A meta-analysis of GFR slope as a surrogate endpoint for kidney failure

Lesley A. Inker, MD[1] , Willem Collier, PhD[2], Tom Greene, PhD[2], Shiyuan Miao, MS[1], Juhi Chaudhari, MPH[1], Gerald B. Appel, MD[3], Sunil V. Badve, PhD[4], Fernando Caravaca-Fontán, PhD[5], Lucia Del Vecchio, MD[6], Jürgen Floege, MD[7], Marian Goicoechea, PhD[8], Benjamin Haaland, PhD[2], William G. Herrington, MD[9], Enyu Imai, MD[10], Tazeen H. Jafar, MBBS[11], Julia B. Lewis, MD[12], Philip K. T. Li, FRCP[13], Bart D. Maes, PhD[14], Brendon L. Neuen, PhD[4], Ronald D. Perrone, MD[1], Giuseppe Remuzzi, MD[15], Francesco P. Schena, MD[16], Christoph Wanner, MD[17], Jack F. M. Wetzels, MD[18], Mark Woodward, PhD[4],[19] , Hiddo J. L. Heerspink, PhD[20], the CKD-EPI Clinical Trials Consortium*

*A list of collaborators and their affiliations is provided at the end of the article.

Corresponding Author
Lesley A. Inker, MD, MS
Tufts Medical Center, 800 Washington Street, Box #391, Boston, MA 02111
Tel: +1617-636-2569
e-mail: Lesley.Inker@TuftsMedicine.org

Affiliations:

[1] Division of Nephrology, Tufts Medical Center, Boston, MA USA
[2] Population Health Sciences, University of Utah School of Medicine, UT, USA
[3] Division of Nephrology, Columbia University Medical Center and the New York Presbyterian Hospital, New York, NY, USA
[4] The George Institute for Global Health, University of New South Wales, Sydney, NSW, Australia
[5] Instituto de Investigación Hospital 12 de octubre (i+12), Madrid, Spain
[6] Department of Nephrology and Dialysis, Sant'Anna Hospital, ASST Lariana, 22042 Como, Italy
[7] Division of Nephrology, RWTH Aachen University, Aachen, Germany
[8] Department of Nephrology, Hospital General Universitario Gregorio Marañón, Madrid, Spain
[9] Medical Research Council Population Health Research Unit, Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
[10] Nakayamadera Imai Clinic, Takarazuka, Japan
[11] Program in Health Services and Systems Research, Duke-NUS Medical School, Singapore
[12] Division of Nephrology, Vanderbilt University Medical Center, Nashville, TN, USA
[13] Division of Nephrology, Prince of Wales Hospital, Chinese University of Hong Kong, Shatin, Hong Kong
[14] Department of Nephrology, AZ Delta, Roeselare, Belgium
[15] Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Bergamo, Italy
[16] Renal, Dialysis and Transplant Unit, Department of Emergency and Organ Transplantation, University of Bari, Bari, Italy
[17] University of Würzburg, Department of Clinical Research and Epidemiology, Renal Research Unit, Comprehensive Heart Failure Center, Würzburg, Germany
[18] Department of Nephrology, Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands
[19] The George Institute for Global Health, School of Public Health, Imperial College London, London, UK
[20] Department of Clinical Pharmacy and Pharmacology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

Abstract (202 words, of 200)

Glomerular filtration rate (GFR) decline is causally associated with kidney failure and is a candidate surrogate endpoint for clinical trials of chronic kidney disease (CKD) progression. Analyses across a diverse spectrum of interventions and populations is required for acceptance of GFR decline as an endpoint. In an analysis of individual participant data, for each of 66 studies(total N participants, 186,312), we estimated treatment effects on the total GFR slope, computed from baseline to 3 years, and chronic slope, starting at 3-months after randomization, and on the clinical endpoint (doubling of serum creatinine, GFR<15 mL/min per 1.73m^2 or kidney failure with replacement therapy). We used a Bayesian mixed effects meta-regression model to relate treatment effects on GFR slope with those on the clinical endpoint across all studies and by disease groups (diabetes, glomerular diseases, CKD, or cardiovascular diseases). Treatment effects on the clinical endpoint were strongly associated with treatment effects on total slope (median R^2=0.97 [95% Bayesian credible interval [BCI] 0.82 to 1.00]), and moderately associated with those on chronic slope (R^2=0.55 [95% BCI 0.25 to 0.77]). There was no evidence of heterogeneity across disease. Our results support use of total slope as a primary endpoint for clinical trials of CKD progression.

Introduction

Global prevalence of Chronic kidney disease (CKD) is common, estimated at 9%, and associated with increased risk off kidney failure, cardiovascular disease and mortality[1]. A critical challenge in evaluation of therapies for CKD is that randomized controlled trials (RCTs) traditionally use kidney failure and doubling of serum creatinine as clinical endpoints, which are often late events in the progression of CKD[2]. In order to obtain sufficient clinical events to achieve adequate statistical power, RCTs often require substantial follow-up or are restricted to patients with rapidly progressive or late-stage disease, yet some interventions may have a greater effect when applied earlier in the disease course[3].

For all kidney diseases, patients must have a decline in glomerular filtration rate (GFR) to reach kidney failure providing strong biological plausibility for GFR decline as a surrogate endpoint for CKD progression in RCTs. There is also robust epidemiological evidence that both a single measurement of GFR and GFR decline over 1-to-3-year periods are strongly associated with subsequent kidney failure across subgroups of age, sex, and diabetes, hypertension, and cardiovascular disease status failure[4-7]. Using a smaller number of RCTs, we previously demonstrated that treatment effects on the GFR slope were highly predictive of treatment effects on clinical endpoints in hypothetical future RCTs[8]. On the basis of this evidence, GFR slope has been used to evaluate efficacy of treatments in less common conditions such as glomerular diseases[9-13] or when treatments with established efficacy and safety in one population are expanded to another population[14]. However, there remains uncertainty as to the use of GFR slope in more common causes of kidney diseases and across severity of kidney disease, further limiting evaluation of new therapies in many kidney disease settings.

RCTs in populations with earlier stages of CKD or who are at risk for CKD demonstrate the efficacy of novel agents for prevention of adverse outcomes including mortality, heart failure events and CKD

progression[15-33]. These studies provide the opportunity to evaluate if the strong associations between treatment effects on GFR slope with those on the clinical endpoint persist with inclusion of more diverse interventions and populations. The expanded set of studies also allow a more robust assessment of consistency of results across disease. Together, these data have the potential to strengthen the evidence for the validity of GFR slope as a surrogate endpoint for kidney failure, providing support for its use as a primary endpoint for trials of CKD progression across a broad series of settings.

Results

Study description

As described previously, we used a systematic search to identify RCTs evaluating interventions for CKD progression (See Protocol)[8,34-36]. Our prior publication included 47 treatment comparisons spanning 49 RCTs[8]. Since then, the updated search allowed identification of 19 new RCTs[15-20,22-26,28-33,37,38], leading to a total of 66 treatment comparisons (herein referred to as studies) for the current analysis (Extended Data Figure 1, Supplementary Tables 1-3)[35]. These 66 studies included 186,312 participants and represented 17 interventions across the four broad disease groups. The 17 interventions included four trials of sodium glucose cotransporter (SGLT)-2 inhibitors[16,23,38] as well as 16 studies of seven treatments not previously included, such as endothelin receptor antagonists[22], glucagon-like peptide-1 receptor agonist (GLP)[19,20], and mineralocorticoid receptor antagonists (MRA)[17] ().

Studies of patients with CKD from other causes or cause not specified (CKD), had lower mean eGFR than the studies of patients with diabetes, glomerular diseases (GN), or cardiovascular disease (CVD), and studies of patients with CVD had lower levels of urine albumin to creatinine ratio (ACR) and slower progression compared to the other disease groups (Table 1, Supplementary Table 4). Compared to the

prior set of studies, the current studies include populations with higher mean levels of GFR and at high risk for cardiovascular disease (CVD) (Extended Data Table 1). Summary characteristics of each individual study are reported in Supplementary Tables 5-7.

Treatment effects on the GFR slope and the clinical endpoint

Patterns of change in GFR (GFR slope) after initiation of an intervention are often nonlinear, with possibly differing direction and rates of changes in early follow-up (herein called acute effect) versus longer term follow- up (herein called chronic slope)[39]. The average rate of decline from beginning to the end of the study incorporates both elements (herein called total slope). We used a shared parameter mixed-effects model to estimate the effects of the treatment on the total GFR slope at three years (herein called 3-year total slope), as this was the approximate mean length of the included studies, and on chronic slope, computed from three months following randomization[8,40]. A benefit on slope was indicated by slower average GFR slope in the treatment versus the control arm. Across the full collection of studies and for studies in diabetes, CKD and GN, but not for CVD, there was an average benefit of the active treatments on the 3-year total slope and on the chronic slope compared with the control arms under a random effects meta-analysis (Figure 1, Extended Data Table 2).

Across the 66 studies, the median follow-up for the clinical endpoint was 35 (25^{th} to 75^{th} percentile 22 to 52) months, with shorter follow-up for studies with CVD compared to the three other disease groups (Extended Data Table 3). Over the full study duration, a total of 11,396 (6.1%) patients reached the primary clinical endpoint defined as the composite of kidney failure with replacement therapy (KFRT) (initiation of chronic treatment with dialysis or kidney transplantation); sustained GFR <15 mL/min per 1.73 m^2; or doubling of serum creatinine (equivalent to 57% decline in GFR). A benefit on the clinical endpoint was indicated by a hazard ratio less than 1 estimated within each study using Cox regression.

Across all of the studies as well as for studies in diabetes, CKD and GN, but not for CVD, an average benefit of treatment was observed (Figure 1, Extended Data Table 3). We also evaluated the treatment effect on the secondary clinical endpoint defined as the composite KFRT or sustained GFR <15 mL/min per 1.73 m^2. In the overall set of studies as well as for studies in diabetes, CKD, GN, and CVD, an average benefit of treatment was observed (Extended Data Table 3).

Trial level analysis in the overall study population

We used Bayesian mixed effects meta-regression analyses to relate the treatment effects on the clinical endpoint to the treatment effects on the total and chronic GFR slopes across the 66 studies[41]. We observed a strong agreement between the treatment effects on the 3-year total slope and on the clinical endpoint with an R^2 of 0.97 [95% BCI 0.82 to 1.00), (Figure 2 and Table 2). The slope of the meta-regression line is -0.35 (95% BCI -0.42 to -0.29) per mL/min per 1.73 m^2/year), indicating that a 0.75 ml/min per 1.73 m^2/year greater beneficial effect of the treatment on the total GFR slope is associated with an average 23.3% lower hazard ratio for the clinical end point (95% BCI 19.3% to 27.2%). The intercept of the meta-regression is -0.04 (95% BCI -0.09 to 0.01), indicating that in the absence of a treatment effect on the 3-year total slope, the average treatment effect on the clinical endpoint is likely to be small (i.e. 95% probability for the hazard ratio to be between 0.91 and 1.01).

In a secondary analysis, we also examined results for the total slope computed over 2 years (2-year total slope). The treatment effect on the 2-year total slope similarly showed a strong association with the treatment effect on the clinical endpoint (R^2 of 0.87 [95% BCI 0.64 to 0.97])(Extended Data Figure 2, Extended Data Table 4). However, the intercept is smaller than 0 (-0.11 [95% BCI -0.17 to -0.06]), indicating that a small average benefit on the clinical endpoint can be expected even in the absence of a treatment effect on the 2-year total slope.

For the chronic slope, the trial-level analysis provided an R^2 of 0.55 (95% BCI 0.25 to 0.77) (Figure 2). The slope of the meta-regression line differs substantially from 0, at -0.33 (95% BCI -0.46 to -0.20) indicating that a 0.75 ml/min per 1.73 m^2 per year greater beneficial treatment effect on the total GFR slope is associated with an average 21.8% lower hazard for the clinical end point (95% BCI 14.1% to 29.4%). The intercept of the regression line is nearly indistinguishable from 0 (-0.01, 95% BCI -0.10 to 0.10), indicating low risk of a false negative conclusion of the absence of a treatment effect on the clinical endpoint when there is no treatment effect on chronic slope (i.e. 95% probability for hazard ratio to be between 0.90 and 1.10.

Trial level analyses for the secondary clinical endpoint

We evaluated associations between the treatment effects on GFR slope with those on the secondary clinical endpoint. For the total slope, compared to the analysis of the primary clinical endpoint the median R^2 decreased from 0.97 (0.82, 1.00) to 0.92 (0.56, 0.99). For the chronic slope, the median R^2 increased from 0.55 (0.25, 0.77) to 0.73 (0.12, 0.98), although BCIs for the secondary clinical endpoint were wider than for the clinical endpoint, indicating reduced precision (Extended Data Table 5, Extended Data Figure 3). The median intercept of the meta-regressions for both chronic slope and total slope were now negative and had credible intervals which did not overlap 0, indicating that on average a modest benefit on the clinical endpoint may be present in the absence of a treatment effect on GFR slope.

Trial level analyses by subgroup

For the total slope, results were consistent according to CKD severity. For the subset of participants with baseline levels of ACR greater than 30 mg/g (3.39 mg/mmol) (Table 2). Similarly, results were consistent

across subgroups for studies with mean baseline GFR below or greater than 60 ml/min per 1.73 m^2, the GFR threshold for chronic kidney disease (Table 2). For chronic slope, there was a small increase in the median R^2 after restricting the patients in the set of studies with ACR available, although the Bayesian credible interval remained quite wide (0.65 [95% BCI 0.30, 0.88]). There was a steeper relationship between treatment effects on the chronic slope and those of the clinical endpoint for studies with higher baseline GFR compared to studies with lower baseline GFR.

We next evaluated whether the results for the meta-regression varied across disease groups by fitting an expanded meta-regression model using a Bayesian partial-pooling modeling approach. For both the 3-year total slope and the chronic slope, for each disease group the BCIs for the meta-regression slopes do not cross 0 and the BCIs for the meta-regression intercepts do cross 0 (Table 2 and Figure 3). For total slope, there was no evidence of heterogeneity in the meta-regression slope, intercept or R^2 across the disease subgroups. For chronic slope, there is a suggestion of modest heterogeneity among the posterior distribution of the meta-regression slopes (range for median slope from -0.24 for diabetes to -0.48 for CVD) and intercepts (range for median intercept from -0.12 for GN to 0.04 for CVD) across the disease groups, but with widely overlapping BCI. The estimated R^2 for the CVD group is 0.47 (95% BCI 0.01 to 0.99), lower than for the other diseases, which ranged from 0.71 (95% BCI 0.23 to 0.99) to 0.96 (95% BCI 0.23 to 1.00). The wide BCI and the small number of CVD studies preclude definitive conclusions. Similarly, in sensitivity analyses, removal of the CVD studies increased the median R^2 to 0.62 but with a wide BCI (0.25 to 0.90) (Extended Data Table 6). There was no impact of removal of the GN studies.

Assessment of model adequacy and outliers

We evaluated model adequacy and identified potential outliers by comparing for each study the observed hazard ratios for the clinical endpoint to the posterior predictive distribution for the estimated treatment effect on the clinical endpoint computed under a meta-regression model fit with that study left out. For the 3 year total slope, 9 (13.6%) of 66 studies had observed hazard ratios that fell outside of the middle 90% intervals of the posterior predictive distribution, and for the chronic slope, 10 (15%) of the observed hazard ratios fell outside of the 90% intervals (Supplementary Table 8). Removal of these studies had negligible effect on the meta-regression coefficients or R^2 values (Supplementary Tables 9-10). Under a perfect model it would be expected that approximately 7 studies would be outside of this range due to chance variation; the small number of additional studies outside of this range are consistent with adequacy of the model.

Prediction intervals and positive predictive value

For application of these results to use of slope as a surrogate endpoint in a future trial, we applied the above meta-regression results to compute Bayesian prediction intervals (BPI) that provide a 95% probability of including the true treatment effect on the clinical endpoint, for varying estimated treatment effects on the slope endpoints for hypothetical large (total sample size (N) of 1600), medium (N of 800) and small trials (N of 400).

As expected, with greater treatment benefit on GFR slope, the estimated treatment benefit on the clinical endpoints also increases (Table 3, Supplementary Table 11). For example, at a treatment effect on total GFR slope of 0.5 ml/min per 1.73 m^2, the estimated HR on the clinical endpoint is 0.80, whereas for a treatment effect on GFR slope of 1 ml/min per 1.73 m^2, it is 0.68. At smaller sample sizes, the estimated treatment effects remain unchanged but the BPI are wider. The BPI are considerably wider for the chronic slope than the 3-year total slope. For example, in a large future trial, the 95% BPI for the HR

on the clinical endpoint associated with an estimated treatment effect of 0.75 ml/min per 1.73m^2/year definitively excludes a HR of 1 or greater for the 3-year total slope (0.60 to 0.89) but not the chronic slope (0.51 to 1.18). Similarly, the threshold for a treatment effect on the GFR slope to assure the posterior probability of a clinical benefit, which we defined as HR for the clinical endpoint of less than 1, of greater than or equal to 97.5% is substantially smaller for the 3-year total slope than for the chronic slope (0.44 versus 1.26 ml/min per 1.73m^2 per year).

Within the different disease groups, the 95% BPI for the HR on the clinical endpoint associated with an estimated treatment effect of 0.75 ml/min per 1.73m^2 per year are also wider for chronic slope than for the 3 year-total slope (Extended Data Table 7). The magnitude of the difference is smaller for the CKD [chronic slope versus 3-year total slope [(0.54 to 0.96) versus. (0.57 to 0.90)] and diabetes [(0.63 to 1.13) versus (0.59 to 0.93)] disease groups compared to that for CVD [(0.30 to 1.52) versus (0.61 to 0.97)] and GN [(0.34 to 1.12) versus (0.52 to 0.95)] groups.

Discussion

CKD is a slowly progressive disease, often taking 10 years or longer from initiation of the disease to development of kidney failure. Therapies initiated earlier in the disease course are often more effective at preventing kidney failure. Validated surrogate endpoints are needed to assess efficacy of therapies in early stages of CKD. GFR decline has been proposed as a possible surrogate endpoint[42]. Evidence to support validity of a surrogate endpoint is derived from three main sources of evidence: biological plausibility, epidemiologic data demonstrating a strong and consistent relationship between the surrogate endpoint and outcome of interest, and trial level analyses. Trial level analyses, as are presented here, are widely regarded as the most important criterion for demonstrating the validity of a surrogate endpoint, and are the most difficult criterion to establish. We found that with analysis of GFR

slope using a robust method for each study, treatment effects on the total slope computed at 3 years accounted for an estimated 97% the variation between studies in treatment effects on the clinical endpoint with similar results across CKD severity and diverse disease groups. The strength of this trial-level association falls in in the range for a strong surrogate based on one proposed classification (e.g., R^2 greater than 0.72) and compares favorably with, and is arguably stronger than, widely used surrogate endpoints in other fields, such as progression-free survival for meta-static breast cancer or LDL cholesterol for major cardiovascular events[43,44]. We define the range of treatment effects on GFR slope that have high confidence for beneficial treatment effects on the clinical endpoint, providing guidance as how to interpret treatment effects on GFR slope in future RCTs. Our results indicate potential concerns with the use of the chronic slope, which ignores the acute effect of a drug on the GFR slope; an indication that the acute reduction in GFR due to treatment initiation might affect the clinical endpoint. The strength of these results derives from their demonstration across a highly heterogeneous interventions, study populations and disease severities. Together, these data provide the necessary evidence to support use of total GFR slope in RCTs evaluating therapies of CKD progression as a valid, fit-for-purpose, and robust surrogate endpoint that can be presented to regulatory agencies for approval for marketing authorization, to payors to support funding of these therapies, and to healthcare professionals and patients to inform of their benefit on slowing CKD progression and preventing kidney failure.

Our results for the trial level associations for chronic slope were considerably less favorable than in our earlier analyses[8]. The earlier analyses demonstrated that treatment effects on chronic slope accounted for a median 96% (95% BCI 63% to 100%) of the variation between studies in treatment effects on the clinical endpoint, similar to that of the total slope in those prior analyses [97% (95% BCI 78%, 100%)][8]. Since the total slope includes both the acute, and chronic GFR slope, the superior performance of total

slope compared to chronic slope in the current analysis of a more diverse set of studies suggests that the acute effect might be correlated with a component of the primary clinical endpoint. Consistent with this hypothesis, treatment effects on the chronic slope predicted treatment effects on the secondary clinical endpoint, which does not include doubling of serum creatinine, more accurately than it predicted treatment effects on the primary clinical endpoint. We note that although the chronic slope exhibited weaker trial level associations compared to the 3-year total slope relative to the primary clinical endpoint, its performance nonetheless falls within the range of R^2 values (0.49 to 0.72) that has been proposed to represent moderate performance for a surrogate endpoint[43]. As such, there might be a role for chronic slope in specific settings. While negative acute effects are most common in our set of studies, positive acute effects are also present among CKD studies[39]. If there is a suspicion that a positive acute effect may not indicate a true lasting impact on preservation of the kidney, use of the total slope may inflate risk of falsely concluding treatment benefit[39]. In such circumstances, the chronic slope may represent the more conservative endpoint[39].

Our trial-level model predicts a modest benefit on the primary clinical endpoint even when there is no effect on the 2-year total slope. This finding may, in part, reflect the average negative direction of the acute effects across the 66 RCTs included in these analyses. The acute effect has a greater impact on the computation of total slope when the total slope is computed over shorter time intervals. For future trials, computation of total GFR slope over shorter durations for interventions that do not have a negative acute effect could be considered.

The generalizability for trial level associations for establishing the validity of a surrogate is increased when based on a heterogeneous set of trials, but an often-asked question is whether the overall results can then be applied to specific subgroups. We found no evidence of heterogeneity in performance of

the 3-year total slope across disease groups. Similarly, we found that results were consistent in more versus less severe kidney disease as reflected by subgroups by level of GFR and ACR, and supported by a recent publication which looked at these same questions using the more rigorous analysis of examining interactions with the meta-regression parameters[45]. In contrast, for chronic slope, we observed modification of these associations based on disease severity[45], and noted possible modest heterogeneity across disease groups.

The observation of weaker associations for chronic slope in the CVD group might suggest that the less favorable results for the chronic slope in the current compared to earlier analyses may be related to inclusion of these studies. The CVD studies were not designed to evaluate efficacy of these interventions on CKD progression specifically and the included populations were at lower risk for progression compared to those included in the other disease groups as indicated by higher GFR and lower ACR. Future studies are required to explore how the associations vary by both disease and CKD.

There are several implications of these results. First, these data provide high confidence in the validity of the 3-year total GFR slope as a surrogate endpoint for kidney failure, thereby allowing it to be considered as a primary endpoint for clinical trials evaluating CKD progression across disease subgroups and CKD severity. The strength of this evidence is indicated in the lower limit for the 95% credible interval for the trial-level R^2 of 0.82, an increase from 0.78 in our prior results, now substantially exceeding the threshold of 0.72 proposed for a strong surrogate[43]. The increased precision was demonstrated while extending our analysis to a substantially more diverse collection of populations and interventions, providing support for widespread applicability. While the validity of the 3-year total slope as a surrogate endpoint appears quite general, its practical use must consider the circumstances of each trial. Decisions regarding the trial duration and the use of the total versus chronic slope should be made

in context of the specific trial. For example, we have previously demonstrated that compared to the clinical endpoint, use of the total slope can substantially reduce the sample size required for adequate power for trials without acute effects, particularly when performed in study populations with high baseline GFR, but may require similar or greater samples sizes in some situations, particularly for interventions with negative acute effects[46]. Second, the consistency in the performance of the total slope across disease severity or disease, suggests that the total slope can be used to explore heterogeneous effects among subgroups within an RCT. Similarly, GFR slope might be an appropriate endpoint for subsequent studies after initial RCTs showed benefit on the clinical endpoint allowing for demonstration of treatment benefit in different populations[9,14].

Strengths of this study include a large collection of RCTs with a diverse range of disease severity, diseases and interventions. Our application of a robust method for analysis of GFR slope that accounts for informative censoring and multiple potential sources of variability in GFR measurements over time allowed us to apply a uniform analysis of GFR slope across all studies[8,40]. Because we analyzed individual patient level data, we were able to characterize agreement between the GFR slope and clinical endpoints after adjusting for spurious correlations in sampling error that resulted from inclusion of the same GFR measurements in the GFR slope and clinical endpoints[41]. Thus, the meta-regression parameters refer to relationships between the true latent treatment effects on different endpoints rather than between the estimated effects which can contain substantial added variability.

There are several limitations. First, we were able to only evaluate the association between the treatment effects on GFR slope and on the clinical endpoints ascertained over the follow-up period for each trial and could not determine if treatment effects on GFR slope would better predict treatment effects on clinical endpoints over a longer time frame. Furthermore, differences in follow-up time

between studies and disease groups may have led to variation in the results. Second, our calculations of the prediction intervals and trial-level positive predictive value for a future trial assume that the relationship between the treatment effects on the slope and clinical endpoints are similar between the future trial and the previously conducted trials. Thus application of these results to future trials with different characteristics, diseases, and interventions than those included here must be done with caution. In particular, our results reflect the directions and magnitudes of the acute effects among previous trials; applications to future trials with especially large acute effects may incur added uncertainty. Finally, GFR slope is computed using estimated GFR from creatinine, which can be affected by factors other than GFR[47]. Thus caution should be applied when evaluating interventions known to effect body composition or dietary intake, two important factors than effect serum creatinine, as changes related to intervention could be another source of acute effects. Cystatin C is an alternative filtration marker that is thought to be more robust to non GFR determinants, and provides more accurate results in combination with creatinine[48,49]. Studies should evaluate it as an indicator of treatment effects prior to widespread use in clinical trials.

In aggregate, the results presented here indicate that GFR slope can be used as a validated surrogate endpoint for CKD progression. The selection of total versus chronic GFR slope, and the duration of time over which the total GFR slope is computed, should be made with consideration of the study design, population, intervention under investigation, and in the context of the specific drug development program.

Acknowledgements

We thank the National Kidney Foundation and its sponsors for funding this research. This work also received support from the Utah Study Design and Biostatistics Center, with funding in part from the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002538. The support and resources from the Center for High Performance Computing at the University of Utah are also gratefully acknowledged. We thank all investigators, study teams, and participants of the studies included in the meta-analysis (Appendix 1). This analysis includes SPRINT and TOPCAT Research Materials obtained from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of the SPRINT or TOPCAT or the NHLBI. The FSGS/FONT, HALT-PKD Study A and B were conducted by the investigators of the respective studies and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The data from these studies reported here were supplied by the NIDDK Central Repository. The manuscript does not necessarily reflect the opinions or views of the studies, the NIDDK Central Repository, or the NIDDK. We are grateful for data from AstraZeneca, Bayer AG, Boehringer Ingelheim, GlaxoSmithKline (made available through Vivli, Inc), Novartis (made available through Clinical Study Data Request), Novo Nordisk A/S, and Takeda (made available through Vivli, Inc). Vivli and CSDR have not contributed to or approved, and are not in any way responsible for, the contents of this publication.

Author Contributions

LAI, TG, and HJHL conceived and designed the study. LAI, HJHL, and JC conducted the literature search and study screening. GBA, SVB, FC-F, LDV, JF, MG, WGH, EI, THJTHZ, JBL, PKTL, BDM, BLN, RDP, GR, FPS, CW, JFMW, and MW collected data in the included studies. WC, TG, SM, and JC analyzed the data. LAI, WC, TG, HJHL, and BH wrote the first draft of the manuscript. All authors contributed to the interpretation of the data, provided critical feedback on paper drafts and approved the final draft.

Declaration of competing interests

LAI reports funding from National Institutes of Health (NIH), National Kidney Foundation (NKF), Omeros, Dialysis Clinics, Inc., and Reata Pharmaceuticals for research and contracts to Tufts Medical Center; consulting agreements to Tufts Medical Center with Tricida and HealthLogistics Interactive; and consulting agreements to Diamerix. WHC reports funding from the NKF to his institute for his graduate research assistantship. TG reports grant support from the NKF, Janssen Pharmaceuticals, Durect Corporation, Pfizer; statistical consulting Astrazeneca, CSL, and Boehringer-Ingleheim. BH consults for the NKF, Proxima Clinical Research, Value Analytics Health, and Guidepoint Global. GBA reports funding for research studies through Columbia U, Calliditas. Travere, Novartis, Equillium, Roche, Sanofi, Reata, Apellis, Nephronet, Goldfinch, and Vertex. Consultations: Sanofi-Genzyme, Roche - Genetech, Apellis, Novartis, Travere, Calliditas, Glaxo, Aurinia, Speaker for Glaxo and Aurinia both on lupus nephritis. SVB reports funding from the National Health and Medical Research Council of Australia, has served on the advisory boards of Bayer, AstraZeneca and Vifor Pharma, has received speaker's honoraria from Bayer, AstraZeneca, Pfizer and Vifor Pharma, all fees paid to the George Institute for Global Health. LDV was member of an Advisory Board for Travere. She is National Leader for the PROTECT Study and of the VISIONARY study. JF reports lecture and consulting honoraria from Alynlam, Amgen, AstraZeneca, Bayer, Boehringer, Calliditas, Fresenius, Novartis, Omeros, Travere, Vifor and member of the data safety

monitoring board in trials of NovoNordisk and Visterra. WGH reports personal funding from UK Medical Research Council (MC_UU_00017/3) and Kidney Research UK (MR/R007764/1), and grants to the University of Oxford from Boehringer Ingelheim and Eli Lilly. EI received research fundings and honorarium for lecture from AstraZeneca, Tanabe Mitsubishi, Boehringer Ingelheim, Daiich Sankyo, Kyowa Kirin, and NovoNordisk. JBL is a consultant for CSL, BIOVIE. PKTL received speaker honorarium from FibroGen, AstraZeneca and Baxter. BDM is the National Leader in the ASCEND-D and ND-trial of GSK, the PROTECT and DUPLEX of TRAVERE, steering committee member in the Novartis APPLAUSE trial and ATACICEPT trial of Vera Therapeutics. BLN has received fees for advisory boards, scientific presentations, continuing education, steering committee roles and travel for AstraZeneca, Bayer, Boehringer and Ingelheim, Cambridge Healthcare, Janssen and Medscape with all honoraria paid to his institution. RDP received research support from Otsuka, Reata, Kadmon, Sanofi Genzyme, and the US Department of Defense (TAME PKD); has been a member of steering committees for Sanofi Genzyme, Otsuka, and TAME PKD with fees paid to employing institutions; and has provided consultancy for Navitor, Palladiobio, Reata, and Otsuka. GR has had consultancy agreement with Alexion Pharmaceuticals Inc, Janssen Pharmaceutical, Akebia Therapeutics, Biocryst Pharmaceuticals, Menarini Ricerche SpA, AstraZeneca; speaker honoraria/travel reimbursement from Boehringer Ingelheim, Novartis. MW reports consulting fees from Amgen and Freeline paid to him. HJLH from the NKF to his institute, is a consultant for AbbVie, AstraZeneca, Bayer, Boehringer Ingelheim, Chinook, CSL Behring, Dimerix, Eli-Lilly, Gilead, GoldFinch, Janssen, Merck, NovoNordisk, Travere Pharmaceuticals. SM, JC, MG, FCF, THJ, FPS, CW, and JFMW have no conflicts to report associated with this manuscript.

References

1. Cockwell, P. & Fisher, L.A. The global burden of chronic kidney disease. *Lancet* **395**, 662-664 (2020).
2. Levey, A.S.*, et al.* GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *Am J Kidney Dis* **64**, 821-835 (2014).
3. Schievink, B.*, et al.* Early renin-angiotensin system intervention is more beneficial than late intervention in delaying end-stage renal disease in patients with type 2 diabetes. *Diabetes, obesity & metabolism* **18**, 64-71 (2016).
4. Grams, M.E.*, et al.* Evaluating Glomerular Filtration Rate Slope as a Surrogate End Point for ESKD in Clinical Trials: An Individual Participant Meta-Analysis of Observational Data. *J Am Soc Nephrol* **30**, 1746-1755 (2019).
5. Kidney Disease: Improving Global Outcomes (KDIGO). KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int Suppl.* **3**, 1-150 (2013).
6. Gutierrez, O.M.*, et al.* Association of Estimated GFR Calculated Using Race-Free Equations With Kidney Failure and Mortality by Black vs Non-Black Race. *JAMA* **327**, 2306-2316 (2022).
7. Coresh, J.*, et al.* Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *JAMA* **311**, 2518-2531 (2014).
8. Inker, L.A.*, et al.* GFR Slope as a Surrogate End Point for Kidney Disease Progression in Clinical Trials: A Meta-Analysis of Treatment Effects of Randomized Controlled Trials. *J Am Soc Nephrol* **30**, 1735-1745 (2019).
9. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). Identifier NCT04573478, Atrasentan in Patients With IgA Nephropathy (ALIGN). (2020).
10. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). Identifier: NCT03608033, Study of the Safety and Efficacy of OMS721 in Patients With Immunoglobulin A (IgA) Nephropathy.
11. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). Identifier: NCT03643965, Efficacy and Safety of Nefecon in Patients With Primary IgA (Immunoglobulin A) Nephropathy (Nefigard).
12. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). ClinicalTrials.gov Identifier: NCT04578834, Study of Efficacy and Safety of LNP023 in Primary IgA Nephropathy Patients (APPLAUSE-IgAN).
13. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). A Study of the Effect and Safety of Sparsentan in the Treatment of Patients With IgA Nephropathy (PROTECT). (2018).
14. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). Identifier NCT00287391 A Trial to Learn How Well Finerenone Works and How Safe it is in Adult Participants With Non-diabetic Chronic Kidney Disease (FIND-CKD). (2021).
15. Jamerson, K.*, et al.* Benazepril plus amlodipine or hydrochlorothiazide for hypertension in high-risk patients. *The New England journal of medicine* **359**, 2417-2428 (2008).
16. Heerspink HJL*, et al.* Dapagliflozin in Patients with Chronic Kidney Disease (DAPA-CKD). *The New England journal of medicine* **383**, 1436-1446 (2020).
17. Bakris, G.L.*, et al.* Effect of Finerenone on Chronic Kidney Disease Outcomes in Type 2 Diabetes. *The New England journal of medicine* **383**, 2219-2229 (2020).
18. Badve, S.V.*, et al.* Effects of Allopurinol on the Progression of Chronic Kidney Disease. *The New England journal of medicine* **382**, 2504-2513 (2020).

19. Mann, J.F.E*., et al.* Liraglutide and Renal Outcomes in Type 2 Diabetes. *The New England journal of medicine* **377**, 839-848 (2017).
20. Hernandez, A.F*., et al.* Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. *Lancet* **392**, 1519-1529 (2018).
21. Perkovic, V*., et al.* Canagliflozin and renal outcomes in type 2 diabetes: results from the CANVAS Program randomised clinical trials. *The lancet. Diabetes & endocrinology* **6**, 691-704 (2018).
22. Heerspink, H.J.L*., et al.* Atrasentan and renal events in patients with type 2 diabetes and chronic kidney disease (SONAR): a double-blind, randomised, placebo-controlled trial. *Lancet* **393**, 1937-1947 (2019).
23. Perkovic, V*., et al.* Canagliflozin and Renal Outcomes in Type 2 Diabetes and Nephropathy. *The New England journal of medicine* **380**, 2295-2306 (2019).
24. Rosenstock, J*., et al.* Effect of Linagliptin vs Glimepiride on Major Adverse Cardiovascular Outcomes in Patients With Type 2 Diabetes: The CAROLINA Randomized Clinical Trial. *JAMA* **322**, 1155-1166 (2019).
25. Rosenstock, J*., et al.* Effect of Linagliptin vs Placebo on Major Cardiovascular Events in Adults With Type 2 Diabetes and High Cardiovascular and Renal Risk: The CARMELINA Randomized Clinical Trial. *JAMA* **321**, 69-79 (2019).
26. Sprint Research Group*, et al.* A Randomized Trial of Intensive versus Standard Blood-Pressure Control. *The New England journal of medicine* **373**, 2103-2116 (2015).
27. Zannad, F*., et al.* Heart failure and mortality outcomes in patients with type 2 diabetes taking alogliptin versus placebo in EXAMINE: a multicentre, randomised, double-blind trial. *Lancet* **385**, 2067-2076 (2015).
28. Bonaca, M.P*., et al.* Long-term use of ticagrelor in patients with prior myocardial infarction. *The New England journal of medicine* **372**, 1791-1800 (2015).
29. Pitt, B*., et al.* Spironolactone for heart failure with preserved ejection fraction. *The New England journal of medicine* **370**, 1383-1392 (2014).
30. McMurray, J.J*., et al.* Angiotensin-neprilysin inhibition versus enalapril in heart failure. *The New England journal of medicine* **371**, 993-1004 (2014).
31. Gipson, D.S*., et al.* Clinical trial of focal segmental glomerulosclerosis in children and young adults. *Kidney international* **80**, 868-878 (2011).
32. Wallentin, L*., et al.* Ticagrelor versus clopidogrel in patients with acute coronary syndromes. *The New England journal of medicine* **361**, 1045-1057 (2009).
33. McMurray, J.J*., et al.* Effects of candesartan in patients with chronic heart failure and reduced left-ventricular systolic function taking angiotensin-converting-enzyme inhibitors: the CHARM-Added trial. *Lancet* **362**, 767-771 (2003).
34. Inker, L.A*., et al.* GFR decline as an alternative end point to kidney failure in clinical trials: a meta-analysis of treatment effects from 37 randomized trials. *Am J Kidney Dis* **64**, 848-859 (2014).
35. Higgins, J. & Green, S. Cochrane Handbook for Systematic Reviews of Interventions. (The Cochrane Collaboration, 2011).
36. Heerspink, H.J.L*., et al.* Change in albuminuria as a surrogate endpoint for progression of kidney disease: a meta-analysis of treatment effects in randomised clinical trials. *The lancet. Diabetes & endocrinology* **7**, 128-139 (2019).
37. Holman, R.R*., et al.* Effects of Once-Weekly Exenatide on Cardiovascular Outcomes in Type 2 Diabetes. *The New England journal of medicine* **377**, 1228-1239 (2017).
38. Neal, B*., et al.* Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. *The New England journal of medicine* **377**, 644-657 (2017).

39. Neuen, B.L.*, et al.* Acute Treatment Effects on GFR in Randomized Clinical Trials of Kidney Disease Progression. *J Am Soc Nephrol* **33**, 291-303 (2022).
40. Vonesh, E.*, et al.* Mixed-effects models for slope-based endpoints in clinical trials of chronic kidney disease. *Statistics in medicine* **38**, 4218-4239 (2019).
41. r Stan Development Team. "RStan: the R interface to Stan." R package version 2.21.5. (2022).
42. Levey, A.S.*, et al.* Change in Albuminuria and GFR as End Points for Clinical Trials in Early Stages of CKD: A Scientific Workshop Sponsored by the National Kidney Foundation in Collaboration With the US Food and Drug Administration and European Medicines Agency. *Am J Kidney Dis* **75**, 84-104 (2020).
43. Prasad, V., Kim, C., Burotto, M. & Vandross, A. The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses. *JAMA internal medicine* **175**, 1389-1398 (2015).
44. Delahoy, P.J., Magliano, D.J., Webb, K., Grobler, M. & Liew, D. The relationship between reduction in low-density lipoprotein cholesterol by statins and reduction in risk of cardiovascular outcomes: An updated meta-analysis. *Clinical Therapeutics* **31**, 236-244 (2009).
45. Collier, W.*, et al.* Evaluation of Variation in the Performance of GFR Slope as a Surrogate End Point for Kidney Failure in Clinical Trials that Differ by Severity of CKD. *Clinical journal of the American Society of Nephrology : CJASN* **18**, 183-192 (2023).
46. Greene, T.*, et al.* Performance of GFR Slope as a Surrogate End Point for Kidney Disease Progression in Clinical Trials: A Statistical Simulation. *J Am Soc Nephrol* **30**, 1756-1769 (2019).
47. Levey, A.S., Coresh, J., Tighiouart, H., Greene, T. & Inker, L.A. Measured and estimated glomerular filtration rate: current status and future directions. *Nature reviews. Nephrology* **16**, 51-64 (2020).
48. Inker, L.A.*, et al.* New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. *The New England journal of medicine* **385**, 1737-1749 (2021).
49. Pottel, H.*, et al.* Cystatin C-Based Equation to Estimate GFR without the Inclusion of Race and Sex. *The New England journal of medicine* **388**, 333-343 (2023).

Figure 1: Treatment effect on 3-year total slope, chronic slope and on the clinical endpoint overall and by disease group

The clinical endpoint is defined as kidney failure with replacement therapy, GFR <15 mL/min per 1.73 m^2 or doubling of serum creatinine. The circles represent the meta-analyzed treatment effects and the horizontal lines represent the associated 95% confidence intervals. N, number; CKD, chronic kidney disease. CKD refers to diseases other than diabetes or glomerular disease or cause not yet specified.

Figure 2: Trial level analyses for the association between treatment effects on GFR slope and treatment effects on the clinical endpoint

Circles represent separate studies with the size of the circle proportional to the number of events (kidney failure with replacement therapy, GFR <15 mL/min per 1.73 m^2, or doubling of serum creatinine). The colors of circles indicate intervention types. The black line is the line of regression through the studies. Blue lines are the 95% pointwise Bayesian confidence band. Pink dashed lines are the 95% pointwise Bayesian prediction bands computed from the model. Triangles indicate studies where the estimated treatment effects are beyond the margins. RASB, renin angiotensin system blockers; CCB, calcium channel blocker; BP, blood pressure; DPP-4, Dipeptidyl peptidase 4; GLP-1, Glucagon-like peptide-1, RMSE, Root mean square error; SGLT2, Sodium glucose cotransporter.

Figure 3: Trial level analyses for the association between treatment effects on GFR slope and treatment effects on the clinical endpoint by disease groups

Each circle represents a separate study with the size of the circle proportional to the number of events. Circles are coloured only if the associated study falls within that disease category; the colors indicate intervention type with the legend shown in Figure 2. The black line is the regression line through the

studies. The blue lines are the 95% pointwise Bayesian confidence bands. The pink dashed lines are the 95% pointwise Bayesian prediction bands computed from the model.

Table 1. Clinical characteristics of the population stratified by disease etiology

Group	Number studies (participants)	Age*	Female*	Diabetes	eGFR	ACR	Rate of progression	Interventions evaluated
All Studies	66 (186,312)	63.4 (10.5)	58,632 (35)	126,439 (68)	68.0 (109.5)	68 (2450)	-3.17 (-3.61, -2.72)	
Disease groups								
CKD	28 (20,149)	57.5 (13.7)	7621 (38)	4856 (24)	39.1 (20.5)	235 (6)	-3.405 (-3.90, -2.90)	RASB v con, RASB v CCB, SGLT-2 I, Allopurinol, Low v Usual Diet, Nurse coord care, Albuminuria Targeted Protocol, Statin + Ezetimibe, Low v Usual BP
Diabetes	21 (101,005)	63.9 (8.7)	35659 (35)	101,005 (100)	69.3 (88.7)	71 (517)	-3.37 (-4.29, -2.44)	RASB v con, RASB v CCB, SGLT-2 I, DPP-4 I, GLP-1 Ag, MRA, ERA, Intensive Glucose, Antiplatelet, Low v Usual BP
Glomerular	10 (1,527)	38.6 (14.3)	562 (37)	5 (1)	78.7 (40.5)	1471 (2)	-3.51 (-4.82, -2.20)	RASB versus Con Immunosuppression
CVD	7 (63,631)	67.1 (9.5)	15052 (33)	20,573 (32)	74.9 (39.2)	16 (18)	-1.19 (-1.61, -0.77)	RASB versus con, RASB + CCB, MRA, Antiplatelet, Low v Usual BP

Rate of progression is indicated by chronic slope in the control arm. Age and eGFR are shown as mean (SD), ACR as geometric mean (geometric SD); Female and diabetes as number of participants (percent), control slope as mean (95% confidence intervals).

*Age was not available for four studies, and sex for one study.

ACR, Urine albumin: creatinine ratio, CKD, chronic kidney disease, CVD, cardiovascular disease, eGFR, estimated glomerular filtration rate, N, sample size; RASB, renin-angiotensin system blockage; v, versus; Con, control; CCB, calcium channel blockers; SGLT2, Sodium glucose cotransporter, DPP-4, Dipeptidyl peptidase 4; GLP-1, Glucagon-like peptide-; inhibitor; Ag, Agonists; MRA, mineral corticoid receptor antagonists; ERA, endothelial receptor antagonists,

Table 2: Trial level analysis by subgroups

Group	Subgroup	N Studies (N Interventions)	Meta-Regression Slope	Intercept	R^2	RMSE
Total Slope computed over 3 years						
Overall		66 (17)	-0.35 (-0.42, -0.29)	-0.04 (-0.09, 0.01)	0.97 (0.82, 1.00)	0.05 (0.02, 0.12)
Disease	CKD	28 (9)	-0.36 (-0.52,-0.25)	-0.05 (-0.14,0.04)	0.91 (0.51,1.00)	0.06 (0.01,0.14)
	Diabetes	21 (10)	-0.32 (-0.42,-0.21)	-0.05 (-0.12,0.01)	0.89 (0.49,1.00)	0.06 (0.01,0.16)
	Glomerular	10 (2)	-0.33 (-0.46,-0.22)	-0.06 (-0.32,0.07)	0.99 (0.85,1.00)	0.06 (0.01,0.23)
	CVD	7 (5)	-0.34 (-0.43,-0.25)	-0.02 (-0.10,0.10)	0.98 (0.82,1.00)	0.05 (0.01,0.15)
GFR	< 60	40 (14)	-0.30 (-0.44, -0.14)	-0.07 (-0.17, 0.01)	0.83 (0.25, 0.99)	0.06 (0.02, 0.13)
	> 60	26 (11)	-0.37 (-0.46, -0.28)	-0.02 (-0.11, 0.07)	0.98 (0.83, 1.00)	0.06 (0.01, 0.19)
Restricted to studies with ACR available		55 (14)	-0.35 (-0.45, -0.27)	-0.07 (-0.14, -0.01)	0.96 (0.75, 1.00)	0.06 (0.02, 0.14)
Restricted to studies with ACR available and participants with ACR >30 mg/g		55 (14)	-0.34 (-0.44, -0.23)	-0.06 (-0.13, 0.02)	0.95 (0.69, 1.00)	0.05 (0.02, 0.13)
Chronic Slope						
Overall		66 (17)	-0.33 (-0.46, -0.20)	-0.01 (-0.10, 0.10)	0.55 (0.25, 0.77)	0.19 (0.13, 0.27)
Disease	CKD	28 (9)	-0.33 (-0.53,-0.15)	-0.06 (-0.19,0.06)	0.74 (0.18,0.99)	0.09 (0.01,0.20)
	Diabetes	21 (10)	-0.24 (-0.37,-0.12)	0.02 (-0.09,0.14)	0.71 (0.23,0.99)	0.11 (0.02,0.22)
	Glomerular	10 (2)	-0.41 (-0.83,-0.20)	-0.12 (-0.50,0.10)	0.96 (0.23,1.00)	0.12 (0.01,0.57)
	CVD	7 (5)	-0.48 (-1.33,-0.07)	0.04 (-0.14,0.27)	0.47 (0.01,0.99)	0.23 (0.05,0.53)
GFR	< 60	40 (14)	-0.15 (-0.27, -0.03)	-0.12 (-0.23, -0.03)	0.47 (0.02, 0.94)	0.09 (0.02, 0.18)
	> 60	26 (11)	-0.56 (-0.80, -0.33)	0.12 (-0.03, 0.29)	0.77 (0.35, 0.95)	0.23 (0.11, 0.38)
Restricted to studies with ACR available		55 (14)	-0.30 (-0.44, -0.18)	-0.04 (-0.14, 0.07)	0.65 (0.30, 0.88)	0.15 (0.08, 0.24)
Restricted to studies with ACR available and participants with ACR >30 mg/g		55 (14)	-0.22 (-0.35, -0.11)	-0.07 (-0.18, 0.03)	0.66 (0.21, 0.94)	0.12 (0.03, 0.21)

CKD, chronic kidney disease; CVD, cardiovascular disease; ACR, urine albumin to creatinine ratio; N, sample size; R^2, coefficient of determination; RMSE, root mean squared error. To convert ACR in mg/g to mg/mmol, multiply by 0.113.

Table 3: Application of GFR slope as Surrogate Endpoint in New RCT: Predicted Treatment effect on clinical endpoint and Positive Predictive Value

GFR slope	Observed Treatment effect on change in GFR slope	Large RCT		Medium RCT		Small RCT	
		Median HR and 95% Bayesian Prediction Interval	PPV_{trial}	Median HR and 95% Bayesian Prediction Interval	PPV_{trial}	Median HR and 95% Bayesian Prediction Interval	PPV_{trial}
Total slope computed at 3 years	**0.5**	0.80 (0.66, 0.98)	0.98	0.80 (0.62, 1.03)	0.96	0.80 (0.58, 1.12)	0.91
	0.75	0.74 (0.60, 0.89)	1.00	0.74 (0.57, 0.94)	0.99	0.74 (0.53, 1.02)	0.97
	1	0.68 (0.54, 0.82)	1.00	0.68 (0.52, 0.86)	1.00	0.68 (0.48, 0.93)	0.99
Threshold for treatment effect on GFR slope associated with 97.5% probability of clinical benefit		0.44		0.58		0.79	
Chronic slope	**0.5**	0.84 (0.55, 1.28)	0.80	0.84 (0.54, 1.32)	0.78	0.84 (0.51, 1.40)	0.75
	0.75	0.78 (0.51, 1.18)	0.89	0.78 (0.49, 1.22)	0.87	0.78 (0.47, 1.29)	0.84
	1	0.72 (0.47, 1.09)	0.94	0.72 (0.45, 1.12)	0.93	0.72 (0.43, 1.18)	0.91
Threshold for treatment effect on GFR slope associated with 97.5% probability of clinical benefit		1.26		1.35		1.51	

GFR, glomerular filtration rate. Units of GFR are mL/min per 1.73 m^2. The treatment effect on GFR slope is expressed as a mean difference in units of mL/min per 1.73 m^2/year. The treatment effect on the clinical endpoint is expressed as a hazard ratio. PPV_{trial} denotes the trial level positive predictive value for clinical benefit, defined as a hazard ratio for the clinical endpoint < 1. A large RCT corresponds to a total sample size (N) of 1600 allocated equally to the treatment and control groups for RCTs whose average follow-up is 3 years with measurements every 6 months. A medium RCT was defined as having sample size of 800 and a small trial, a sample size of 400. The PPV_{trial} should be interpreted in relation to estimated probabilities of clinical benefit of 0.67, 0.63, or 0.60 for large, median and small RCTs, respectively, when there is no estimated treatment effect on the 3-year total slope, or 0.51, 0.51, and 0.52 when there is no estimated treatment effect on the chronic slope.

Methods

The research complies with all relevant ethical regulations. Protocols of individual trials were reviewed by the participating centers' institutional review boards. The analyses in this study were deemed exempt from review by the Tufts Medical Center Institutional Review Board.

Overview

Our meta-analytic approach followed seven main steps: (1) Search and identify data from published randomized control trials (RCTs) using a standardized approach; (2) Obtain agreement for use of individual data and transfer to Data Coordinating Center at Tufts Medical center or identify methods for cloud based data access; (3) Within each RCT, use intent-to-treat analyses to estimate a) the effect of the treatment on GFR slope under a shared parameter mixed effect model, and b) the effect of the treatment on clinical endpoint (defined as a composite of kidney failure, sustained GFR <15 mL/min per 1.73 m^2; or doubling of serum creatinine) under a Cox proportional hazards regression model; (4) Quantify the association between treatment effects on the GFR slope and treatment effects on clinical endpoints across RCTs using trial-level Bayesian meta-regression analysis; (5) Identify outliers and evaluate model adequacy across the individual trials by using a cross-validation approach to compare the observed treatment effect on the clinical endpoint in each trial to the posterior predictive distribution for that trial derived from the trial-level model fit to the remaining trials; (6) Quantify the consistency in these associations across disease subgroups using Bayesian meta-regression with partial pooling; (7) Assess the utility of using slope as a surrogate endpoint in a future trial by providing 95% Bayesian prediction intervals and trial level positive predictive values corresponding with a range of possible treatment effects on the slope endpoint in a hypothetical future trial.

The analytic approach is based on the causal association framework in which the validity of surrogate endpoints is evaluated based on the relationship between the average causal effect of the treatment on the surrogate endpoint and the average causal effect of the treatment on the clinical endpoint across a population of RCTs. This approach takes advantage of the fact that the average causal effects on the surrogate and clinical endpoints can be estimated with little bias within each RCT by applying intent-to-treat analyses[50-53]. Confidence in the use of these results for a future new trial is increased when the previously conducted trials provide evidence that the treatment effect on the clinical endpoint is consistently predicted from the treatment effect on the surrogate across a broad collection of RCTs evaluating treatments which operate through diverse mechanisms and across diverse patient populations. As a result, the empirical support for a surrogate endpoint is enhanced when the trial-level meta-regression is conducted over a heterogeneous rather than a homogenous collection of previously conducted RCTs.

Datasets and analytical groups

As previously described, we conduct a systematic review of the literature which we update periodically to identify randomized controlled trials (RCTs) of CKD progression (Extended Data Figure 1, Protocol)[8,34-36]. For trials that evaluated more than one intervention, we included a separate group for each independent treatment comparison, such that some participants were included in more than one analytical comparison[54-58]. In the prior analyses, we included 47 treatment comparisons spanning 49 RCTs[8]. Since then, from a systematic review conducted in August 2018 to April 2020, we were able to add 19 new RCTs, including three trials of SGLT-2 inhibitors and 16 studies of seven novel treatments not previously included, such as endothelin receptor antagonists, glucagon-like peptide-1 receptor agonist, and mineralocorticoid receptor antagonists (Extended Data Table 1). This led to 66 treatment comparisons (herein referred to as studies) for the current analysis. In comparison to the prior analyses,

the current set of studies include populations with higher mean levels of GFR and at high risk for cardiovascular disease.

For each study, we defined the active treatment as the treatment hypothesized to produce the greater reduction in the risk of the clinical endpoint. We categorized the studies by intervention type: renin angiotensin system blockade (RASB) versus control, RASB versus. calcium channel blocker (CCB), RASB + CCB versus placebo; intensive blood pressure (BP) control, low protein diet, antiplatelet, allopurinol, nurse coordinated care, albuminuria targeted therapy, endothelin receptor antagonist, mineralcorticoid receptor antagonist, DPP-4, Dipeptidyl peptidase 4; GLP-1, Glucagon-like peptide-1, SGLT2, Sodium glucose cotransporter; intensive glucose, statin+ezetimibe, immunosuppressive therapy (including steroid, azathioprine, tacrolimus, fish oil, plasmapheresis). We grouped the studies according to diseases which defined enrollment, categorized as diabetes, glomerular diseases (GN), CKD from other causes or cause not specified (CKD), or cardiovascular diseases (CVD) (Supplementary Table 4).

Clinical Endpoint

The clinical endpoint was defined as a composite of kidney failure with replacement therapy (KFRT) (initiation of chronic treatment with dialysis or kidney transplantation); a sustained GFR <15 mL/min per 1.73 m^2 for participants with baseline GFR $\geq$ 25 ml/min per 1.73 m^2 (herein referred to as GFR < 15 mL/min per 1.73 m^2); or doubling of serum creatinine, over full study duration. This is the definition that is standard practice of clinical trials for CKD progression. The secondary clinical endpoint was defined as a composite of KFRT or GFR <15 mL/min per 1.73 m^2 over full study duration. The exclusion of doubling of serum creatinine from the secondary clinical endpoint assures that the remaining components of the composite endpoint signify the occurrence or near-occurrence of a kidney failure, considered to be the true clinical event, and reduces the chance that endpoint would be affected by large acute effects.

Compared to the primary clinical endpoint, the secondary clinical endpoint poses a number of challenges for our trial level analyses, including: 1) Treatment effects are estimated with reduced precision due to fewer events, reducing statistical power, particularly for subgroup analyses; 2) Trial-level analyses are weighted even more heavily towards the subgroup of trials with lower levels of baseline GFR as KFRT and GFR <15 mL/min per 1.73 m^2 are rare events for trials with high baseline GFR; 3) the competing risk of death will have a greater impact. For these reasons the composite of KFRT and GFR <15 mL/min per 1.73 m^2 is treated as the secondary clinical endpoint and the broader composite which includes doubling of serum creatinine in addition to these two events is designated as the primary clinical endpoint in these analyses.

GFR

GFR was estimated using the CKD-EPI 2009 creatinine equation[59]. Creatinine was standardized to isotope dilution mass spectroscopy (IDMS) traceable reference methods using direct comparison or was reduced by 5% from non-IDMS traceable methods as has previously been described (Supplementary Table 6)[60].

Analyses of the total and chronic GFR slope (surrogate endpoints*).*

Changes in GFR after intervention are often nonlinear, with possibly differing direction and rates of change in early follow-up (herein called acute slope) and longer-term follow-up (herein called chronic slope). The total overall average rate of decline from randomization to a defined time point late in follow-up incorporates both elements (herein called total slope). We used a shared parameter mixed-effects model that we had previously described[8,40,61]. The model assumes a linear mean GFR slope starting at three months post randomization. Differences between the randomized groups in the mean

GFR levels at 3 months follow-up, the mean slopes from 3 months onward, and the estimated mean changes from baseline to either 2- or 3-year follow-up factored by the follow-up duration represented the treatment effects on the acute, chronic, and total slopes, respectively. Our primary analyses focused on the total GFR slope computed at three years (herein called 3-year total slope), the approximate mean length of the included studies, and on the chronic slope. In a secondary analysis, we considered the total GFR slope at two years (2-year total slope).

The model accounts for between-subject variability in GFR trajectories by inclusion of random slopes and intercepts. Random residuals account for variation in individual GFR measurements from the underlying GFR trajectories. A power of the mean (POM) model for the residual variance is used to allow for greater variation in individual GFR measurements at higher GFR. The model allows for non-uniform treatment effects in which treatments slow progression by a greater extent among patients with faster GFR decline than for patients with slower GFR decline by allowing different between-patient slope variances in the treatment and control groups[8,40]. In studies in which at least 15 subjects died or reached KFRT, the model accounts for informative censoring by these events by incorporating the mixed model for the GFR measurements within a shared parameter model in which the risk of KFRT or death was assumed to be related to the random slopes and intercepts of the GFR part of the model[62,63]. Simplified models were used in cases where convergence could not be obtained with the full model. The full shared parameter mixed effects models was fit using the SAS (version 9.4) nonlinear mixed-effects regression procedure, NLMIXED.

Trial level analysis.

Estimation of treatment effects As described above, we applied the mixed effects models to estimate the treatment effects on GFR slope within each study by treatment arm, with treatment effects

expressed as differences in the mean GFR slopes between the treatment versus control groups in units of ml/min/1.73m^2/year. We estimated treatment effects on the clinical endpoint by performing separate Cox proportional hazard regressions to estimate log hazard ratios for the treatment in each trial. We summarized the mean treatment effects on each individual endpoint using restricted maximum likelihood with study as a random effect (rma.uni function in R metafor package)[64].

Bayesian meta-regression We next applied a trial-level Bayesian mixed effects meta-regression model to relate the treatment effects on the clinical endpoint to the treatment effects on GFR slope with study as the unit of analysis[41,50]. The model relates the treatment effects on the two endpoints after accounting for the standard deviations of the random errors in the estimated effects in each study and the correlation of these errors with each other. This approach takes advantage of the fact that the average causal effects on the surrogate and clinical endpoints can be estimated with little bias within each randomized trial by applying intent-to-treat analyses[51-53]. The model includes two stages, where the first stage relates the estimated treatment effects to the true latent treatment effects within each trial, and the second stage describes the relationships between the true latent treatment effects across the different trials.

To express the model precisely, let i = 1, 2, ..., denote the 66 randomized treatment comparisons included in the analysis. For simplicity, as most trials included a single treatment comparison, we abuse the notation slightly and write that the index *i* refers to the i^{th} trial. We let θ_i and γ_i denote the true latent treatment effects on the clinical endpoint and on the GFR slope in the i^{th} trial, and use $\hat{\theta}_i$ and $\hat{\gamma}_i$ to indicate the estimated effects obtained as described above. The first stage of the model relates the estimated and true latent treatment effects in the i^{th} trial by:

$$\begin{bmatrix}\hat{\theta}_i\\ \hat{\gamma}_i\end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix}\theta_i\\ \gamma_i\end{bmatrix}, \begin{bmatrix}\sigma_i^2 & r_i\sigma_i\delta_i\\ r_i\sigma_i\delta_i & \delta_i^2\end{bmatrix}\right).$$

Here, σ_i is the standard error of the estimated treatment effect on the clinical endpoint and δ_i is the standard error of the estimated treatment effect on GFR slope in the i^{th} trial, and r_i is the correlation between these estimated treatment effects. We used robust sandwich estimates to estimate the correlations r_i within each trial. For two trials which included two different comparisons of active treatments to control, we also accounted for the correlations between the two treatment contrasts with the same control group. The notation MVN() indicates that the estimated treatment effects are assumed to follow a bivariate normal distribution given the true treatment effects within each trial; this assumption is satisfied to a high degree of accuracy due to the central limit theorem.

The second stage of the model characterizes the variation in the true latent treatment effects on GFR slope and on the clinical endpoint across the trials. This second stage is expressed as

$$\begin{bmatrix}\theta_i\\ \gamma_i\end{bmatrix} \sim N\left(\begin{bmatrix}\mu_\theta\\ \mu_\gamma\end{bmatrix}, \begin{bmatrix}\sigma_\theta^2 & \mathrm{R}\sigma_\theta\sigma_\gamma\\ \mathrm{R}\sigma_\theta\sigma_\gamma & \sigma_\gamma^2\end{bmatrix}\right)$$

where μ_θ and μ_γ are respectively the means of the true latent treatment effects on the clinical endpoint and on GFR slope in the population of trials represented by this meta-regression, σ_θ and σ_γ are the standard deviations of the true latent treatment effects across the population of trials, and R is the correlation between the true latent treatment effects on the two endpoints. Based on this 2-stage model, the slope and intercept of the meta-regression line predicting the true latent treatment effect on the clinical endpoint from the true latent treatment effect on the surrogate endpoint are given by $\beta = \mathrm{R}\sigma_\theta/\sigma_\gamma$ and $\alpha = \mu_\theta - \beta\mu_\gamma$, respectively, and the residual standard deviation or root mean square error, which defines the uncertainty in the treatment effect on the clinical endpoint given a particular treatment effect on the surrogate endpoint, is RMSE = $\sigma_\theta \times (1 - R^2)^{\frac{1}{2}}$.

The 2-stage model was fit using a Bayesian Monte-Carlo Markov Chain sampling, using diffuse prior distributions for the model parameters that we selected so that the final results would depend primarily on the data with little influence of the prior distributions. The priors on the meta-regression slope were uniform with range of -5 to 5. The priors for the mean treatment effects on the clinical endpoint (expressed as a log hazard ratio) and on each GFR slope endpoint (expressed as difference between treatment arms in ml/min/1.73m^2/year) were taken to be normal distributions each with mean 0 and variance 10,000; the priors for the variances of the treatment effects on the clinical endpoint and on the GFR slope endpoints were each taken to be inverse gamma distributions with shape parameter 0.261. The scale parameter was 0.000408 for the clinical endpoint and 0.005 for the slope endpoints. The prior distribution for the clinical endpoint was selected by the investigators to assign 1/3 prior probabilities each to low treatment effect heterogeneity (which we defined as a treatment effect SDs on the log scale ≤ 0.05), medium treatment effect heterogeneity (defined as a treatment effect SD on the log scale between 0.05 and 0.20), and high treatment effect heterogeneity (defined as a treatment effect SD on the log scale > 0.20). For slope, the prior assigns a 1/3 prior probability to slope SDs ≤0.175 ml/min/1.73m^2/year, 1/3 to a slope SD between 0.175 and 0.70 ml/min/1.73m^2/year, and 1/3 to a slope SD > 0.70 ml/min/1.73m^2/year, respectively. We checked that the prior distributions had only a small influence on the results by verifying that the results of each analysis were similar under alternative inverse gamma (0.001, 0.001) prior distributions for the variances for the treatment effects on the clinical endpoint and on GFR slope.

Interpretation The meta-regression supports validity of GFR slope as a surrogate endpoint if 1) the slope of the meta-regression line has a large magnitude with a Bayesian credible interval that does not cross 0; 2) the intercept is close to 0 and with a Bayesian credible interval that crosses 0, implying absence of a substantial average effect on the clinical endpoint when the treatment does not affect GFR slope; 3) the

R^2 of the meta-regression is large, indicating strong associations (e.g., >0.72)[43], so that treatment effects on GFR slope account for most of the variation in treatment effects on the clinical endpoint.

The meta-regression slope provides the mean difference in the log hazard ratio for the clinical endpoint associated with each 1 ml/min/1.73m^2/year increment in the treatment effect on the GFR slope endpoint. The meta-regression intercepts provide the estimated mean log hazard ratio for the clinical endpoint when the mean treatment effect on the GFR slope endpoint is equal to 0. To assist with interpretation, we express the impact under the meta-regression model of a 0.75 ml/min per 1.73 m^2 difference in the treatment effect on GFR slope on the hazard ratio for the treatment effect on the clinical endpoint using the formula:

$$\%\ \text{Difference in HR for clinical endpoint} = 100 \times (1 - \exp(0.75 \times (\text{meta} - \text{regression slope}))).$$

For example, a 0.75 ml/min per 1.73 m^2/year greater beneficial effect of the treatment on the 3-year total GFR slope is associated with an average 23% = 100 × (1- exp (0.75*-0.35)) = 1 – 0.77 = 0.23 lower hazard ratio for the clinical end point.

Subgroup analyses. We performed the following subgroup analyses.

1. GFR of 60 ml/min per 1.73 m^2 is the threshold for the definition for CKD. We wished to evaluate whether the strength of the associations differed for studies with higher versus lower levels of GFR. We repeated the trial level analyses as described above according to studies with mean baseline GFR above and below 60 ml/min per 1.73 m^2
2. Urine albumin greater than 30 mg/g (3.39 mg/mmol is the threshold for diagnosis of CKD, regardless of the level of GFR. It is also a critical value for determining risk of progression, even at high levels of GFR. We wished to evaluate whether the strength of the associations differed

for participants with albuminuria. We repeated the trial level analyses after restricting to the subset of participants who had urine albumin to creatinine ratio greater than 30 mg/g (3.39 mg/mmol) within each study.

3. We evaluated whether the results for the meta-regression varied across disease groups by fitting an expanded meta-regression model using a Bayesian partial-pooling modeling approach to evaluate subgroups by disease. Because there were a limited number of studies with varying sizes within subgroups, there was low statistical power to estimate meta-regression parameters for analyses which are fully restricted to specific subgroups of trials. The partial-pooling model allows subgroup-specific meta-regressions but induces a data-adaptive information sharing across subgroups to improve the precision in estimation.

   This approach adds a third stage to the basic 2-stage mixed effects meta-regression model described for the analysis of the full cohort of 66 studies described above. In the extended partial-pooling model, the first stage again describes the relationship between the estimated treatment effects and the true latent treatment effects on the slope and clinical endpoints within each trial. The second stage includes separate meta-regressions that describe the relationship between the true latent treatment effects on the slope and clinical endpoints across the studies *within* each of the designated subgroups of trials. The third stage describes the variation in the meta-regression terms of the second stage *between* the designated subgroups. When the data suggest a small amount of variation in a particular meta-regression term (such as the meta-regression slope or intercept) between the subgroups, then the subgroup-specific posterior distributions for that term will be more distinct between subgroups due to a lesser degree of information sharing. Alternatively, if the data does not suggest a large degree of heterogeneity in any given term across subgroups, then there will be more information sharing

across subgroups and the posterior distributions for that term will be more similar. In this way, the partial pooling approach allows subgroup specific meta-regressions, while also implementing data-adaptive sharing of information across subgroups to provide improved precision in estimating parameters in subgroups containing limited numbers of studies of varying sizes.

4. *Sensitivity assessment.* We refit the model removing GN and CVD studies, hypothesizing that these diseases would behave differently than diabetes and other causes of CKD.

Model adequacy and outlier assessment. For each study, we produced a posterior predictive distribution (PPD) for the estimated treatment effect on the clinical endpoint based on the estimated treatment effect on GFR slope in that study and the meta-regression model fit to the remaining trials with that study held out. The PPD takes into account uncertainty in the meta-regression parameter estimates as well as the sampling error of the estimated treatment effects within the held-out trial. We graphically displayed the actual observed effect on the clinical endpoint to an interval extending from the 5^{th} to the 95^{th} percentile of the PPD for each trial to display how well the trial level model predicts the treatment effect in a "new trial" based on a model developed from the remaining trials. We defined potential outliers as trials where the observed effects fell in the top 5% or in the bottom 5% of the PPD, recognizing that on average 10% of trials would be flagged as outliers by chance under a perfect model. We additionally re-fit the trial level model to datasets where individual trial flagged as an outlier was left-out of the analysis. We computed posterior medians and 95% credible intervals for the meta-regression intercept, slope, RMSE, and R^2 to identify whether, after an individual trial was removed, the interpretation of results meaningfully changed.

Prediction intervals and trial-level positive predictive value. We obtained 95% prediction intervals for the treatment effect on the clinical endpoint given a particular value for the true latent treatment effect on GFR slope by simulating the posterior distribution of $\alpha + \beta \times \mathrm{True.Eff}_{slope} + \Delta_0$, where $\mathrm{True.Eff}_{slope}$ is the true latent treatment effect on the GFR slope endpoint, $\alpha + \beta \times \mathrm{True.Eff}_{slope}$ represents the predicted mean true latent treatment effect on the clinical endpoint based on the meta-regression model, and Δ_0 is normally distributed with mean 0 and standard deviation given by the RMSE from the meta-regression. Here Δ_0 represents the variation in the true latent treatment effects on the clinical endpoint across different trials with the same treatment effect on GFR slope. This prediction interval accounts for uncertainty in the estimation of α, β, and in the RMSE that define the meta-regression, as well as uncertainty due to variation in the treatment effects on the clinical endpoint about the regression line for different trials.

When the trial level meta-regression is applied to a newly conducted randomized trial, there is an additional source of uncertainty that results from imprecision in the estimation of the treatment effect on GFR slope in the new trial. This added uncertainty depends on the sample size, and is smaller when the sample size for the new trial is large. We obtained 95% prediction intervals for the true latent treatment effect on the clinical endpoint in a new trial that take into account this additional uncertainty by again sampling from the posterior distribution of $\alpha + \beta \times \mathrm{True.Eff}_{slope} + \Delta_0$, but now assume that $\mathrm{True.Eff}_{slope}$ has a random distribution to reflect the uncertainty in its estimation in the new trial instead of taking $\mathrm{True.Eff}_{slope}$ to be a fixed value. Specifically, we assumed that the posterior distribution of $\mathrm{True.Eff}_{slope}$ is normally distributed with mean equal to the estimated treatment effect on GFR slope and standard deviation given by the standard error for the estimated treatment effect on GFR slope based on the sample size. We considered sample size of 1600, 800 and 400 to reflect a large, modest and small sized RCT respectively. In each case we assumed follow-up times of 3 years and

assumed two measurements of serum creatinine at baseline, one measurement at 3 months follow-up, and additional measurements at months 6, 12, and every 6 months thereafter. We performed this calculation for a trial with a between-patient standard deviation in chronic GFR slopes of 4.0 ml/min/1.73m^2/year, with residual GFR variance equal to 0.67 x [mean baseline GFR], where the mean baseline GFR is 40 ml/min/1.73m^2.

We used a similar sampling approach for the posterior distribution of $\alpha + \beta \times \mathrm{True.Eff}_{slope} + \Delta_0$ to estimate the probability that the treatment effect of the clinical endpoint in the new trial, expressed as a log hazard ratio, would fall below 0 (corresponding to a non-zero treatment benefit with a hazard ratio for the clinical endpoint less than 1) given either the true or the estimated treatment effects on GFR slope in the new trial. We refer to the probability of a clinical benefit in the new trial associated with a particular treatment effect on GFR slope as the trial level positive predictive value, which we denoted by PPV_{trial}. By considering the positive predictive value as a function of the estimated treatment effect on GFR slope, we determined the size of the smallest treatment effect on GFR slope that would be required to assure a positive predictive value of at least 0.975 for a benefit on the clinical endpoint.

Statistical analysis

Analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC) and R 4.1.0 (2021-05-18 R Project for Statistical Computing www.r-project.org)[64]. For Bayesian model-fitting within R, we used the RStan package (2.21.3)[41]. For treatment effects displayed in Figure 1 we used restricted maximum likelihoods with study as a random effect with the rma.uni function in R metafor package (3.0-2)[64].

Statistics and Reproducibility

All studies included in the analyses presented in this manuscript were randomized clinical trials. The majority of the clinical trials included in the analyses were double blind placebo-controlled trials. However, trials of blood pressure control, dietary intervention and other lifestyle interventions were unblinded in that the participants were aware of their randomly assigned treatment. The majority of the 66 participating studies included sample size calculations which are described in the protocols of the respective studies. We used a pre-defined protocol to identify studies for inclusion in the analyses of this report using systematic review of the literature with predefined inclusion and exclusion criteria (Protocol and Extended Data Figure 1). Our study design was to use all studies conducted that met the entry criteria and for which we received agreement to use the data. Because our analyses were applied to all available studies that met the entry criteria, we did not perform sample size calculations for the number of studies to be included in the trial level analyses.

Data Availability

All data used in the analysis were obtained by the CKD-EPI CT group through third parties. Data use agreements prohibit CKD-EPI CT from sharing data with parties external to the agreement. See Supplement Table 1 for identity of third party providers. The following datasets could be requested through data sharing platforms: Vivli: CANVAS (NCT01032629), CANVAS-R (NCT01989754), CREDENCE (NCT02065791), EMPA-REG Outcome (NCT01131676), EXAMINE (NCT00968708), Harmony Outcome (NCT02465515), FIDELIO-DKD (NCT02540993); NIDDK: AASK (NCT04364139), FSGS/FONT (NCT00135811), HALT-PKD A and B (NCT00283686), MDRD (NCT03202914); NHLBI BioLINCC: TOPCAT (NCT00094302), SPRINT (NCT01206062); Clinical Study Data Request: PARADIGM-HF (NCT01035255), ACCOMPLISH (NCT00170950); Sponsors' website: LEADER (NCT01179048)

Code Availability

The statistical code we used for the primary analysis can be found at

https://github.com/UofUEpiBio/ckdepict.

The CKD-EPI Clinical Trials Consortium Collaborators

Raymond O. Estacio[21], Rebecca Hanratty[21], Mark Woodward[4,19], John Chalmers[4], Pietro Canetta[3], Brendan Barrett[22], Bruce Neal[4], Vlado Perkovic[4], Kenneth W Mahaffey[23] , David Johnson[4], Meg Jardine[4], Maximilian von Eynatten[24], Eduardo Verde[8], Ursula Verdalles[8], David Arroyo[8], Arlene Chapman[25], Vicente Torres[26], Alan Yu[27], Godela Brosnahan[28], Thierry Hannedouche[29], Kai-Ming Chow[13], Cheuk-Chun Szeto[13], Chi-Bon Leung[13], Di Xie[Error! Bookmark not defined.], Fan Fan Hou[Error! Bookmark not defined.], Jamie Dwyer[12], Marc A. Pohl[30], Itamar Raz[31], Lawrence G. Hunsicker[32], An Vanacker[33], Thomas Malfait[34], Giuseppe Maschio[29], Francesco Locatelli[29], Peter J. Blankestijn[35], Arjan van Zuilen[35], Fumiaki Kobayashi[36], Hirofumi Makino[37], Juliana C.N. Chan[38], Simeone Andrulli[29], Claudio Pozzi[29], Donatella Casartelli[29], Manuel Praga[39], Hernando Trujillo[5], Teresa Cavero[5], Angel

[21] Denver Health, Denver, CO, USA
[22] Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada
[23] Stanford Center for Clinical Research, Department of Medicine, Stanford University School of Medicine, Stanford, CA
[24] Boehringer Ingelheim International, Ingelheim, Germany
[25] Section of Nephrology, University of Chicago School of Medicine, Chicago, Illinois, USA
[26] Division of Nephrology and Hypertension, Mayo Clinic, Rochester, Minnesota, USA
[27] Division of Nephrology and Hypertension, Department of Medicine, University of Kansas Medical Center, Kansas City, Kansas, USA
[28] Division of Diseases and Hypertension, University of Colorado Denver Anschutz Medical Campus, Aurora, Colorado, USA
[29] Department of Nephrology, Dialysis and Renal Transplantation, A. Manzoni Hospital, Lecco, Italy
[30] Cleveland Clinic Foundation, Cleveland, Ohio, USA
[31] Diabetes Unit, Department of Endocrinology and Metabolism, Hadassah Medical Center, Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem, Israel
[32] Department of Internal Medicine, University of Iowa College of Medicine, Iowa City, Iowa, USA
[33] Department of Nephrology, Heilig Hartziekenhuis Roeselare-Menen, Roeselare, Belgium
[34] Renal Division, Department of Internal Medicine, Ghent University Hospital, Ghent, Belgium
[35] Department of Nephrology & Hypertension, University Medical Center Utrecht, Utrecht, The Netherlands
[36] Daiichi Sankyo, Tokyo, Japan
[37] Department of Medicine, Clinical Science, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan
[38] Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Prince of Wales Hospital, 30-32 Ngan Shing Street, Shatin, Hong Kong, SAR China
[39] Nephrology Department, Hospital Universitario 12 de Octubre. Department of Medicine, Complutense University, Madrid, Spain

Sevillano[5], Piero Ruggenenti[15], Annalisa Perna[15], Fabiola Carrara[15], William F Keane[40], Carlo Manno[16], Richard Haynes[41], William G. Herrington[41], Colin Baigent[41], Martin Landray[41], Thomas Rauen[7], Claudia Seikrit[42], Stefanie Wied[43], Robert D. Toto[44], Paul E. de Jong[45], Mauro Saddelli[46]

[40] Merck Research Laboratories, Merck & Co., Inc., Upper Gwynedd, Pennsylvania, USA
[41] Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
[42] Division of Nephrology and Clinical Immunology, RWTH Aachen University Hospital, Aachen, Germany
[43] Department of Medical Statistics, RWTH Aachen University, Aachen, Germany
[44] Department of Internal Medicine, UT Southwestern Medical Center, Dallas, TX, USA
[45] Departments of Nephrology, University of Groningen, Groningen, The Netherlands
[46] Department of Nephrology, St. Donato Hospital, Arezzo, Italy

Methods-only references

8. Inker, L.A.*, et al.* GFR Slope as a Surrogate End Point for Kidney Disease Progression in Clinical Trials: A Meta-Analysis of Treatment Effects of Randomized Controlled Trials. *J Am Soc Nephrol* **30**, 1735-1745 (2019).
34. Inker, L.A.*, et al.* GFR decline as an alternative end point to kidney failure in clinical trials: a meta-analysis of treatment effects from 37 randomized trials. *Am J Kidney Dis* **64**, 848-859 (2014).
35. Higgins, J. & Green, S. Cochrane Handbook for Systematic Reviews of Interventions. (The Cochrane Collaboration, 2011).
36. Heerspink, H.J.L.*, et al.* Change in albuminuria as a surrogate endpoint for progression of kidney disease: a meta-analysis of treatment effects in randomised clinical trials. *The lancet. Diabetes & endocrinology* **7**, 128-139 (2019).
40. Vonesh, E.*, et al.* Mixed-effects models for slope-based endpoints in clinical trials of chronic kidney disease. *Statistics in medicine* **38**, 4218-4239 (2019).
41. r Stan Development Team. "RStan: the R interface to Stan." R package version 2.21.5. (2022).
43. Prasad, V., Kim, C., Burotto, M. & Vandross, A. The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses. *JAMA internal medicine* **175**, 1389-1398 (2015).
50. Joffe, M.M. & Greene, T. Related causal frameworks for surrogate outcomes. *Biometrics* **65**, 530-538 (2009).
51. Daniels, M.J. & Hughes, M.D. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in medicine* **16**, 1965-1982 (1997).
52. Burzykowski, T. & Buyse, M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical statistics* **5**, 173-186 (2006).
53. Burzykowski, T., Molenberghs, G. & Buyse, M. (eds.). *The Evaluation of Surrogate Endpoints*, (Springer, New York, 2005).
54. Lewis, E.J.*, et al.* Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. *The New England journal of medicine* **345**, 851-860 (2001).
55. Wright, J.T., Jr.*, et al.* Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: results from the AASK trial. *JAMA* **288**, 2421-2431 (2002).
56. Klahr, S.*, et al.* The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. Modification of Diet in Renal Disease Study Group. *The New England journal of medicine* **330**, 877-884 (1994).
57. Estacio, R.O., Jeffers, B.W., Gifford, N. & Schrier, R.W. Effect of blood pressure control on diabetic microvascular complications in patients with hypertension and type 2 diabetes. *Diabetes care* **23 Suppl 2**, B54-64 (2000).
58. Torres, V.E.*, et al.* Angiotensin blockade in late autosomal dominant polycystic kidney disease. *New England Journal of Medicine* **371**, 2267-2276 (2014).
59. Levey, A.S.*, et al.* A new equation to estimate glomerular filtration rate. *Annals of internal medicine* **150**, 604-612 (2009).
60. Skali, H.*, et al.* Prognostic assessment of estimated glomerular filtration rate by the new Chronic Kidney Disease Epidemiology Collaboration equation in comparison with the Modification of Diet in Renal Disease Study equation. *American heart journal* **162**, 548-554 (2011).
61. Vonesh, E.F. & Greene, T. Biased estimation with shared parameter models in the presence of competing dropout mechanisms. *Biometrics* **78**, 399-406 (2022).

62. Vonesh, E.F., Greene, T. & Schluchter, M.D. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in medicine* **25**, 143-163 (2006).
63. Rizopoulos, D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*, (Chapman and Hall/CRC, New York, 2012).
64. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* **36**, 1 - 48 (2010).