Figure S1. Relationship between model sparsity and model fit. For the complete GCKD cohort, the partial likelihood deviance (PLD) assessed in an inner five-fold cross-validation is shown in relationship to the hyperparameter λ (bottom) and the number of predictive features included in the respective model (top). The left dotted line indicates the λ value for which the smallest average PLD is obtained, whereas the right dotted line indicates the largest value of λ for which the PLD remains within one standard deviation of the minimal PLD. The former one was chosen for the Z14 model and the latter one for the sparser, more regularized Z6 model.

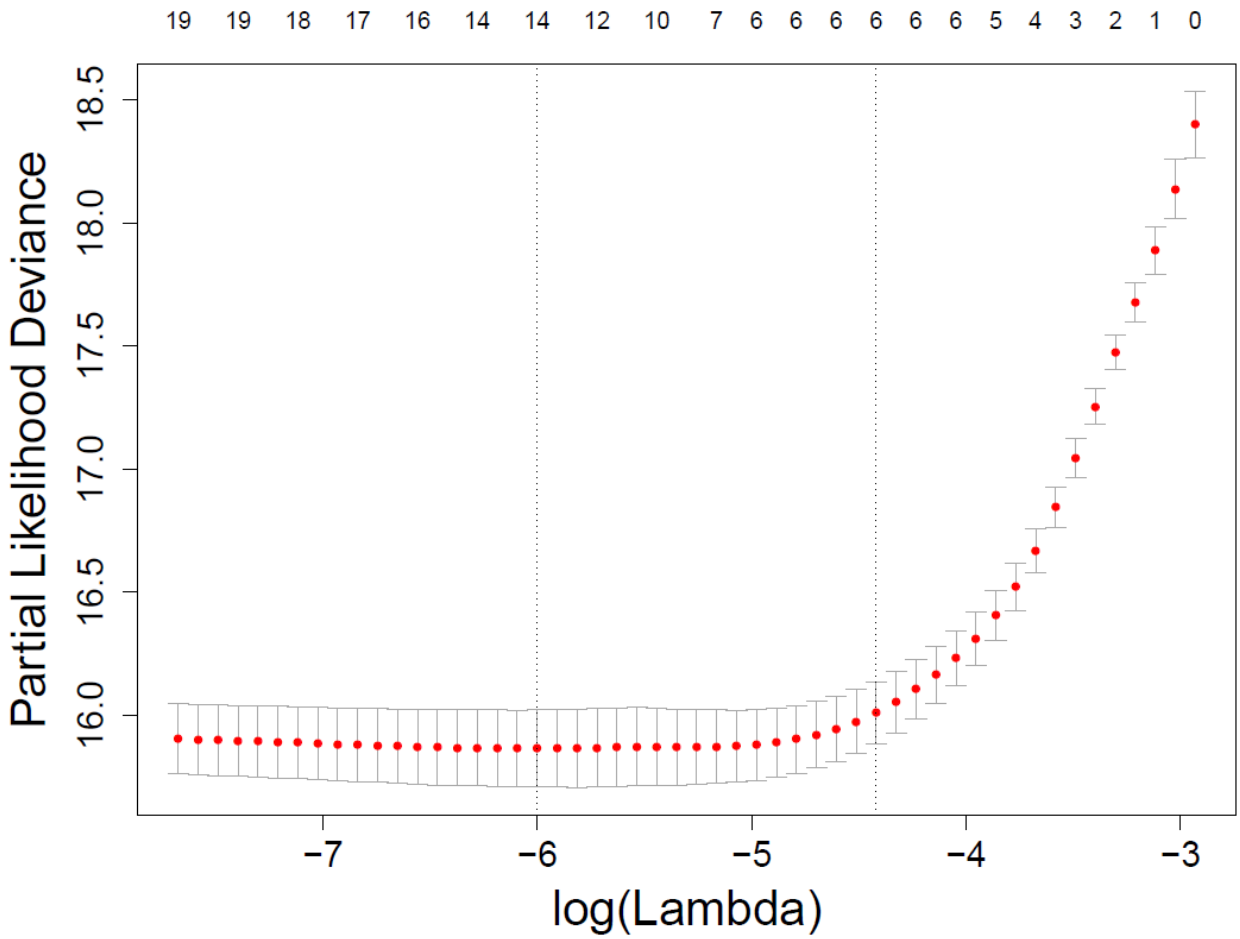


Figure S2. Kaplan-Meier curve for time to first ESKD event or death and cumulative incidence curves of ESKD and death events for the 4,915 GCKD patients included in the present study. (A) Kaplan-Meier curve for time to first ESKD event (initiation of dialysis or renal transplantation) or death, whichever occurred first. The lower table gives the number of individuals at risk and the number of incident events in brackets for each follow-up year. Note that patients were censored at last available follow-up. (B) Cumulative incidence curves for ESKD and death events, respectively.

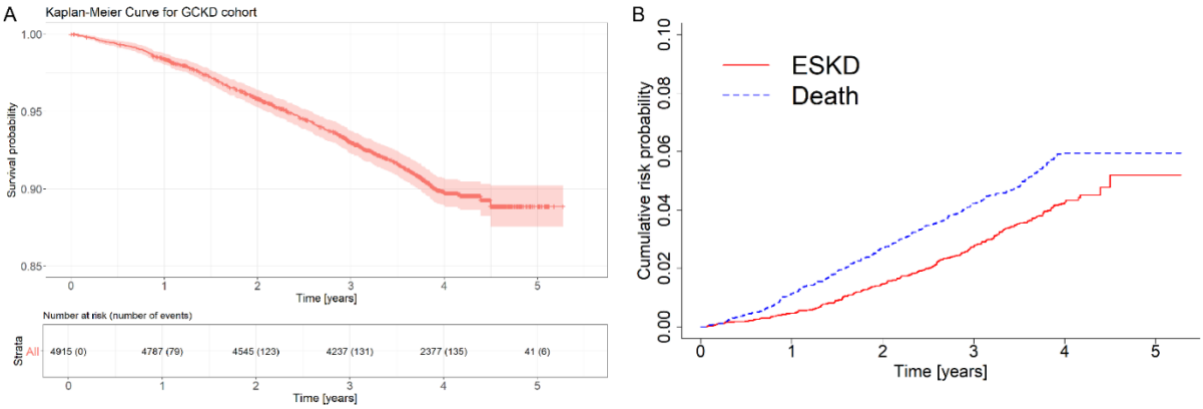


Figure S3. Estimated LASSO coefficients of the (A,B) $\lambda_{PLD\text{-}min}$ and (C,D) $\lambda_{PLD\text{-}1sd}$ risk scores summarized across 100 subsampling runs for (A,C) unstandardized and (B,D) standardized $\log_2$-transformed variables, respectively. Percentages indicate the selection frequency for the respective variable over 100 subsampling runs. Variables are ordered from left to right according to their appearance in the LASSO variable path.

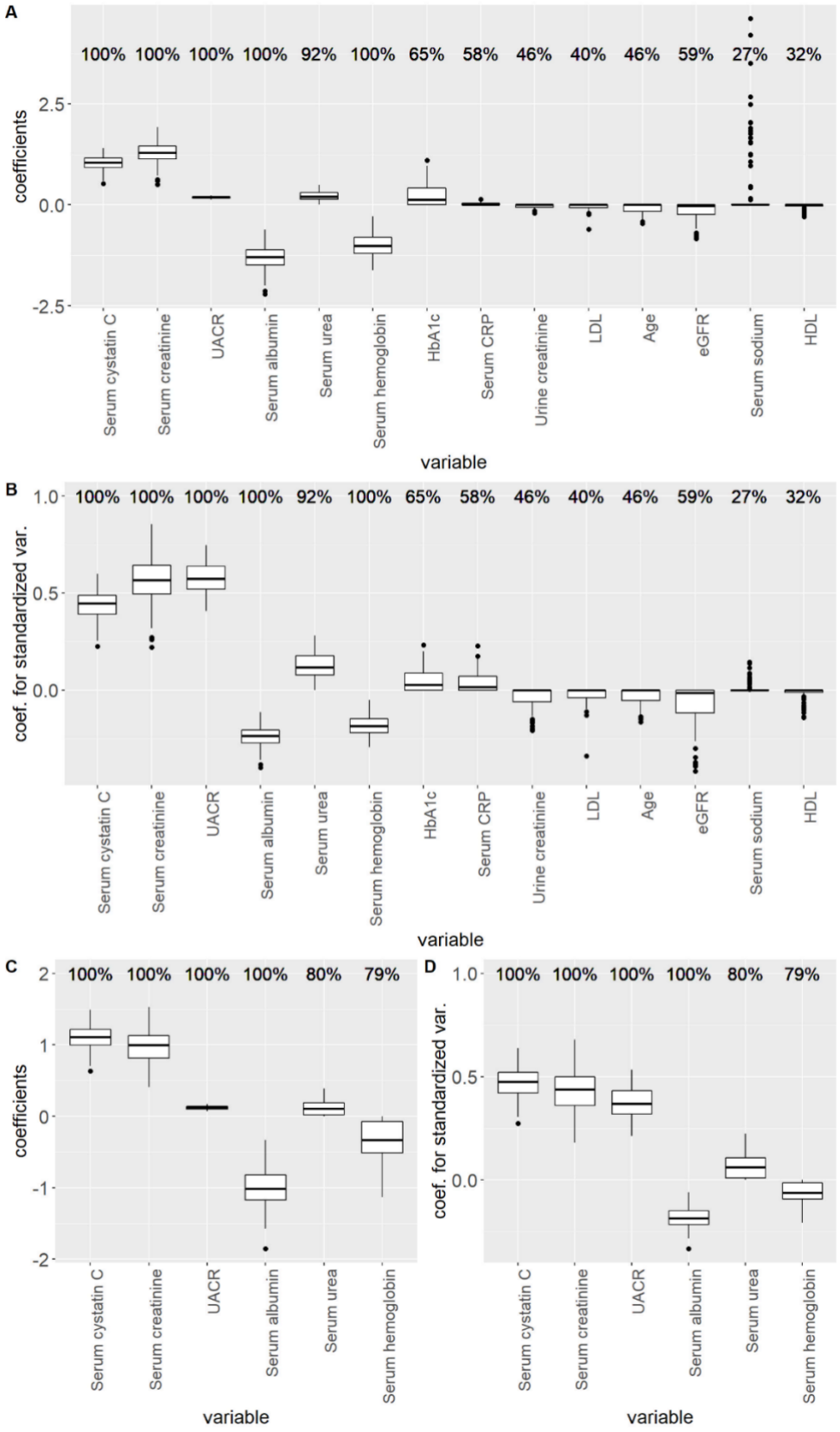


Figure S4. (A) Predictive performances of the 4-variable Tangri risk model without (T4) and with coefficient refitting by Cox PH regression on the GCKD training data sets (T4-coef-refit) and the newly developed risk models after regularization parameter optimization within an inner 5-fold cross-validation ($\lambda_{PLD\text{-}min}$ and $\lambda_{PLD\text{-}1sd}$), an inner 10-fold cross-validation ($\lambda_{PLD\text{-}min\text{-}10foldCV}$ and $\lambda_{PLD\text{-}1sd\text{-}10foldCV}$), as well as after regularization parameter optimization with respect to the *C* statistic within an inner 5-fold cross-validation ($\lambda_{C\text{-}max}$ and $\lambda_{C\text{-}1sd}$) in the GCKD resampling approach. The ESKD-specific *C* statistic treating death as a competing risk was computed for the test sets only and displayed in boxplots. (B) *C* index difference distributions in comparison to the T4. (C) *C* index difference distributions in comparison to the T4-coef-refit.

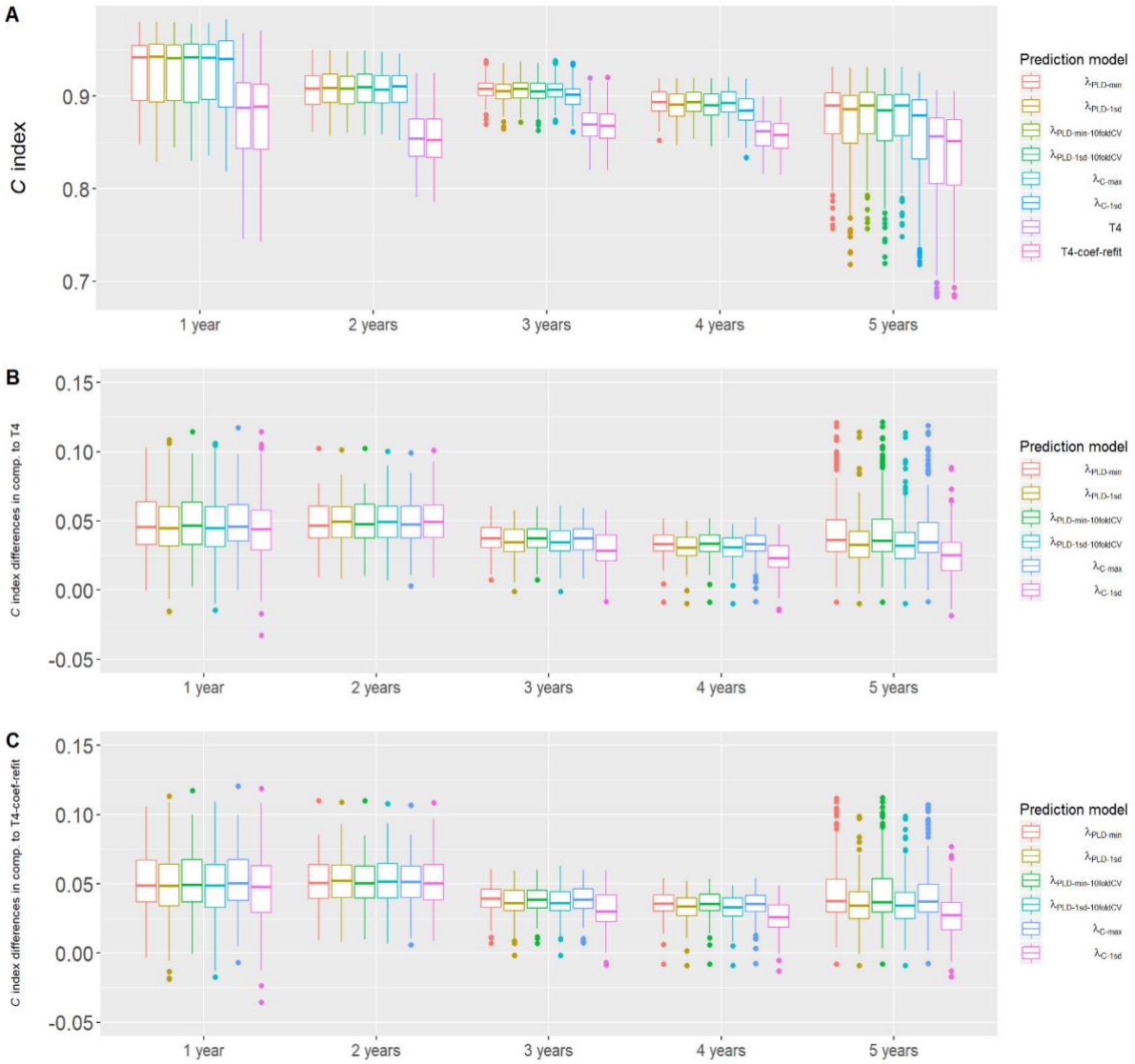


Figure S5. CKD stage-dependent performance of the $\lambda_{PLD\text{-}min}$, $\lambda_{PLD\text{-}1sd}$, and T4 models in the GCKD resampling approach. $\lambda_{PLD\text{-}min}$ and $\lambda_{PLD\text{-}1sd}$ models were trained on each complete GCKD training data set, and the ESKD-specific *C* statistic treating death as a competing risk was computed separately for CKD stage G1-2 (eGFR ≥ 60mL/min per 1.73m^2) and G3-5 (eGFR<60mL/min per 1.73m^2) patients in each test set and summarized across 100 resampling runs. Patients were assigned to CKD stages G1-2 and G3-5, respectively, based on estimates of GFR using the CKD-EPI equation. Of note, for CKD stages G1-2, *C* indices could only be evaluated for 13, 47, 70, and 98 out of 100 runs at one, two, three, and four years after the baseline visit, as no patients experiencing an ESKD event were available for the remaining test runs.

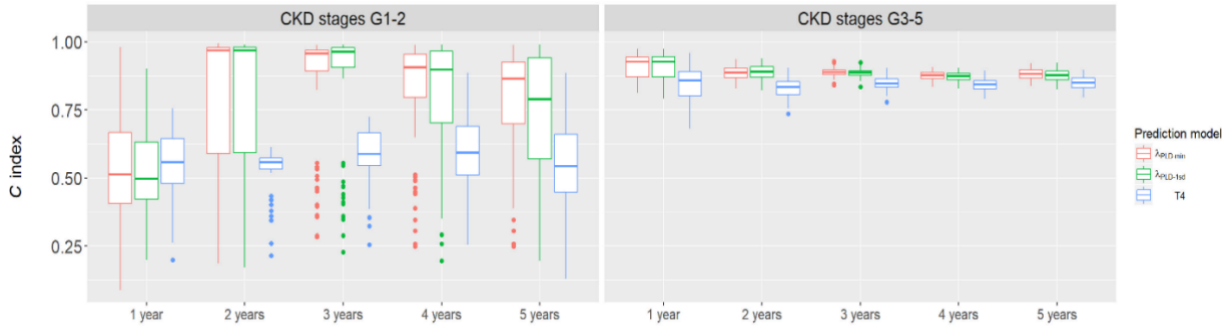


Figure S6. Distribution of categorical net reclassification improvement values in the GCKD test data sets comparing from left to right: the $\lambda_{PLD\text{-}min}$ *vs.* the GCKD-recalibrated T4 (T4-surv-recal), the $\lambda_{PLD\text{-}1sd}$ *vs.* the T4-surv-recal, the $\lambda_{PLD\text{-}min}$ *vs.* the original T4, the $\lambda_{PLD\text{-}1sd}$ *vs.* the original T4, and the $\lambda_{PLD\text{-}1sd}$ *vs.* the $\lambda_{PLD\text{-}min}$ risk equation evaluated (A) one, (B) two, (C) three, (D) four, and (E) five years after the baseline visit. Please note that at five years after the baseline visit, a high censoring rate is apparent in the Kaplan-Meier curve in Figure S2A, which impairs accurate estimation of absolute risk probabilities. Boxplots display the distribution of NRI values evaluated for the test sets over the 100 subsampling runs. The predicted risk probabilities were divided into 3 categories: 0% - <3%, 3% - <10%, 10% - 100%, respectively. Abbreviations: $\lambda_{PLD\text{-}min}$, LASSO Cox PH risk equation with λ parameter optimization to yield the minimum partial likelihood deviance (PLD) and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; $\lambda_{PLD\text{-}1sd}$, LASSO Cox PH risk equation with λ parameter optimization to yield maximum penalty while simultaneously keeping the PLD within one standard deviation of the minimum PLD and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; NRI, net reclassification improvement; NRI+, net reclassification improvement considering only patients with an event during the observation period; NRI-, net reclassification improvement considering only patients without an event during the observation period; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets.

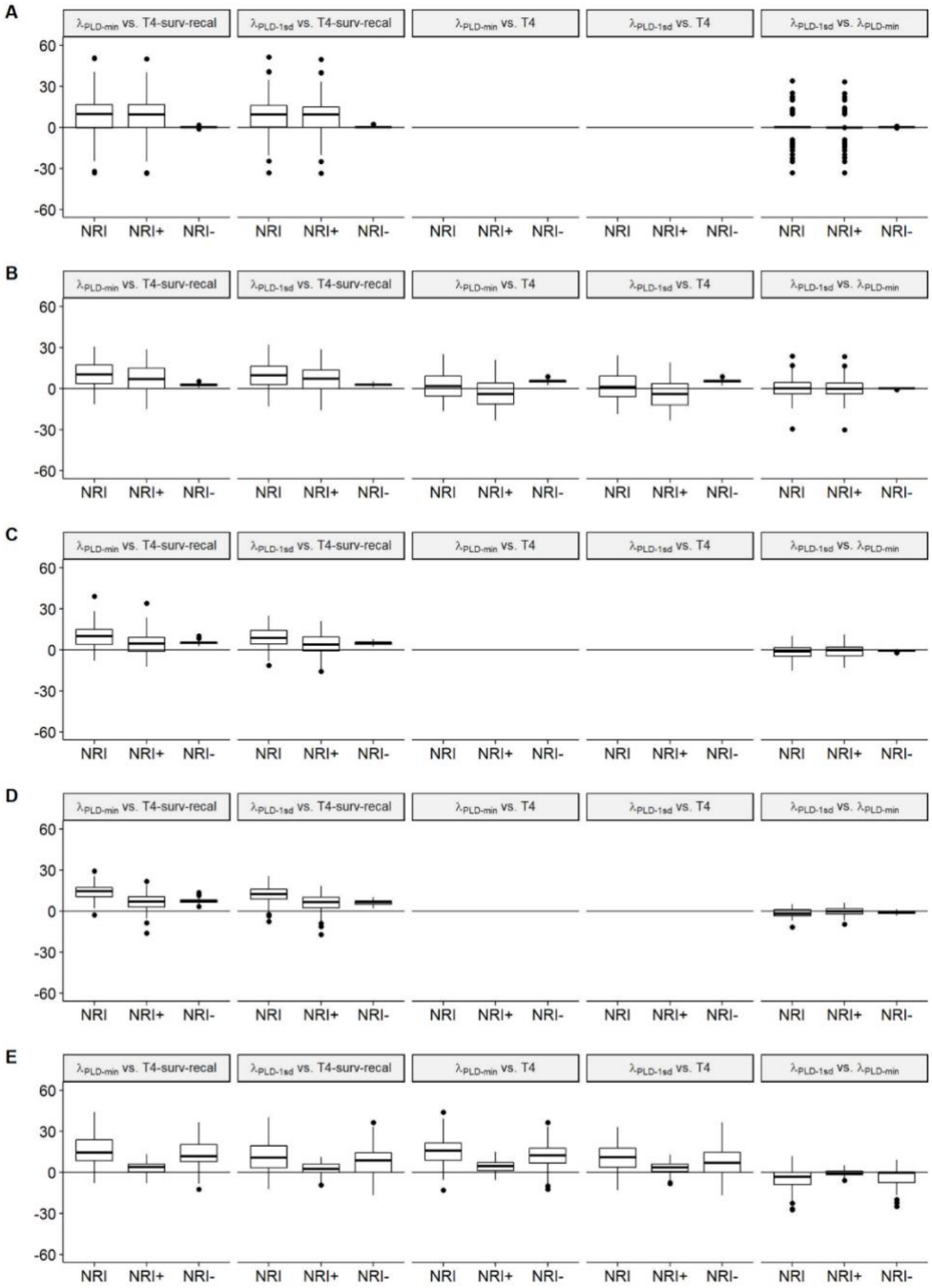


Figure S7. Distribution of continuous (category-free) net reclassification improvement (cNRI) values in the GCKD test data sets comparing from left to right: the $\lambda_{PLD\text{-}min}$ *vs.* the GCKD-recalibrated T4 (T4-surv-recal), the $\lambda_{PLD\text{-}1sd}$ *vs.* the T4-surv-recal, the $\lambda_{PLD\text{-}min}$ *vs.* the original T4, the $\lambda_{PLD\text{-}1sd}$ *vs.* the original T4 and the $\lambda_{PLD\text{-}1sd}$ *vs.* the $\lambda_{PLD\text{-}min}$ risk equation evaluated (A) one, (B) two, (C) three, (D) four, and (E) five years after the baseline visit. Please note that at five years after the baseline visit, a high censoring rate is apparent in the Kaplan-Meier curve in Figure S2A, which impairs accurate estimation of absolute risk probabilities. Boxplots display the distribution of cNRI values evaluated for the test sets over 100 subsampling runs. Abbreviations: cNRI, continuous net reclassification improvement; cNRI+, continuous net reclassification improvement considering only patients with an event during the observation period; cNRI-, continuous net reclassification improvement considering only patients without an event during the observation period; $\lambda_{PLD\text{-}min}$, LASSO Cox PH risk equation with λ parameter optimization to yield the minimum partial likelihood deviance (PLD) and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; $\lambda_{PLD\text{-}1sd}$, LASSO Cox PH risk equation with λ parameter optimization to yield maximum penalty while simultaneously keeping the PLD within one standard deviation of the minimum PLD and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets.

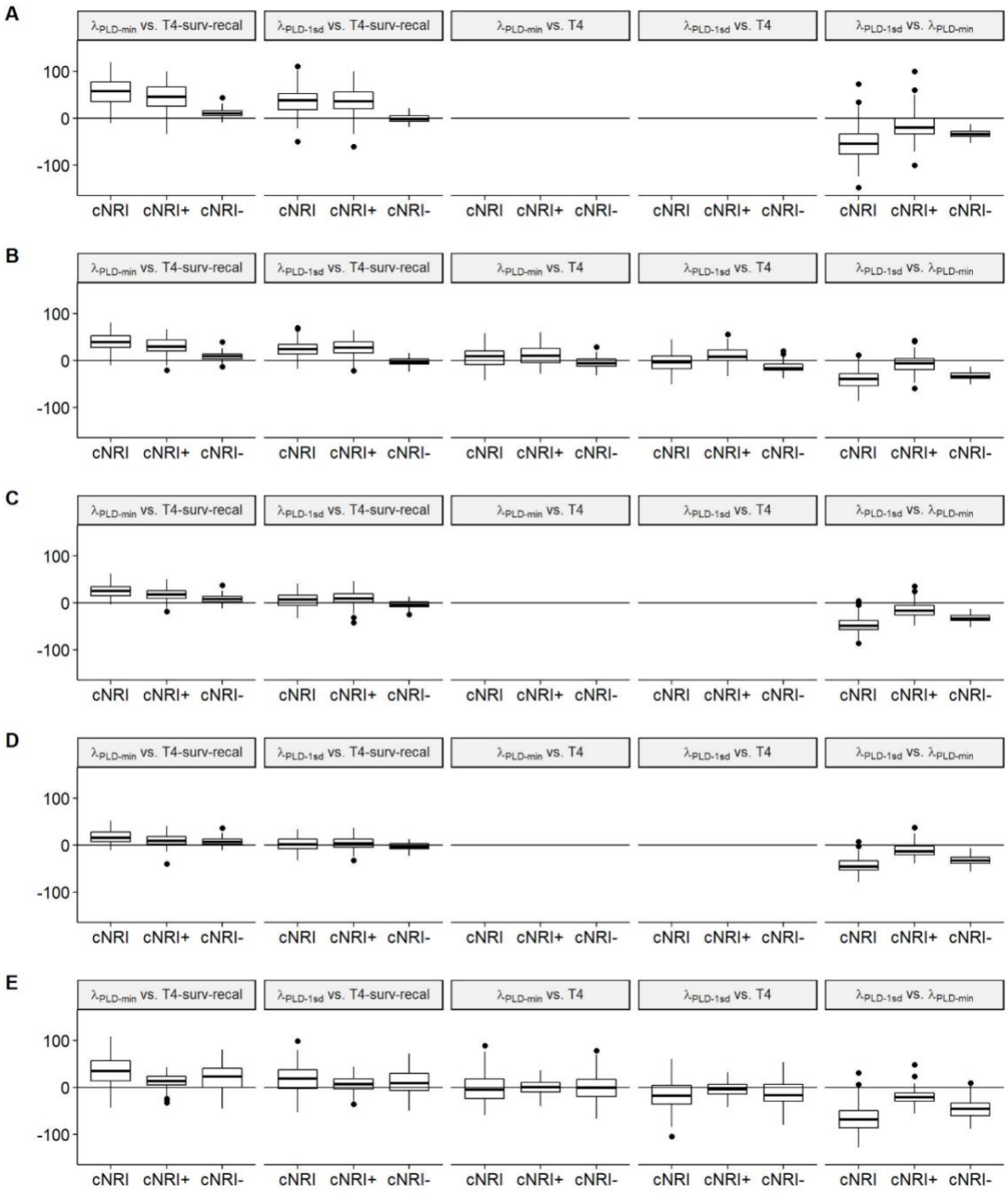


Figure S8. Receiver operating characteristic (ROC) curves for the $\lambda_{PLD\text{-}min}$ (red), the $\lambda_{PLD\text{-}1sd}$ (green), and the T4 score (blue) evaluated on the GCKD test data sets at (A,B) one, (C,D) two, (E,F) three, (G,H) four, and (I,J) five years after the baseline visit. The "best" and the "Youden" cut-off were individually optimized for each observation period, risk score, and test set, respectively, and then averaged across 100 randomly selected test sets. The means ± standard deviations of the "best" cut-off are displayed in (A), (C), (E), (G), and (I), whereas the corresponding values of the "Youden" cut-off are displayed in (B), (D), (F), (H), and (J). Medians and 95% confidence intervals of the areas under the ROC curves for the $\lambda_{PLD\text{-}min}$ (red), $\lambda_{PLD\text{-}1sd}$ (green), and T4 score (blue) were determined empirically on 100 GCKD test sets and are given in the lower right corners of the plots. Please note that at five years after the baseline visit, a high censoring rate is apparent in the Kaplan-Meier curve in Figure S2A, which affects the ROC curve evaluation. On average, 33, 99, 182, 773, and 1560 GCKD study participants had to be excluded from the ROC curve analyses due to censoring or death one, two, three, four, and five years after the baseline visit, respectively.

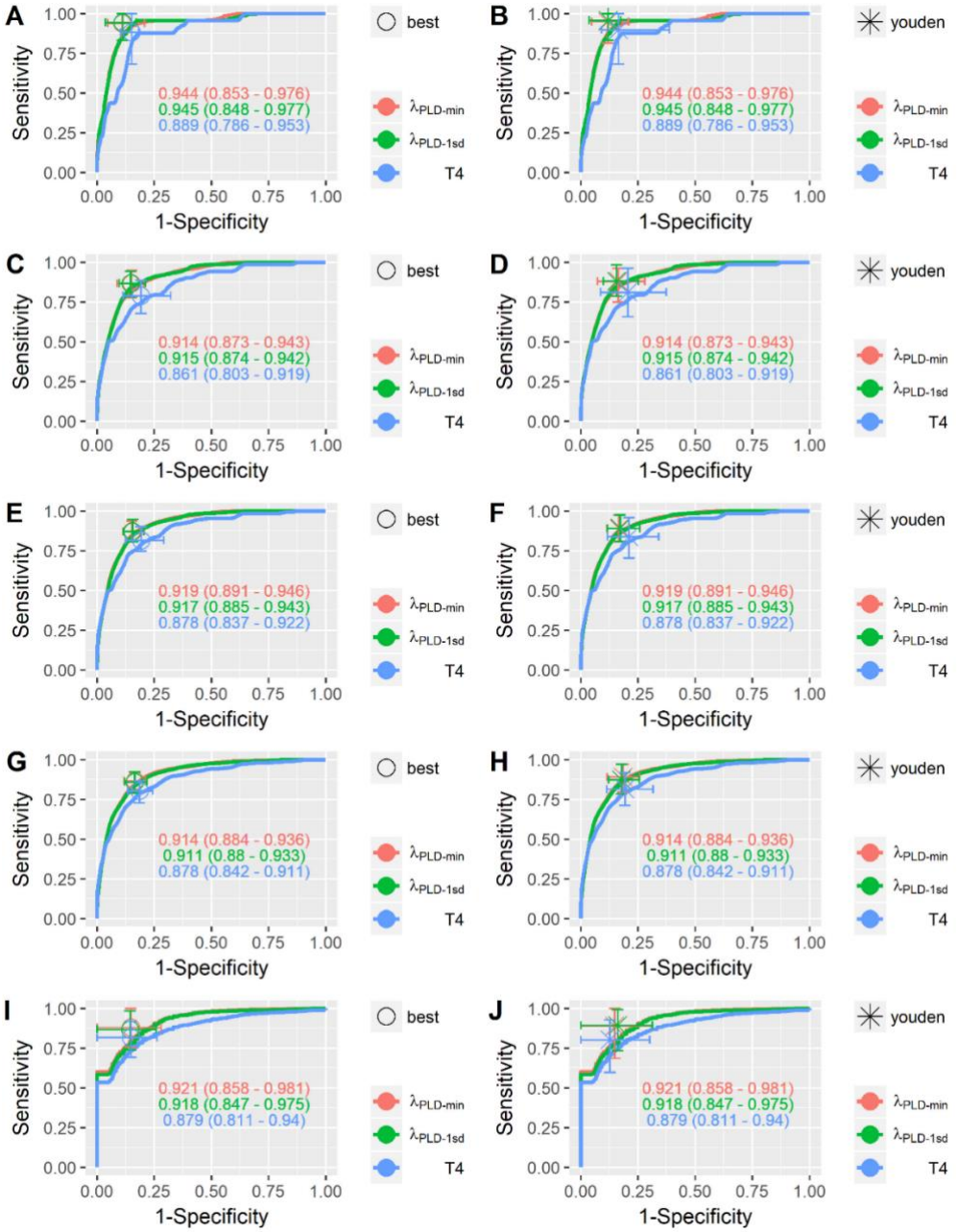


Figure S9. Predictive performances of the $\lambda_{PLD\text{-}min}$, $\lambda_{PLD\text{-}1sd}$, and T4 models in terms of ESKD-specific *C* statistics treating death as a competing risk upon additionally censoring at the date of non-fatal cardiovascular events in the GCKD subsampling approach. To this end, the following non-fatal cardiovascular events were censored in the test cohort: myocardial infarction, non-hemorrhagic stroke, and coronary revascularization procedures. The *C* statistic was computed for the test sets only and displayed in boxplots.

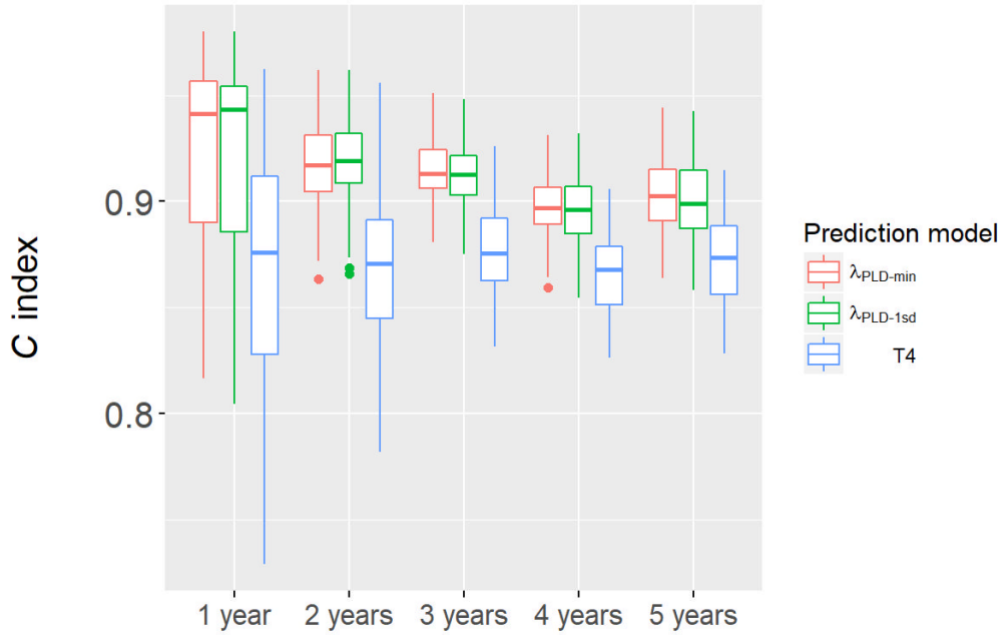


Figure S10. Distribution of categorical net reclassification improvement values in the GCKD test data sets additionally censoring at the date of non-fatal cardiovascular events comparing from left to right: the $\lambda_{PLD\text{-}min}$ *vs.* the GCKD-recalibrated T4 (T4-surv-recal), the $\lambda_{PLD\text{-}1sd}$ *vs.* the T4-surv-recal, the $\lambda_{PLD\text{-}min}$ *vs.* the original T4, the $\lambda_{PLD\text{-}1sd}$ *vs.* the original T4, and the $\lambda_{PLD\text{-}1sd}$ *vs.* the $\lambda_{PLD\text{-}min}$ risk equation evaluated (A) one, (B) two, (C) three, (D) four, and (E) five years after the baseline visit. Please note that the results at five years after the baseline visit are strongly compromised by the high censoring rate at that time-point. Boxplots display the distribution of NRI values evaluated for the test sets across the 100 subsampling runs. The predicted risk probabilities were divided into 3 categories: 0% - <3%, 3% - <10%, 10% - 100%, respectively. This analysis did not comprise a training of new predictive models, but only involved a re-evaluation of the test data and a re-calibration of the risk equations on the respective training set now additionally censoring at the date of non-fatal cardiovascular events. The following non-fatal cardiovascular events were censored: myocardial infarction, non-hemorrhagic stroke, and coronary revascularization procedures. Abbreviations: $\lambda_{PLD\text{-}min}$, LASSO Cox PH risk equation with λ parameter optimization to yield the minimum partial likelihood deviance (PLD) and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; $\lambda_{PLD\text{-}1sd}$, LASSO Cox PH risk equation with λ parameter optimization to yield maximum penalty while simultaneously keeping the PLD within one standard deviation of the minimum PLD and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; NRI, net reclassification improvement; NRI+, net reclassification improvement considering only patients with an event during the observation period; NRI-, net reclassification improvement considering only patients without an event during the observation period; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets.

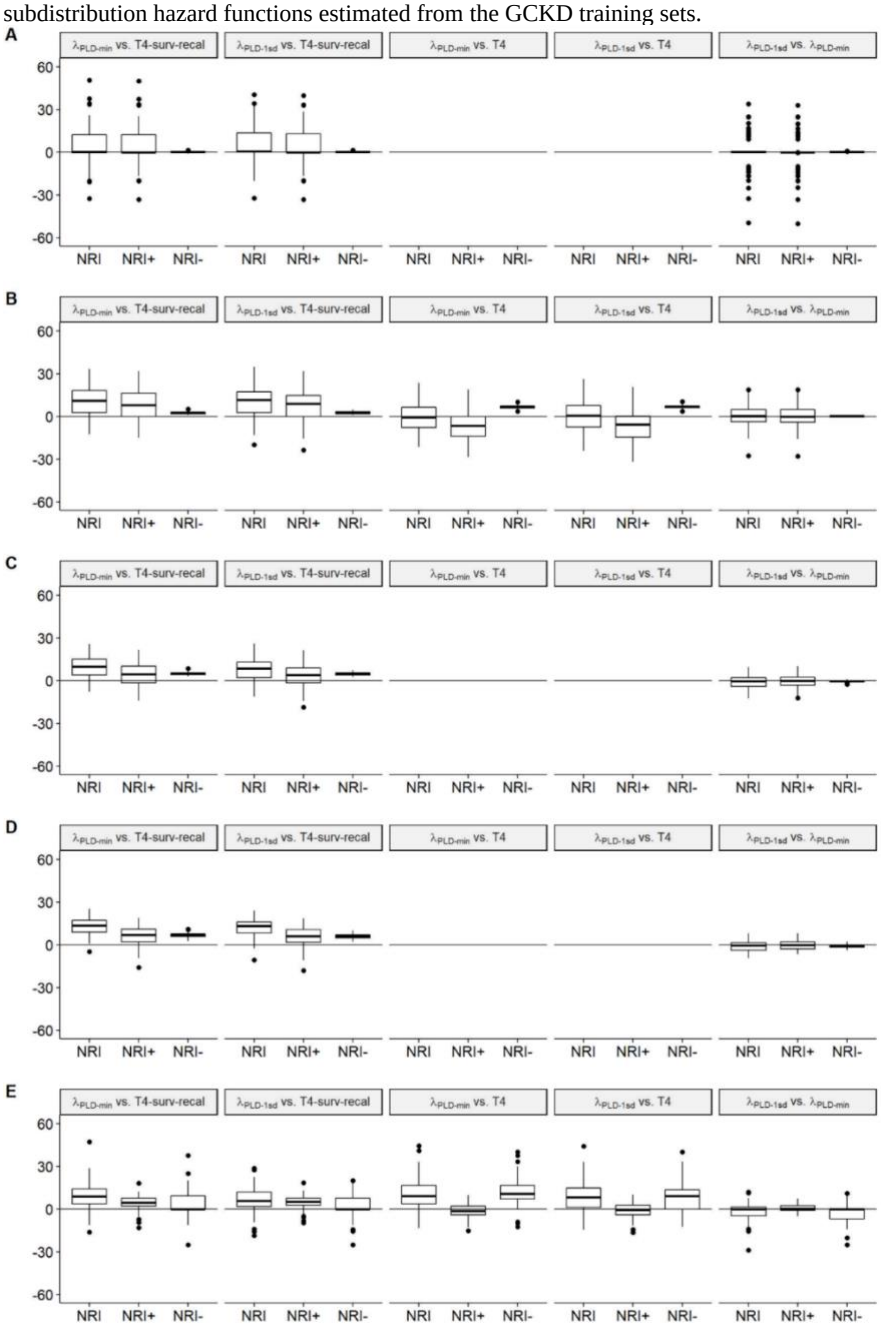


Figure S11. Distribution of continuous (category-free) net reclassification improvement (cNRI) values in the GCKD test data sets additionally censoring at the date of non-fatal cardiovascular events comparing from left to right: the $\lambda_{PLD\text{-}min}$ *vs.* the GCKD-recalibrated T4 (T4-surv-recal), the $\lambda_{PLD\text{-}1sd}$ *vs.* the T4-surv-recal, the $\lambda_{PLD\text{-}min}$ *vs.* the original T4, the $\lambda_{PLD\text{-}1sd}$ *vs.* the original T4, and the $\lambda_{PLD\text{-}1sd}$ *vs.* the $\lambda_{PLD\text{-}min}$ risk equation evaluated (A) one, (B) two, (C) three, (D) four, and (E) five years after the baseline visit. Please note that the results at five years after the baseline visit are strongly compromised by the high censoring rate at that time-point. Boxplots display the distribution of cNRI values evaluated for the test sets across the 100 subsampling runs. This analysis did not comprise a training of new predictive models, but only involved a re-evaluation of the test data and a re-calibration of the risk equations on the respective training set now additionally censoring at the date of non-fatal cardiovascular events. The following non-fatal cardiovascular events were censored: myocardial infarction, non-hemorrhagic stroke, and coronary revascularization procedures. Abbreviations: $\lambda_{PLD\text{-}min}$, LASSO Cox PH risk equation with λ parameter optimization to yield the minimum partial likelihood deviance (PLD) and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; $\lambda_{PLD\text{-}1sd}$, LASSO Cox PH risk equation with λ parameter optimization to yield maximum penalty while simultaneously keeping the PLD within one standard deviation of the minimum PLD and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; cNRI, continuous net reclassification improvement; cNRI+, continuous net reclassification improvement considering only patients with an event during the observation period; cNRI-, continuous net reclassification improvement considering only patients without an event during the observation period; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets.

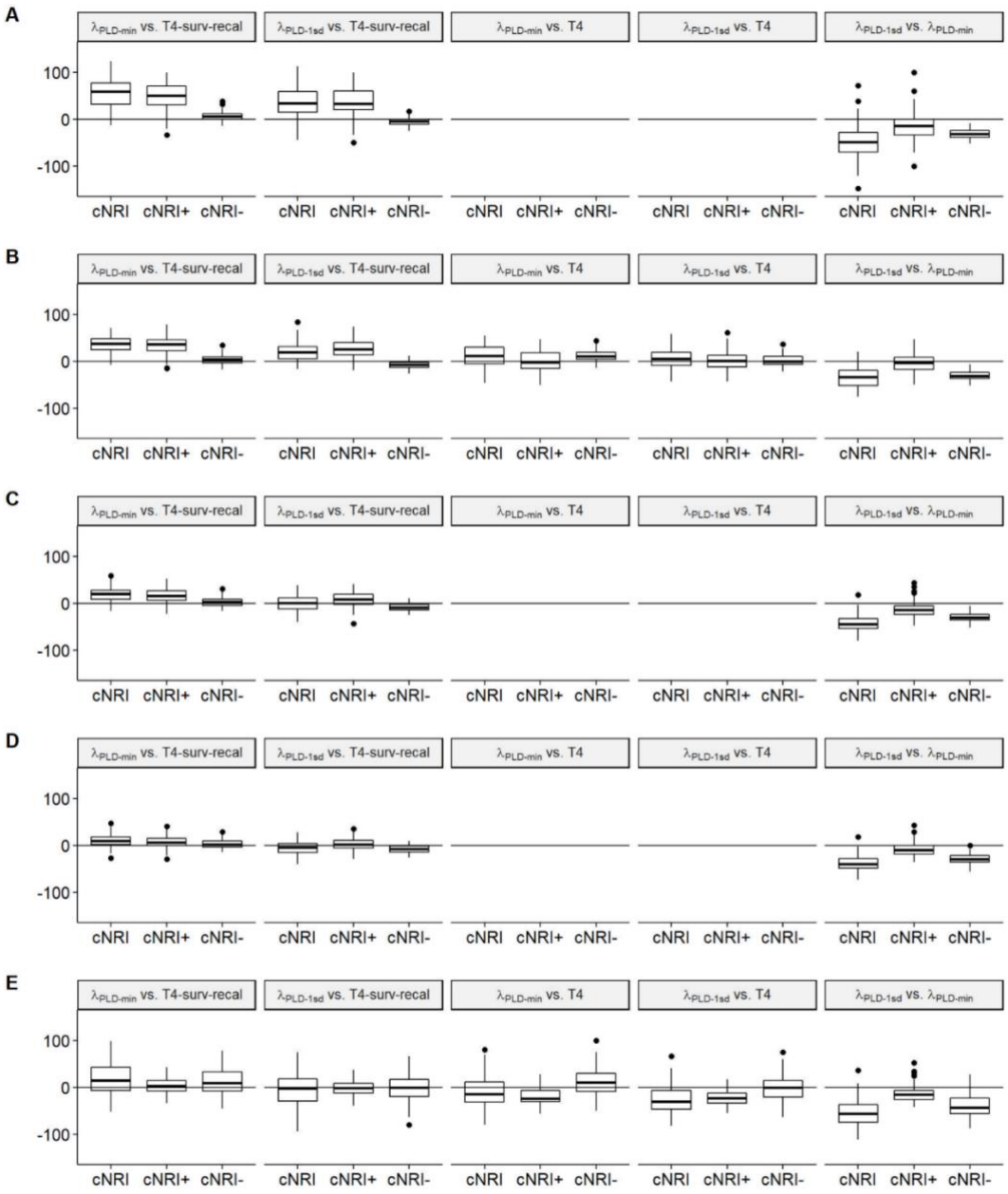


Figure S12. Receiver operating characteristic (ROC) curves for the $\lambda_{PLD\text{-}min}$ (red), $\lambda_{PLD\text{-}1sd}$ (green), and T4 score (blue) evaluated on GCKD test data sets at (A,B) one, (C,D) two, (E,F) three, (G,H) four, and (I,J) five years, after the baseline visit additionally censoring at the date of non-fatal cardiovascular events. We additionally excluded from the ROC curve evaluation those patients, who had been censored due to a cardiovascular event before the end of the respective observation period. The following non-fatal cardiovascular events were censored: myocardial infarction, non-hemorrhagic stroke, and coronary revascularization procedures. The "best" and the "Youden" cut-off were individually optimized for each observation period, risk score, and test set, respectively, and then averaged across 100 test sets. The means ± standard deviations of the "best" cut-off are displayed in (A), (C), (E), (G), and (I), whereas the corresponding values of the "Youden" cut-off are displayed in (B), (D), (F), (H), and (J). Medians and 95% confidence intervals of the areas under the ROC curves for the $\lambda_{PLD\text{-}min}$ (red), $\lambda_{PLD\text{-}1sd}$ (green), and T4 score (blue) were determined empirically on 100 GCKD test sets and are given in the lower right corners of the plots. Please note that at five years after the baseline visit, a high censoring rate affects the ROC curve evaluation. On average, 58, 145, 248, 827, and 1566 GCKD study participants had to be excluded from the ROC curve analyses due to censoring or death one, two, three, four, and five years after the baseline visit, respectively.

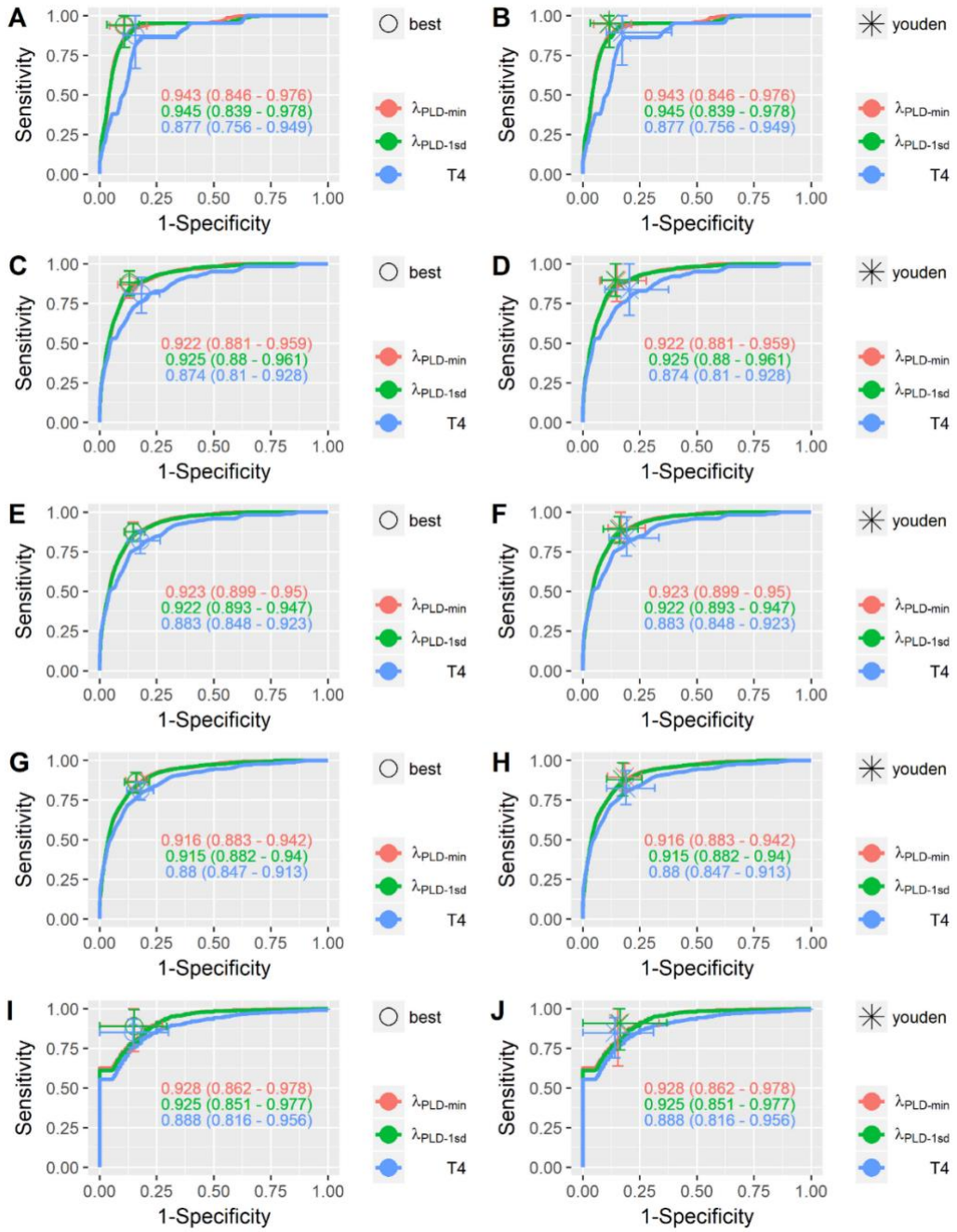


Figure S13. Distribution of Brier scores evaluated on the GCKD test data sets over 100 subsampling runs for the $\lambda_{PLD\text{-}min}$ (olive green), the $\lambda_{PLD\text{-}1sd}$ (green), the T4 risk equation calibrated on the GCKD training sets (T4-surv-recal, blue), and the original T4 risk equation (violet) in comparison to the null model (red) at one, two, three, four, and five years after the baseline visit (from left to right). The null model refers to the Kaplan-Meier estimator, and the inverse probability of censoring weights (IPCW) method employing the Kaplan-Meier estimator was used to deal with censored individuals.

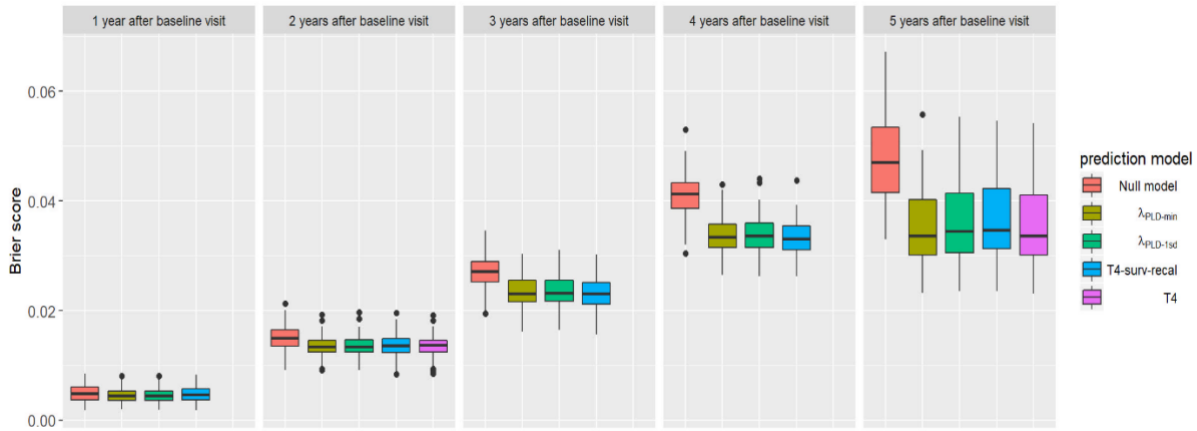


Figure S14. Averaged calibration curves evaluated on the GCKD test data sets over 100 subsampling runs for the $\lambda_{PLD\text{-}min}$ (black), the $\lambda_{PLD\text{-}1sd}$ (red), the T4 risk equation calibrated on the GCKD training sets (T4-surv-recal, green), and the original T4 risk equation (blue) at (A) two, (B) three, (E) four, and (F) five years after the baseline visit. An evaluation at one year after the baseline visit was omitted due to the low number of ESKD events at that time-point. For estimating the calibration curves, the expected ESKD event status was obtained in the nearest neighborhood around the predicted ESKD event probabilities. (C), (D), (G), and (H) show the corresponding averaged distributions of predicted kidney failure risk probabilities at two, three, four, and five years after the baseline visit.

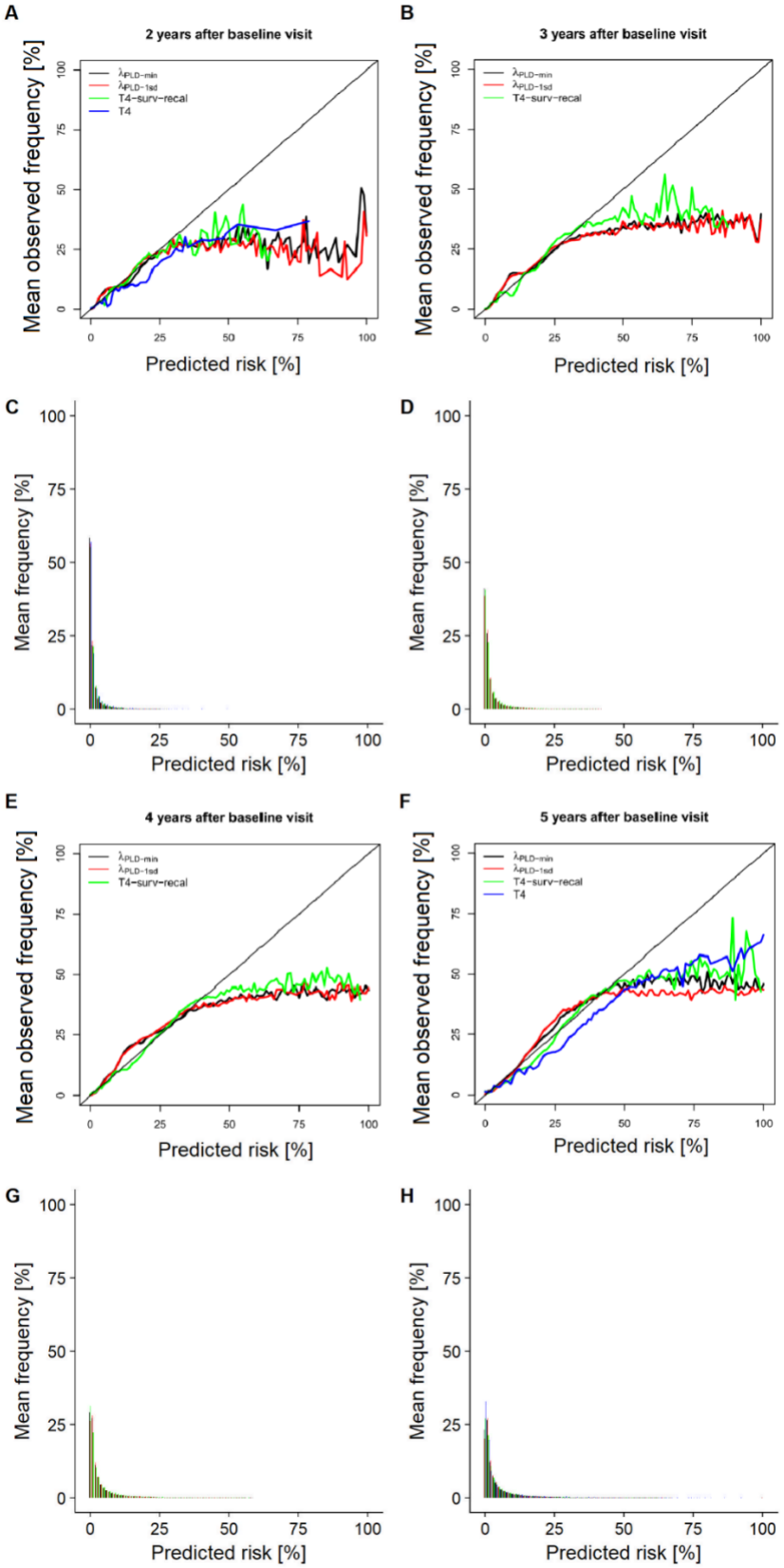


Figure S15. Observed vs. predicted ESKD risk probability estimates according to the $\lambda_{PLD\text{-}min}$, the $\lambda_{PLD\text{-}1sd}$, the T4 risk equation calibrated on the GCKD training sets (T4-surv-recal), and the original T4 risk equation (from left to right) at one, two, three, four, and five years after the baseline visit evaluated for one exemplary GCKD test set. Observed and predicted ESKD risk probability estimates are divided into deciles of predicted risk probability. Please note that at five years after the baseline visit, a high censoring rate is apparent in the Kaplan-Meier curve in Figure S2A, which impairs accurate estimation of absolute risk probabilities.

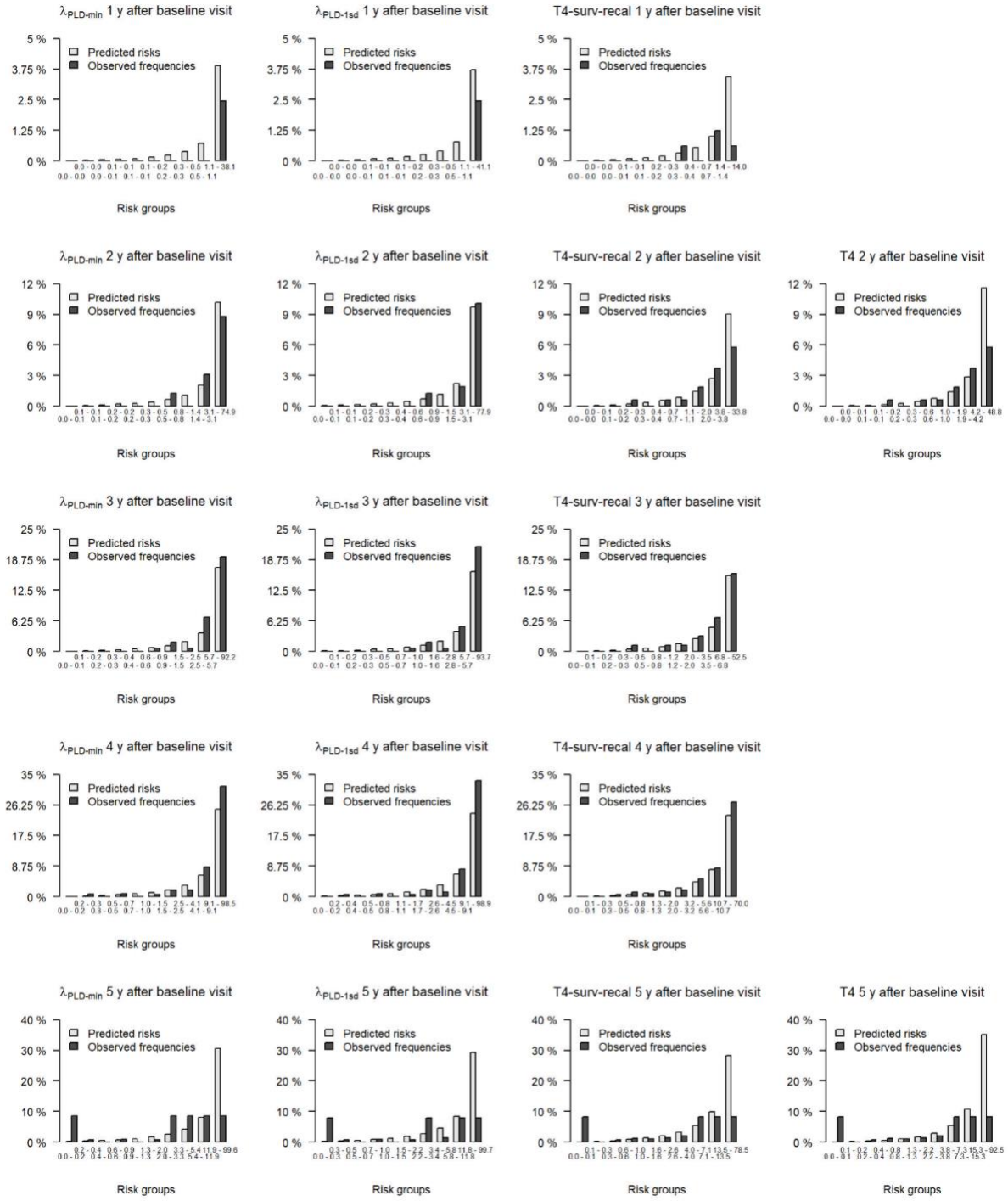


Figure S16. Calibration curves evaluated in the CKD-REIN cohort for the Z6 (black), the T4 risk equation calibrated on the complete GCKD cohort (T4-surv-recal, red), and the original T4 risk equation (blue) at (A) one, (B) two, (C) three, (D) four, and (E) five years after the baseline visit. For estimating the calibration curves, the expected ESKD event status was obtained in the nearest neighborhood around the predicted ESKD event probabilities.

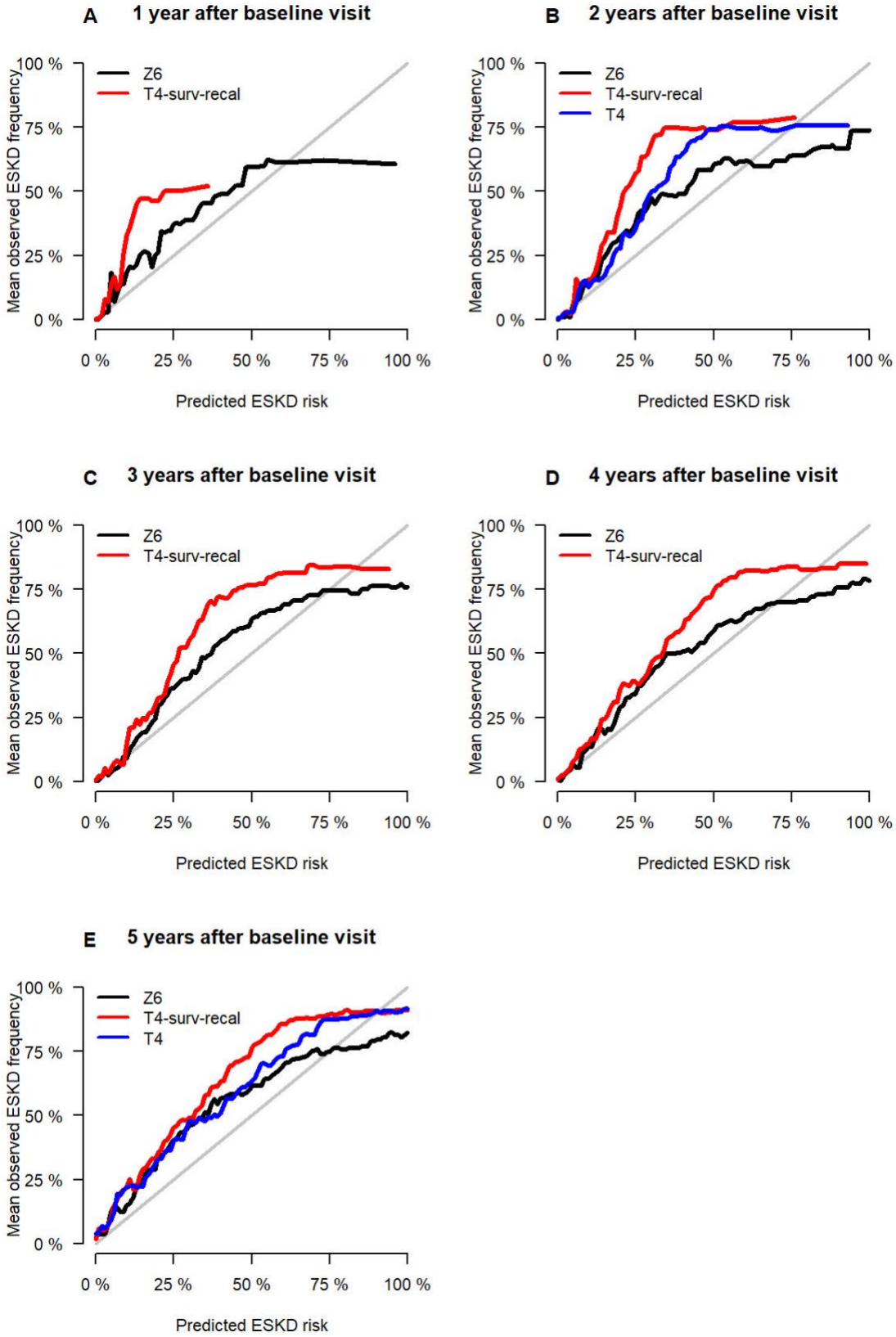


Figure S17. Observed vs. predicted ESKD risk probability estimates according to the Z6, the T4 risk equation calibrated on the complete GCKD cohort (T4-surv-recal), and the original T4 risk equation (from left to right) at one, two, three, four, and five years after the baseline visit evaluated for the CKD-REIN cohort. Observed and predicted ESKD risk probability estimates are divided into deciles of predicted risk probability.

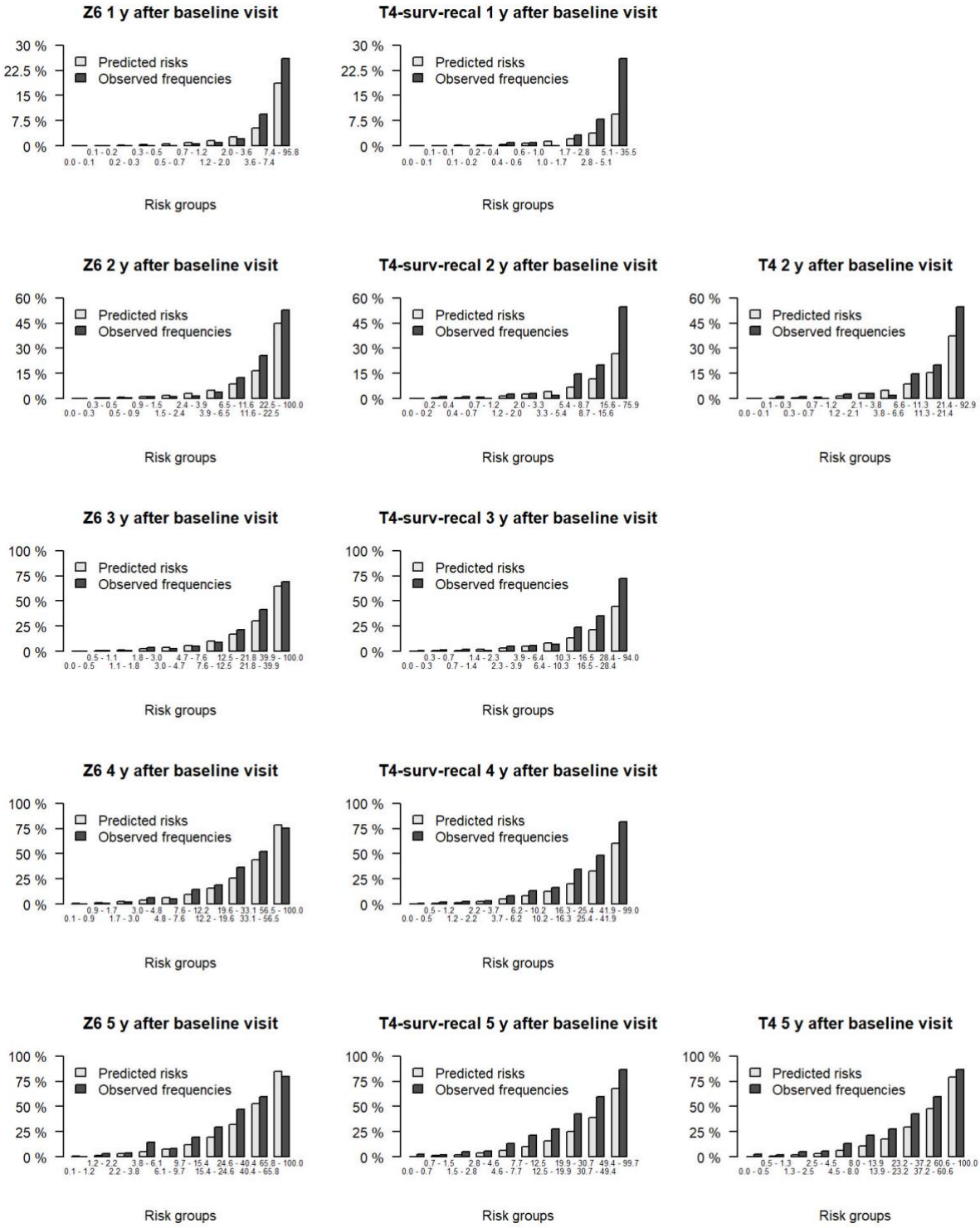


Figure S18. Calibration curves evaluated on the SKS cohort for the Z6 (black), the T4 risk equation calibrated on the complete GCKD cohort (T4-surv-recal, red), and the original T4 risk equation (blue) at (A) one, (B) two, (C) three, (D) four, and (E) five years after the baseline visit. For estimating the calibration curves, the expected ESKD event status was obtained in the nearest neighborhood around the predicted ESKD event probabilities.

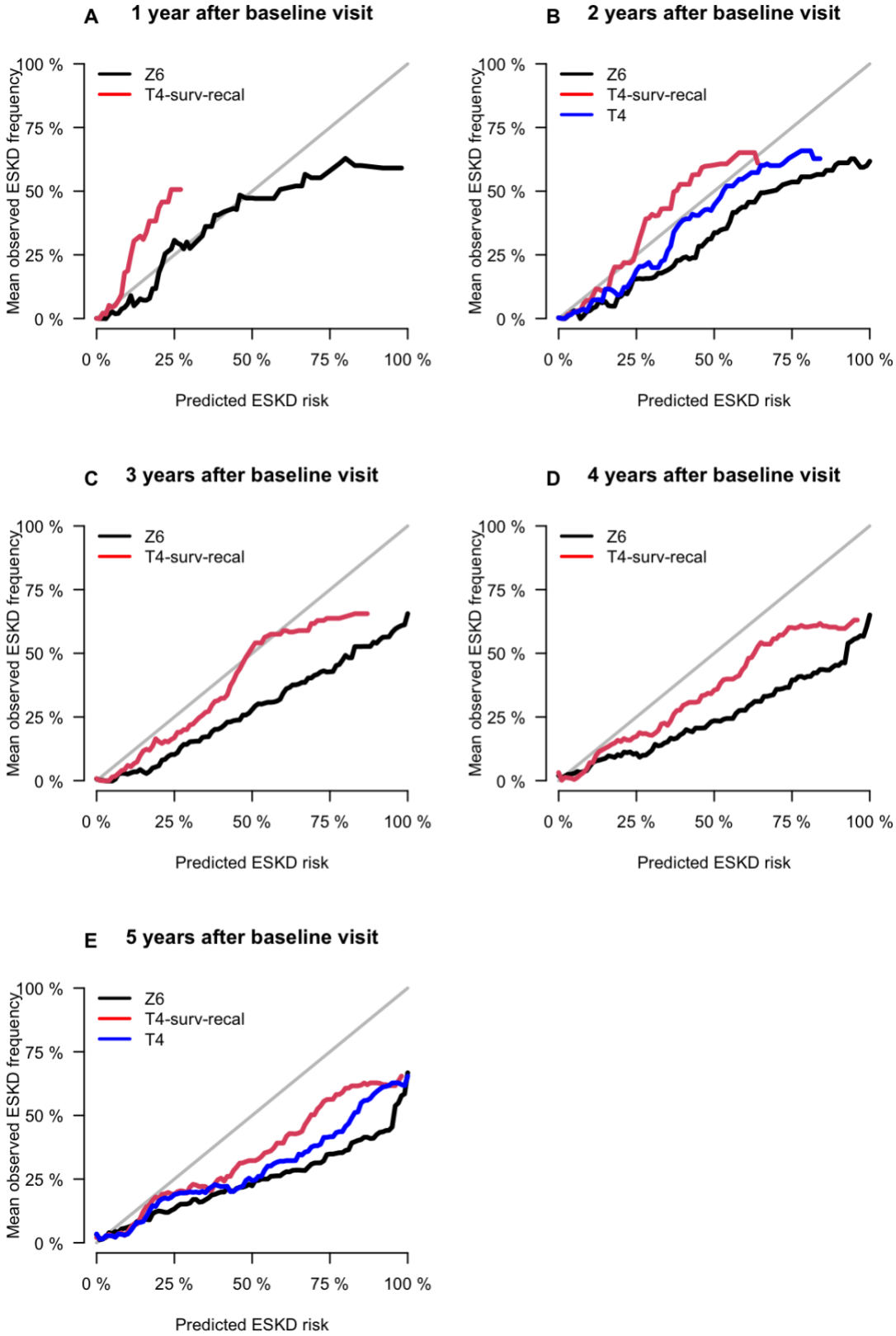


Figure S19. Observed vs. predicted ESKD risk probability estimates according to the Z6, the T4 risk equation calibrated on the complete GCKD cohort (T4-surv-recal), and the original T4 risk equation (from left to right) at one, two, three, four, and five years after the baseline visit evaluated for the SKS cohort. Observed and predicted ESKD risk probability estimates are divided into deciles of predicted risk probability.

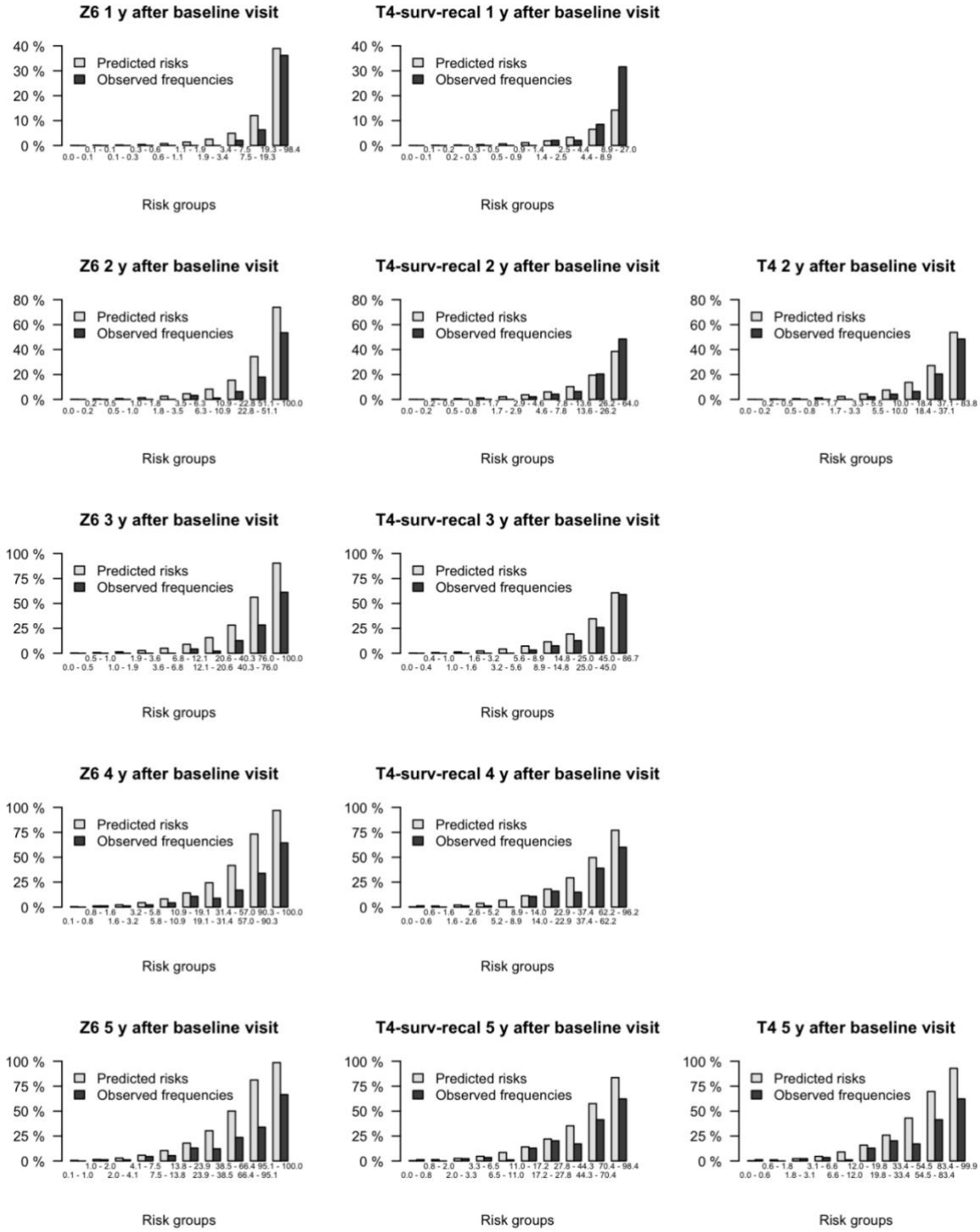


Figure S20. Calibration curves evaluated on the MMKD cohort for the Z6 (black), the T4 risk equation calibrated on the complete GCKD cohort (T4-surv-recal, red), and the original T4 risk equation (blue) at (A) two, (B) three, (C) four, and (D) five years after the baseline visit. An evaluation at one year after the baseline visit was omitted due to the low number of ESKD events at that time-point. For estimating the calibration curves, the expected ESKD event status was obtained in the nearest neighborhood around the predicted ESKD event probabilities.

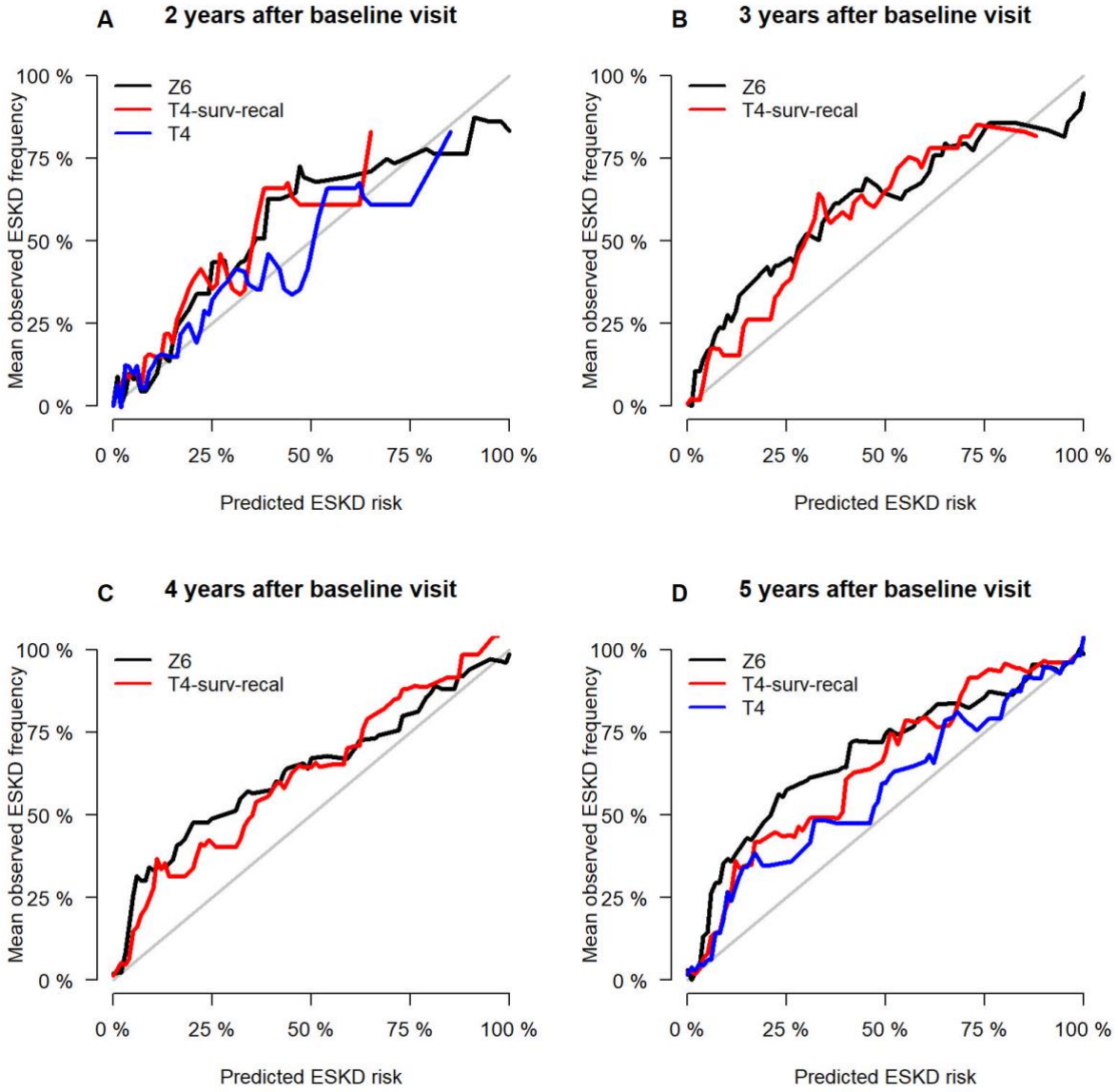


Figure S21. Observed vs. predicted ESKD risk probability estimates according to the Z6, the T4 risk equation calibrated on the complete GCKD cohort (T4-surv-recal), and the original T4 risk equation (from left to right) at one, two, three, four, and five years after the baseline visit evaluated for the MMKD cohort. Observed and predicted ESKD risk probability estimates are divided into deciles of predicted risk probability.

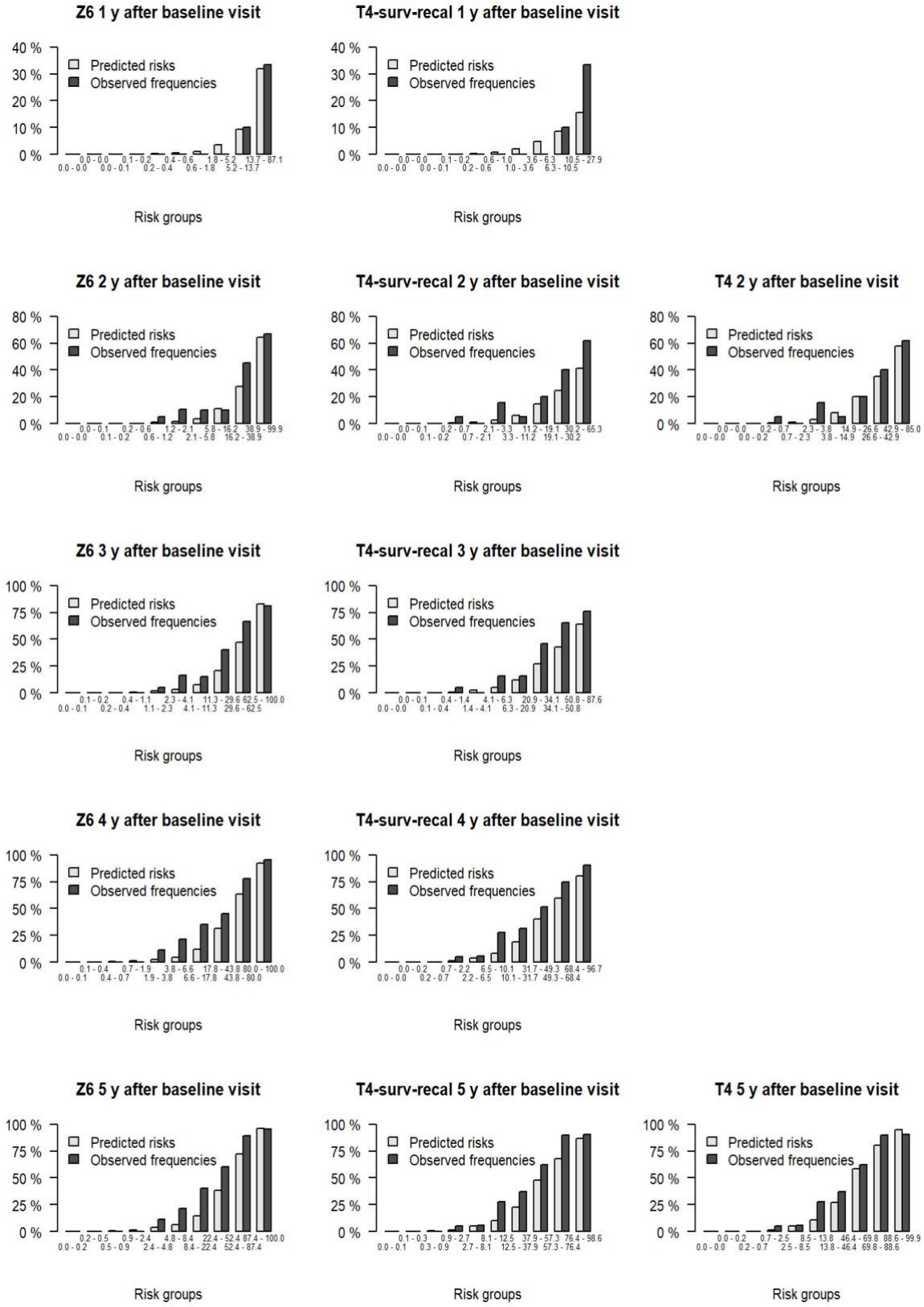


Figure S22. Calibration curves evaluated on the CKD-REIN subcohort including only study participants with measured urine albumin-to-creatinine ratio values for the Z6 (black), the T4 risk equation calibrated on the complete GCKD cohort (T4-surv-recal, red), and the original T4 risk equation (blue) at (A) one, (B) two, (C) three, (D) four, and (E) five years after the baseline visit. For estimating the calibration curves, the expected ESKD event status was obtained in the nearest neighborhood around the predicted ESKD event probabilities.

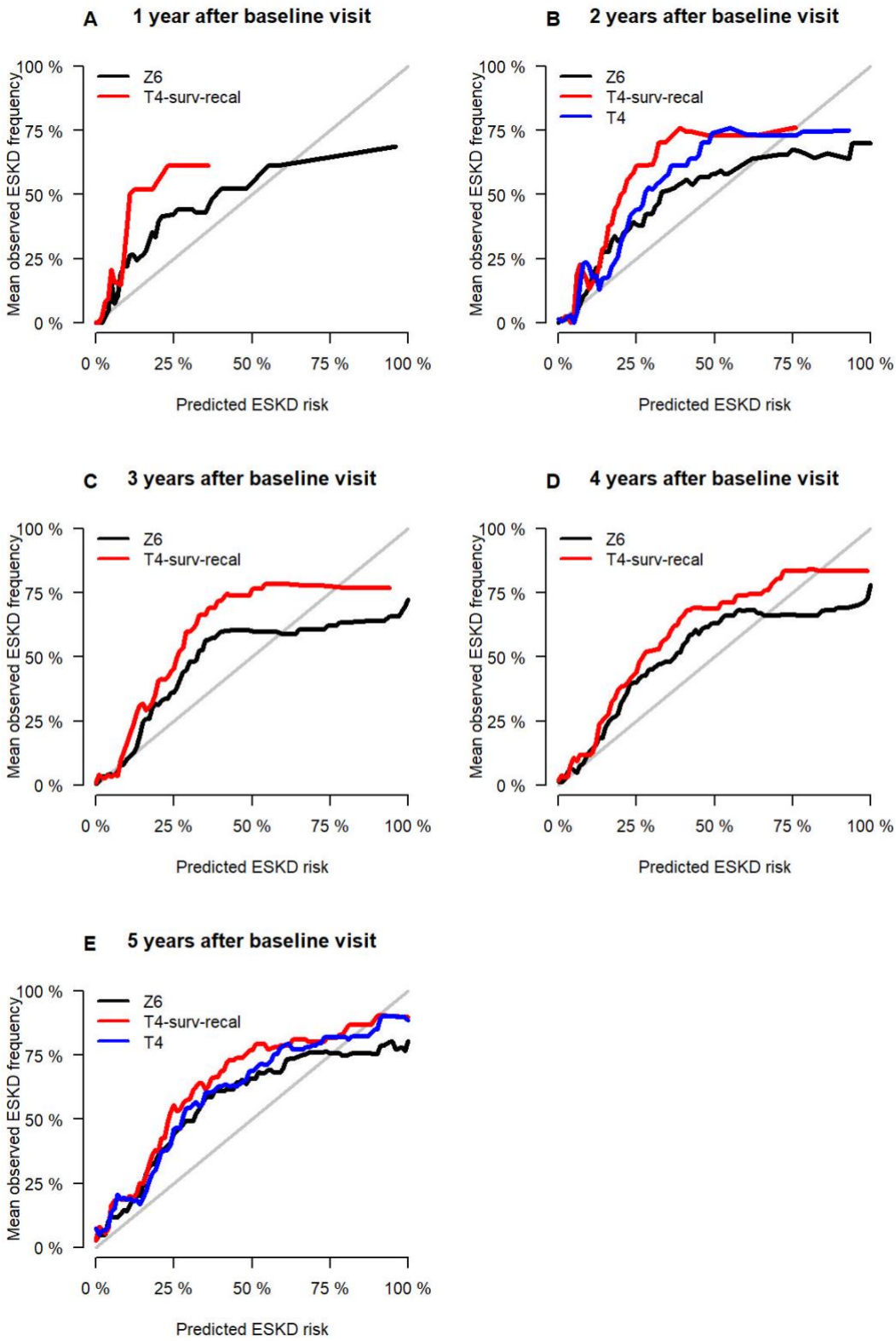


Figure S23. Observed vs. predicted ESKD risk probability estimates according to the Z6, the T4 risk equation calibrated on the complete GCKD cohort (T4-surv-recal), and the original T4 risk equation (from left to right) at one, two, three, four, and five years after the baseline visit evaluated for the CKD-REIN subcohort including only study participants with measured urine albumin-to-creatinine ratio values. Observed and predicted ESKD risk probability estimates are divided into deciles of predicted risk probability.

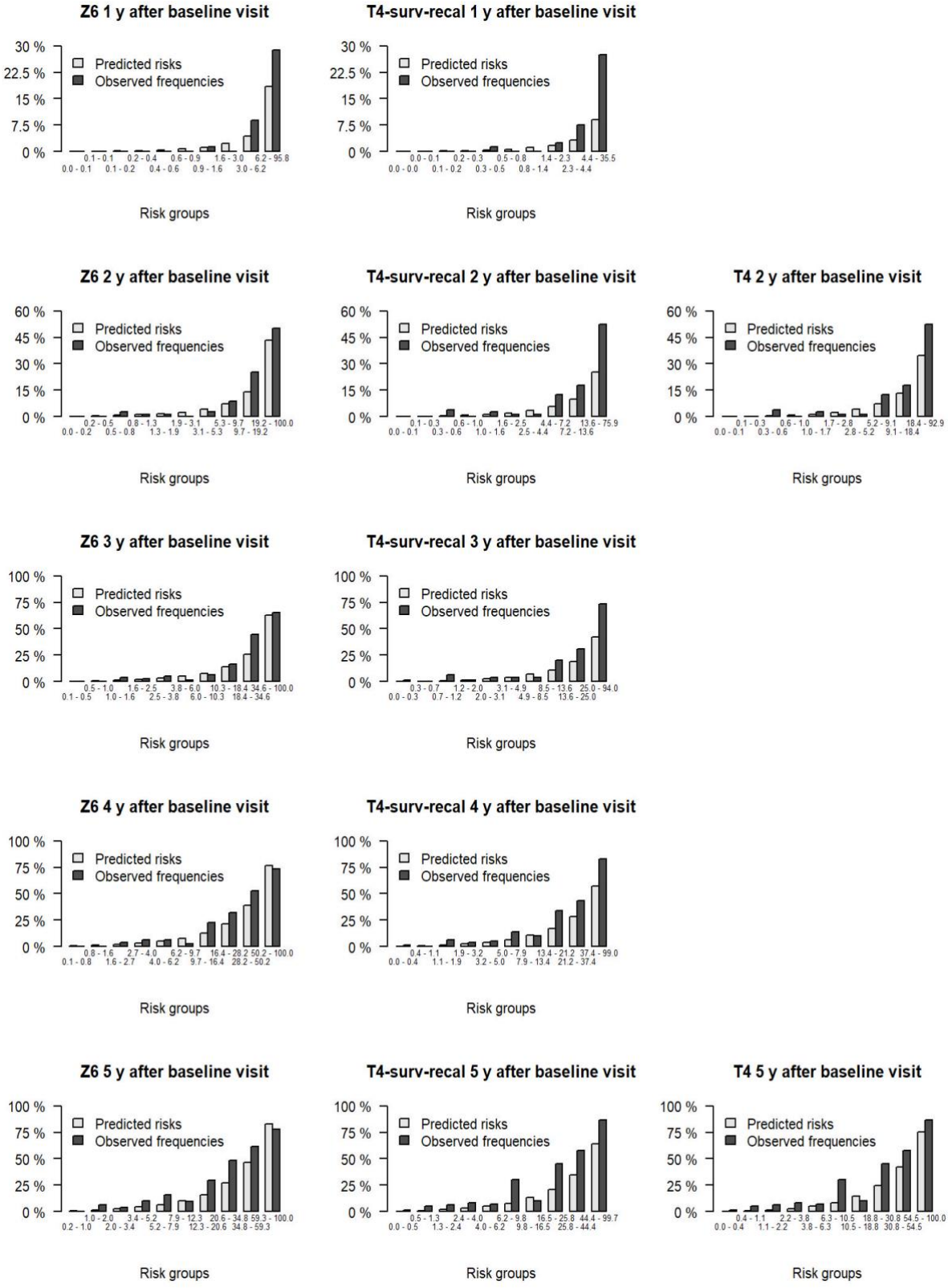


Figure S24. Calibration curves evaluated on the CKD-REIN subcohort including only study participants with estimated urine albumin-to-creatinine ratio values for the Z6 (black), the T4 risk equation calibrated on the complete GCKD cohort (T4-surv-recal, red), and the original T4 risk equation (blue) at (A) one, (B) two, (C) three, (D) four, and (E) five years after the baseline visit. For estimating the calibration curves, the expected ESKD event status was obtained in the nearest neighborhood around the predicted ESKD event probabilities.

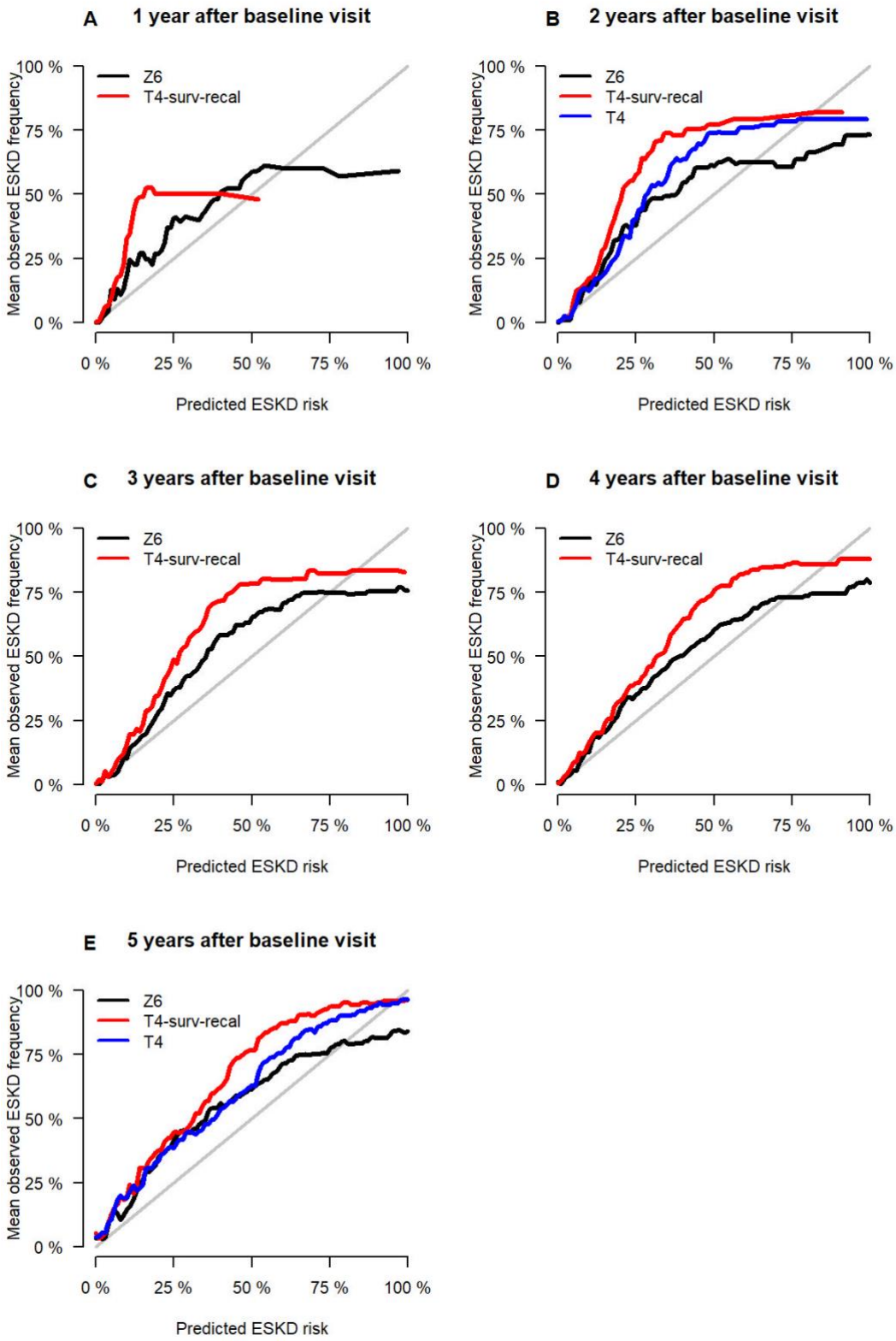


Figure S25. Observed vs. predicted ESKD risk probability estimates according to the Z6, the T4 risk equation calibrated on the complete GCKD cohort (T4-surv-recal), and the original T4 risk equation (from left to right) at one, two, three, four, and five years after the baseline visit evaluated for the CKD-REIN subcohort including only study participants with estimated urine albumin-to-creatinine ratio values. Observed and predicted ESKD risk probability estimates are divided into deciles of predicted risk probability.

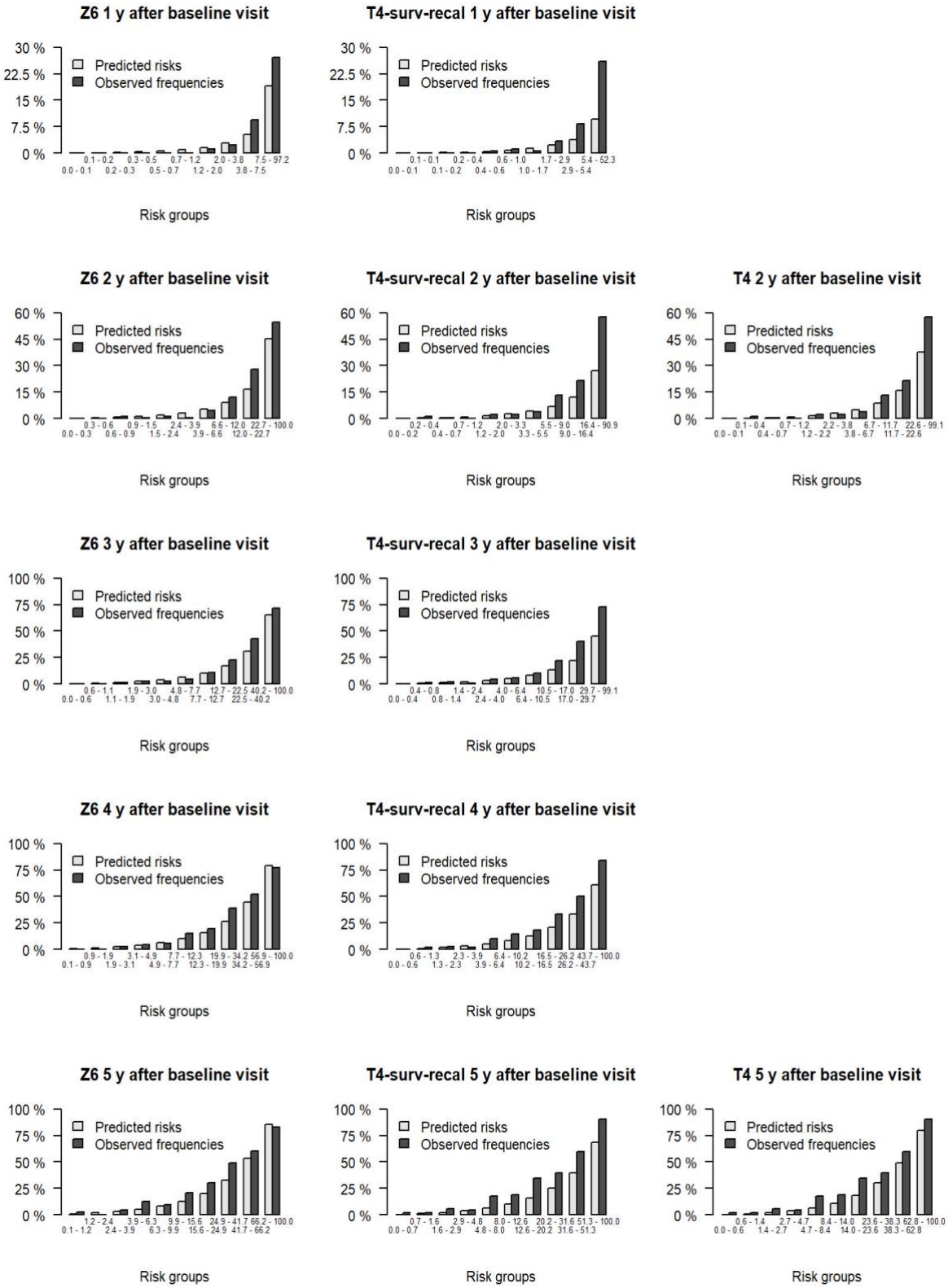
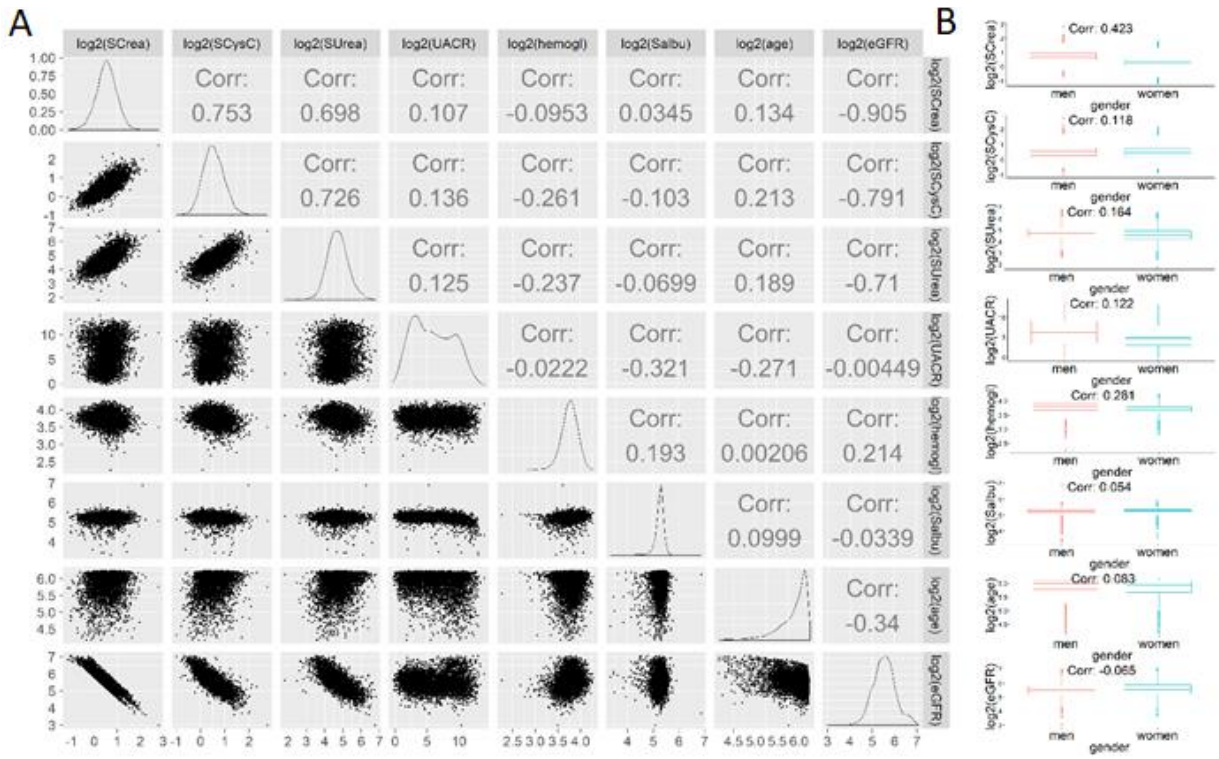


Figure S26. Correlation analyses between the Z6 and the T4 predictor variables in the GCKD cohort. (A) Correlation analysis between all log_2-transformed continuous variables. The upper triangle panels give the Pearson correlation for each correlation analysis, the lower triangle panels show respective scatter plots. The diagonal panels show the respective data distribution histogram for each continuous variable. (B) Boxplots and correlation analyses between gender and all log_2-transformed variables. The respective point biserial correlations are given above the boxes.



Item S1. Detailed patient selection procedures

Development cohort: The German Chronic Kidney Disease Study

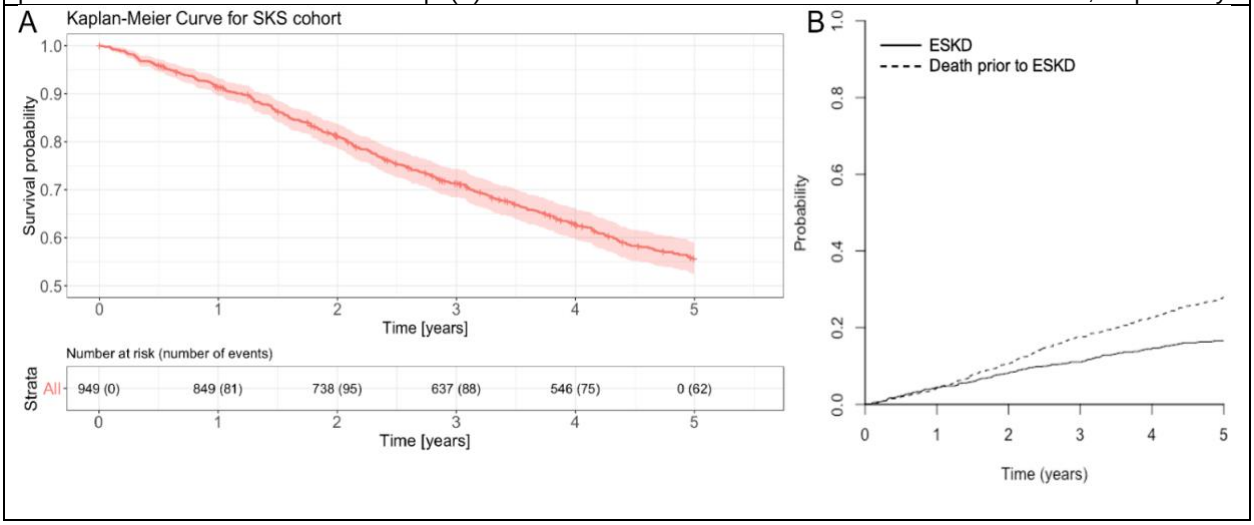
The *German Chronic Kidney Disease (GCKD)* study was registered with the national registry for clinical studies (DRKS 00003971) and carried out in accordance with relevant guidelines and regulations upon approval by the local ethics committees of the RWTH Aachen University, Aachen, Germany, the Charité – University-Medicine, Berlin, Germany, the Friedrich-Alexander-University, Erlangen, Germany (coordination center), the Albert-Ludwigs-University, Freiburg, Germany, the Friedrich-Schiller-University, Jena, Germany, the Hannover Medical School, Hannover, Germany, the medical faculty of the Ruprecht-Karls-University, Heidelberg, Germany, the medical faculty of the Ludwig-Maximilians-University, Munich, Germany, and the Julius-Maximilians-University, Würzburg, Germany. Inclusion criteria, based on the most recent health record data available, were an eGFR of 30-60 mL/min/1.73m^2 (G3-5) or overt proteinuria with an eGFR of >60 mL/min/1.73m^2 (G1-2 and A3). During the original GCKD study recruitment phase from March 2010 to March 2012, patients who did not self-report "white ancestry" were excluded, as they represented at that time a relatively small and heterogeneous group in Germany. Additionally excluded were patients with a history of solid organ or bone marrow transplantation, active cancer within the last 24 months, and severe heart failure[S1]. Reasons for exclusion of a total of 302 patients from the present study, which only included 4,915 of the original 5,217 GCKD study participants, were missing clinical chemistry and demographic data, as well as missing censoring time (n=1). Baseline characteristics of GCKD study participants included vs. those excluded from the present study are compared in Table S1. Of note, centralized laboratory analysis conducted at baseline resulted in restaging of patients and, thus, inclusion of patients with an eGFR of >60 mL/min/1.73m^2 even in the absence of overt proteinuria. In the GCKD study, endpoints are continuously adjudicated from hospital discharge letters, nephrologist out-patient letters and death certificates collected by trained study personal. Letters from all regional centers are sent to the coordinating center (Erlangen) of the GCKD study for scanning and upload to the database "Confluence". Letter collection and processing as well as quality

control steps after adjudication is time consuming and leads to a significant time lag between follow-up visits, letter collection adjudication and ultimate data availability. Analyses in this manuscript are based on a data freeze from June 2017 concordant with endpoints of follow-up 4. A median follow-up time of less than four years results from patients, who were not followed-up for the whole time period due to death or loss to follow-up.

Validation cohort 1: The Salford Kidney Study

The Salford Kidney Study (SKS) is a prospective observational single center cohort study in the United Kingdom that has been recruiting patients since 2002, aged ≥18 years old with moderate-to-severe CKD.[S2] Information on race and ethnicity was not available for individual patients.[S2] Approximately 3,200 patients were recruited into SKS as of 2018. Blood and urine collection for routine clinical chemistry testing was performed at baseline and at subsequent clinic visits and results are stored in the hospital's electronic patient files. Endpoints include death, initiation of kidney replacement therapy (chronic dialysis or transplantation), loss to follow-up, discharge from renal clinic or withdrawal of consent. The study complies with the declaration of Helsinki and ethical approval has been obtained from the regional ethics committee (current REC reference 15/NW/0818). Serum cystatin C measurements were available for 982 SKS participants, of whom 31 had to be excluded due to missing data and 2 due to acute kidney injury. Until 31st March 2018, 150 (15.8%) out of the 949 patients included here had reached ESKD defined as long-term dialysis (n=149) or kidney transplantation (n=1) within a mean observation time of 2.32 ± 1.37 years. For patients that experienced more than one event during the observation period, only the earliest event was considered. Baseline clinical and demographic characteristics are given in Table 1. A Kaplan-Meier curve for ESKD events or death and cumulative incidence curves for ESKD events and death are depicted in Figure *a*.

Figure a. Kaplan-Meier curve of time to first ESKD event or death and cumulative incidence curves of ESKD and death events for the 949 SKS patients included in the present study. (A) Kaplan-Meier curve for the time to first ESKD event (initiation of dialysis or renal transplantation) or death, whichever occurred first. The lower table gives the number of individuals at risk and the number of incident events in brackets for each follow-up year. Note that patients were censored at last follow-up. (B) Cumulative incidence curves for ESKD and death events, respectively.

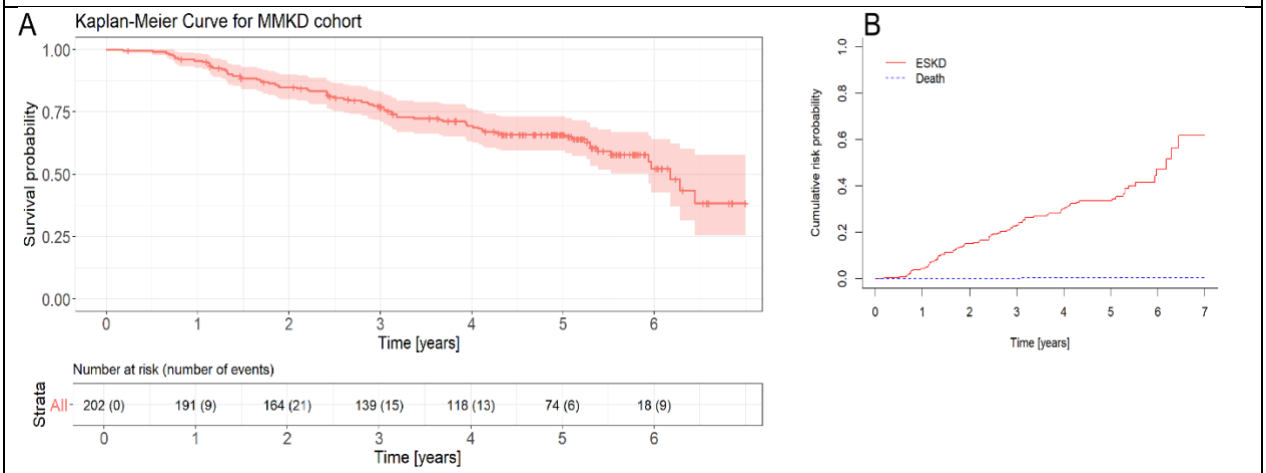


Validation cohort 2: The Mild to Moderate Kidney Disease Study

The Mild to Moderate Kidney Disease (MMKD) Study comprised 254 white patients between 18-65 years of age with non-diabetic CKD (stages G1-5) and various degrees of renal impairment from 8 nephrology clinics in Germany, Austria, and South Tyrol.[S3] White ethnicity was not self-reported, but observed by the recruiting physician. The study was approved by the respective institutional ethics committees, and all subjects gave written informed consent. They had stable renal function for at least 3 months before entry into the study. Exclusion criteria were treatment with immunosuppressive agents, fish oil or erythropoietin, serum creatinine >6 mg/dL (>0.53 mmol/L), diabetes mellitus of any type, malignancy, liver, thyroid or infectious disease, organ transplantation, allergy to ionic contrast media, and pregnancy. To avoid inter-observer differences, all patients were recruited by one physician, who visited all participating centers. Patient history, including smoking habits and antihypertensive treatment at baseline, was recorded by interview and confirmed by checking patient records. This was complemented by clinical examination including assessment of BMI and blood pressure. Hypertension was defined as blood pressure above 140/90 mmHg and/or the use of antihypertensive medication, which was withheld on days of measurement

of the GFR. After the baseline investigation, patients were followed prospectively until the primary study endpoint or the end of the observation period. The primary endpoint was defined as terminal renal failure necessitating kidney replacement therapy by dialysis or transplantation. For the present analysis, a total of 202 patients (89%) with complete follow-up information available were assessed. Seventy-five (37.1%) patients reached ESKD defined as long-term dialysis (*n*=69) or kidney transplantation (*n*=6) within a mean observation time of 3.97 ± 1.75 years. If a patient experienced more than one ESKD event, only the earliest event was considered. Baseline clinical and demographic characteristics are given in Table 1. A Kaplan-Meier curve for ESKD events or death and cumulative incidence curves for ESKD events and death are depicted in Figure *b*.

Figure *b*. Kaplan-Meier curve of time to first ESKD event or death and cumulative incidence curves of ESKD and death events for the 202 MMKD patients included in the present study. (A) Kaplan-Meier curve for the time to first ESKD event (initiation of dialysis or renal transplantation) or death, whichever occurred first. The lower table gives the number of individuals at risk and the number of incident events in brackets for each follow-up year. Note that patients were censored at last follow-up. (B) Cumulative incidence curves for ESKD and death events, respectively.

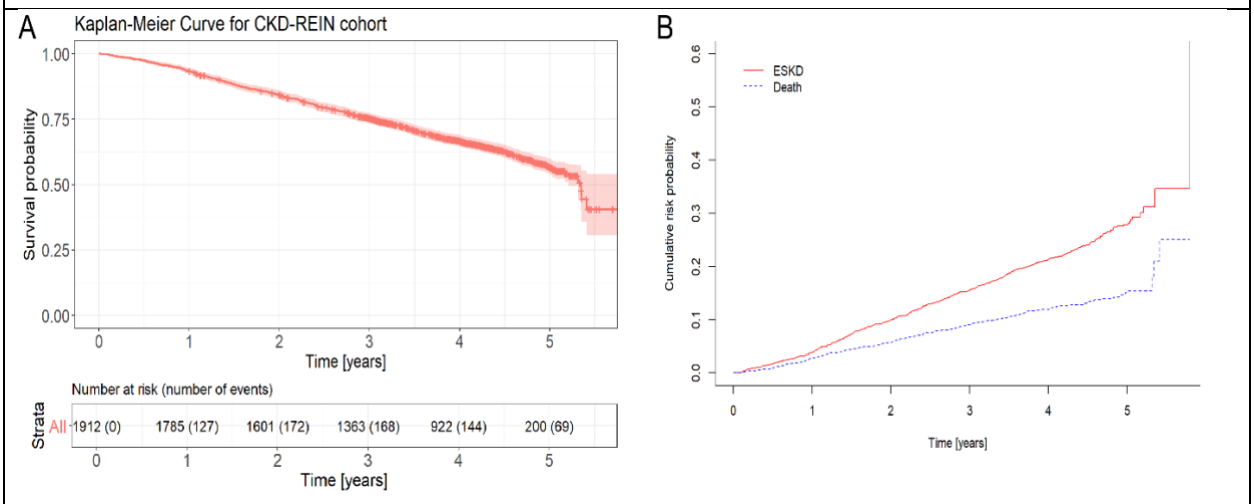


Validation cohort 3: The Chronic Kidney Disease-Renal Epidemiology and Information Network Study

The Chronic Kidney Disease-Renal Epidemiology and Information Network (CKD-REIN) study enrolled a total of 3,033 CKD patients between 2013 and 2016.[S4] For the present study, we excluded 230 patients that did not meet the inclusion criteria: 78 patients of sub-Saharan African or West Indies origin, 86 with a history of cancer within the last 24 months, and 71 with a history of solid organ transplantation other than renal transplantation. Patient origin was determined based on self-reported origin of the patient's parents

and "sub-Saharan African/West Indies origin" was determined if the patient reported at least one parent originating from sub-Saharan Africa or West Indies. All patients provided written informed consent. The study protocol had been approved by the Institut national de la santé et de la recherche médicale (Inserm) Institutional Review Board (IRB00003888) and registered on ClinicalTrials.gov (Identifier: NCT03381950). Cystatin C and serum creatinine were measured from blood specimens collected and stored at baseline, but were missing for 333 patients. Other parameters including hemoglobin, serum urea, serum albumin, and urinary albumin (or protein) values were requested at baseline per study protocol and measured at the patients' usual laboratory within 6 months before or after study enrolment. One or the other of these measurements was missing in 558 patients, leaving a sample of 1,912 patients for this analysis. Follow-up duration was calculated from the date of blood sample collection until initiation of renal replacement therapy (dialysis or pre-emptive transplantation), death or the date of the latest patient contact, whichever came first. After a median observation period of 4.0 years, 445 (23.3%) patients reached ESKD defined as long-term dialysis (*n*=379) or kidney transplantation (*n*=66). Baseline clinical and demographic characteristics are given in Table 1. Kaplan-Meier and cumulative incidence curves for ESKD events and/or death are depicted in Figure c.

Figure c. Kaplan-Meier curve for time to first ESKD event or death and cumulative incidence curves of ESKD and death events for the 1,912 CKD-REIN patients included in the present study. (A) Kaplan-Meier curve for the time to first ESKD event (initiation of dialysis or renal transplantation) or death, whichever occurred first. The lower table gives the number of individuals at risk and the number of incident events in brackets for each follow-up year. Note that patients were censored at last follow-up. (B) Cumulative incidence curves for ESKD and death events, respectively.



Item S2. Clinical variables measurement information

Development cohort: The German Chronic Kidney Disease Study

Baseline clinical variables employed for the development of a novel risk score in the GCKD study-included age, sex, body mass index (BMI), serum and urine creatinine (cobas® CREA plus, Roche Diagnostics GmbH, Mannheim, Germany), serum and urine albumin (cobas® ALBT2, Roche), high-sensitivity C-reactive protein (cobas® CRPHS, Roche), total cholesterol (cobas® CHOL2, Roche), high-density lipoprotein cholesterol (cobas® HDL-C plus 3rd generation, Roche), low-density lipoprotein cholesterol (cobas® LDL_C, Roche), triglycerides (cobas® TRIGL, Roche), cystatin C (CYSC, ADVIA®, Siemens Healthcare Diagnostics Ltd., Camberly, UK), calcium (cobas® Ca, Roche), sodium (ISE indirect Na for Gen.2, Roche), phosphate (cobas® PHOS, Roche), urea (cobas® UREA/BUN, Roche), uric acid (cobas® UA2, Roche), hemoglobin and glycated hemoglobin (COBAS INTEGRA Tina-quant Hemoglobin A1c Gen.2, Roche), UACR, and eGFR, which was estimated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation. All laboratory parameters other than hemoglobin and HbA1c, which were determined at the central laboratory of the University Hospital Erlangen, were measured by SYNLAB International GmbH, Munich, Germany.

Validation cohort 1: The SKS Study

In the SKS validation cohort, serum cystatin C was not measured in all patients upon entry into SKS but instead measured in a random selection of patients between 2004 and 2013 employing an Architect Ci8200 with reagents from Gentian (Moss, Norway). Demographic and other laboratory data were therefore collected on the date serum cystatin C had been measured for all patients within this validation cohort, and this date was classed as 'baseline'. In this cohort, hemoglobin (Siemens ADVIA 2120i, Siemens Healthcare Diagnostics, Munich, Germany), serum and urine creatinine (Siemens ADVIA 1800/2400, Siemens ADVIA Chemistry Systems Creatinine), serum urea (Siemens ADVIA 1800/2400, Siemens ADVIA Chemistry Urea Nitrogen), serum albumin (Siemens ADVIA 1800/2400, Siemens ADVIA Chemistry Systems Albumin), and urine protein (Siemens ADVIA 1800/2400, Siemens ADVIA Chemistry Systems

Total Protein_2) were available for the complete SKS validation cohort. Estimated UACR (eUACR) was calculated from UPCR using the following adjusted equation from Sumida et al.[S5]:

eUACR = exp(0.2445 × log(min (UPCR/50, 1)) + 1.5531 × log(max(min(UPCR/500, 1) , 0.1)) + 1.1057 × log(max(UPCR/500, 1)) + 5.2562 – 0.0793 × (if female) + 0.0802 × (if diabetic) + 0.1339 × (if hypertensive)).

Validation cohort 2: The MMKD Study

Blood specimens were drawn after an overnight fast of at least 12 hours. Measurement of routine chemistry such as hemoglobin and serum urea, albumin and high sensitivity C-reactive protein (hsCRP) were done by the local hospitals. Further blood specimens were collected and centrifuged at 1.500 *g* and 4°C for 10 minutes. The supernatants were stored in aliquots at -80°C for later measurement of creatinine and cystatin C by a kinetic Jaffe method traceable to IDMS (Roche Diagnostics) and an automated nephelometric immunoassay on a BN ProSpec analyzer (Dade Behring), respectively, in the clinical chemistry laboratory of the University Hospital of Zürich.

24-hour urine collection was used to determine urine total protein values reported as mg/24 h per 1.73 m². Estimated UACR (eUACR) was calculated from UPCR using the following adjusted equation from Sumida et al.[S5]:

eUACR = exp(0.2445 × log(min (UPCR/50, 1)) + 1.5531 × log(max(min(UPCR/500, 1) , 0.1)) + 1.1057 × log(max(UPCR/500, 1)) + 5.2562 – 0.0793 × (if female) + 0.0802 × (if diabetic) + 0.1339 × (if hypertensive)).

Validation cohort 3: The CKD-REIN Study

Serum creatinine and cystatin C were measured centrally by an IDMS traceable enzymatic method and immunoturbidimetry, respectively, using an Architect C1600 ABBOTT analyzer. GFR was estimated using the CKD-EPI equation. The other laboratory values were measured at patients' usual laboratory (assays used are not known). The urine albumin-to-creatinine (UACR) and urine protein-to-creatinine (UPCR) ratios were calculated using spot urine measurements of albumin, protein, and creatinine. However, when

these measurements were missing, we used data obtained for 24-hour urine specimens. Estimated UACR (eUACR) was calculated from UPCR using the following adjusted equation from Sumida et al.[S5]:

eUACR = exp(0.2445 × log(min (UPCR/50, 1)) + 1.5531 × log(max(min(UPCR/500, 1) , 0.1)) + 1.1057 × log(max(UPCR/500, 1)) + 5.2562 – 0.0793 × (if female) + 0.0802 × (if diabetic) + 0.1339 × (if hypertensive)).

Item S3. Multivariable data analysis

LASSO Cox proportional hazards (PH) regression was used for variable selection and prediction modeling in the GCKD development cohort. To this end, all available predictor variables except sex were $\log_2$ transformed to account for skewed data and extreme values. The *R* package *glmnet*[S6] was used to fit the LASSO Cox PH model.[S7] The hyperparameter λ, which adjusts the trade-off between model fit and model sparsity was optimized to yield the minimum cross-validated partial likelihood deviance. This parameter was estimated using inner five-fold cross-validation.[S6] Next, to obtain a more regularized and, thus, sparser model estimate, the penalty parameter λ was increased as long as the five-fold-cross-validated error remained within one standard deviation of the minimum deviance. An example obtained for the training data of the complete GCKD cohort is given in Figure S1. To obtain the final risk equations reported together with their respective coefficients in Box 1 and Item S5, LASSO Cox models were fitted to the complete data. To assess the robustness of the derived LASSO Cox PH models with respect to different regularization parameter optimization procedures, we additionally performed all analyses in the development cohort employing inner ten-fold cross-validation. Likewise, we repeated all analyses in the development cohort optimizing λ to yield the maximum cross-validated concordance statistic *C* in the inner five-fold cross-validation.[S8]

The performance of the estimated risk scores in distinguishing patients that subsequently required KRT was assessed by the concordance statistic *C,* accounting for death as a competing event.[S9] The *C* statistic or index represents the probability that, for a pair of randomly chosen patients, the patient with the higher risk prediction will experience an ESKD event before the other patient. Thus, a *C* index of 1.0 indicates the perfect discrimination of two groups, whereas a *C* index of 0.5 indicates a model that does not perform better than chance at predicting membership to a group. The ESKD-specific *C* statistic was computed with the *R* package *pec*.[S10] Employing the Kaplan-Meier estimator for the censoring times, the inverse-probability-of-censoring weighted estimator was used to deal with right-censored data.[S10] Confidence intervals (CIs) in the GCKD resampling approach were determined from the empirical 2.5% and 97.5%

quantiles of the 100 different test sets. CIs in the three external validation cohorts were determined by ordinary nonparametric percentile bootstrap with 10,000 bootstrap replicates employing the *R* package *boot*[S11], and one-sided *p*-values testing whether *C* statistics differences between the Z6 and the T4 were significantly greater than zero were determined accordingly. Meta-analysis of *p*-values across the three validation cohorts was performed according to the sum of logs method. A one-sided *p*-value < 0.05 was termed statistically significant.

Both categorical and category-free net reclassification improvements (NRI)[S12,S13] as well as time-specific receiver operating characteristic (ROC) curves were used to assess the relative ability of the various risk equations to distinguish between low- and high-risk CKD patients. A reclassification improvement is defined as assignment to a higher or lower risk category dependent on whether an individual did or did not experience an event during a given observation period. The NRI quantifies how well a new model reclassifies dichotomous outcomes as compared to an old model. The categorical NRI is only comparable across time-points if the same risk categories are applied.[S13] Motivated by Tangri et al.[S14], we applied the following three risk categories: 0% - <3%, 3% - <10%, 10% - 100%. Besides the categorical NRI, we also evaluated category-free NRIs for all our comparisons.[S13] The NRIs were computed employing the *R* package *nricens*.[S15] An NRI was considered significant, if the corresponding 95% CIs, which had been determined empirically across 100 subsampling runs in the GCKD development cohort, excluded zero. The 95% CIs in the validation cohorts were determined by percentile bootstrap with 10,000 bootstrap replicates.

To calculate the absolute risk of experiencing an ESKD event during a given observation period in the presence of death as a competing risk, we translated the predicted risk scores into risk probabilities according to

$$P(t) = 1 - \exp\{-\Lambda_0(t)\exp[\alpha\,(f(x) - \overline{f(x)})]\}, \quad (1)$$

where $\Lambda_0(t)$ is the cumulative ESKD subdistribution hazard function, *f(x)* is either the Z14 ($\lambda_{PLD\text{-}min}$ in the GCKD resampling approach) risk score as provided in Item S5 or the Z6 risk score ($\lambda_{PLD\text{-}1sd}$ in the GCKD resampling approach), $\overline{f(x)}$ is the respective mean of either the Z14 or Z6 scores evaluated for the complete

GCKD data set (or for the respective training set in the GCKD resampling approach), and α is the coefficient of *f(x)* estimated in a Fine-Gray proportional subdistribution hazards regression model[S16] with ESKD as the event of interest and death as the competing event evaluated for the complete GCKD data set (or for the respective training set in the GCKD resampling approach). Likewise, $\Lambda_0(t)$ was estimated by a Breslow-type estimator using a Fine-Gray proportional subdistribution hazards regression model with *f(x)* as predictor variable evaluated for the complete GCKD cohort (or for the respective training set in the GCKD resampling approach). The Fine-Gray proportional subdistribution hazards regression models were fitted employing the *R* package *cmprsk*.[S17]

For the evaluation of time-specific ROC curves, we considered all patients, who experienced an ESKD event during a given observation period, as true positives, and all patients, who did not experience an ESKD event during the same observation period, as true negatives. Patients, who were censored or died before the end of the respective observation period, were excluded from ROC curve evaluation. For illustration, ROC curves were averaged across 100 test sets employing the *R* package *precrec*.[S18] Sensitivities, specificities, balanced accuracies, as well as positive and negative predictive values were evaluated for two specific cut-offs, individually optimized for each observation period, risk score, and test set, respectively, and then averaged over 100 test sets. The first cut-off, termed "best", minimizes $(1 - \textit{specificity})^2 + (\textit{sensitivity} - 1)^2$, whereas the "Youden" cut-off maximizes $\textit{sensitivity} + \textit{specificity} - 1$.[S19] We further determined the areas under the ROC curves (AUC-ROCs), which are diagnostic performance measures independent of the chosen cut-off values and range between 0 and 1, where a perfect score would obtain an AUC-ROC of 1.

In addition, to assess whether cardiovascular events impacted model performance, we repeated all model performance assessments in the GCKD study upon additionally censoring at the date of non-fatal cardiovascular events, including coronary revascularization procedures, non-hemorrhagic stroke and myocardial infarction. Please note that this performance analysis did not involve the training of new predictive models, just the recalibration of the risk equations and the re-evaluation of the test data. We now

additionally excluded from the ROC curve evaluation patients, who had been censored at the date of a cardiovascular event before the end of the respective observation period.

Calibration of the $\lambda_{PLD\text{-}min}$, $\lambda_{PLD\text{-}1sd}$, and the GCKD-recalibrated T4 risk equations at one to five years after the baseline visit, as well as the T4 at two and five years after the baseline visit was assessed on each test set in the GCKD resampling approach by evaluating Brier scores[S20] as well as by visual inspection of calibration graphs plotting the predicted ESKD risks *vs.* the observed ESKD frequencies in calibration curves, and by plotting observed and predicted ESKD risk probability estimates divided into deciles of predicted risk probability. Calibration of the Z6, T4, and the GCKD-recalibrated T4 risk equations was assessed in the independent validation cohorts in the same manner. Calculation of Brier scores and plotting of calibration graphs was performed with the *R* package *pec*,[S10] whereas ESKD represented the event of interest and death the competing event.

The Kaplan-Meier curve analyses were carried out using the *R* packages *survival*[S21,S22] and *survminer*.[S23] The cumulative incidence curves were generated with the *R* package *cmprsk*.[S17] All statistical analyses, except for optimization of the regularization parameter with respect to the *C* statistic in the GCKD cohort and statistical analyses in the independent validation cohorts, were carried out in *R* version 3.4.3.[S24] Optimization of the regularization parameter with respect to the *C* statistic was done using *R* version 3.6.3.[S24] For the SKS cohort, all statistical analyses were performed with *R* version 4.0.2. For the MMKD cohort, statistical analyses were performed using either SPSS for Windows, version 25.0 (IBM Corp., Armonk, New York, NY, USA) or *R* for Windows, versions 3.4.3 and 3.6.2 (Vienna, Austria). For the CKD-REIN cohort, statistical analyses were performed with *R* versions 3.4.3 and 3.6.3.[S24]

Item S4. Z14 Risk Equation

To predict the probability of experiencing an ESKD event at four years after the baseline visit employing the Z14 risk equation, use the following parameter settings:

$$P_{\mathrm{Z14}}(t = 4y) = 1 - \exp\{-0.01288 \exp[1.030\ (f_{\text{14-var}}(x) + 6.450)]\} \qquad (2)$$

with $f_{\text{14-var}}(x) = +1.436 \log_2$ (serum creatinine [mg/dL]) $+ 0.980 \log_2$ (serum cystatin C [mg/L])

$+ 0.496 \log_2$ (serum sodium [mmol/L]) $+ 0.415 \log_2$ (HbA1c [%]) $+ 0.217 \log_2$ (serum urea [mg/dL])

$+ 0.204 \log_2$ (UACR [mg/g]) $+ 0.054 \log_2$ (CRP [mg/L]) $- 0.012 \log_2$ (HDL [mg/dL])

$- 0.087 \log_2$ (urine creatinine [mg/dL]) $- 0.091 \log_2$ (LDL [mg/dL])

$- 0.150 \log_2$ (eGFR (CKD-EPI) [mL/min/1.73m^2]) $- 0.187 \log_2$ (age [years])

$- 1.160 \log_2$ (hemoglobin [g/dL]) $- 1.368 \log_2$ (serum albumin [g/L]).

The Z14 and the Z6 risk equations are available as an online web service tool at https://ckdn.app/tools/eskdcalc/. This web service tool also provides risk probability predictions for one, two, and three years after the baseline visit.

Item S5. LASSO Cox PH ESKD risk models after regularization parameter optimization in an inner 10-fold cross-validation

LASSO Cox PH ESKD risk models were fitted on the complete GCKD cohort including an optimization of the hyperparameter λ to yield (1) the minimum partial likelihood deviance (termed $Z14_{10foldCV}$), and (2) the maximum penalty while keeping the partial likelihood deviance (PLD) within one standard deviation of the minimum PLD (termed $Z6_{10foldCV}$) within an inner 10-fold cross-validation.

$Z14_{10foldCV}$ model = +1.340 $\log_2$ (serum creatinine [mg/dL]) + 0.967 $\log_2$ (serum cystatin C [mg/L]) + 0.901 $\log_2$ (serum sodium [mmol/L]) + 0.466 $\log_2$ (HbA1c [%]) + 0.221 $\log_2$ (serum urea [mg/dL]) + 0.208 $\log_2$ (UACR [mg/g]) + 0.062 $\log_2$ (CRP [mg/L]) – 0.029 $\log_2$ (HDL [mg/dL]) – 0.103 $\log_2$ (urine creatinine [mg/dL]) – 0.108 $\log_2$ (LDL [mg/dL]) – 0.212 $\log_2$ (eGFR (CKD-EPI) [mL/min/1.73m^2]) – 0.245 $\log_2$ (age [years]) – 1.176 $\log_2$ (hemoglobin [g/dL]) – 1.400 $\log_2$ (serum albumin [g/L])

$Z6_{10foldCV}$ model = +1.128 $\log_2$ (serum creatinine [mg/dL]) + 1.108 $\log_2$ (serum cystatin C [mg/L]) + 0.135 $\log_2$ (UACR [mg/g]) + 0.125 $\log_2$ (serum urea [mg/dL]) – 0.523 $\log_2$ (hemoglobin [g/dL]) - 1.070 $\log_2$ (serum albumin [g/L])

Item S6. LASSO Cox PH ESKD risk models after regularization parameter optimization with respect to the *C* statistic

LASSO Cox PH ESKD risk models were fitted on the complete GCKD cohort including an optimization of the hyperparameter λ to yield (1) the maximum *C* statistic (termed $Z6_{C\text{-max}}$), and (2) the maximum penalty while keeping the *C* statistic within one standard deviation of the maximum *C* statistic (termed $Z4_{C\text{-1sd}}$) within an inner 5-fold cross-validation.

$Z6_{C\text{-max}}$ model = + 1.213 $\log_2$ (serum creatinine [mg/dL]) + 1.095 $\log_2$ (serum cystatin C [mg/L]) + 0.148 $\log_2$ (UACR [mg/g]) + 0.150 $\log_2$ (serum urea [mg/dL]) – 0.676 $\log_2$ (hemoglobin [g/dL]) – 1.121 $\log_2$ (serum albumin [g/L])

$Z4_{C\text{-1sd}}$ model = + 0.692 $\log_2$ (serum creatinine [mg/dL]) + 1.116 $\log_2$ (serum cystatin C [mg/L]) + 0.074 $\log_2$ (UACR [mg/g]) – 0.613 $\log_2$ (serum albumin [g/L])

Item S7. MMKD and CKD-REIN study investigators

MMKD study: The following members of the "Mild to Moderate Kidney Disease" (MMKD) Study Group collaborated with the authors of this project: Erich Kuen, Institute of Genetic Epidemiology, Innsbruck Medical University (Innsbruck, Austria); Paul König, Innsbruck University Hospital (Innsbruck, Austria); Günter Kraatz, Ernst-Moritz-Arndt-University (Greifswald, Germany); Johannes F. E. Mann, München Klinik Schwabing (Munich, Germany); Gerhard A. Müller, Georg-August-University (Göttingen, Germany); Ulrich Neyer, Feldkirch Hospital (Feldkirch, Austria); Hans Köhler, Medizinische Universitätskliniken des Saarlandes (Homburg/Saar, Germany); Peter Riegler, Bozen Hospital (Bozen, Italy).

CKD-REIN study investigators: Carole Ayav, Serge Briançon, Christian Combe, Denis Fouque, Luc Frimat, Yves-Edouard Herpe, Christian Jacquelinet, Maurice Laville, Sophie Liabeuf, Ziad A. Massy, Christophe Pascal, Bruce M. Robinson, Elodie Speyer, Bénédicte Stengel.

Table S1. Comparison of the baseline characteristics between included and excluded GCKD study participants.

Cohort	**Included GCKD study sample**	**Excluded GCKD study sample**	**Number of missing values in excluded GCKD study sample**	***p*-value**
Total number of participants	4,915	302	-	-
Health parameters	Mean ± SD (min - max)	Mean ± SD (min - max)	*n* (%)	
Age (years)	60.07 ± 11.94 (18-76)	59.90 ± 12.37 (18-75)	1 (0.33)	0.813[d]
Serum albumin (g/L)	38.37 ± 4.42 (10.10-117.46)	37.87 ± 4.64 (15.01-51.61)	58 (19.21)	0.100[d]
Total cholesterol (mg/dL)	211.11 ± 52.49 (72.26-771.58)	213.90 ± 62.74 (91.76-537.18)	63 (20.86)	0.501[d]
HDL cholesterol (mg/dL)	51.94 ± 18.11 (7.27-152.20)	52.46 ± 18.88 (18.47-134.45)	67 (22.19)	0.675[d]
LDL cholesterol (mg/dL)	118.29 ± 43.24 (4.67-601.78)	117.89 ± 49.69 (7.56-408.21)	68 (22.52)	0.904[d]
Serum creatinine (mg/dL)	1.51 ± 0.48 (0.45-7.01)	1.47 ± 0.47 (0.49-3.28)	58 (19.21)	0.134[d]
Urine creatinine (mg/dL)	82.06 ± 55.47 (0.95-513.27)	83.25 ± 58.36 (1.64-323.84)	89 (29.47)	0.770[d]
Hemoglobin (g/dL)	13.60 ± 1.67 (4.80-19.30)	13.52 ± 1.69 (8.60-18.20)	132 (43.71)	0.544[d]
HbA1c (%)	6.33 ± 1.02 (4.60-16.30)	6.61 ± 1.34 (5.00-14.20)	121 (40.07)	0.006[d]*
HbA1c (mmol/mol)	45.71 ± 11.08 (27.00-154.20)	48.72 ± 14.63 (31.10-131.80)	121 (40.07)	0.007[d]*
Serum urea (mg/dL)	28.46 ± 12.25 (3.54-111.55)	27.60 ± 12.84 (10.46-105.46)	57 (18.87)	0.308[d]
Serum uric acid (mg/dL)	7.21 ± 1.91 (0.72-26.77)	7.11 ± 1.92 (2.79-14.00)	56 (18.54)	0.426[d]
Serum cystatin C (mg/L)	1.52 ± 0.48 (0.50-6.71)	1.55 ± 0.59 (0.53-7.01)	55 (18.21)	0.469[d]
Serum calcium (mmol/L)	2.27 ± 0.15 (0.24-6.18)	2.26 ± 0.17 (1.57-2.98)	55 (18.21)	0.444[d]
Serum phosphate (mmol/L)	1.11 ± 0.20 (0.43-3.14)	1.09 ± 0.22 (0.53-2.25)	56 (18.54)	0.267[d]
Serum sodium (mmol/L)	139.71 ± 3.08 (94-172)	139.16 ± 3.93 (100-148)	54 (17.88)	0.031[d]*
eGFR (CKD-EPI; mL/min/1.73 m^2)	49.40 ± 18.17 (8-134)	50.93 ± 19.28 (20-136)	58 (19.21)	0.226[d]
eGFR (MDRD4; mL/min/1.73m^2)	46.98 ± 16.50 (8-151)	48.65 ± 18.19 (20-141)	58 (19.21)	0.162[d]
Systolic blood pressure (mmHg)	139.40 ± 20.28 (77-232)	140.89 ± 21.58 (87-212)	7 (2.32)	0.250[d]
Diastolic blood pressure (mmHg)	79.25 ± 11.75 (30-151)	79.73 ± 11.88 (48-114)	7 (2.32)	0.503[d]
Pulse (bpm)	70.43 ± 12.12 (29-135)	71.97 ± 13.51 (43-127)	7 (2.32)	0.057[d]
Waste-hip ratio	0.94 ± 0.09 (0.64-1.30)	0.95 ± 0.10 (0.66-1.25)	21 (6.95)	0.075[d]
Weight (kg)	87.33 ± 18.86 (38-121)	89.24 ± 20.19 (49-158.30)	55 (18.21)	0.149[d]
BMI (kg/m^2)	29.77 ± 5.96 (15.50-69.70)	30.41 ± 5.96 (17-48.60)	55 (18.21)	0.103[d]
Follow-up time (days)	1,354.05 ± 320.42 (1-1,925)	1,272.21 ± 410.58 (1-1,869)	1 (0.33)	0.867[e]
	Median; IQR (min - max)	Median; IQR (min - max)	*n* (%)	
CRP (mg/L)	2.26; 3.967 (0.07-216.64)	3.02; 4.38 (0.07-92.42)	57 (18.87)	0.098[d]
Triglycerides (mg/dL)	168.34; 121.49 (33.3-1,962.0)	166.84; 127.98 (54.05-948.77)	68 (22.52)	0.632[d]
Urine albumin (mg/L)	34.49; 247.20 (2.12-17,445.60)	46.57; 404.39 (2.12-9,062.90)	88 (29.14)	0.118[d]
UACR (mg/g)	50.74; 376.70 (0.84-15,783.59)	56.38; 488.73 (1.68-189,512.20)	90 (29.80)	0.280[d]
Sex	*n* (%)	*n* (%)	1 (0.33)	0.380[f]

Female	1,956 (39.80)	128 (42.38)	-	-
Male	2,959 (60.20)	173 (57.29)	-	-
Smoking	*n* (%)	*n* (%)	*n* (%)	0.412[f]
Former smoker	2,124 (43.22)	117 (38.74)	-	-
Non-smoker	2,002 (40.73)	129 (42.72)	-	-
Smoker	777 (15.81)	51 (16.89)	-	-
Unknown	12 (0.24)	5 (1.65)	-	-
Smoker or former smoker	2,901 (59.0)	168 (55.63)	-	-
Proteinuria	*n* (%)	*n* (%)	*n* (%)	0.175[f]
< 30 mg/g creatinine	2,368 (48.18)	95 (31.46)	-	-
30-300 mg/g creatinine	1,416 (28.81)	58 (19.20)	-	-
> 300 mg/g creatinine	1,131 (23.01)	61 (20.20)	-	-
Unknown	0 (0)	88 (29.14)	-	-
Comorbidities	*n* (%)	*n* (%)	*n* (%)	
Diabetes[a]	1,296 (26.4)	121 (40.07)	1 (0.33)	0.092[f]
Vascular disease[b]			-	0.044[f]
Positive disease history	1,116 (22.7)	83 (27.48)		
Negative disease history	3,799 (77.3)	214 (70.68)		
Unknown	0 (0)	5 (1.66)		
Hypertension[c]	4,725 (96.1)	290 (96.03)	2 (0.66)	0.756[f]
Outcome	*n* (%)	*n* (%)	*n* (%)	
Kidney failure events, total	200 (4.07)	12 (3.97)	1 (0.33)	1[f]
Dialysis	194 (3.95)	12 (3.97)	-	-
Transplantation	6 (0.12)	0 (0)	-	-

Abbreviations: BMI, body mass index; CRP, C-reactive protein; eGFR, estimated glomerular filtration rate; GCKD, German Chronic Kidney Disease; HbA1c, glycated hemoglobin; HDL, high density lipoprotein; IQR, interquartile range; LDL, low density lipoprotein; sd, standard deviation; UACR, urine albumin-to-creatinine ratio. **P*-value < 0.05. [a]Positive comorbidity of diabetes has been defined as presence of either type-1 or type-2 diabetes mellitus in the GCKD study. [b]Positive comorbidity of vascular disease has been defined as presence of coronary heart disease as well as past carotid artery surgery, carotid artery angioplasty or stent placement, or catheter angiography of peripheral arteries including angioplasty of a peripheral artery in the GCKD study. [c]Hypertension has been defined as systolic blood pressure ≥ 140mmHg, or diastolic blood pressure ≥ 90mmHg, or intake of anti-hypertensive medication in the GCKD study. [d]*P*-values for continuous variables have been determined by a Student's *t*-test. [e]*P*-values comparing two different survival curves have been determined by a log-rank test. [f]*P*-values for categorical variables have been determined by Pearson's Chi-squared test with Yates' continuity correction.

Table S2. Standard deviations of log_2-transformed variables, estimated coefficients, and coefficients of standardized log_2-transformed variables (original coefficient * standard deviation of log_2-transformed variable) in the new Z6 and Z14 risk equations ranked from top to bottom according to their appearance in the LASSO variable path.

Variable	**Standard deviation after log_2-transformation**	**Coefficients of the Z6 risk equation**	**Coefficients of the Z6 risk equation for standardized log_2-transformed variables**	**Coefficients of the Z14 risk equation**	**Coefficients of the Z14 risk equation for standardized log_2-transformed variables**
Serum cystatin C	0.428	+ 1.108	+ 0.474	+ 0.980	+ 0.419
Serum creatinine	0.441	+ 1.128	+ 0.497	+ 1.436	+ 0.633
UACR	3.123	+ 0.135	+ 0.422	+ 0.204	+ 0.637
Serum albumin	0.183	– 1.070	– 0.196	– 1.368	- 0.250
Serum urea	0.573	+ 0.125	+ 0.072	+ 0.217	+ 0.124
Hemoglobin	0.183	– 0.523	– 0.096	– 1.160	– 0.212
HbA1c	0.207	-	-	+ 0.415	+ 0.086
Serum C-reactive protein	1.676	-	-	+ 0.054	+ 0.091
Urine creatinine	1.003	-	-	– 0.087	– 0.087
LDL	0.552	-	-	– 0.091	– 0.050
Age	0.345	-	-	– 0.187	– 0.065
eGFR (CKD-EPI)	0.502	-	-	– 0.150	– 0.075
Serum sodium	0.032	-	-	+ 0.496	+ 0.016
HDL	0.477	-	-	– 0.012	– 0.006

Table S3. (a) Cause-specific concordance (*C*) statistics of the $\lambda_{PLD\text{-}min}$, $\lambda_{PLD\text{-}1sd}$, and T4 risk models for experiencing an ESKD event in the presence of death as a competing risk evaluated for 100 random subsample test sets in the GCKD resampling approach. (b) Corresponding ESKD-specific *C* statistic differences (C_{diff}) between the $\lambda_{PLD\text{-}min}$ and $\lambda_{PLD\text{-}1sd}$ risk models, respectively, and the T4, and number of resampling runs with C_{diff} larger than 0. The inverse probability of censoring weighted estimator employing the Kaplan-Meier estimator for the censoring times was used to deal with right-censored data.

(a) ESKD-specific *C* statistics					
	1 year Median (95% CI) (Mean ± SD)	**2 years** Median (95% CI) (Mean ± SD)	**3 years** Median (95% CI) (Mean ± SD)	**4 years** Median (95% CI) (Mean ± SD)	**5 years** Median (95% CI) (Mean ± SD)
$\lambda_{PLD\text{-}min}$	0.942 (0.851-0.974) (0.928±0.037)	0.909 (0.868-0.939) (0.906±0.020)	0.908 (0.880-0.936) (0.908±0.014)	0.894 (0.868-0.918) (0.894±0.014)	0.890 (0.764-0.926) (0.874±0.045)
$\lambda_{PLD\text{-}1sd}$	0.943 (0.846-0.975) (0.926±0.040)	0.909 (0.868-0.937) (0.907±0.020)	0.906 (0.875-0.933) (0.905±0.014)	0.891 (0.860-0.916) (0.891±0.015)	0.886 (0.740-0.927) (0.865±0.055)
T4	0.887 (0.785-0.952) (0.879±0.048)	0.855 (0.799-0.915) (0.857±0.031)	0.870 (0.828-0.914) (0.870±0.021)	0.862 (0.829-0.895) (0.860±0.018)	0.857 (0.689-0.903) (0.829±0.069)
(b) ESKD-specific *C* statistic differences in comparison to T4					
	1 year Median (95% CI) (Mean ± SD) # runs $C_{diff} > 0$	**2 years** Median (95% CI) (Mean ± SD) # runs $C_{diff} > 0$	**3 years** Median (95% CI) (Mean ± SD) # runs $C_{diff} > 0$	**4 years** Median (95% CI) (Mean ± SD) # runs $C_{diff} > 0$	**5 years** Median (95% CI) (Mean ± SD) # runs $C_{diff} > 0$
$\lambda_{PLD\text{-}min}$	0.046 (0.009-0.097) (0.049±0.022) 99	0.047 (0.012-0.077) (0.049±0.018) 100	0.038 (0.013-0.058) (0.038±0.011) 100	0.033 (0.014-0.051) (0.033±0.010) 99	0.036 (0.010-0.110) (0.045±0.028) 99
$\lambda_{PLD\text{-}1sd}$	0.045 (-0.004-0.100) (0.047±0.025) 96	0.050 (0.015-0.081) (0.050±0.018) 100	0.035 (0.009-0.056) (0.035±0.012) 99	0.031 (0.011-0.046) (0.030±0.011) 98	0.033 (0.006-0.086) (0.036±0.020) 98

Abbreviations: C_{diff}, concordance statistic difference in comparison to T4; $\lambda_{PLD\text{-}min}$, LASSO Cox PH model with λ parameter optimization in an internal 5-fold cross-validation to yield the minimum partial likelihood deviance (PLD); $\lambda_{PLD\text{-}1sd}$, LASSO Cox PH model with λ parameter optimization in an internal 5-fold cross-validation to yield maximum penalty while keeping the PLD within one standard deviation of the minimum PLD; T4, original four-variable ESKD risk equation developed by Tangri et al.[S25] without coefficient recalibration to the GCKD cohort.

Table S4. Categorical net reclassification improvement values for the GCKD test data sets averaged over 100 subsampling runs comparing from left to right: the $\lambda_{PLD\text{-}min}$ *vs.* the GCKD-recalibrated T4 (T4-surv-recal), the $\lambda_{PLD\text{-}1sd}$ *vs.* the T4-surv-recal, the $\lambda_{PLD\text{-}min}$ *vs.* the original T4, the $\lambda_{PLD\text{-}1sd}$ *vs.* the original T4, and the $\lambda_{PLD\text{-}1sd}$ *vs.* the $\lambda_{PLD\text{-}min}$ risk equation evaluated (a) one, (b) two, (c) three, (d) four, and (e) five years after the baseline visit. Please note that at five years after the baseline visit, a high censoring rate is apparent in the Kaplan-Meier curve in Figure S2A, which impairs accurate estimation of absolute risk probabilities. The predicted risk probabilities were divided into 3 categories: 0% - <3%, 3% - <10%, 10% - 100%, respectively.

(a) Risk probability prediction one year after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
NRI	9.7 (-22.9 – 40.5) (8.2 ± 16.1)	9.6 (-20.2 – 40.5) (8.6 ± 15.3)	-	-	-0.1 (-21.2 – 21.6) (-0.03 ± 10.1)
NRI^+	9.6 (-22.6 – 40.0) (8.0 ± 16.0)	9.5 (-20.0 – 39.9) (8.3 ± 15.2)	-	-	0.0 (-21.1 – 21.1) (-0.2 ± 10.1)
NRI^-	0.2 (-0.7 – 1.1) (0.2 ± 0.5)	0.4 (-0.6 – 1.4) (0.3 ± 0.5)	-	-	0.1 (-0.3 – 0.7) (0.1 ± 0.2)
(b) Risk probability prediction two years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
NRI	10.5 (-8.0 – 10.2) (10.2 ± 9.2)	9.8 (-7.9 – 26.2) (9.6 ± 9.2)	1.9 (-14.8 – 20.8) (2.3 ± 9.6)	1.5 (-16.0 – 19.2) (1.6 ± 9.8)	0.2 (-12.2 – 14.3) (0.6 ± 6.9)
NRI^+	7.0 (-11.5 – 24.8) (7.3 ± 9.5)	7.3 (-10.3 – 23.4) (6.7 ± 9.4)	-3.9 (-20.4 – 16.3) (-3.1 ± 9.9)	-3.8 (-21.8 – 14.6) (-3.9 ± 10.0)	0.0 (-12.3 – 14.5) (0.5 ± 6.9)
NRI^-	2.8 (1.2 – 4.6) (2.8 ± 0.9)*	2.9 (1.3 – 4.8) (2.9 ± 0.9)*	5.4 (3.2 – 7.8) (5.4 ± 1.2)*	5.4 (3.3 – 7.7) (5.4 ± 1.2)*	0.1 (-0.6 – 0.6) (0.1 ± 0.3)
(c) Risk probability prediction three years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
NRI	10.2 (-3.3 – 24.6) (10.4 ± 8.3)	8.7 (-7.4 – 22.6) (8.6 ± 7.8)	-	-	-1.2 (-13.8 – 6.5) (-1.9 ± 5.3)
NRI^+	4.7 (-10.0 – 20.7) (5.0 ± 8.3)	4.0 (-12.1 – 18.5) (3.7 ± 7.9)	-	-	-0.2 (-12.6 – 7.3) (-1.3 ± 5.2)
NRI^-	5.3 (3.1 – 7.4) (5.3 ± 1.2)*	4.7 (2.6 – 7.1) (4.8 ± 1.2)*	-	-	-0.5 (-2.0 – 0.4) (-0.6 ± 0.7)
(d) Risk probability prediction four years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
NRI	14.5 (3.5 – 24.1) (13.9 ± 5.6)*	12.7 (-2.1 – 22.3) (12.1 ± 6.2)	-	-	-1.8 (-6.7 – 4.8) (-1.4 ± 3.2)
NRI^+	7.1 (-5.4 – 16.4) (6.4 ± 6.1)	6.6 (-8.3 – 15.5) (5.6 ± 6.5)	-	-	0.0 (-6.2 – 5.1) (-0.3 ± 3.1)
NRI^-	7.4 (4.7 – 11.2) (7.5 ± 1.7)*	6.6 (3.2 – 9.8) (6.4 ± 1.7)*	-	-	-0.9 (-3.2 – 0.6) (-1.1 ± 1.0)
(e) Risk probability prediction five years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal	$\lambda_{PLD\text{-}min}$ *vs.* T4	$\lambda_{PLD\text{-}1sd}$ *vs.* T4	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$

	Median (95% CI)[a] (Mean ± SD)	Median (95% CI)[a] (Mean ± SD)	Median (95% CI)[a] (Mean ± SD)	Median (95% CI)[a] (Mean ± SD)	Median (95% CI)[a] (Mean ± SD)
NRI	14.6 (-1.6 – 39.4) (16.8 ± 11.1)	10.7 (-7.2 – 31.6) (11.7 ± 10.7)	15.9 (-2.0 – 35.4) (15.9 ± 10.2)	11.2 (-6.2 – 31.1) (10.8 ± 10.3)	-3.0 (-22.0 – 5.1) (-4.8 ± 6.9)
NRI^+	3.9 (-4.7 – 10.9) (3.5 ± 4.1)	2.8 (-6.8 – 9.8) (2.8 ± 4.3)	4.6 (-4.4 – 12.6) (4.1 ± 4.3)	3.8 (-5.5 – 11.6) (3.4 ± 4.4)	-0.7 (-5.0 – 4.2) (-0.5 ± 2.3)
NRI^-	11.8 (0.0 – 34.4) (13.3 ± 10.4)*	8.7 (-7.2 – 30.4) (9.0 ± 10.0)	12.5 (-8.7 – 29.0) (11.8 ± 9.7)	7.1 (-11.8 – 24.3) (7.4 ± 10.0)	0.0 (-21.2 – 3.3) (-4.3 ± 6.2)

Abbreviations: CI, confidence interval; $\lambda_{PLD\text{-}min}$, LASSO Cox PH risk equation with λ parameter optimization in an internal 5-fold cross-validation to yield the minimum partial likelihood deviance (PLD) and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; $\lambda_{PLD\text{-}1sd}$, LASSO Cox PH risk equation with λ parameter optimization in an internal 5-fold cross-validation to yield maximum penalty while simultaneously keeping the PLD within one standard deviation of the minimum PLD and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; NRI, net reclassification improvement; NRI^+, net reclassification improvement considering only patients with an event during the observation period; NRI^-, net reclassification improvement considering only patients without an event during the observation period; SD, standard deviation; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets. [a]The 95% CIs were determined empirically over 100 subsampling runs. [b]Please note, that Tangri et al. provided only survival rates for two and five years after the baseline visit.[S14] *significant NRIs, i.e. 95% CIs excluding zero.

Table S5. Continuous (category-free) net reclassification improvement values for the GCKD test data sets averaged across 100 subsampling runs comparing from left to right: the $\lambda_{PLD\text{-}min}$ *vs.* the GCKD-recalibrated T4 (T4-surv-recal), the $\lambda_{PLD\text{-}1sd}$ *vs.* the T4-surv-recal, the $\lambda_{PLD\text{-}min}$ *vs.* the original T4, the $\lambda_{PLD\text{-}1sd}$ *vs.* the original T4, and the $\lambda_{PLD\text{-}1sd}$ *vs.* the $\lambda_{PLD\text{-}min}$ risk equations evaluated (a) one, (b) two, (c) three, (d) four, and (e) five years after the baseline visit. Please note that at five years after the baseline visit, a high censoring rate is apparent in the Kaplan-Meier curve in Figure S2A, which impairs accurate estimation of absolute risk probabilities.

(a) Risk probability prediction one year after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
cNRI	57.8 (-4.2 – 110.5) (56.6 ± 28.6)	38.2 (-20.7 – 99.1) (36.2 ± 29.9)	-	-	-53.6 (-120.2 – 17.9) (-52.2 ± 35.4)
$cNRI^+$	45.5 (-10.6 – 100.0) (45.8 ± 28.6)	36.6 (-33.5 – 100.0) (36.8 ± 30.8)	-	-	-20.0 (-86.4 – 46.6) (-18.6 ± 33.7)
$cNRI^-$	10.6 (-4.1 – 26.5) (10.8 ± 8.6)	-1.4 (-15.4 – 15.3) (-0.6 ± 8.4)	-	-	-33.4 (-48.7 – -19.4) (-33.6 ± 8.0)*
(b) Risk probability prediction two years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
cNRI	39.5 (11.0 – 71.2) (39.8 ± 16.8)*	24.1 (-11.7 – 61.1) (23.9 ± 17.3)	9.0 (-38.9 – 40.7) (5.6 ± 20.6)	-3.3 (-48.0 – 35.2) (-4.2 ± 21.5)	-38.4 (-83.2 – -0.1) (-40.3 ± 20.6)*
$cNRI^+$	30.3 (-4.6 – 62.4) (31.3 ± 16.8)	27.8 (-14.2 – 53.2) (26.8 ± 17.0)	10.0 (-20.7 – 39.4) (10.0 ± 18.6)	8.2 (-24.7 – 44.0) (9.8 ± 18.2)	-5.9 (-40.3 – 27.7) (-7.3 ± 18.4)
$cNRI^-$	9.3 (-6.3 – 24.6) (8.5 ± 8.5)	-3.7 (-18.7 – 11.2) (-2.9 ± 8.3)	-6.1 (-23.1 – 17.9) (-4.4 ± 11.4)	-15.9 (-31.5 – 10.8) (-14.0 ± 11.3)	-33.0 (-48.9 – -19.3) (-33.1 ± 7.8)*
(c) Risk probability prediction three years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
cNRI	25.1 (-2.6 – 58.5) (25.0 ± 15.1)	6.8 (-25.8 – 32.5) (5.0 ± 15.6)	-	-	-48.6 (-73.9 – -13.0) (-47.8 ± 16.5)*
$cNRI^+$	18.0 (-10.6 – 44.4) (16.9 ± 14.5)	10.0 (-22.1 – 30.4) (8.5 ± 15.1)	-	-	-15.8 (-40.0 – 23.5) (-14.8 ± 16.1)
$cNRI^-$	8.6 (-6.0 – 24.4) (8.1 ± 8.5)	-4.0 (-18.2 – 11.3) (-3.6 ± 8.0)	-	-	-33.3 (-49.4 – -18.6) (-33.0 ± 8.1)*
(d) Risk probability prediction four years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD
cNRI	16.4 (-3.6 – 41.9) (17.6 ± 13.1)	2.3 (-28.7 – 25.1) (1.1 ± 14.3)	-	-	-45.1 (-69.6 – -12.3) (-43.1 ± 15.1)*
$cNRI^+$	9.6 (-12.5 – 32.9) (9.7 ± 12.8)	4.5 (-23.5 – 30.2) (4.3 ± 13.3)	-	-	-13.5 (-35.0 – 20.2) (-11.1 ± 14.2)
$cNRI^-$	7.5 (-4.9 – 23.9) (7.9 ± 8.2)	-3.2 (-19.2 – 10.1) (-3.2 ± 7.9)	-	-	-32.1 (-48.8 – -14.0) (-32.1 ± 9.4)*
(e) Risk probability prediction five years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD
cNRI	35.0 (-14.3 – 91.0) (35.9 ± 30.3)	19.3 (-36.4 – 72.7) (18.4 ± 29.8)	-4.6 (-56.0 – 71.7) (-1.0 ± 31.0)	-17.7 (-66.2 – 45.4) (-15.6 ± 30.3)	-67.9 (-103.3 – -4.1) (-65.2 ± 27.4)*
$cNRI^+$	13.6 (-23.3 – 40.0) (13.1 ± 15.6)	7.8 (-22.9 – 35.6) (7.2 ± 15.8)	0.5 (-38.7 – 31.7) (-0.7 ± 17.3)	-3.4 (-37.3 – 26.9) (-3.8 ± 16.5)	-20.4 (-51.0 – 7.8) (-20.4 ± 16.8)

cNRI$^-$	23.6 (-26.0 – 74.8) (22.8 ± 26.0)	9.1 (-39.7 – 57.1) (11.2 ± 25.0)	0.0 (-42.9 – 57.1) (-0.3 ± 28.0)	-16.7 (-52.9 – 40.8) (-11.7 ± 26.0)	-45.5 (-84.0 – 0.0) (-44.8 ± 22.3)

Abbreviations: CI, confidence interval; $\lambda_{PLD\text{-}min}$, LASSO Cox PH risk equation with λ parameter optimization in an internal 5-fold cross-validation to yield the minimum partial likelihood deviance (PLD) and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; $\lambda_{PLD\text{-}1sd}$, LASSO Cox PH risk equation with λ parameter optimization in an internal 5-fold cross-validation to yield maximum penalty while simultaneously keeping the PLD within one standard deviation of the minimum PLD and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; cNRI, continuous net reclassification improvement; cNRI$^+$, continuous net reclassification improvement considering only patients with an event during the observation period; cNRI$^-$, continuous net reclassification improvement considering only patients without an event during the observation period; SD, standard deviation; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets. [a]The 95% CIs were determined empirically over 100 subsampling runs. [b]Please note, that Tangri et al. provided only survival rates for two and five years after the baseline visit.[S14] *significant cNRIs, i.e. 95% CIs excluding zero.

Table S6. Sensitivities, specificities, balanced accuracies, positive predictive values, and negative predictive values for the new $\lambda_{PLD\text{-}min}$ and $\lambda_{PLD\text{-}1sd}$ as well as the T4 score evaluated on the GCKD test data sets at (a) one, (b) two, (c) three, (d) four, and (e) five years after the baseline visit and each averaged over 100 subsampling runs. All performance measures were evaluated for both the "best" and the "Youden" cut-off, individually optimized for each observation time period, risk score, and test set, respectively, and then averaged over 100 test sets. Corresponding averaged receiver operating characteristic curves and cut-off points are displayed in Figure S8. Additionally, area under the ROC curve values averaged across 100 subsampling runs are given. Of note, at five years after the baseline visit, a high censoring rate is apparent in the Kaplan-Meier curve in Figure S2A, which affects the ROC curve evaluation. On average, 33, 99, 182, 773, and 1560 GCKD study participants had to be excluded from the ROC curve analyses due to censoring or death one, two, three, four, and five years after the baseline visit, respectively.

(a) ROC curve evaluation one year after baseline visit						
Cut-off	**"Best"**			**"Youden"**		
	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)
Sensitivity (%)	100.0 (83.3 – 100.0) (94.0 ± 6.7)	100.0 (83.3 – 100.0) (94.4 ± 6.4)	88.9 (68.3 – 100.0) (88.2 ± 10.1)	100.0 (81.6 – 100.0) (94.9 ± 6.8)	100.0 (83.3 – 100.0) (95.7 ± 6.3)	90.0 (68.3 – 100.0) (89.8 ± 9.9)
Specificity (%)	89.0 (79.3 – 95.3) (89.0 ± 4.3)	88.5 (82.9 – 96.3) (88.9 ± 3.8)	85.0 (81.9 – 89.8) (84.8 ± 3.9)	88.4 (79.1 – 95.3) (88.3 ± 4.4)	86.9 (82.7 – 96.3) (88.1 ± 3.9)	85.0 (61.2 – 91.0) (83.5 ± 6.9)
Balanced accuracy (%)	91.6 (83.6 – 97.6) (91.5 ± 4.2)	91.7 (84.0 – 98.0) (91.7 ± 3.8)	86.9 (75.5 – 93.7) (86.5 ± 5.4)	92.2 (83.9 – 97.6) (91.6 ± 4.2)	92.5 (84.0 – 98.0) (91.9 ± 3.7)	86.9 (75.9 – 93.7) (86.9 ± 5.2)
PPV (%)	4.2 (1.8 – 9.4) (4.4 ± 1.9)	4.1 (1.9 – 8.1) (4.3 ± 1.7)	2.7 (1.1 – 4.5) (2.9 ± 1.6)	4.0 (1.8 – 8.5) (4.2 ± 1.7)	3.7 (1.9 – 7.8) (4.1 ± 1.5)	2.7 (1.1 – 4.5) (2.9 ± 1.7)
NPV (%)	100.0 (99.9 – 100.0) (100.0 ± 0.0)	100.0 (99.9 – 100.0) (100.0 ± 0.0)	99.9 (99.8 – 100.0) (99.9 ± 0.1)	100.0 (99.9 – 100.0) (100.0 ± 0.0)	100.0 (99.9 – 100.0) (100.0 ± 0.0)	99.9 (99.9 – 100.0) (99.9 ± 0.1)
AUC-ROC	0.944 (0.853 – 0.976) (0.930 ± 0.037)	0.945 (0.848 – 0.977) (0.928 ± 0.040)	0.889 (0.786 – 0.953) (0.880 ± 0.048)	0.944 (0.853 – 0.976) (0.930 ± 0.037)	0.945 (0.848 – 0.977) (0.928 ± 0.040)	0.889 (0.786 – 0.953) (0.880 ± 0.048)
(b) ROC curve evaluation two years after baseline visit						
Cut-off	**"Best"**			**"Youden"**		
	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)
Sensitivity (%)	87.0 (77.7 – 95.0) (86.6 ± 4.6)	87.0 (78.3 – 94.3) (86.8 ± 4.4)	79.2 (67.9 – 88.6) (79.0 ± 5.4)	89.3 (75.0 – 96.3) (88.1 ± 6.2)	88.1 (78.6 – 98.4) (88.2 ± 5.9)	80.0 (65.9 – 96.3) (81.1 ± 8.4)
Specificity (%)	84.6 (79.5 – 91.2) (84.9 ± 3.4)	85.3 (78.9 – 90.4) (85.3 ± 3.0)	80.3 (67.8 – 88.6) (80.8 ± 4.8)	84.4 (72.1 – 92.7) (83.7 ± 5.4)	85.3 (74.9 – 90.2) (84.4 ± 4.7)	80.3 (62.7 – 91.5) (79.4 ± 8.1)
Balanced accuracy (%)	85.9 (80.2 – 90.9) (85.7 ± 2.8)	86.2 (81.2 – 90.7) (86.1 ± 2.7)	79.9 (74.9 – 86.6) (79.9 ± 3.2)	86.0 (80.9 – 90.9) (85.9 ± 2.7)	86.6 (81.2 – 90.9) (86.3 ± 2.6)	80.2 (75.2 – 87.0) (80.3 ± 3.0)
PPV (%)	8.2 (4.6 – 13.2) (8.5 ± 1.2)	8.8 (5.3 – 12.1) (8.7 ± 2.0)	6.0 (3.6 – 10.0) (6.3 ± 1.6)	8.0 (4.5 – 13.2) (8.0 ± 2.2)	8.4 (5.3 – 12.0) (8.5 ± 2.1)	6.0 (3.6 – 11.6) (6.4 ± 2.2)
NPV (%)	99.8 (99.6 – 99.9) (99.8 ± 0.1)	99.8 (99.6 – 99.9) (99.8 ± 0.1)	99.6 (99.3 – 99.8) (99.6 ± 0.1)	99.8 (99.6 – 99.9) (99.8 ± 0.1)	99.8 (99.6 – 100.0) (99.8 ± 0.1)	99.6 (99.3 – 99.9) (99.6 ± 0.2)
AUC-ROC	0.914 (0.873 – 0.943) (0.911 ± 0.020)	0.915 (0.874 – 0.943) (0.913 ± 0.020)	0.861 (0.803 - 0.919) (0.861 ± 0.031)	0.914 (0.873 – 0.943) (0.911 ± 0.020)	0.915 (0.874 – 0.943) (0.913 ± 0.020)	0.861 (0.803 - 0.919) (0.861 ± 0.031)
(c) ROC curve evaluation three years after baseline visit						
Cut-off	**"Best"**			**"Youden"**		
	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)
Sensitivity (%)	87.0 (80.9 – 93.6) (87.1 ± 3.4)	87.4 (80.6 – 94.7) (87.1 ± 3.6)	81.3 (74.7 – 90.2) (81.6 ± 3.9)	89.5 (80.7 – 97.3) (89.1 ± 4.7)	89.2 (81.0 – 97.6) (89.1 ± 4.5)	83.8 (70.3 – 95.9) (83.9 ± 7.1)
Specificity (%)	84.8 (79.8 – 88.6) (84.9 ± 2.2)	84.7 (79.5 – 88.6) (84.4 ± 2.5)	80.7 (70.9 – 87.6) (80.7 ± 3.8)	83.7 (74.2 – 88.6) (83.3 ± 3.7)	83.6 (74.3 – 88.4) (82.8 ± 3.9)	80.1 (66.0 – 88.3) (79.0 ± 6.8)
Balanced accuracy (%)	86.1 (82.4 – 89.3) (86.0 ± 1.9)	85.8 (82.0 – 89.1) (85.8 ± 1.8)	81.0 (77.1 – 85.9) (81.2 ± 2.4)	86.3 (82.4 – 89.8) (86.2 ± 1.9)	86.0 (82.0 – 89.5) (86.0 ± 1.8)	81.3 (77.4 – 85.9) (81.4 ± 2.3)
PPV (%)	14.7 (10.7 – 19.3) (14.7 ± 2.0)	14.5 (9.8 – 18.4) (14.3 ± 2.1)	11.5 (7.0 – 15.9) (11.3 ± 2.1)	14.2 (9.6 – 18.9) (13.9 ± 2.3)	13.5 (9.0 – 17.7) (13.6 ± 2.3)	11.5 (6.8 – 16.2) (11.2 ± 2.6)
NPV (%)	99.5 (99.3 – 99.8) (99.6 ± 0.1)	99.5 (99.3 – 99.8) (99.6 ± 0.1)	99.3 (99.0 – 99.7) (99.3 ± 0.2)	99.6 (99.3 – 99.9) (99.6 ± 0.2)	99.6 (99.3 – 99.9) (99.6 ± 0.2)	99.4 (99.0 – 99.8) (99.4 ± 0.2)
AUC-ROC	0.919 (0.891 – 0.946) (0.919 ± 0.014)	0.917 (0.885 – 0.943) (0.917 ± 0.014)	0.878 (0.837 – 0.922) (0.879 ± 0.021)	0.919 (0.891 – 0.946) (0.919 ± 0.014)	0.917 (0.885 – 0.943) (0.917 ± 0.014)	0.878 (0.837 – 0.922) (0.879 ± 0.021)
(d) ROC curve evaluation four years after baseline visit						
Cut-off	**"Best"**			**"Youden"**		
	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)
Sensitivity (%)	86.8 (79.5 – 91.9)	85.9 (79.1 – 92.3)	80.3 (72.9 – 86.7)	89.2 (79.5 – 97.2)	87.6 (78.5 – 97.2)	81.5 (71.3 – 92.0)

	(86.8 ± 3.2)	(85.9 ± 3.5)	(80.5 ± 3.6)	(89.1 ± 4.3)	(87.5 ± 4.9)	(81.5 ± 5.3)
Specificity (%)	83.4 (78.7 – 88.3) (83.4 ± 2.6)	83.8 (78.2 – 87.9) (83.5 ± 2.5)	82.0 (75.7 – 86.6) (81.4 ± 2.7)	81.1 (74.3 – 88.6) (81.5 ± 3.8)	82.8 (74.4 – 88.0) (82.2 ± 3.7)	81.9 (68.5 – 88.7) (80.7 ± 5.2)
Balanced accuracy (%)	85.1 (81.9 – 88.3) (85.1 ± 1.6)	84.8 (81.2 – 88.0) (84.7 ± 1.8)	81.0 (77.5 – 84.5) (80.9 ± 2.0)	85.3 (82.0 – 88.4) (85.3 ± 1.6)	85.0 (81.3 – 88.1) (84.9 ± 1.8)	81.2 (77.6 – 84.7) (81.1 ± 1.9)
PPV (%)	19.9 (14.7 – 25.3) (19.8 ± 3.0)	19.6 (14.0 – 26.1) (19.7 ± 3.0)	17.0 (12.2 – 21.8) (16.9 ± 2.4)	18.3 (14.1 – 25.5) (18.7 ± 3.3)	18.8 (13.2 – 26.1) (19.0 ± 3.1)	17.1 (11.0 – 23.0) (17.0 ± 3.2)
NPV (%)	99.3 (98.9 – 99.6) (99.3 ± 0.2)	99.2 (98.9 – 99.6) (99.2 ± 0.2)	98.9 (98.5 – 99.3) (98.9 ± 0.2)	99.4 (98.9 – 99.8) (99.4 ± 0.2)	99.3 (98.9 – 99.8) (99.3 ± 0.3)	99.0 (98.5 – 99.5) (99.0 ± 0.3)
AUC-ROC	0.914 (0.884 – 0.936) (0.913 ± 0.014)	0.911 (0.880 – 0.933) (0.910 ± 0.015)	0.878 (0.842 – 0.911) (0.876 ± 0.018)	0.914 (0.884 – 0.936) (0.913 ± 0.014)	0.911 (0.880 – 0.933) (0.910 ± 0.015)	0.878 (0.842 – 0.911) (0.876 ± 0.018)
(e) ROC curve evaluation five years after baseline visit						
Cut-off	**"Best"**			**"Youden"**		
	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)
Sensitivity (%)	87.2 (75.6 – 100.0) (87.9 ± 7.3)	86.4 (74.0 – 98.5) (86.8 ± 6.4)	82.2 (69.4 – 92.0) (81.7 ± 6.2)	91.7 (68.7 – 100.0) (89.0 ± 8.8)	90.4 (73.4 – 99.4) (89.4 ± 7.4)	81.3 (59.8 – 92.7) (80.3 ± 8.7)
Specificity (%)	85.7 (72.1 – 100.0) (85.3 ± 6.9)	85.2 (74.1 – 100.0) (85.6 ± 6.4)	85.7 (73.8 – 100.0) (85.3 ± 6.8)	85.2 (68.8 – 100.0) (85.1 ± 9.3)	83.3 (68.8 – 100.0) (83.9 ± 9.3)	88.2 (70.0 – 100.0) (87.3 ± 8.3)
Balanced accuracy (%)	86.2 (78.2 – 95.8) (86.6 ± 4.6)	86.6 (78.1 – 94.6) (86.2 ± 4.6)	83.4 (74.9 – 93.1) (83.5 ± 4.8)	86.5 (78.6 – 95.8) (87.1 ± 4.4)	86.6 (78.3 – 94.6) (86.6 ± 4.4)	83.6 (75.1 – 93.1) (83.8 ± 4.7)
PPV (%)	25.6 (13.7 – 100.0) (29.9 ± 19.4)	24.5 (14.4 – 100.0) (30.0 ± 19.6)	22.1 (13.0 – 100.0) (28.6 ± 19.9)	25.6 (13.1 – 100.0) (36.3 ± 29.0)	23.1 (13.0 – 100.0) (32.0 ± 24.5)	26.6 (13.2 – 100.0) (36.5 ± 28.0)
NPV (%)	99.3 (98.2 – 100.0) (99.2 ± 0.5)	99.2 (98.2 – 99.9) (99.2 ± 0.5)	98.9 (97.7 – 99.6) (98.8 ± 0.5)	99.4 (98.0 – 100.0) (99.3 ± 0.5)	99.4 (98.3 – 100.0) (99.3 ± 0.5)	98.8 (97.6 – 99.6) (98.8 ± 0.5)
AUC-ROC	0.921 (0.858 – 0.981) (0.921 ± 0.032)	0.918 (0.847 – 0.975) (0.916 ± 0.036)	0.880 (0.811 – 0.940) (0.876 – 0.039)	0.921 (0.858 – 0.981) (0.921 ± 0.032)	0.918 (0.847 – 0.975) (0.916 ± 0.036)	0.880 (0.811 – 0.940) (0.876 – 0.039)

Abbreviations: AUC-ROC, area under the receiver operating characteristic curve; CI, confidence interval; $\lambda_{PLD\text{-}min}$, LASSO Cox PH risk model newly derived by fitting a LASSO Cox PH regression on a GCKD training cohort in the resampling approach, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield the minimum partial likelihood deviance (PLD); $\lambda_{PLD\text{-}1sd}$, LASSO Cox PH risk model newly derived by fitting a LASSO Cox PH regression on a GCKD training cohort in the resampling approach, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield the most penalized model such that the PLD remained within one standard deviation of the minimum PLD; NPV, negative predictive value; PPV, positive predictive value; SD, standard deviation; T4, original 4-variable Tangri risk equation. [a]The 95% CIs were determined empirically over 100 subsampling runs.

Table S7. Categorical net reclassification improvement values additionally censoring at the date of non-fatal cardiovascular events for the GCKD test data sets averaged over 100 subsampling runs comparing from left to right: the $\lambda_{PLD\text{-}min}$ *vs.* the GCKD-recalibrated T4 (T4-surv-recal), the $\lambda_{PLD\text{-}1sd}$ *vs.* the T4-surv-recal, the $\lambda_{PLD\text{-}min}$ *vs.* the original T4, the $\lambda_{PLD\text{-}1sd}$ *vs.* the original T4, and the $\lambda_{PLD\text{-}1sd}$ *vs.* the $\lambda_{PLD\text{-}min}$ risk equation evaluated (a) one, (b) two, (c) three, (d) four, and (e) five years after the baseline visit. The results at the 5-year time point are strongly affected by the high censoring rate at that time point. The predicted risk probabilities were divided into 3 categories: 0% - <3%, 3% - <10%, 10% - 100%, respectively. This analysis did not comprise a training of new predictive models, but only involved a re-evaluation of the test data and a re-calibration of the risk equations on the respective training set now additionally censoring at the date of non-fatal cardiovascular events. The following non-fatal cardiovascular events were censored: myocardial infarction, non-hemorrhagic stroke, and coronary revascularization procedures.

(a) Risk probability prediction one year after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
NRI	0.4 (-19.3 – 34.2) (4.1 ± 14.5)	0.7 (-19.1 – 34.0) (5.7 ± 14.0)	-	-	0.2 (-22.6 – 24.9) (1.3 ± 11.3)
NRI^+	0.04 (-19.9 – 33.4) (3.9 ± 14.5)	0.1 (-19.9 – 33.1) (5.6 ± 13.9)	-	-	0.0 (-22.5 – 24.8) (1.2 ± 11.3)
NRI^-	0.2 (-0.7 – 1.0) (0.2 ± 0.5)	0.4 (-0.5 – 1.2) (0.3 ± 0.5)	-	-	0.1 (-0.2 – 0.6) (0.2 ± 0.2)
(b) Risk probability prediction two years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
NRI	11.0 (-9.1 – 26.7) (10.7 ± 10.3)	11.5 (-9.6 – 29.9) (10.7 ± 10.4)	-0.5 (-17.9 – 19.7) (-0.1 ± 10.9)	0.7 (-18.2 – 20.1) (0.3 ± 10.9)	0.3 (-14.5 – 13.1) (0.5 ± 7.2)
NRI^+	8.1 (-12.5 – 25.1) (8.2 ± 10.5)	9.0 (-11.5 – 27.2) (8.1 ± 10.5)	-6.6 (-25.9 – 13.3) (-6.7 ± 11.2)	-5.6 (-25.6 – 14 2) (-6.4 ± 11.2)	0.0 (-14.8 – 13.2) (0.4 ± 7.2)
NRI^-	2.5 (1.1 – 3.8) (2.5 ± 0.8)*	2.5 (1.3 – 4.1) (2.6 ± 0.8)*	6.6 (4.5 – 9.2) (6.7 ± 1.2)*	6.8 (4.7 – 9.1) (6.8 ± 1.1)*	0.1 (-0.4 – 0.8) (0.2 ± 0.3)
(c) Risk probability prediction three years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
NRI	9.7 (-5.7 – 21.7) (9.3 ± 7.6)	8.5 (-7.5 – 20.5) (7.8 ± 7.5)	-	-	-0.3 (-10.9 – 7.7) (-1.0 ± 5.0)
NRI^+	4.7 (-11.2 – 16.7) (4.3 ± 7.7)	3.9 (-13.1 – 17.0) (3.2 ± 7.7)	-	-	0.0 (-10.8 – 8.0) (-6.5 ± 5.0)
NRI^-	4.9 (3.3 – 6.9) (5.0 ± 1.0)*	4.5 (2.6 – 6.9) (4.6 ± 1.1)*	-	-	-0.4 (-1.8 – 0.7) (-0.4 ± 0.6)
(d) Risk probability prediction four years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
NRI	13.6 (1.6 – 22.6) (13.0 ± 5.9)*	13.4 (-1.7 – 22.1) (11.7 ± 6.8)	-	-	-0.5 (-6.8 – 5.7) (-0.8 ± 3.5)
NRI^+	6.9 (-7.1 – 15.2) (6.1 ± 6.4)	6.2 (-8.4 – 17.1) (5.7 ± 7.0)	-	-	0.1 (-6.0 – 6.5) (0.0 ± 3.2)
NRI^-	6.7 (4.0 – 10.8)	6.2 (2.6 – 9.3)	-	-	-0.8 (-2.9 – 1.1)

	(6.8 ± 1.8)*	(6.0 ± 1.8)*			(-0.9 ± 1.1)
(e) Risk probability prediction five years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
NRI	9.0 (-6.3 – 27.6) (9.0 ± 9.0)	5.8 (-11.8 – 25.1) (6.8 ± 8.9)	9.3 (-7.3 – 31.5) (10.2 ± 10.1)	8.2 (-8.6 – 27.3) (8.1 ± 9.8)	-0.1 (-14.6 – 7.6) (-1.9 ± 6.2)
NRI^+	4.7 (-6.4 – 11.8) (4.5 ± 4.9)	5.2 (-6.8 – 12.6) (4.9 ± 5.0)	-1.4 (-12.3 – 8.1) (-1.4 ± 4.9)	-0.6 (-11.0 – 8.5) (-0.9 ± 5.0)	0.3 (-3.9 – 5.7) (0.7 ± 2.5)
NRI^-	0.0 (-9.6 – 20.0) (4.5 ± 8.7)	0.0 (-12.8 – 19.4) (1.9 ± 8.0)	10.8 (-4.8 – 33.3) (11.6 ± 9.5)	9.1 (-8.2 – 27.6) (9.0 ± 9.2)	0.0 (-14.3 – 8.4) (-2.6 ± 5.9)

Abbreviations: CI, confidence interval; $\lambda_{PLD\text{-}min}$, LASSO Cox PH risk equation with λ parameter optimization to yield the minimum partial likelihood deviance (PLD) and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; $\lambda_{PLD\text{-}1sd}$, LASSO Cox PH risk equation with λ parameter optimization to yield maximum penalty while simultaneously keeping the PLD within one standard deviation of the minimum PLD and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; NRI, net reclassification improvement; NRI+, net reclassification improvement considering only patients with an event during the observation period; NRI-, net reclassification improvement considering only patients without an event during the observation period; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets. [a]The 95% CIs were determined empirically over 100 subsampling runs. [b]Please note, that Tangri et al. provided only survival rates for two and five years after the baseline visit.[S14] *significant NRIs, i.e. 95% CIs excluding zero.

Table S8. Continuous (category-free) net reclassification improvement values additionally censoring at the date of non-fatal cardiovascular events for the GCKD test data sets averaged over 100 subsampling runs comparing from left to right: the $\lambda_{PLD\text{-}min}$ *vs.* the GCKD-recalibrated T4 (T4-surv-recal), the $\lambda_{PLD\text{-}1sd}$ *vs.* the T4-surv-recal, the $\lambda_{PLD\text{-}min}$ *vs.* the original T4, the $\lambda_{PLD\text{-}1sd}$ *vs.* the original T4, and the $\lambda_{PLD\text{-}1sd}$ *vs.* the $\lambda_{PLD\text{-}min}$ risk equations evaluated (a) one, (b) two, (c) three, (d) four, and (e) five years after the baseline visit. The results at the 5-year time point are strongly affected by the high censoring rate at that time point. This analysis did not comprise a training of new predictive models, but only involved a re-evaluation of the test data and a re-calibration of the risk equations on the respective training set now additionally censoring at the date of non-fatal cardiovascular events. The following non-fatal cardiovascular events were censored: myocardial infarction, non-hemorrhagic stroke, and coronary revascularization procedures.

(a) Risk probability prediction one year after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
cNRI	59.3 (3.6 – 109.6) (55.9 ± 29.6)*	34.4 (-19.6 – 97.5) (36.0 ± 31.0)	-	-	-48.2 (-118.9 – 17.1) (-49.3 ± 35.5)
$cNRI^+$	50.2 (-16.2 – 100.0) (49.5 ± 30.1)	33.6 (-27.3 – 100.0) (40.3 ± 31.4)	-	-	-14.5 (-100.0 – 42.8) (-18.0 ± 34.3)
$cNRI^-$	6.4 (-9.8 – 24.0) (6.4 ± 9.5)	-4.8 (-20.1 – 13.2) (-4.3 ± 9.1)	-	-	-31.4 (-48.9 – -14.7) (-31.4 ± 9.3)*
(b) Risk probability prediction two years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
cNRI	36.9 (5.9 – 66.7) (36.7 ± 17.0)*	19.4 (-13.7 – 59.1) (19.6 ± 19.1)	11.4 (-34.9 – 48.2) (10.4 ± 23.7)	5.5 (-32.5 – 49.6) (4.7 ± 21.7)	-33.5 (-73.5 – 5.1) (-35.7 ± 20.6)
$cNRI^+$	36.3 (-3.7 – 60.3) (33.0 ± 17.5)	25.4 (-10.5 – 57.4) (26.7 ± 18.4)	-2.4 (-42.2 – 30.0) (-1.6 ± 21.1)	0.8 (-30.2 – 47.1) (2.4 ± 20.1)	-3.7 (-37.9 – 33.9) (-4.7 ± 18.8)
$cNRI^-$	3.8 (-11.5 – 21.4) (3.6 ± 9.3)	-7.5 (-23.4 – 9.4) (-7.1 ± 8.7)	10.9 (-7.7 – 32.0) (11.9 ± 11.3)	-0.1 (-18.3 – 25.7) (2.3 ± 11.8)	-31.0 (-49.2 – -14.9) (-31.0 ± 9.1)*
(c) Risk probability prediction three years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
cNRI	19.8 (-8.6 – 51.0) (19.1 ± 16.1)	1.3 (-28.5 – 29.1) (-0.2 ± 16.2)	-	-	-44.4 (-71.0 – -3.6) (-43.2 ±17.1)*
$cNRI^+$	15.7 (-14.3 – 43.7) (15.9 ± 15.2)	8.7 (-23.4 – 31.1) (7.6 ± 15.2)	-	-	-13.7 (-41.1 – 24.7) (-12.7 ±17.1)
$cNRI^-$	3.5 (-10.9 – 21.4) (3.2 ± 9.1)	-8.7 (-23.8 – 8.4) (-7.8 ± 8.4)	-	-	-30.6 (-48.9 – -13.9) (-30.5 ±9.3)*
(d) Risk probability prediction four years after baseline visit					
	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4[b] Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
cNRI	9.4 (-15.3 – 36.2) (10.1 ± 13.7)	-4.0 (-34.6 – 19.5) (-5.6 ± 14.7)	-	-	-39.8 (-64.0 – -6.5) (-38.0 ± 15.6)*
$cNRI^+$	6.4 (-17.3 – 31.4) (7.0 ± 12.9)	2.2 (-23.4 – 26.0) (2.0 ± 13.4)	-	-	-10.1 (-32.2 – 25.5) (-8.6 ± 14.7)
$cNRI^-$	2.1 (-11.4 – 20.8) (3.0 ± 8.7)	-7.4 (-22.8 – 7.6) (-7.6 ± 8.4)	-	-	-29.2 (-48.4 – -10.9) (-29.4 ± 10.4)*
(e) Risk probability prediction five years after baseline visit					

	$\lambda_{PLD\text{-}min}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4-surv-recal Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* T4 Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ *vs.* $\lambda_{PLD\text{-}min}$ Median (95% CI)[a] (Mean ± SD)
cNRI	14.5 (-42.2 – 80.2) (16.7 ± 33.0)	-2.6 (-61.7 – 58.9) (-3.3 ± 32.0)	-13.9 (-71.4 – 58.9) (-10.5 ± 32.5)	-30.0 (-74.3 – 37.8) (-25.3 ± 30.1)	-56.4 (-99.6 – 2.8) (-53.8 ± 28.4)
$cNRI^+$	3.1 (-28.8 – 33.6) (3.6 ± 16.0)	-1.0 (-30.3 – 31.5) (-1.7 ± 15.4)	-23.3 (-52.3 – 15.3) (-20.0 ± 17.6)	-23.2 (-50.9 – 10.1) (-22.8 ± 16.4)	-15.7 (-38.6 – 28.2) (-14.1 ± 16.9)
$cNRI^-$	9.1 (-40.0 – 71.0) (13.0 ± 28.7)	0.0 (-55.3 – 47.8) (-1.6 ± 27.9)	10.1 (-45.5 – 65.2) (9.5 ± 29.0)	0.0 (-55.3 – 50.0) (-2.5 ± 27.6)	-42.9 (-77.6 – 3.5) (-39.8 ± 23.4)

Abbreviations: CI, confidence interval; $\lambda_{PLD\text{-}min}$, LASSO Cox PH risk equation with λ parameter optimization in an internal 5-fold cross-validation to yield the minimum partial likelihood deviance (PLD) and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; $\lambda_{PLD\text{-}1sd}$, LASSO Cox PH risk equation with λ parameter optimization in an internal 5-fold cross-validation to yield maximum penalty while simultaneously keeping the PLD within one standard deviation of the minimum PLD and cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets; cNRI, continuous net reclassification improvement; $cNRI^+$, continuous net reclassification improvement considering only patients with an event during the observation period; $cNRI^-$, continuous net reclassification improvement considering only patients without an event during the observation period; SD, standard deviation; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the GCKD training sets. [a]The 95% CIs were determined empirically over 100 subsampling runs. [b]Please note, that Tangri et al. provided only survival rates for two and five years after the baseline visit.[S14] *significant cNRIs, i.e. 95% CIs excluding zero.

Table S9. Sensitivities, specificities, balanced accuracies, positive predictive values, and negative predictive values for the new $\lambda_{PLD\text{-}min}$ and $\lambda_{PLD\text{-}1sd}$ as well as the T4 score evaluated on the GCKD test data sets at (a) one, (b) two, (c) three, (d) four, and (e) five years after the baseline visit additionally censoring at the date of non-fatal cardiovascular events and each averaged across 100 subsampling runs. We excluded patients that had been censored due to a cardiovascular event before the end of the respective observation time period from the ROC curve evaluation. In addition to death from any cause or loss-to-follow-up, the following cardiovascular events were censored: non-fatal myocardial infarction, non-fatal non-hemorrhagic stroke, and coronary revascularization procedures. All performance measures were evaluated for both the "best" and the "Youden" cut-off, individually optimized for each observation time period, risk score, and test set, respectively, and then averaged across 100 test sets. Corresponding averaged ROC curves and cut-off points are displayed in Figure S12. Additionally, area under the ROC curve values averaged over 100 subsampling runs are given. Please note that at five years after the baseline visit, a high censoring rate affects the ROC curve evaluation. On average, 58, 145, 248, 827, and 1566 GCKD study participants had to be excluded from the ROC curve analyses due to censoring or death one, two, three, four, and five years after the baseline visit, respectively.

(a) ROC curve evaluation one year after baseline visit						
Cut-off	**"Best"**			**"Youden"**		
	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)
Sensitivity (%)	100.0 (80.0 – 100.0) (93.5 ± 7.4)	100.0 (80.0 – 100.0) (94.2 ± 7.2)	87.5 (66.7 – 100.0) (87.5 ± 10.8)	100.0 (80.0 – 100.0) (94.5 ± 7.3)	100.0 (80.0 – 100.0) (95.2 ± 7.1)	88.9 (68.9 – 100.0) (89.4 ± 10.2)
Specificity (%)	90.2 (79.5 – 95.7) (89.4 ± 4.4)	88.7 (82.8 – 96.9) (89.1 ± 4.0)	85.0 (66.3 – 89.7) (84.4 ± 4.9)	89.5 (78.8 – 95.3) (88.6 ± 4.6)	87.1 (82.7 – 96.9) (88.5 ± 4.1)	84.9 (61.2 – 89.7) (82.8 ± 7.6)
Balanced accuracy (%)	91.6 (82.4 – 97.6) (91.4 ± 4.6)	91.9 (83.0 – 98.1) (91.7 ± 4.1)	86.0 (75.0 – 93.6) (86.0 ± 5.8)	92.3 (82.4 – 97.6) (91.6 ± 4.5)	92.6 (83.0 – 98.1) (91.8 ± 4.0)	86.0 (75.1 – 93.6) (86.1 ± 5.7)
PPV (%)	3.9 (1.6 – 9.2) (4.3 ± 1.9)	3.8 (1.7 – 8.2) (4.2 ± 1.8)	2.5 (0.9 – 4.0) (2.6 ± 1.3)	3.7 (1.6 – 8.3) (4.0 ± 1.8)	3.5 (1.7 – 7.8) (3.9 ± 1.7)	2.4 (0.9 – 4.0) (2.5 ± 1.4)
NPV (%)	100.0 (99.9 – 100.0) (100.0 ± 0.4)	100.0 (99.9 – 100.0) (100.0 ± 0.0)	99.9 (99.9 – 100.0) (99.9 ± 0.1)	100.0 (99.9 – 100.0) (100.0 ± 0.4)	100.0 (99.9 – 100.0) (100.0 ± 0.0)	99.9 (99.9 – 100.0) (99.9 ± 0.1)
AUC-ROC	0.943 (0.846 – 0.976) (0.927 ± 0.041)	0.945 (0.839 – 0.978) (0.925 ± 0.044)	0.877 (0.756 – 0.949) (0.870 ± 0.053)	0.943 (0.846 – 0.976) (0.927 ± 0.041)	0.945 (0.839 – 0.978) (0.925 ± 0.044)	0.877 (0.756 – 0.949) (0.870 ± 0.053)
(b) ROC curve evaluation two years after baseline visit						
Cut-off	**"Best"**			**"Youden"**		
	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a]	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a]	T4 risk score Median (95% CI)[a]	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a]	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a]	T4 risk score Median (95% CI)[a]

	(Mean ± SD)	(Mean ± SD)	(Mean ± SD)	(Mean ± SD)	(Mean ± SD)	(Mean ± SD)
Sensitivity (%)	87.0 (78.4 – 95.1) (86.9 ± 4.6)	87.8 (80.0 – 95.7) (88.3 ± 5.0)	81.3 (68.8 – 91.5) (81.2 ± 5.9)	88.9 (76.3 – 100.0) (89.3 ± 6.4)	89.6 (79.5 – 100.0) (89.8 ± 6.1)	83.3 (67.5 – 100.0) (83.7 ± 8.2)
Specificity (%)	87.0 (81.5 – 92.2) (87.0 ± 3.0)	86.9 (83.0 – 90.9) (87.0 ± 2.2)	80.9 (73.7 – 89.2) (81.7 ± 4.5)	85.7 (72.1 – 92.7) (85.2 ± 5.2)	86.5 (75.8 – 91.6) (85.8 ± 4.1)	79.9 (62.6 – 90.4) (79.7 ± 7.4)
Balanced accuracy (%)	86.8 (82.1 – 92.7) (87.0 ± 2.8)	87.2 (83.4 – 92.8) (87.6 ± 2.8)	81.5 (74.6 – 87.1) (81.4 ± 3.4)	87.2 (83.2 – 92.8) (87.3 ± 2.8)	87.3 (83.4 – 93.1) (87.8 ± 2.8)	81.6 (75.3 – 88.4) (81.7 ± 3.4)
PPV (%)	8.4 (5.4 – 13.3) (8.8 ± 2.3)	8.6 (5.5 – 12.3) (8.7 ± 1.8)	5.8 (3.4 – 10.1) (6.0 ± 1.7)	8.1 (4.0 – 13.2) (8.2 ± 2.4)	8.4 (4.7 – 12.4) (8.4 ± 2.2)	5.5 (3.1 – 10.7) (5.9 ± 1.9)
NPV (%)	99.8 (99.7 – 99.9) (99.8 ± 0.1)	99.8 (99.7 – 99.9) (99.8 ± 0.1)	99.7 (99.5 – 99.9) (99.7 ± 0.1)	99.8 (99.7 – 100.0) (99.8 ± 0.1)	99.9 (99.7 – 100.0) (99.8 ± 0.1)	99.7 (99.5 – 100.0) (99.7 ± 0.1)
AUC-ROC	0.923 (0.881 – 0.959) (0.922 ± 0.021)	0.925 (0.880 – 0.961) (0.924 ± 0.021)	0.874 (0.810 – 0.928) (0.872 ± 0.033)	0.923 (0.881 – 0.959) (0.922 ± 0.021)	0.925 (0.880 – 0.961) (0.924 ± 0.021)	0.874 (0.810 – 0.928) (0.872 ± 0.033)
(c) ROC curve evaluation three years after baseline visit						
Cut-off	**"Best"**			**"Youden"**		
	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)
Sensitivity (%)	87.9 (81.7 – 93.8) (87.6 ± 3.4)	87.5 (81.5 – 92.9) (87.6 ± 3.1)	81.6 (73.9 – 88.8) (81.9 ± 4.1)	89.5 (81.3 – 100.0) (89.9 ± 4.9)	89.0 (80.4 – 97.2) (89.3 ± 4.4)	83.7 (72.5 – 96.9) (83.8 ± 6.7)
Specificity (%)	85.3 (80.2 – 88.7) (85.2 ± 2.2)	85.6 (80.4 – 89.2) (85.3 ± 2.4)	81.8 (73.5 – 87.9) (82.2 ± 3.6)	84.2 (72.7 – 89.0) (83.4 ± 4.4)	85.0 (75.8 – 91.2) (84.1 ± 3.9)	81.2 (66.8 – 88.5) (80.9 ± 6.3)
Balanced accuracy (%)	86.3 (83.1 – 90.5) (86.4 ± 1.9)	86.5 (82.8 – 89.8) (86.4 ± 1.9)	82.2 (77.9 – 86.4) (82.0 ± 2.4)	86.5 (83.5 – 90.5) (86.7 ± 1.9)	86.7 (83.4 – 89.8) (86.7 ± 1.8)	82.5 (78.6 – 86.4) (82.3 ± 2.3)
PPV (%)	13.9 (10.2 – 17.8) (13.9 ± 2.0)	13.8 (10.0 – 18.3) (14.0 ± 2.2)	10.9 (7.7 – 16.8) (11.3 ± 2.3)	13.4 (8.8 – 18.1) (13.2 ± 2.7)	13.4 (8.7 – 18.6) (13.6 ± 2.6)	11.2 (6.4 – 16.8) (11.2 ± 2.7)
NPV (%)	99.6 (99.4 – 99.8) (99.6 ± 0.1)	99.6 (99.4 – 99.8) (99.6 ± 0.1)	99.4 (99.2 – 99.7) (99.4 ± 0.1)	99.7 (99.4 – 100.0) (99.7 ± 0.2)	99.7 (99.4 – 99.9) (99.7 ± 0.1)	99.5 (99.1 – 99.9) (99.5 ± 0.2)
AUC-ROC	0.923 (0.899 – 0.950) (0.924 ± 0.014)	0.922 (0.893 – 0.947) (0.923 ± 0.014)	0.883 (0.848 – 0.923) (0.885 ± 0.021)	0.923 (0.899 – 0.950) (0.924 ± 0.014)	0.922 (0.893 – 0.947) (0.923 ± 0.014)	0.883 (0.848 – 0.923) (0.885 ± 0.021)
(d) ROC curve evaluation four years after baseline visit						
Cut-off	**"Best"**			**"Youden"**		
	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)
Sensitivity (%)	87.3 (80.2 – 92.2) (87.0 ± 3.2)	86.3 (79.7 – 92.4) (86.3 ± 3.4)	80.9 (75.0 – 86.5) (81.0 ± 3.4)	90.2 (81.0 – 98.3) (89.6 ± 4.3)	88.3 (77.5 – 98.4) (88.1 ± 5.4)	82.4 (72.3 – 93.8) (82.5 ± 5.9)

Specificity (%)	83.7 (78.9 – 88.6) (83.8 ± 2.7)	84.3 (78.2 – 89.2) (84.1 ± 2.9)	82.3 (76.5 – 88.1) (82.2 ± 2.6)	81.4 (74.5 – 89.2) (81.7 ± 3.9)	82.7 (74.1 – 89.6) (82.6 ± 4.4)	82.3 (68.6 – 89.5) (81.1 ± 5.8)
Balanced accuracy (%)	85.4 (82.6 – 88.1) (85.4 ± 1.6)	85.1 (81.9 – 88.8) (85.2 ± 1.8)	81.4 (77.5 – 85.5) (81.6 ± 2.0)	85.7 (82.6 – 88.6) (85.7 ± 1.6)	85.3 (81.9 – 89.0) (85.3 ± 1.9)	81.6 (78.6 – 85.5) (81.8 ± 1.9)
PPV (%)	19.4 (14.4 – 25.0) (19.4 ± 3.0)	19.5 (13.0 – 26.0) (19.7 ± 3.3)	16.6 (12.3 – 22.5) (16.9 ± 2.6)	18.0 (13.2 – 25.6) (18.2 ± 3.4)	18.0 (12.8 – 27.0) (19.0 ± 3.8)	16.5 (10.5 – 24.1) (17.0 ± 3.6)
NPV (%)	99.4 (99.0 – 99.6) (99.3 ± 0.2)	99.2 (99.0 – 99.6) (99.3 ± 0.2)	99.0 (98.6 – 99.4) (99.0 ± 0.2)	99.5 (99.0 – 99.9) (99.5 ± 0.2)	99.4 (99.0 – 99.9) (99.4 ± 0.3)	99.1 (98.6 – 99.6) (99.1 ± 0.3)
AUC-ROC	0.916 (0.883 – 0.942) (0.916 ± 0.014)	0.915 (0.882 – 0.940) (0.914 ± 0.015)	0.880 (0.847 – 0.913) (0.880 ± 0.018)	0.916 (0.883 – 0.942) (0.916 ± 0.014)	0.915 (0.882 – 0.940) (0.914 ± 0.015)	0.880 (0.847 – 0.913) (0.880 ± 0.018)
(e) ROC curve evaluation five years after baseline visit						
Cut-off	**"Best"**			**"Youden"**		
	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}min}$ risk score Median (95% CI)[a] (Mean ± SD)	$\lambda_{PLD\text{-}1sd}$ risk score Median (95% CI)[a] (Mean ± SD)	T4 risk score Median (95% CI)[a] (Mean ± SD)
Sensitivity (%)	90.6 (73.1 – 100.0) (89.2 ± 7.5)	88.6 (78.5 – 99.3) (88.9 ± 5.8)	85.4 (75.7 – 93.4) (85.0 ± 5.1)	93.2 (64.0 – 100.0) (90.7 ± 9.0)	92.7 (74.1 – 100.0) (90.9 ± 8.1)	85.7 (69.1 – 94.3) (84.9 ± 6.7)
Specificity (%)	84.6 (72.5 – 100.0) (85.1 ± 6.8)	84.6 (70.7 – 100.0) (84.9 ± 6.9)	86.2 (70.0 – 100.0) (85.4 ± 7.6)	84.0 (66.6 – 100.0) (84.5 ± 9.3)	83.3 (63.4 – 100.0) (83.8 ± 9.6)	86.7 (69.0 – 100.0) (86.0 ± 9.0)
Balanced accuracy (%)	87.4 (77.7 – 95.9) (87.2 ± 4.6)	86.9 (77.6 – 95.3) (86.9 ± 4.6)	85.5 (76.0 – 94.5) (85.2 ± 4.9)	87.8 (78.9 – 95.9) (87.6 ± 4.3)	87.5 (78.9 – 95.3) (87.3 ± 4.4)	85.8 (76.6 – 94.5) (85.4 ± 4.8)
PPV (%)	22.6 (14.1 – 100.0) (28.6 ± 20.8)	22.2 (13.2 – 100.0) (27.9 ± 19.8)	23.2 (12.6 – 100.0) (29.7 ± 23.4)	22.6 (11.5 – 100.0) (33.1 ± 28.0)	21.7 (11.1 – 100.0) (31.5 ± 26.6)	23.7 (13.0 – 100.0) (33.7 ± 27.9)
NPV (%)	99.4 (98.5 – 100.0) (99.4 ± 0.4)	99.4 (98.7 – 100.0) (99.4 ± 0.4)	99.1 (98.5 – 99.7) (99.1 ± 0.3)	99.6 (98.1 – 100.0) (99.5 ± 0.5)	99.6 (98.6 – 100.0) (99.5 ± 0.5)	99.2 (98.2 – 99.7) (99.1 ± 0.4)
AUC-ROC	0.928 (0.862 – 0.978) (0.925 ± 0.032)	0.925 (0.851 – 0.977) (0.922 ± 0.035)	0.888 (0.816 – 0.956) (0.888 ± 0.039)	0.928 (0.862 – 0.978) (0.925 ± 0.032)	0.925 (0.851 – 0.977) (0.922 ± 0.035)	0.888 (0.816 – 0.956) (0.888 ± 0.039)

Abbreviations: AUC-ROC, area under the receiver operating characteristic curve; CI, confidence interval; $\lambda_{PLD\text{-}min}$, end-stage kidney disease risk model newly derived by fitting a Cox LASSO PH regression on a GCKD training cohort in the resampling approach, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield the minimum partial likelihood deviance; $\lambda_{PLD\text{-}1sd}$, end-stage kidney disease risk model newly derived by fitting a Cox LASSO PH regression on a GCKD training cohort in the resampling approach, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield the most penalized model such that the partial likelihood deviance remained within one standard deviation of the minimum deviance; NPV, negative predictive value; PPV, positive predictive value; SD, standard deviation; T4, original 4-variable Tangri risk equation. [a]The 95% CIs were determined empirically over 100 subsampling runs.

Table S10. Cause-specific *C* statistic differences for experiencing an ESKD event, *i.e.* long-term dialysis or renal transplantation, in the presence of death as a competing risk comparing the Z6 *vs.* the T4 in three independent validation cohorts evaluated in an ordinary nonparametric bootstrap resampling analysis with 10,000 bootstrap replicates. The inverse probability of censoring weighted estimator, employing the Kaplan-Meier estimator for the censoring times, was used to deal with right-censored data. In case of several ESKD events, only the first event was considered.

CKD-REIN (n_{total} = 1,912; n_{events} = 445)					
	1 year	**2 years**	**3 years**	**4 years**	**5 years**
Median ESKD-specific C_{diff} (95% CI)[a]	0.019 (0.001 - 0.038)	0.012 (0.002 - 0.022)	0.011 (0.002 - 0.019)	0.003 (-0.005 - 0.012)	-0.001 (-0.010 – 0.008)
***p*-value[b]**	0.018*	0.011*	0.0088**	0.2159	0.6083
SKS (n_{total} = 949; n_{events} = 150)					
	1 year	**2 years**	**3 years**	**4 years**	**5 years**
Median ESKD-specific C_{diff} (95% CI)[a]	0.018 (0.003 - 0.036)	0.008 (-0.005 - 0.021)	0.009 (-0.003 - 0.021)	-0.004 (-0.018 - 0.009)	-0.002 (-0.016 - 0.013)
***p*-value[b]**	0.0124*	0.1058	0.0642	0.7301	0.6021
MMKD (n_{total} = 202; n_{events} = 75)					
	1 year	**2 years**	**3 years**	**4 years**	**5 years**
Median ESKD-specific C_{diff} (95% CI)[a]	0.004 (-0.034 - 0.045)	0.028 (0.009 – 0.050)	0.018 (0.003 - 0.035)	0.016 (0.002 – 0.031)	0.015 (0.002 - 0.029)
***p*-value[b]**	0.4065	0.0019**	0.0092**	0.0113*	0.0133*

Abbreviations: C_{diff}, *C* statistic differences comparing the Z6 vs. the T4; CI, confidence interval; CKD-REIN, Chronic Kidney Disease-Renal Epidemiology and Information Network; MMKD, Mild to Moderate Kidney Disease; n_{events}, number of study participants experiencing an ESKD event included in analyses; n_{total}, total number of study participants included in analyses; SKS, Salford Kidney Study; T4, original four-variable ESKD risk model developed by Tangri et al. without coefficient recalibration to the GCKD cohort; Z6, final ESKD risk model derived by fitting a LASSO Cox PH regression on the complete GCKD cohort, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield maximum penalty while keeping the partial likelihood deviance (PLD) within one standard deviation of the minimum PLD. [a]CIs were determined by ordinary nonparametric percentile bootstrap with 10,000 bootstrap replicates. [b]One-sided *p*-value testing $C_{diff} > 0$. *significant *p*-value < 0.05; **significant *p*-value < 0.01.

Table S11. Continuous (category-free) net reclassification improvement values comparing the Z6 *vs.* the GCKD-recalibrated T4 (T4-surv-recal) or the T4 risk equation in the three independent validation cohorts evaluated one, two, three, four, and five years after the baseline visit.

	1 year		2 years		3 years		4 years		5 years	
(a) CKD-REIN										
	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4
cNRI (95% CI)[a]	26.9 (8.2 – 45.5)*	-	22.1 (9.5 – 34.6)*	-3.9 (-18.7 – 10.5)	17.7 (6.8 – 28.3)*	-	5.8 (-4.9 – 16.4)	-	11.0 (-4.4 – 27.0)	-23.8 (-40.0 – -7.2)*
cNRI$^+$ (95% CI)[a]	62.6 (44.3 – 79.7)*	-	58.8 (47.0 – 70.1)*	23.9 (9.8 – 37.5)*	54.2 (44.6 – 63.5)*	-	45.3 (36.3 –54.3)*	-	48.1 (38.4 – 57.5)*	13.3 (2.9 – 23.6)*
cNRI$^-$ (95% CI)[a]	-35.7 (-40.0 – -31.3)*	-	-36.7 (-41.1 – -32.2)*	-27.8 (-32.5 – -23.2)*	-36.5 (-41.4 – -31.6)*	-	-39.5 (-45.5 – -33.6)*	-	-37.1 (-49.3 – -24.3)*	-37.1 (-49.3 – -24.5)*
(b) SKS										
	1 year		**2 years**		**3 years**		**4 years**		**5 years**	
	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4
cNRI (95% CI)[a]	64.0 (48.1 – 76.8)*	-	65.6 (52.6 – 77.2)*	38.0 (17.1 – 57.9)*	64.7 (51.3 – 77.0)*	-	48.0 (32.3 – 63.1)*	-	51.5 (36.6 – 66.1)*	24.5 (7.2 – 41.5)*
cNRI$^+$ (95% CI)[a]	90.3 (75.3 – 100.0)*	-	89.4 (78.2 – 97.7)*	54.8 (35.3 – 73.1)*	85.6 (74.2 – 95.3)*	-	68.5 (55.0 – 81.0)*	-	69.4 (57.0 – 80.9)*	44.1 (29.5 – 58.6)*
cNRI$^-$ (95% CI)[a]	-26.3 (-32.8 – -19.7)*	-	-23.9 (-30.9 – -16.8)*	-16.8 (-23.9 – -9.7)*	-20.9 (-28.5 – -13.4)*	-	-20.5 (-28.7 – -12.4)*	-	-17.9 (-26.5 – -9.0)*	-19.6 (-28.5 – -10.5)*
(c) MMKD										
	1 year		**2 years**		**3 years**		**4 years**		**5 years**	
	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4
cNRI (95% CI)[a]	62.4 (-2.6 – 113.2)	-	58.9 (22.3 – 93.2)*	38.1 (-1.3 – 75.7)	38.7 (5.5 – 71.1)*	-	22.1 (-9.7 – 53.4)	-	25.3 (-8.2 – 58.2)	-10.4 (-42.1 – 23.0)
cNRI$^+$ (95% CI)[a]	55.6 (-9.1 – 100.0)	-	46.7 (13.7 – 77.7)*	19.8 (-16.8 – 54.7)	27.9 (-0.7 – 55.5)	-	13.6 (-12.5 – 39.4)	-	11.7 (-12.8 – 36.2)	-26.6 (-50.1 – -2.5)*
cNRI$^-$ (95% CI)[a]	6.8 (-7.2 – 20.8)	-	12.2 (-3.0 – 27.6)	18.3 (3.2 – 33.0)*	10.8 (-5.6 – 27.1)	-	8.5 (-9.9 – 26.2)	-	13.5 (-9.6 – 35.4)	16.2 (-6.5 – 38.7)

Abbreviations: CI, confidence interval; CKD-REIN, Chronic Kidney Disease-Renal Epidemiology and Information Network; cNRI, continuous net reclassification improvement; cNRI$^+$, continuous net reclassification improvement considering only patients with an event during the observation period; cNRI$^-$, continuous net reclassification improvement considering only patients without an event during the observation period; MMKD, Mild to Moderate Kidney Disease; SKS, Salford Kidney Study; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the complete GCKD cohort; Z6, final ESKD risk equation derived by fitting and calibrating a LASSO Cox PH regression on the complete GCKD cohort, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield maximum penalty while simultaneously keeping the partial likelihood deviance (PLD) within one standard deviation of the minimum PLD. [a]CIs were determined by percentile bootstrap with 10,000 bootstrap replicates. [b]Please note, that Tangri et al. provided only survival rates for two and five years after the baseline visit.[S14] *significant NRIs, i.e. 95% CIs excluding zero.

Table S12. Brier scores for the Z6, the GCKD-recalibrated T4 (T4-surv-recal), and the T4 risk equation in comparison to the null model at one, two, three, four, and five years after the baseline visit evaluated in the three independent validation cohorts. The null model refers to the Kaplan-Meier estimator, and the inverse probability of censoring weights (IPCW) method employing the Kaplan-Meier estimator was used to deal with censored individuals.

(a) CKD-REIN					
	1 year	**2 years**	**3 years**	**4 years**	**5 years**
null model	0.0377	0.0897	0.1313	0.1679	0.2010
Z6	0.0298	0.0640	0.0856	0.1130	0.1495
T4-surv-recal	0.0336	0.0701	0.0932	0.1146	0.1488
T4[a]	-	0.0643	-	-	0.1422
(b) SKS					
	1 year	**2 years**	**3 years**	**4 years**	**5 years**
null model	0.0429	0.0760	0.0983	0.1252	0.1228
Z6	0.0301	0.0593	0.0844	0.1246	0.0969
T4-surv-recal	0.0357	0.0525	0.0647	0.0928	0.0765
T4[a]	-	0.0525	-	-	0.0794
(c) MMKD					
	1 year	**2 years**	**3 years**	**4 years**	**5 years**
null model	0.0428	0.1288	0.1782	0.2102	0.2231
Z6	0.0323	0.0789	0.1027	0.1143	0.1156
T4-surv-recal	0.0351	0.0948	0.1103	0.1192	0.1179
T4[a]	-	0.0885	-	-	0.1142

Abbreviations: CKD-REIN, Chronic Kidney Disease-Renal Epidemiology and Information Network; MMKD, Mild to Moderate Kidney Disease; SKS, Salford Kidney Study; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the complete GCKD cohort; Z6, final ESKD risk equation derived by fitting and calibrating a LASSO Cox PH regression on the complete GCKD cohort, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield maximum penalty while simultaneously keeping the partial likelihood deviance (PLD) within one standard deviation of the minimum PLD. [a]Please note, that Tangri et al. provided only survival rates for two and five years after the baseline visit.

Table S13. Cause-specific *C* statistics and *C* statistic differences of experiencing an ESKD event, *i.e.* dialysis or renal transplantation, in the presence of death as a competing risk comparing the Z6 *vs.* the T4 for CKD-REIN study participants with (a) measured or (b) estimated urine albumin-to-creatinine ratio values, evaluated in an ordinary nonparametric bootstrap resampling analysis with 10,000 bootstrap replicates. The inverse probability of censoring weighted estimator employing the Kaplan-Meier estimator for the censoring times was used to deal with right-censored data. In case of several ESKD events, only the first event was considered.

(a) CKD-REIN mUACR (n_{total} = 796; n_{events} = 177)					
	1 year	**2 years**	**3 years**	**4 years**	**5 years**
Z6 ESKD-specific *C* statistics (95% CI)[a]	0.943 (0.915 - 0.966)	0.895 (0.857 - 0.927)	0.878 (0.847 - 0.907)	0.862 (0.834 - 0.888)	0.833 (0.796 - 0.867)
T4 ESKD-specific *C* statistics (95% CI)[a]	0.926 (0.884 - 0.959)	0.883 (0.841 - 0.919)	0.865 (0.828 - 0.898)	0.853 (0.823 - 0.881)	0.828 (0.790 - 0.863)
Z6 *vs.* T4 Median ESKD-specific C_{diff} (95% CI)[a] *p*-value[b]	0.017 (-0.005 - 0.042) 0.0666	0.012 (-0.002 - 0.026) 0.052	0.013 (0.0 - 0.026) 0.0281*	0.009 (-0.004 - 0.022) 0.0799	0.006 (-0.009 – 0.020) 0.2209
(b) CKD-REIN eUACR (n_{total} = 1,802; n_{events} = 428)					
	1 year	**2 years**	**3 years**	**4 years**	**5 years**
Z6 ESKD-specific *C* statistics (95% CI)[a]	0.928 (0.908 - 0.946)	0.899 (0.880 - 0.916)	0.884 (0.868 - 0.899)	0.861 (0.845 - 0.877)	0.835 (0.814 - 0.855)
T4 ESKD-specific *C* statistics (95% CI)[a]	0.911 (0.883 - 0.936)	0.890 (0.869 - 0.910)	0.877 (0.860 - 0.894)	0.860 (0.843 - 0.876)	0.838 (0.817 - 0.857)
Z6 *vs.* T4 Median ESKD-specific C_{diff} (95% CI)[a] *p*-value[b]	0.017 (-0.002 - 0.036) 0.0376*	0.008 (-0.002 - 0.019) 0.0543	0.007 (-0.002 - 0.016) 0.0599	0.001 (-0.007 – 0.010) 0.3701	-0.003 (-0.012 - 0.006) 0.7569

Abbreviations: C_{diff}, *C* statistic differences comparing the Z6 vs. the T4; CI, confidence interval; CKD-REIN, Chronic Kidney Disease-Renal Epidemiology and Information Network; eUACR, estimated urine albumin-to-creatinine ratio; mUACR, measured urine albumin-to-creatinine ratio; n_{events}, number of study participants experiencing an ESKD event included in analyses; n_{total}, total number of study participants included in analyses; T4, original four-variable ESKD risk model developed by Tangri et al. without coefficient recalibration to the GCKD cohort; Z6, final ESKD risk model derived by fitting a LASSO Cox PH regression on the complete GCKD cohort, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield maximum penalty while simultaneously keeping the partial likelihood deviance (PLD) within one standard deviation of the minimum PLD. [a]CIs were determined by ordinary nonparametric percentile bootstrap with 10,000 bootstrap replicates. [b]One-sided *p*-value testing $C_{diff} > 0$; *significant *p*-value < 0.05; **significant *p*-value < 0.01.

Table S14. Categorical net reclassification improvement values comparing the Z6 *vs.* the GCKD-recalibrated T4 (T4-surv-recal) or the T4 risk equation for CKD-REIN study participants with (a) measured or (b) estimated urine albumin-to-creatinine ratio values, evaluated one, two, three, four, and five years after the baseline visit. The predicted risk probabilities were divided into 3 categories: 0% - <3%, 3% - <10%, 10% - 100%, respectively.

(a) CKD-REIN mUACR										
	1 year		**2 years**		**3 years**		**4 years**		**5 years**	
	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4
NRI (95% CI)[a]	26.8 (7.7 – 46.1)*	-	14.0 (4.2 – 24.5)*	11.6 (3.3 – 20.4)*	-5.3 (-10.7 – 0.2)	-	1.2 (-4.8 – 7.6)	-	1.5 (-9.6 – 12.7)	-2.3 (-14.8 – 10.5)
NRI$^+$ (95% CI)[a]	32.3 (13.4 – 51.5)*	-	19.2 (9.8 – 29.1)*	11.0 (3.1 – 19.4)*	1.7 (-2.6 – 6.1)	-	7.9 (3.4 – 12.8)*	-	11.2 (5.7 – 17.5)*	8.4 (0.5 – 16.1)*
NRI$^-$ (95% CI)[a]	-5.5 (-7.7 – -3.3)*	-	-5.2 (-8.1 – -2.4)*	0.6 (-2.3 – 3.4)	-7.0 (-10.4 – -3.5)*	-	-6.7 (-10.8 – -2.6)*	-	-9.7 (-19.2 – 0.0)	-10.8 (-20.8 – -0.9)*
(b) CKD-REIN eUACR										
	1 year		**2 years**		**3 years**		**4 years**		**5 years**	
	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4
NRI (95% CI)[a]	21.2 (8.5 – 34.2)*	-	4.6 (-1.2 – 10.7)	5.0 (-0.5 – 10.7)	0.1 (-3.8 – 4.3)	-	-3.8 (-8.1 – 0.8)	-	-3.5 (-10.4 – 3.4)	-5.3 (-12.7 – 2.2)
NRI$^+$ (95% CI)[a]	27.8 (15.3 – 40.7)*	-	11.4 (6.0 – 17.0)*	5.4 (0.5 – 10.6)*	6.0 (2.6 – 9.4)*	-	5.8 (2.8 – 9.0)*	-	5.9 (2.7 – 9.3)*	4.6 (1.2 – 8.1)*
NRI$^-$ (95% CI)[a]	-6.6 (-8.1 – -5.1)*	-	-6.7 (-8.8 – -4.7)*	-0.4 (-2.5 – 1.7)	-5.8 (-8.0 – -3.5)*	-	-9.6 (-12.7 – -6.5)*	-	-9.4 (-15.6 – -3.3)*	-9.9 (-16.6 – -3.2)*

Abbreviations: CKD-REIN, Chronic Kidney Disease-Renal Epidemiology and Information Network; eUACR, estimated urine albumin-to-creatinine ratio; mUACR, measured urine albumin-to-creatinine ratio; NRI, net reclassification improvement; NRI$^+$, net reclassification improvement considering only patients with an event during the observation period; NRI$^-$, net reclassification improvement considering only patients without an event during the observation period; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the complete GCKD cohort; Z6, final ESKD risk equation derived by fitting and calibrating a LASSO Cox PH regression on the complete GCKD cohort, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield maximum penalty while simultaneously keeping the partial likelihood deviance (PLD) within one standard deviation of the minimum PLD. [a]CIs were determined by percentile bootstrap with 10,000 bootstrap replicates. [b]Please note, that Tangri et al. provided only survival rates for two and five years after the baseline visit. *significant NRIs, i.e. 95% CIs excluding zero.

Table S15. Continuous (category-free) net reclassification improvement values comparing the Z6 *vs.* the GCKD-recalibrated T4 (T4-surv-recal) and the T4 risk equation, respectively, for CKD-REIN study participants with (a) measured or (b) estimated urine albumin-to-creatinine ratio values, evaluated one, two, three, four, and five years after the baseline visit.

(a) CKD-REIN mUACR										
	1 year		**2 years**		**3 years**		**4 years**		**5 years**	
	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4
cNRI (95% CI)[a]	29.0 (-0.2 – 54.3)	-	23.3 (4.1 – 42.1)*	-15.7 (-39.6 – 7.6)	17.3 (0.2 – 33.6)*	-	5.8 (-11.1 – 22.1)	-	7.4 (-15.6 – 30.5)	-18.8 (-43.5 – 6.1)
cNRI$^+$ (95% CI)[a]	67.7 (39.4 – 92.5)*	-	64.3 (45.5 – 81.4)*	20.2 (-2.5 – 42.7)	58.0 (42.4 – 72.5)*	-	49.7 (35.2 – 63.1)*	-	55.8 (40.8 – 69.4)*	23.2 (6.1 – 39.8)*
cNRI$^-$ (95% CI)[a]	-38.7 (-45.3 – -32.2)*	-	-40.9 (-47.7 – -33.8)*	-35.9 (-42.9 – -28.9)*	-40.7 (-48.0 – -33.1)*	-	-43.9 (-52.5 – -35.0)*	-	-48.4 (-66.3 – -30.0)*	-41.9 (-60.0 – -23.7)*
(b) CKD-REIN eUACR										
	1 year		**2 years**		**3 years**		**4 years**		**5 years**	
	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4[b]	Z6 *vs.* T4-surv-recal	Z6 *vs.* T4
cNRI (95% CI)[a]	31.2 (12.5 – 48.3)*	-	20.2 (7.1 – 32.9)*	-2.7 (-17.7 – 12.1)	18.9 (7.7 – 29.6)*	-	8.0 (-3.2 – 19.0)	-	12.4 (-3.4 – 28.6)	-23.1 (-40.2 – -6.7)*
cNRI$^+$ (95% CI)[a]	66.6 (48.4 – 83.1)*	-	57.6 (45.4 – 69.1)*	22.9 (8.6 – 36.8)*	55.7 (45.8 – 65.1)*	-	47.3 (38.2 – 56.3)*	-	48.3 (38.6 – 58.0)*	13.8 (3.2 – 24.5)*
cNRI$^-$ (95% CI)[a]	-35.4 (-39.8 – -30.9)*	-	-37.4 (-42.1 – -32.7)*	-25.6 (-30.4 – -20.6)*	-36.8 (-41.9 – -31.6)*	-	-39.4 (-45.4 – -33.1)*	-	-36.0 (-48.6 – -22.9)*	-37.0 (-50.0 – -24.5)*

Abbreviations: CKD-REIN, Chronic Kidney Disease-Renal Epidemiology and Information Network; cNRI, continuous net reclassification improvement; cNRI$^+$, continuous net reclassification improvement considering only patients with an event during the observation period; cNRI$^-$, continuous net reclassification improvement considering only patients without an event during the observation period; eUACR, estimated urine albumin-to-creatinine ratio; mUACR, measured urine albumin-to-creatinine ratio; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the complete GCKD cohort; Z6, final ESKD risk equation derived by fitting and calibrating a LASSO Cox PH regression on the complete GCKD cohort, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield maximum penalty while simultaneously keeping the partial likelihood deviance (PLD) within one standard deviation of the minimum PLD. [a]CIs were determined by percentile bootstrap with 10,000 bootstrap replicates. [b]Please note, that Tangri et al. provided only survival rates for two and five years after the baseline visit.[S14] *significant NRIs, i.e. 95% CIs excluding zero.

Table S16. Brier scores evaluated for CKD-REIN study participants with (a) measured or (b) estimated urine albumin-to-creatinine ratio values for the Z6, the GCKD-recalibrated T4 (T4-surv-recal), and the T4 risk equation in comparison to the null model at one, two, three, four, and five years after the baseline visit. The null model refers to the Kaplan-Meier estimator, and the inverse probability of censoring weights (IPCW) method employing the Kaplan-Meier estimator was used to deal with censored individuals.

(a) CKD-REIN mUACR						
	1 year	**2 years**	**3 years**	**4 years**	**5 years**	
null model	0.0374	0.0836	0.1246	0.1616	0.1969	
Z6	0.0278	0.0590	0.0785	0.1054	0.1512	
T4-surv-recal	0.0325	0.0653	0.0876	0.1104	0.1505	
T4[a]	-	0.0601	-	-	0.1447	
(b) CKD-REIN eUACR						
	1 year	**2 years**	**3 years**	**4 years**	**5 years**	
null model	0.0384	0.0924	0.1347	0.1706	0.2039	
Z6	0.0300	0.0647	0.0864	0.1127	0.1482	
T4-surv-recal	0.0339	0.0714	0.0947	0.1143	0.1478	
T4[a]	-	0.0650	-	-	0.1398	

Abbreviations: CKD-REIN, Chronic Kidney Disease-Renal Epidemiology and Information Network; eUACR, estimated urine albumin-to-creatinine ratio; mUACR, measured urine albumin-to-creatinine ratio; T4, original four-variable ESKD risk equation for non-North American cohorts developed by Tangri et al. without coefficient or survival recalibration to the GCKD cohort; T4-surv-recal, four-variable ESKD risk equation developed by Tangri et al. without coefficient recalibration but employing cumulative ESKD subdistribution hazard functions estimated from the complete GCKD cohort; Z6, final ESKD risk equation derived by fitting and calibrating a LASSO Cox PH regression on the complete GCKD cohort, where the hyperparameter λ was optimized in an internal 5-fold cross-validation to yield maximum penalty while simultaneously keeping the partial likelihood deviance (PLD) within one standard deviation of the minimum PLD. [a]Please note, that Tangri et al. provided only survival rates for two and five years after the baseline visit.[S14]

Supplementary References

S1. Titze S, Schmid M, Köttgen A, Busch M, Floege J, Wanner C, Kronenberg F, Eckardt KU; GCKD study investigators. Disease burden and risk profile in referred patients with moderate chronic kidney disease: composition of the German Chronic Kidney Disease (GCKD) cohort. Nephrol Dial Transplant. 2015;3083):441-451.

S2. Hoefield RA, Kalra PA, Baker PG, Sousa I, Diggle PJ, Gibson MJ, O'Donoghue DJ, Middleton RJ, New JP. The use of eGFR and ACR to predict decline in renal function in people with diabetes. *Nephrol Dial Transplant.* 2011;26(3):887-892.

S3. Boes E, Fliser D, Ritz E, König P, Kraatz G, Lhotta K, Mann JFE, Müller GA, Neyer U, Riegel W, Riegler P, Kronenberg F; for the MMKD Study Group. Apolipoprotein A-IV predicts progression of chronic kidney disease: The Mild to Moderate Kidney Disease (MMKD) Study. *J Am Soc Nephrol.* 2006;17(2):528-536.

S4. Stengel B, Combe C, Jacquelinet C, Briançon S, Fouque D, Laville M, Frimat L, Pascal C, Herpe YE, Deleuze JF, Schanstra J, Pisoni RL, Robinson BM, Massy ZA. The French chronic kidney disease-renal epidemiology and information network (CKD-REIN) cohort study. *Nephrol Dial Transplant.* 2014;29(8):1500-1507.

S5. Sumida K, Nadkarni GN, Grams ME, Sang Y, Ballew SH, Coresh J, Matsushita K, Surapaneni A, Brunskill N, Chadban SJ, Chang AR, Cirillo M, Daratha KB, Gansevoort RT, Garg AX, Iacoviello L, Kayama T, Konta T, Kovesdy CP, Lash J, Lee BJ, Major RW, Metzger M, Miura K, Naimark DMJ, Nelson RG, Sawhney S, Stempniewicz N, Tang M, Townsend RR, Traynor JP, Valdivielso JM, Wetzels J, Polkinghorne KR, Heerspink HJL; Chronic Kidney Disease Prognosis Consortium. Conversion of urine protein-creatinine ratio or urine dipstick protein to urine albumin-creatinine ratio for use in chronic kidney disease screening and prognosis: An individual participant-based meta-analysis. *Ann Intern Med.* 2020;173(6):426-435.

S6. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1–22.

S7. Tibshirani R. The LASSO method for variable selection in the COX model. *Stat Med.* 1997;16(4):385–395.

S8. Harrel Jr FE, Lee KL, Mark DB. Tutorial in biostatistics: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing error. *Stat Med.* 1996;15(4):361–387.

S9. Wolbers M, Blanche P, Koller MT, Witteman JCM, Gerds TA. Concordance for prognostic models with competing risks. *Biostatistics* 2014; 15(3):526–539.

S10. Mogensen UB, Ishwaran H, Gerds TA. Evaluating random forests for survival analysis using prediction error curves. *J Stat Softw*. 2012;50(11):1-23.

S11. Canty A, Ripley B. *boot*: Bootstrap R (S-Plus) Functions. R package version 1.3-20. https://CRAN.R-project.org/package=boot; 2017

S12. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* 2008;27(2):157-172.

S13. Pencina MJ, D'Agostino Sr RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* 2011; 30(1):11-21.

S14. Tangri N, Grams ME, Levey AS, Coresh J, Appel LJ, Astor BC, Chodick G, Collins AJ, Djurdjev O, Elley CR, Evans M, Garg AX, Hallan SI, Inker LA, Ito S, Jee SH, Kovesdy CP, Kronenberg F, Heerspink HJ, Marks A, Nadkarni GN, Navaneethan SD, Nelson RG, Titze S, Sarnak MJ, Stengel B, Woodward M, Iseki K; CKD Prognosis Consortium. Multinational assessment of accuracy of equations for predicting risk of kidney failure: A meta-analysis. *JAMA.* 2016;315(2):164-174.

S15. Inoue E. *nricens*: NRI for risk prediction models with time to event and binary response data. R package version 1.5. https://CRAN.R-project.org/package=nricens; 2017.

S16. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *JASA*. 1999; 94(446):496-509.

S17. Gray B. *cmprsk*: Subdistribution Analysis of Competing Risks. *R* package version 2.2-7. https://CRAN.R-project.org/package=cmprsk; 2014.

S18. Saito T, Rehmsmeier M. Precrec: fast and accurate precision-recall and ROC curve calculations in R. *Bioinformatics.* 2017;33(1):145-147.

S19. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3(1):32-35.

S20. Brier GW. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev*. 1950;78(1):1-3.

S21. Therneau TM. *Survival*, A Package for Survival Analysis in S, version 2.38, https://CRAN.R-project.org/package=survival; 2015.

S22. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Springer, New York. ISBN 0-387-98784-3; 2000.

S23. Kassambara A, Kosinski M. *Survminer*: Drawing Survival Curves using 'ggplot2'. *R* package version 0.4.1. https://CRAN.R-project.org/package=survminer

S24. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org; 2019.

S25. Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, Levin A, Levey AS. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA.* 2011;305(15):1553–1559.