease, chronic kidney disease and end-stage renal
Timing	At least 1 year follow-up
Setting	Applicable to people with type 2 diabetes treated in primary care

Table S3 Search strings used for the systematic review.

PubMed
"Diabetic Nephropathies"[Mesh] OR "Albuminuria"[Mesh] OR "Renal Insufficiency"[Mesh] OR nephropath*[tiab] OR albuminuria*[tiab] OR microalbuminuria*[tiab] OR macroalbuminuria*[tiab] OR kidney disease*[tiab] OR kidney disorder*[tiab] OR kidney failure*[tiab] OR kidney insufficien*[tiab] OR renal disorder*[tiab] OR renal failure*[tiab] OR renal insufficien*[tiab] AND "Diabetes Mellitus, Type 2"[Mesh] OR diabetes[tiab] OR (diabetic*[tiab] AND (non insulin depend*[tiab] OR noninsulin depend*[tiab] OR noninsulindepend*[tiab] OR non insulindepend*[tiab])) OR dm2[tiab] OR niddm[tiab] OR dm 2[tiab] OR t2d*[tiab] OR dm type 2[tiab] OR type 2 diabet*[tiab] OR type two diabet*[tiab] OR type II diabet*[tiab] OR dm type II[tiab] AND "Decision Support Techniques"[Mesh] OR "Prognosis"[Mesh] OR ((predict*[tiab] OR prognos*[tiab]) AND (model*[tiab] OR rule*[tiab] OR score*[tiab] OR tool[tiab])) OR risk score*[tiab] OR risk assessment*[tiab] OR risk algorithm*[tiab] OR risk engine*[tiab] OR risk equation*[tiab] OR risk prediction*[tiab] OR risk calculation*[tiab] NOT ("Animals"[Mesh] NOT "Humans"[Mesh])
Embase
'diabetic nephropathy'/exp OR 'albuminuria'/exp OR 'kidney failure'/de OR 'chronic kidney failure'/exp OR nephropath*:ab,ti OR albuminuria*:ab,ti OR microalbuminuria*:ab,ti OR macroalbuminuria*:ab,ti OR (kidney NEAR/3 disease*):ab,ti OR (kidney NEAR/3 disorder*):ab,ti OR (kidney NEAR/3 failure*):ab,ti OR (kidney NEAR/3 insufficien*):ab,ti OR ‘renal disorder*’:ab,ti OR ‘renal failure*’:ab,ti OR ‘renal insufficien*’:ab,ti AND 'non insulin dependent diabetes mellitus'/exp OR diabetes:ab,ti OR (diabetic* NEAR/3 ('non insulin depend*' OR 'noninsulin depend*' OR noninsulindepend* OR 'non insulindepend*')):ab,ti OR dm2:ab,ti OR niddm:ab,ti OR 'dm 2':ab,ti OR t2d*:ab,ti OR 'dm type 2':ab,ti OR 'type 2 diabet*':ab,ti OR 'type two diabet*':ab,ti OR 'type ii diabet*':ab,ti OR 'dm type ii':ab,ti AND 'decision support system'/exp OR 'prognosis'/de OR (predict* NEAR/3 model*):ab,ti OR (predict* NEAR/3 rule*):ab,ti OR (predict* NEAR/3 score*):ab,ti OR (predict* NEAR/3 tool*):ab,ti OR (prognos* NEAR/3 model*):ab,ti OR (prognos* NEAR/3 rule*):ab,ti OR (prognos* NEAR/3 score*):ab,ti OR (prognos* NEAR/3 tool*):ab,ti OR (risk NEAR/3 score*):ab,ti OR (risk NEAR/3 assessment*):ab,ti OR (risk NEAR/3 algorithm*):ab,ti OR (risk NEAR/3 engine*):ab,ti OR (risk NEAR/3 equation*):ab,ti OR (risk* NEAR/3 prediction*):ab,ti OR (risk NEAR/3 calculation*):ab,ti NOT ([animals]/lim NOT [humans]/lim) NOT ('conference abstract'/it OR 'conference review'/it OR 'editorial'/it OR 'letter'/it OR 'note'/it)

Table S4. Characteristics of prediction models for the risk of nephropathy developed in the type 2 diabetes population

Study	Development population	Country	Design	Type of model	Outcome	n events/ n total	Predicted time (years)	Model	Number of predictors	Discrimination (AUC)	Calibration (Hosmer-Lemeshow p-value)	Method of internal validation
Afghahi 2011[1]	T2D	Sweden	Registry	Logistic	Albuminuria (microalbuminuria urine albumin excretion 20-200 µg/min or macroalbuminuria > 200 µg/min)	729/3,667	5	-	8	0.67-0.87	p-value = not significant	N.R.
Afghahi 2011[1]	T2D	Sweden	Registry	Logistic	Renal impairment (MDRD eGFR < 60 mL/min/1.73 m^2)	407/3,667	5	-	6	0.67-0.87	p-value = not significant	N.R.
Afghahi 2011[1]	T2D	Sweden	Registry	Logistic	Renal impairment (eCrCl < 60 mL/min)	407/3,667	5	-	6	0.67-0.87	p-value = not significant	N.R.
Altemtam 2012[2]	T2D with kidney damage or CKD	United Kingdom	Registry	Logistic	Progression of diabetic kidney disease (> 2 mL/min decline in MDRD eGFR)	94/270	5	-	5	N.R.	N.R.	N.R.
Basu 2017[3]	T2D	USA	Cohort	Cox	Micro-albuminuria (UACR ≥ 30 mg/g)	1551/9635	4.7	-	14	0.62	0.77*	External validation set
					Macro-albuminuria (UACR ≥ 300 mg/g)	627/9635				0.84	<0.001*	
					Renal failure or end-stage renal disease (Serum creatinine ≥ 3.3 mg/L)	292/9635				0.60	<0.001*	
					Doubling serum creatinine or >20 mL/min/1.73m^2 decrease in eGFR	5910/9635				0.76	<0.001*	
					Macro-albuminuria, renal failure, end stage renal disease, doubling of serum creatinine or >20 mL/min/1.73m^2 decrease in eGFR	6195/9635				0.73	<0.001*	
					Macroalbuminuria, microalbuminuria, renal failure, or end-stage renal disease	2321/9635				0.61	0.87*	
Cheng 2020[4]	T2D with CKD	China	Cohort	Logistic	Kidney failure (Dialysis, renal transplantation)	272/641	3	Clinical	3	0.626	0.315	Discovery and validation set
								Laboratory	5	0.986	0.755	
								Lab-medication	6	0.986	0.438	
								Full	8	0.986	0.540	
Clarke 2004[5]	T2D	United Kingdom	RCT	Weibull	Renal failure (creatinine level > 250 µmol/l or death caused by renal failure)	24/3,642	N.R.	-	2	N.R.	N.R.	N.R.
Dagliati 2018[6]	T2D	Italy	Cohort	Logistic	Nephropathy (<60 mL/min/1.73m^2 or UACR≥30 mg/g)	121/943	3, 5, 7	-	4	3Y: 0.647 5Y: 0.693 7Y: 0.686	N.R.	Leave-one-out
Dunkler 2015[7]	T2D with vascular disease or organ damage	40 countries	RCT	Multinominal logistic	CKD (UACR > 30 mg/g and 300 mg/g or doubling creatinine) or ESRD (CKD-EPI eGFR < 15 mL/min/1.73 m^2)	1,079/6,766	5.5	Laboratory model	5	0.68	Calibration slope = 0.98	Bootstrap
								Clinical model	14	0.69	Calibration slope = 0.88 to 0.90	
Elley 2013[8]	T2D	New Zealand	Registry	Cox	ESRD (RRT or death by CKD or ESRD)	637/25,736	5	-	10	0.89	N.R.	N.R.

Table S4. Continued

Study	Development population	Country	Design	Type of model	Outcome	n events/ n total	Predicted time (years)	Model	Number of predictors	Discrimination (AUC)	Calibration (Hosmer-Lemeshow p-value)	Method of internal validation
Fernandez-Fernandez 2020	T1D and T2D	Spain	Cohort	Logistic	Rapid progression (>5 mL/min/1.73 m^2/year)	All: 48/261 Men: 33 Women: 15	2.5	-	All: 5 Men: 6 Women:4	All: 0.81 Men: 0.92 Women: 0.90	N.R.	N.R.
Goldfarb-Rumyantzev 2002[9]	T2D	United States	Cohort	Artificial neural network	Renal insufficiency (GFR < 71 mL/min)	40/86	4	-	13	0.91	N.R.	Random split sample
Hu 2020[10]	T2D	China	Cohort	Logistic	Diabetic nephropathy (UACR ≥ 30 µg/mg)	701/3489	4	-	8	0.744	Medium calibration	Bootstrap
Imbroll 2017[11]	T2D	Malta	Cohort		Albuminuria progression (Normo to micro, micro to macro)	59/243	8	-	4	N.R.	N.R.	N.R.
Jardine 2012[12]	T2D	20 countries	RCT	Cox	Major kidney event (doubling serum creatinine level to ≥ 200 µmol/l)	166/11,140	5	-	7	0.85	0.90	Bootstrap
Jardine 2012[12]	T2D	20 countries	RCT	Cox	New-onset albuminuria (UACR ≥ 30 µg/mg)	2,715/7,377	5	-	8	0.65	0.06	Bootstrap
Keane 2006[13]	T2D and nephropathy	Asia, Europe, Latin America, New Zealand and North America	RCT	Cox	ESRD (RRT)	341/1513	N.R.	-	4	N.R.	Calibration table	Naïve and jackknife
Li 2016[14]	T2D	Taiwan	Cohort	Cox	ESRD (dialysis)	22/604	Variable	-	4	All: 0.94 CKD Stage 3-5:0.91	N.R.	Leave-one-out
Lin 2017[15]	T2D	Chinese	Cohort	Cox	ESRD (MDRD eGFR < 15 mL/min/1.73 m^2 or RRT)	Derivation: 813/16,070 Validation 402/8,034	3, 5 and 8 years	-	11	3 years: 0.90 5 years: 0.86 8 years: 0.81	p > 0.05	Random split sample
Low 2017[16]	T2D	Malaysia	Cohort	Logistic	CKD progression (worsening in MDRD eGFR category coupled with ≥ 25% reduction in eGFR from baseline)	679/1,582	6	-	6	Training: 0.80 Test: 0.83	0.93	Random split sample
Miao 2017[17]	T2D	China	Cohort	Cox	Diabetic nephropathy Persistent albuminuria and diabetic retinopathy in absence other kidney or urinary tract diseases	Training: 45/5,705 Validation: 32/6,066	20	-	Women: 6 Men: 8	Women: 0.84 Men: 0.80	Calibration plot	Split sample
Nelson 2019[18]	T2D	Multinational	Cohort	Multivariable competing risk	CKD (CKD-EPI eGFR < 60 mL/min/1.73 m^2)	GP: 313,646/ 660,856	5	-	12	0.801	Calibration plot	Discovery and validation cohort
Parrinello 2016[19]	T1D and T2D	United States	Cohort	Fine and Gray	CKD (eGFR < 60 mL/min/1.73 m^2 and ≥ 25% decline in eGFR since visit 2, hospitalization due to kidney disease, RRT, or death from kidney disease)	152/654	10	-	22	0.716	Well calibrated	Bootstrap

Table S4. Continued

Study	Development population	Country	Design	Type of model	Outcome	n events/ n total	Predicted time (years)	Model	Number of predictors	Discrimination (AUC)	Calibration (Hosmer-Lemeshow p-value)	Method of internal validation
Riphagen 2015[20]	T2D	The Netherlands	Cohort	Cox and Fine and Gray	(Micro)albuminuria (UACR > 2.5 mg/mmol in men and > 3.5 mg/mmol in women)	183/640	10	-	7	0.69 (both standard and competing risk)	Calibration plot = moderate (standard) Calibration plot = good (competing risk)	N.R.
Riphagen 2015[20]	T2D	The Netherlands	Cohort	Cox and Fine and Gray	Progressive renal function loss (60% increase of baseline SCr)	79/1,143	10	-	5	0.73 (standard) 0.74 (competing risk)	Calibration plot. Standard: over-estimation; competing risk: good	N.R.
Rodriguez-Romero 2019[21]	T2D	United States and Canada	RCT	Random Forest	Nephropathy (doubling of baseline SCr, >20 mL/minute/1.73-m2 decline in eGFR, UACR ≥ 30 mg/g, RRT or a rise of SCr > 291.72 µmol/L)	6,777/10,251	7	-	18	0.768-0.840	N.R.	Cross-validation
Sun 2020[22]	T2D	China	Cohort	Logistic	ESRD (renal death, hospitalization due to non-fatal renal failure, or eGFR < 15 mL/min/1.73 m^2)	225/968	3	Clinical	4	0.86	0.558	Random split sample
Sun 2020[22]	T2D	China	Cohort	Logistic	ESRD (renal death, hospitalization due to non-fatal renal failure, or eGFR < 15 mL/min/1.73 m^2)	225/968	3	Clinical-pathological	5	0.87	0.909	Random split sample
Sun 2020[22]	T2D	China	Cohort	Logistic	ESRD (renal death, hospitalization due to non-fatal renal failure, or eGFR < 15 mL/min/1.73 m^2)	225/968	3	Clinical-medical	5	0.84	0.418	Random split sample
Sun 2020[22]	T2D	China	Cohort	Logistic	ESRD (renal death, hospitalization due to non-fatal renal failure, or eGFR < 15 mL/min/1.73 m^2)	225/968	3	Full model	7	0.87	0.623	Random split sample
Tanaka 2013[23]	T2D	Japan	RCT	Cox	Nephropathy (spot urinary albumin excretion > 33.9 mg/mmol creatinine)	71/1,748	5	-	5	0.77	0.11	Cross-validation
Wada 2014[24]	T2D	Japan	Cohort	Cox	Renal event (RRT or half reduction in eGFR)	419/4,328	N.R.	-	5	N.R.	Calibration table	N.R.
Wan 2017[25]	T2D	China	Cohort	Cox	ESRD (ICD-9: 250.3x, 585.x and 586.x or eGFR <15 ml/min/1.73m^2)	239/116,509	5	Male Female	7	0.87 0.86	Calibration plot	Random split sample
Woodward 2016[26]	T2D and at least 1 risk factor for micro- or macrovascular disease	20 countries	RCT	Cox	Major renal event (renal death or RRT)	84/7,301	10	-	13	0.81	0.13	N.R.
Wysham 2020[27]	T2D	United States	Registry	Logistic	Moderate to severe DKD (>2 measurements < 60 mL/min/1.73 m^2 or CKD stage 3/4)	6,219/ 160,031	6-months windows (0->60 months)	-	13	0.71	Calibration plot =good	Random split sample

Table S4 Continued

Study	Development population	Country	Design	Type of model	Outcome	n events/ n total	Predicted time (years)	Model	Number of predictors	Discrimination (AUC)	Calibration (Hosmer-Lemeshow p-value)	Method of internal validation
Wysham 2020[27]	T2D	United States	Registry	Logistic	ESRD (CKD stage 5, ESRD medical claim, RRT)	1,619/ 160,031		-	12	0.82		Random split sample
Wysham 2020[27]	T2D	United States	Registry	Logistic	Renal death (Death <30 days after CKD diagnosis)	2,135 / 160,031		-	17	0.85		Random split sample
Yang 2006[28]	T2D	China	Registry	Cox	ESRD (renal death, hospitalization due to non-fatal renal failure, or MDRD eGFR < 15 mL/min/1.73 m^2)	159/4,438	4	-	8	0.97	> 0.10	Random split sample

* The study of Base et al. uses the Greenwood-D'Agostino-Nam test instead of the Hosmer-L

Table S5. Predictors included in the prediction models developed in the type 2 diabetes population

	Afghahi 2011[1] Albuminuria	Afghahi 2011[1] Renal impairment (MDRD)	Afghahi 2011[1] Renal impairment (eGFR)	Altemtam 2012[2]	Basu 2017[3]	Cheng 2020[4] Clinical	Cheng 2020[4] Laboratory	Cheng 2020[4] Lab-medical	Cheng 2020[4] Full model	Clarke 2004[5]	Dagliati 2018[6]	Dunkler 2015[7] Laboratory	Dunkler 2015[7] Clinical	Elley 2013[8]	Fernandez-Fernandez 2020[29]	Goldfarb-Rumyantzev 2020[9]	Hu 2020[10]	Imbroll 2017[11]	Jardine 2012[12] Kidney event	Jardine 2012[12] Albuminuria	Keane 2006[13]	Li 2016[14]	Lin 2017[15]	Low 2017[16]	Miao 2017[17]	Nelson 2019[10]	Parinello 2016[19]	Riphagen 2015[20] (Micro)albuminuria	Riphagen 2015[20] Renal function loss	Rodriguez-Romero 2019[30]	Sun 2020[21] Clinical	Sun 2020[21] Clinical-pathological	Sun 2020[22] Clinical-medical	Sun 2020[22] Full model	Tanaka 2013[23]	Wada 2014[24]	Wan 2017[25]	Woodward 2016[26]	Wyshsam 2020[27] DKD	Wyshsam 2020 ESRD	Wyshsam 2020 Renal death	Yang 2006[28]	Total
Demographic variables																																											
Age	●	●	●		●	●			●			●	●			●							●	●	●	●	●	●	●					●		●	●	●	●			●		22
Sex	●	●	●		●	●			●			●	●	●					●				●			●	●	●										●	●			●		17
Ethnicity					●								●	●						●						●	●															●	●	8
Geographic location																									●																	●	●	3
Education																			●								●																	2
Age at completion education																																							●						1
Lifestyle variables																																												
Smoking	●				●						●			●				●								●	●	●						●									9
Physical activity																											●												●					2
Alcohol																											●																	1
Diet control or physical activity																									●																			1
DM-related variables																																												
(Baseline) HbA1C	●			●	●						●			●			●	●	●	●		●			●	●	●							●	●	●	●						18
Glucose-lowering medication					●	●		●	●														●				●										●						7
Duration DM													●	●		●		●					●															●				●		6
DRP (history)																●			●	●			●															●				●		6
Glucose													●				●													●													3
Diabetes severity index																																										●	●	●		3
Macrovascular complications																												●	●															2
Age onset DM														●									●																					2
>1 T2D-related hospitalization																																											●	●	2
HbA1c variation																							●																					1
Recent DM																											●																	1
Laser therapy DRP													●																															1
Insulin use																										●																		1
No DM medication use																										●																		1
>2 T2D outpatient visits																																											●		1
Anthropometric variables																																												
BMI	●	●	●								●						●								●	●			●														8
Waist circumference													●		●					●																									3
Weight																●																													1

Table S5. Continued

	Afghahi 2011 Albuminuria	**Afghahi 2011** Renal impairment (MDRD)	**Afghahi 2011** Renal impairment (eCrCl)	**Altemtam 2012**	**Basu 2017**	**Cheng 2020**[4] Clinical	**Cheng 2020**[4] Laboratory	**Cheng 2020**[4] Lab-medical	**Cheng 2020**[4] Full model	**Clarke 2004**	**Dagliati 2018**	**Dunkler 2015** Laboratory	**Dunkler 2015** Clinical	**Elley 2013**	**Fernandez-Fernandez 2020**	**Goldfarb-Rumyantzev 2002**	**Hu 2020**	**Imbroll 2017**	**Jardine 2012** Kidney event	**Jardine 2012** Albuminuria	**Keane 2006**	**Li 2016**	**Lin 2017**[45]	**Low 2017**	**Miao 2017**	**Nelson 2019**	**Parrinello 2016**	**Riphagen 2015** (Micro)albuminuria	**Riphagen 2015** Renal function loss	**Rodriguez-Romero 2019**	**Sun 2020** Clinical	**Sun 2020** Clinical-pathological	**Sun 2020** Clinical-medical	**Sun 2020** Full model	**Tanaka 2013**	**Wada 2014**	**Wan 2017**	**Woodward 2016**	**Wysham 2020** DKD	**Wysham 2020** ESRD	**Wysham 2020** Renal death	**Yang 2006**	**Total**
Physical examination variables																																									
SBP	●	●		●	●					●				●		●	●		●	●				●			●	●	●					●	●		●			●	18
Artrial fibrillation (current or previous)																																			●			●				2
DBP																	●																									1
Variation in BP																							●																			1
Pulse pressure			●																																							1
Laboratory variables																																									
(Baseline) eGFR							●	●	●			●	●			●			●	●		●		●		●				●	●	●	●	●		●	●	●			●	20
UACR					●							●	●		●				●	●	●			●						●								●	●			11
ACR																●										●		●		●	●		●	●	●						●	10
Serum creatinine		●	●		●									●			●				●		●	●	●																	9
Triglycerides	●	●	●												●		●										●			●													7
Cystatin C							●	●	●																						●	●	●	●								7
HDL cholesterol	●				●																				●		●			●													5
LDL cholesterol													●			●								●			●			●													5
Haemoglobin							●	●	●													●																					4
B-type natriuretic peptide																															●	●	●	●								4
Neutrophil:lymphocyte ratio							●	●	●																																	3
24-h urine protein							●	●	●																																	3
Albuminuria														●									●																			2
Proteinuria				●																			●																			2
Albuminuria stage												●																															2
Cholesterol					●																									●												2
Serum uric acid				●											●																												2
Second ACR screening																●																											1
Mean of ACR screenings																●																											1
Serum albumin																					●																					1
Urinary albumin																													●													1
eGFR-Cr																											●															1
Urinary creatinine																													●													1
VLDL cholesterol																													●													1
TC/HDL ratio																																									●	1
Haematocrit																																									●	1
Serum VAP-1																							●																			1
1/β2 microglobin																											●															1
IgG-creatinine ratio																●																											1

Table S5. Continued

	Afghahi 2011 Albuminuria	**Afghahi 2011** Renal impairment	**Afghahi 2011** Renal impairment	**Altemtam 2012**	**Basu 2017**	**Cheng 2020**[4] Clinical	**Cheng 2020**[4] Laboratory	**Cheng 2020**[4] Lab-medical	**Cheng 2020**[4] Full model	**Clarke 2004**	**Dagliati 2018**	**Dunkler 2015** Laboratory	**Dunkler 2015** Clinical	**Elley 2013**	**Fernandez-Fernandez 2020**	**Goldfarb-Rumyantzev 2002**	**Hu 2020**	**Imbroll 2017**	**Jardine 2012** Kidney event	**Jardine 2012** Albuminuria	**Keane 2006**	**Li 2016**	**Lin 2017**	**Low 2017**	**Miao 2017**	**Nelson 2019**	**Parrinello 2016**	**Riphagen 2015** (Micro)albuminuria	**Riphagen 2015** Renal function loss	**Rodriguez-Romero 2019**	**Sun 2020** Clinical	**Sun 2020** Clinical-pathological	**Sun 2020** Clinical-medical	**Sun 2020** Full model	**Tanaka 2013**	**Wada 2014**	**Wan 2017**	**Woodward 2016**	**Wysham 2020** DKD	**Wysham 2020** ESRD	**Wysham 2020** Renal death	**Yang 2006**	**Total**
FE phosphate															●																												1
Vitamin B12															●																												1
Fractional clearance IgG																●																											1
Plasma renin activity																●																											1
Blood urea nitrogen																	●																										1
White blood cell count																		●																									1
Potassium																													●													1
Medical (history) variables																																										
Baseline anti-hypertensive medication					●															●			●				●						●	●				●				7
Hypertension											●															●														●	●		4
CVD					●									●												●																	3
Diabetic nephropathy																																							●	●	●		3
Chronic kidney disease																																							●	●	●		3
Heart failure																																							●	●	●		3
Anaemia																																							●	●	●		3
Pathological grade																																●			●							2
Cholesterol lowering medication																							●				●															2
Familial CVD																											●															1
Peripheral artery disease													●																													1
Vascular comorbidities				●																																						1
Blindness										●																																1
No of antihypertensive medications													●																													1
Anticoagulants					●																																					1
Atherosclerotic cardiac events													●																													1
Stroke/ transient ischemic attack													●																													1
Months using ACEIs																																									●	1
Hypertension or dyslipidaemia																									●																	1
Retinopathy																									●																	1
Cardiac arrhythmias																																									●	1
Gall stones																																									●	1

Table S5. Continued

	Afghahi 2011 Albuminuria	Afghahi 2011 Renal impairment (eGFR)	Afghahi 2011 Renal impairment (eCrCl)	Altemtam 2012	Cheng 2020[a] Clinical	Cheng 2020[a] Laboratory	Cheng 2020[a] Lab-medical	Cheng 2020[a] Full model	Clarke 2004	Cummings 2011 Linear	Cummings 2011 Logistic	Dagliati 2018	Dunkler 2015 Laboratory	Dunkler 2015 Clinical	Elley 2013	Fernandez-Fernandez 2020	Goldfarb-Rumyantzev 2002	Hu 2020	Imbroll 2017	Jardine 2012 Kidney event	Jardine 2012 Albuminuria	Keane 2006	Li 2016	Lin 2017	Low 2017	Miao 2017	Nelson 2019	Parrinello 2016	Riphagen 2015 (Micro)albuminuria	Riphagen 2015 Renal function loss	Rodriguez-Romero 2019	Sun 2020 Clinical	Sun 2020 Clinical-pathological	Sun 2020 Clinical-medical	Sun 2020 Full model	Tanaka 2013	Wada 2014	Wan 2017	Woodward 2016	Wysham 2020 DKD	Wysham 2020 ESRD	Wysham 2020 Renal death	Yang 2006	Total
Interaction terms																																											
Male * glucose lowering medication																												●															1
White * antihypertensive medication																											●																1
Male * triglycerides																											●																1
eGFR * sex																																									●				1
Age completion education * sex																																									●				1
Other																																												
End of baseline period																																										●	●	●	3
Insurance type																																										●	●	●	3
Payer type																																										●		●	2
Time interval																																												●	1

Table S6. Characteristics of prediction models for the risk of nephropathy developed in the general population

Study	Development population	Country	Design	Type of model	Outcome	n events/ n total	Predicted time (years)	Model	Number of predictors	Discrimination (AUC)	Calibration (Hosmer-Lemeshow p-value)	Method of internal validation
Chien 2010[31]	GP	Taiwan	Cohort	Cox	CKD stage 3 (MDRD eGFR < 60 mL/min/1.73 m^2)	190/5,168	4	Clinical	5	0.77 (point-based model) 0.78 (coefficient-based model)	0.21 (point-based model) 0.01 (coefficient-based model)	Cross-validation
Chang 2019[32]	GP with CKD stage 1-5	Taiwan	Cohort	Cox	Initiation of dialysis	1,017/1,549	1.8	-	4	N.R.	N.R.	N.R.
Hanratty 2010[33]	GP with hypertension	United States	Registry	Logistic	CKD (MDRD eGFR < 60 mL/min/1.73 m^2)	429/10,420	N.R.	-	5	0.81	0.01	N.R.
Hemmelgarn 2007[34]	GP	Canada	Registry	Logistic	Rapid progression of kidney dysfunction (> 25% decline in MDRD eGFR)	1,216/10,184	2	-	5	0.59	HL goodness of fit X^2 = 0.77	Random split sample
Hippisley-Cox 2010[35]	GP	England and Wales	Registry	Cox	Moderate/severe CKD (RRT, diagnosis of nephropathy, MDRD eGFR < 45 mL/min/1.73 m^2 or proteinuria)	23,786/775,091	5	CKD model women	14	0.88	N.R.	Random split sample
						17,333/799,658		CKD model men	12	0.88		
Hippisley-Cox 2010[35]	GP	England and Wales	Registry	Cox	ESRD (RRT or MDRD eGFR < 15 mL/min/1.73 m^2)	1,266/775,091	5	ESRD model women	13	0.84	N.R.	Random split sample
						1,543/799,658		ESRD model men	11	0.85		
Johnson 2008[36]	GP with CKD stage 3-4	United States and Canada	Registry	Cox	RRT	323/9,782	5	-	6	0.89	0.99	Bootstrap

Table S6. Continued

Study	Development population	Country	Design	Type of model	Outcome	n events/ n total	Predicted time (years)	Model	Number of predictors	Discrimination (AUC)	Calibration (Hosmer-Lemeshow p-value)	Method of internal validation
Kshirsagar 2008[37]	GP	United States	Cohort	Logistic	CKD (MDRD eGFR < 60 mL/min/1.73 m^2)	1,605/14,155	< 10	Best-fitting categorical model	10	0.70	> 0.20	Random split sample
								Simplified categorical model	8	0.69		
O'Seaghdha 2012[38]	GP	United States	Cohort	Logistic	CKD (MDRD eGFR < 60 mL/min/1.73 m^2)	229/2,490	10	-	5	0.79	0.60	Bootstrap
Saranburut 2017[39]	GP	Thailand	Cohort	Logistic	CKD (CKD-EPI eGFR < 60 mL/min/1.73 m^2)	271/3,186	10	Clinical	5	0.72	0.35	Discovery and validation cohort
Saranburut 2017[39]	GP	Thailand	Cohort	Logistic	CKD (CKD-EPI eGFR < 60 mL/min/1.73 m^2)	271/3,186	10	Clinical+ Limited laboratory	6	0.79	0.21	Discovery and validation cohort
Saranburut 2017[39]	GP	Thailand	Cohort	Logistic	CKD (CKD-EPI eGFR < 60 mL/min/1.73 m^2)	271/3,186	10	Clinical+ Full laboratory	9	0.80	0.41	Discovery and validation cohort
Schroeder 2017[40]	GP with CKD stage 3-4	United States and Canada	Registry	Cox	RRT	737/22,460	5	-	8	0.96	High calibration	Bootstrap
Umesawa 2018[41]	GP	Japan	Cohort	Logistic	CKD (eGFR < 60 mL/min/1.73m^2 and/or proteinuria 2+/3+)	16,464/135,007	10	Simple risk model	4	Men: 0.827 Women: 0.814	Calibration plot	External set
Umesawa 2018[41]	GP	Japan	Cohort	Logistic	CKD (eGFR < 60 mL/min/1.73m^2 and/or proteinuria 2+/3+)	16,464/135,007	10	Full risk model	11	Men: 0.823 Women: 0.824	Calibration plot	External set
Wen 2020 [42]	GP	China	Cohort	Logistic	CKD (eGFR < 60 mL/min/1.73m2 or UACR ≥30 mg/g)	590/3,266	5.6	Simple risk model	5	0.717	0.769	Random split sample
Wen 2020 [42]	GP	China	Cohort	Logistic	CKD (eGFR < 60 mL/min/1.73m2 or UACR ≥30 mg/g)	590/3,266	5.6	Best-fit risk score	7	0.721	0.961	Random split sample

Table S7. Predictors included in the prediction models developed in the general population

	Chien 2010[31] Clinical	Chang 2019[32]	Hanratty 2010[33]	Hemmelgarn 2007[34]	Hippisley-Cox 2010[35] CKD women	Hippisley-Cox 2010[35] CKD men	Hippisley-Cox 2010[35] ESRD women	Hippisley-Cox 2010[35] ESRD men	Johnson 2008[36]	Kshirsagar 2008[37] Best-fitting	Kshirsagar 2008[37] Simplified	O' Seaghdha 2012[38]	Saranburut 2017[39] Clinical	Saranburut 2017[39] Clinical + partial	Saranburut 2017[39] Clinical + Full laboratory	Schroeder 2017[40]	Umesawa 2018[41] Simple model	Umesawa 2018[41] Full model	Wen 2020[42] Simple risk score	Wen 2020[42] Best-fit risk score	Total
Study variables																					
Period of observation			●																		1
Demographic variables																					
Age	●	●	●	●					●	●	●	●	●	●	●	●	●	●			14
Sex									●	●	●		●	●	●	●			●	●	9
Ethnicity					●	●	●	●		●											5
Townsend deprivation					●	●	●	●													4
Education																			●	●	2
Lifestyle variables																					
Smoking					●	●	●	●										●			5
Alcohol intake																		●			1
DM-related variables																					
DM			●	●					●	●	●	●	●	●	●	●			●	●	12
DM type 1					●	●	●	●													4
DM type 2	●	●			●	●	●	●													6
Glucose tolerance																		●			1
Glucose-lowering drugs																		●			1
Anthropometric variables																					
Waist circumference													●	●	●				●		4
BMI	●																	●	●		3
Physical examination variables																					
SBP													●	●	●	●		●	●	●	7
DBP	●																				1
Laboratory variables																					
eGFR			●									●		●	●	●	●	●			7
HDL cholesterol										●					●						2
Haemoglobin															●	●					2
Proteinuria																	●	●			2
Albuminuria/proteinuria																●					1
Uric acid															●						1
Kidney function									●												1
Albuminuria												●									1
Creatinine		●																			1
Urea nitrogen		●																			1
UACR																				●	1
Triglycerides																				●	1
C-reactive protein																				●	1
Medical (history) variables																					
Hypertension					●	●	●	●	●	●	●	●									8
(Peripheral) vascular disease			●		●	●	●	●		●	●										7
Cardiovascular disease					●	●	●	●		●	●										6

Table S7. Continued

	Chien 2010[31] Clinical	**Chang 2019**[32]	**Hanratty 2010**[33]	**Hemmelgarn 2007**[34]	**Hippisley-Cox 2010**[35] CKD women	**Hippisley-Cox 2010**[35] CKD men	**Hippisley-Cox 2010**[35] ESRD women	**Hippisley-Cox 2010**[35] ESRD men	**Johnson 2008**[36]	**Kshirsagar 2008**[37] Best-fitting	**Kshirsagar 2008**[37] Simplified	**O' Seaghdha 2017**[38]	**Saranburut 2017**[39] Clinical	**Saranburut 2017**[39] Clinical + partial laboratory	**Saranburut 2017**[39] Clinical + Full laboratory	**Schroeder 2017**[40]	**Umesawa 2018**[41] Simple model	**Umesawa 2018**[41] Full model	**Wen 2020**[42] Simple risk score	**Wen 2020**[42] Best-fit risk score	**Total**
Heart failure					●	●	●	●		●	●										6
Familial kidney disease					●	●	●	●													4
Rheumatoid arthritis					●	●	●	●													4
Anaemia									●	●	●										3
Systemic lupus erythematosis					●		●														2
Kidney stones					●		●														2
Antihypertensive medication																●		●			2
NSAID use					●	●															2
Hematuria																	●	●			2
Gout				●																	1
Cardiac disease				●																	1
Anti-emetic medications				●																	1
Stroke	●																				1

Table S8. Reasons for excluding studies in people with type 2 diabetes for external validation.

Study	Exclusion reason
Altemtam et al. 2011[43]	The model contains variables that are not available in DCS including uric acid levels
Basu et al. 2017[3]	The frequency of cases in the renal failure outcome was too low to perform external validation
Cheng et al. 2020[4]	The frequency of cases was too low to perform external validation
Clarke et al. 2004[44]	The algorithm for retinopathy is part of larger simulation model, and could not be applied to the Diabetes Care System cohort
Elley et al. 2013[45]	The frequency of cases was too low to perform external validation
Fernandez-Fernandez et al. 2020[29]	Multiple variables that are not available in DCS, including FE phosphate and uric acid
Goldfarb-Rumyantzev et al. 2002[9]	The model contains variables that are not available in DCS including IgG/Cr ratio
Hu et al. 2020[10]	The model contains variables that are not available in DCS including blood urea nitrogen
Imbroll et al. 2017[11]	The model contains variables that are not available in DCS including white cell count
Jardine et al. 2012[46]	The frequency of cases in the major kidney-related events outcome was too low to perform external validation
Keane et al. 2006[47]	The frequency of ESRD cases was too low in DCS to perform external validation
Li et al. 2016[14]	The frequency of ESRD cases was too low in DCS to perform external validation
Lin et al. 2017[15]	The frequency of ESRD cases was too low in DCS to perform external validation
Parrinello et al. 2016[19]	The model contains variables that are not available in DCS including alcohol consumption and physical activity
Riphagen et al. 2015[20]	The frequency of progressive renal function loss outcome was too low in DCS to perform external validation
Rodriguez-Romero et al. 2019[21]	The models contain variables that are not available in DCS
Sun et al. 2020[22]	The frequency of ESRD cases was too low in DCS to perform external validation
Tanaka et al. 2013[23]	The models contain variables that are not available in DCS
Wada et al. 2014[24]	The frequency of ESRD cases was too low in DCS to perform external validation
Wan et al. 2017[25]	The frequency of ESRD cases was too low in DCS to perform external validation
Woodward et al. 2016[26]	The frequency of ESRD cases was too low in DCS to perform external validation
Wysham et al. 2020[27]	The models contain variables that are not available in DCS
Yang et al. 2006[28]	The frequency of ESRD cases was too low in DCS to perform external validation

Table S9 Reasons for excluding studies in the general population for external validation.

Study	Exclusion reason
Chang et al. 2019[32]	The frequency of cases was too low in DCS to perform external validation
Hemmelgarn et al. 2007[48]	The models contain variables that are not available in DCS
Hippisley-Cox et al. 2010[49]	The frequency of ESRD cases was too low in DCS to perform external validation
Johnson et al. 2008[50]	The frequency of RRT cases was too low in DCS to perform external validation
Kshirsagar et al. 2008[51]	The models contain variables that are not available in DCS
Saranburut et al. 2017[39]	One of the three models could not be validated because uric acid was not available in DCS
Schroeder et al. 2017[52]	The frequency of RRT cases was too low in DCS to perform external validation
Umesawa et al. 2018[41]	The models contain variables that are not available in DCS
Wen et al. 2020[42]	Only one of the two models could be validated. The best fit model contain variables that are not available in DCS, the simple clinical model was valdiated

Table S10 Number of cases and controls used in the external validation in DCS.

Author	Type	Horizon	N_{total}	N_{event}	Definition	Group
Type 2 diabetes						
Afghahi[1]	Logistic	2	9015	1474	UACR 20-200 µg/min or UACR > 200 µg/min	Albuminuria
	Logistic	5	6649	1413		
	Logistic	10	2977	827		
Basu (Micro)[3]	Cox	2	10063	1691	UACR ≥ 300mg/g	Albuminuria
	Cox	5	7126	1760		
	Cox	10	3106	1119		
Basu (Macro)[3]	Cox	2	10063	263	UACR ≥ 300mg/g	Albuminuria
	Cox	5	7126	267		
	Cox	10	3106	173		
Jardine[12]	Cox	2	8204	221	UACR at baseline ≤30 ug/mg UACR ≥ 30ug/mg	Albuminuria
	Cox	5	6108	440		
	Cox	10	2754	421		
Riphagen[20]	Cox	2	8799	395	UACR > 2.5 mg/mmol in men and > 3.5 mg/mmol in women	Albuminuria
	Cox	5	6362	749		
	Cox	10	2796	627		
Afghahi (eCrCl)[1]	Logistic	2	9411	1765	eCrCl < 60 mL/min	DKD
	Logistic	5	6840	1657		
	Logistic	10	3022	925		
	Logistic	GoDARTS	7403	1827		
Afghahi (MDRD)[1]	Logistic	2	9603	2159	MDRD < 60 mL/min/1.73m^2	DKD
	Logistic	5	6912	2036		
	Logistic	10	3034	1205		
	Logistic	GoDARTS	7403	2883		
Basu (Composite 1)[3]	Cox	2	10063	614	Doubling serum creatinine or >20 mL/min/1.73m^2 decrease in eGFR	DKD
	Cox	5	7126	1068		
	Cox	10	3106	721		
Basu (Composite 2)[3]	Cox	2	10063	861	Macro-albuminuria, renal failure, end stage renal disease, doubling of serum creatinine or >20 mL/min/1.73m^2 decrease in eGFR	DKD
	Cox	5	7126	1262		
	Cox	10	3106	825		
Basu (Composite 3)[3]	Cox	2	10063	1693	Macroalbuminuria, microalbuminuria, renal failure, or end-stage renal disease	DKD
	Cox	5	7126	1761		
	Cox	10	3106	1120		
Dunkler (Clinical)[7]	Logistic	2	9858	1676	UACR > 30 mg/g and 300 mg/g, doubling creatinine, CKD-EPI eGFR < 15 mL/min/1.73 m^2	DKD
	Logistic	5	7061	1779		
	Logistic	10	3088	1141		
Dunkler (Laboratory)[7]	Logistic	2	9858	1676	UACR > 30 mg/g and 300 mg/g, doubling creatinine, CKD-EPI eGFR < 15 mL/min/1.73 m^2	DKD
	Logistic	5	7061	1779		
	Logistic	10	3088	1141		
Low[16]	Logistic	2	9888	278	MDRD eGFR category coupled with ≥ 25% reduction in eGFR from baseline	DKD
	Logistic	5	6980	705		
	Logistic	10	3054	613		
Nelson[18]	Cox	2	8577	810	CKD-EPI eGFR > 60 mL/min/1.73 m^2 at baseline CKD-EPI eGFR < 60 mL/min/1.73 m^2	DKD
	Cox	5	6294	1098		
	Cox	10	2810	802		
Dagliati[6]	Logistic	2	9879	3871	eGFR<60 mL/min/1.73m^2 or UACR≥30 mg/g	DKD
	Logistic	5	6978	3342		
	Logistic	10	3054	1761		
General population						
Saranburut (Clinical)[39]	Logistic	2	8577	810	CKD-EPI eGFR < 60 mL/min/1.73 m^2	CKD
	Logistic	5	6294	1098		
	Logistic	10	2810	802		
Saranburut (Laboratory)[39]	Logistic	2	8577	810	CKD-EPI eGFR < 60 mL/min/1.73 m^2	CKD
	Logistic	5	6294	1098		
	Logistic	10	2810	802		
	Logistic	GoDARTS	8552	3199		
Hanratty[33]	Cox	2	9603	2159	MDRD eGFR < 60 mL/min/1.73 m^2	CKD
	Cox	5	6912	2036		
	Cox	10	3034	1205		
	Cox	GoDARTS	8607	3210		
O'Seaghdha[38]	Cox	2	9603	2159	MDRD eGFR < 60 mL/min/1.73 m^2	CKD
	Cox	5	6912	2036		
	Cox	10	3034	1205		
	Cox	GoDARTS	5322	2883		
Wen[42]	Logistic	2	8577	810	eGFR < 60 mL/min/1.73m2 or UACR ≥30 mg/g	CKD
	Logistic	5	6294	1098		
	Logistic	10	2810	802		
Chien[31]	Cox	2	9603	2159	MDRD eGFR < 60 mL/min/1.73 m^2	CKD
	Cox	5	6912	2036		
	Cox	10	3034	1205		

In Basu et al composite outcomes were used. Composite 1, doubling serum creatinine or >20 mL/min/1.73m2 decrease in eGFR; composite 2, macro-albuminuria, renal failure, end stage renal disease, doubling of serum creatinine or >20 mL/min/1.73m2 decrease in eGFR; Composite 3, Macroalbuminuria, microalbuminuria, renal failure, or end-stage renal disease.CKD, chronic kidney disease; ESRD, end-stage renal disease.

Table S11 Comparison of discovery and external validation

Discovery	Model	Population	Horizon	Outcome	C_{Disc}	C_{Val}	C_{DCS} (2Y)	C_{DCS} (5Y)	C_{DCS} (10Y)	Reporting study
Nelson[18]	Diabetes	Diabetes	5	DKD	0.81	0.81	0.87	0.81	0.76	Nelson[18]
Dunkler[7]	Laboratory	Diabetes	5	DKD	0.68	0.68	0.65	0.68	0.69	Dunkler[7]
Dunkler[7]	Clinical	Diabetes	5	DKD	0.69	0.69	0.66	0.70	0.70	Dunkler[7]
O'Seaghdha[38]		Diabetes	5	CKD	0.79	0.76	0.74	0.76	0.77	Nelson[18]
O'Seaghdha[38]		GP	5	CKD	0.79	0.91	0.74	0.76	0.77	Fraccaro[53]
Chien[31]		Diabetes	5	CKD	0.78	0.69	0.69	0.66	0.65	Nelson[18]
Chien[31]		GP	5	CKD	0.78	0.90	0.69	0.66	0.65	Fraccaro[53]

C_{Disc}, C-statistic observed in the development study; C_{Val}, C-statistic observed in the validation study, C_{DCS}, C-statistic observed in the current study.

Table S12 Best performing models across outcomes and horizons

Outcome	2-year horizon	5-year horizon	10-year horizon
Macroalbuminuria	Basu[1]	Basu[1]	Basu[1]
DKD	Afghahi[1]	Afghahi[1]	Afghahi[1]
CKD*	Saranburut[39] Hanratty[33]	Saranburut[39] Hanratty[33]	Saranburut[39] Hanratty[33]

***** The model of Saranburut showed better calibration, while the model of Hanratty showed better discrimination.

References

1 Afghahi, H. *et al.* Risk factors for the development of albuminuria and renal impairment in type 2 diabetes--the Swedish National Diabetes Register (NDR). *Nephrol Dial Transplant* **26**, 1236-1243, doi:10.1093/ndt/gfq535 (2011).

2 Altemtam, N., Russell, J. & El Nahas, M. A study of the natural history of diabetic kidney disease (DKD). *Nephrol Dial Transplant* **27**, 1847-1854, doi:10.1093/ndt/gfr561 (2012).

3 Basu, S., Sussman, J. B., Berkowitz, S. A., Hayward, R. A. & Yudkin, J. S. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODe) using individual participant data from randomised trials. *Lancet Diabetes Endocrinol* **5**, 788-798, doi:10.1016/s2213-8587(17)30221-8 (2017).

4 Cheng, Y., Shang, J., Liu, D., Xiao, J. & Zhao, Z. Development and validation of a predictive model for the progression of diabetic kidney disease to kidney failure. *Ren Fail* **42**, 550-559, doi:10.1080/0886022x.2020.1772294 (2020).

5 Clarke, P. M. *et al.* A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia* **47**, 1747-1759, doi:10.1007/s00125-004-1527-z (2004).

6 Dagliati, A. *et al.* Machine Learning Methods to Predict Diabetes Complications. *J Diabetes Sci Technol* **12**, 295-302, doi:10.1177/1932296817706375 (2018).

7 Dunkler, D. *et al.* Risk Prediction for Early CKD in Type 2 Diabetes. *Clin J Am Soc Nephrol* **10**, 1371-1379, doi:10.2215/CJN.10321014 (2015).

8 Elley, C. R. *et al.* Derivation and validation of a renal risk score for people with type 2 diabetes. *Diabetes Care* **36**, 3113-3120, doi:10.2337/dc13-0190 (2013).

9 Goldfarb-Rumyantzev, A. S. & Pappas, L. Prediction of renal insufficiency in Pima Indians with nephropathy of type 2 diabetes mellitus. *Am J Kidney Dis* **40**, 252-264, doi:10.1053/ajkd.2002.34503 (2002).

10 Hu, Y., Shi, R., Mo, R. & Hu, F. Nomogram for the prediction of diabetic nephropathy risk among patients with type 2 diabetes mellitus based on a questionnaire and biochemical indicators: a retrospective study. *Aging (Albany NY)* **12**, 10317-10336, doi:10.18632/aging.103259 (2020).

11 Giordano Imbroll, M., Agius Lauretta, D., Tabone, T. & Fava, S. Predictors and generation of risk equations for albuminuria progression in type 2 diabetes. *Clin Nephrol* **88**, 33-39, doi:10.5414/cn109010 (2017).

12 Jardine, M. J. *et al.* Prediction of kidney-related outcomes in patients with type 2 diabetes. *Am J Kidney Dis* **60**, 770-778, doi:10.1053/j.ajkd.2012.04.025 (2012).

13 Keane, W. F. *et al.* Risk scores for predicting outcomes in patients with type 2 diabetes and nephropathy: the RENAAL study. *Clin J Am Soc Nephrol* **1**, 761-767, doi:10.2215/CJN.01381005 (2006).

14 Li, H. Y. *et al.* Serum Vascular Adhesion Protein-1 Predicts End-Stage Renal Disease in Patients with Type 2 Diabetes. *PLoS One* **11**, e0147981, doi:10.1371/journal.pone.0147981 (2016).

15 Lin, C. C. *et al.* Development and validation of a risk prediction model for end-stage renal disease in patients with type 2 diabetes. *Sci Rep* **7**, 10177, doi:10.1038/s41598-017-09243-9 (2017).

16 Low, S. *et al.* Development and validation of a predictive model for Chronic Kidney Disease progression in Type 2 Diabetes Mellitus based on a 13-year study in Singapore. *Diabetes Res Clin Pract* **123**, 49-54, doi:10.1016/j.diabres.2016.11.008 (2017).

17 Miao, D. D. *et al.* Development and Validation of a Model for Predicting Diabetic Nephropathy in Chinese People. *Biomed Environ Sci* **30**, 106-112, doi:10.3967/bes2017.014 (2017).

18 Nelson, R. G. *et al.* Development of Risk Prediction Equations for Incident Chronic Kidney Disease. *Jama* **322**, 2104-2114, doi:10.1001/jama.2019.17379 (2019).

19 Parrinello, C. M. *et al.* Risk prediction of major complications in individuals with diabetes: the Atherosclerosis Risk in Communities Study. *Diabetes Obes Metab* **18**, 899-906, doi:10.1111/dom.12686 (2016).

20 Riphagen, I. J. *et al.* Comparison of methods for renal risk prediction in patients with type 2 diabetes (ZODIAC-36). *PLoS One* **10**, e0120477, doi:10.1371/journal.pone.0120477 (2015).

21 Rodriguez-Romero, V. *et al.* Prediction of Nephropathy in Type 2 Diabetes: An Analysis of the ACCORD Trial Applying Machine Learning Techniques. *Clin Transl Sci* **12**, 519-528, doi:10.1111/cts.12647 (2019).

22 Sun, L., Shang, J., Xiao, J. & Zhao, Z. Development and validation of a predictive model for end-stage renal disease risk in patients with diabetic nephropathy confirmed by renal biopsy. *PeerJ* **8**, e8499, doi:10.7717/peerj.8499 (2020).
23 Tanaka, S. *et al.* Predicting macro- and microvascular complications in type 2 diabetes: the Japan Diabetes Complications Study/the Japanese Elderly Diabetes Intervention Trial risk engine. *Diabetes Care* **36**, 1193-1199, doi:10.2337/dc12-0958 (2013).
24 Wada, T. *et al.* Clinical impact of albuminuria and glomerular filtration rate on renal and cardiovascular events, and all-cause mortality in Japanese patients with type 2 diabetes. *Clin Exp Nephrol* **18**, 613-620, doi:10.1007/s10157-013-0879-4 (2014).
25 Wan, E. Y. F. *et al.* Prediction of new onset of end stage renal disease in Chinese patients with type 2 diabetes mellitus - a population-based retrospective cohort study. *BMC Nephrol* **18**, 257, doi:10.1186/s12882-017-0671-x (2017).
26 Woodward, M. *et al.* Prediction of 10-year vascular risk in patients with diabetes: the AD-ON risk score. *Diabetes Obes Metab* **18**, 289-294, doi:10.1111/dom.12614 (2016).
27 Wysham, C. H. *et al.* Development of risk models for major adverse chronic renal outcomes among patients with type 2 diabetes mellitus using insurance claims: a retrospective observational study. *Curr Med Res Opin* **36**, 219-227, doi:10.1080/03007995.2019.1682981 (2020).
28 Yang, X. L. *et al.* End-stage renal disease risk equations for Hong Kong Chinese patients with type 2 diabetes: Hong Kong Diabetes Registry. *Diabetologia* **49**, 2299-2308, doi:10.1007/s00125-006-0376-3 (2006).
29 Fernandez-Fernandez, B. *et al.* Gender, Albuminuria and Chronic Kidney Disease Progression in Treated Diabetic Kidney Disease. *J Clin Med* **9**, doi:10.3390/jcm9061611 (2020).
30 Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet* **49**, 1385-1391, doi:10.1038/ng.3913 (2017).
31 Chien, K. L. *et al.* A prediction model for the risk of incident chronic kidney disease. *Am J Med* **123**, 836-846 e832, doi:10.1016/j.amjmed.2010.05.010 (2010).
32 Chang, H. L. *et al.* A predictive model for progression of CKD. *Medicine (Baltimore)* **98**, e16186, doi:10.1097/md.0000000000016186 (2019).
33 Hanratty, R. *et al.* Incident chronic kidney disease and the rate of kidney function decline in individuals with hypertension. *Nephrol Dial Transplant* **25**, 801-807, doi:10.1093/ndt/gfp534 (2010).
34 Hemmelgarn, B. R., Culleton, B. F. & Ghali, W. A. Derivation and validation of a clinical index for prediction of rapid progression of kidney dysfunction. *QJM* **100**, 87-92, doi:10.1093/qjmed/hcm001 (2007).
35 Hippisley-Cox, J. & Coupland, C. Predicting the risk of chronic Kidney Disease in men and women in England and Wales: prospective derivation and external validation of the QKidney Scores. *BMC Fam Pract* **11**, 49, doi:10.1186/1471-2296-11-49 (2010).
36 Johnson, E. S., Thorp, M. L., Platt, R. W. & Smith, D. H. Predicting the risk of dialysis and transplant among patients with CKD: a retrospective cohort study. *Am J Kidney Dis* **52**, 653-660, doi:10.1053/j.ajkd.2008.04.026 (2008).
37 Kshirsagar, A. V. *et al.* A simple algorithm to predict incident kidney disease. *Arch Intern Med* **168**, 2466-2473, doi:10.1001/archinte.168.22.2466 (2008).
38 O'Seaghdha, C. M. *et al.* A risk score for chronic kidney disease in the general population. *Am J Med* **125**, 270-277, doi:10.1016/j.amjmed.2011.09.009 (2012).
39 Saranburut, K. *et al.* Risk scores to predict decreased glomerular filtration rate at 10 years in an Asian general population. *BMC Nephrol* **18**, 240, doi:10.1186/s12882-017-0653-z (2017).
40 Schroeder, E. B. *et al.* Predicting 5-Year Risk of RRT in Stage 3 or 4 CKD: Development and External Validation. *Clin J Am Soc Nephrol* **12**, 87-94, doi:10.2215/CJN.01290216 (2017).
41 Umesawa, M. *et al.* Validity of a Risk Prediction Equation for CKD After 10 Years of Follow-up in a Japanese Population: The Ibaraki Prefectural Health Study. *Am J Kidney Dis* **71**, 842-850, doi:10.1053/j.ajkd.2017.09.013 (2018).
42 Wen, J. *et al.* Risk scores for predicting incident chronic kidney disease among rural Chinese people: a village-based cohort study. *BMC Nephrol* **21**, 120, doi:10.1186/s12882-020-01787-9 (2020).

43 Altemtam, N., Russell, J. & El Nahas, M. A study of the natural history of diabetic kidney disease (DKD). *Nephrology Dialysis Transplantation* **27**, 1847-1854 (2011).
44 Clarke, P. *et al.* A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia* **47**, 1747-1759 (2004).
45 Elley, C. R. *et al.* Derivation and validation of a renal risk score for people with type 2 diabetes. *Diabetes Care* **36**, 3113-3120 (2013).
46 Jardine, M. J. *et al.* Prediction of kidney-related outcomes in patients with type 2 diabetes. *American journal of kidney diseases* **60**, 770-778 (2012).
47 Keane, W. F. *et al.* Risk scores for predicting outcomes in patients with type 2 diabetes and nephropathy: the RENAAL study. *Clin J Am Soc Nephrol* **1**, 761-767, doi:10.2215/cjn.01381005 (2006).
48 Hemmelgarn, B. R., Culleton, B. F. & Ghali, W. A. Derivation and validation of a clinical index for prediction of rapid progression of kidney dysfunction. *QJM : monthly journal of the Association of Physicians* **100**, 87-92, doi:10.1093/qjmed/hcm001 (2007).
49 Hippisley-Cox, J. & Coupland, C. Predicting the risk of chronic Kidney Disease in men and women in England and Wales: prospective derivation and external validation of the QKidney Scores. *BMC family practice* **11**, 49-49, doi:10.1186/1471-2296-11-49 (2010).
50 Johnson, E. S., Thorp, M. L., Platt, R. W. & Smith, D. H. J. A. j. o. k. d. Predicting the risk of dialysis and transplant among patients with CKD: a retrospective cohort study. **52**, 653-660 (2008).
51 Kshirsagar, A. V. *et al.* A simple algorithm to predict incident kidney disease. **168**, 2466-2473 (2008).
52 Schroeder, E. B. *et al.* Predicting 5-year risk of RRT in stage 3 or 4 CKD: development and external validation. **12**, 87-94 (2017).
53 Fraccaro, P. *et al.* An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK. *BMC Med* **14**, 104, doi:10.1186/s12916-016-0650-2 (2016).