University of Groningen

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

Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI)*; Collier, Willem; Inker, Lesley A.; Haaland, Benjamin; Appel, Gerald B.; Badve, Sunil V.; Caravaca-Fontán, Fernando; Chalmers, John; Floege, Jürgen; Goicoechea, Marian

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

Link to publication in University of Groningen/UMCG research database

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

Willem Collier,[1,2] Lesley A. Inker,[3] Benjamin Haaland,[1] Gerald B. Appel,[4] Sunil V. Badve,[5] Fernando Caravaca-Fontán ⓘ ,[6] John Chalmers ⓘ ,[5] Jürgen Floege,[7] Marian Goicoechea,[8] Enyu Imai,[9] Tazeen H. Jafar,[10] Julia B. Lewis,[11] Philip K.T. Li ⓘ ,[12] Francesco Locatelli,[13] Bart D. Maes,[14] Brendon L. Neuen ⓘ ,[5] Ronald D. Perrone ⓘ ,[3] Giuseppe Remuzzi,[15] Francesco P. Schena,[16] Christoph Wanner,[17] Hiddo J.L. Heerspink ⓘ ,[18] and Tom Greene,[1] and the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI)*

Abstract

Background The GFR slope has been evaluated as a surrogate end point for kidney failure in meta-analyses on a broad collection of randomized controlled trials (RCTs) in CKD. These analyses evaluate how accurately a treatment effect on GFR slope predicts a treatment effect on kidney failure. We sought to determine whether severity of CKD in the patient population modifies the performance of GFR slope.

Methods We performed Bayesian meta-regression analyses on 66 CKD RCTs to evaluate associations between effects on GFR slope (the chronic slope and the total slope over 3 years, expressed as mean differences in ml/min per 1.73 m^2/yr) and those of the clinical end point (doubling of serum creatinine, GFR <15 ml/min per 1.73 m^2, or kidney failure, expressed as a log-hazard ratio), where models allow interaction with variables defining disease severity. We evaluated three measures (baseline GFR in 10 ml/min per 1.73 m^2, baseline urine albumin-to-creatinine ratio [UACR] per doubling in mg/g, and CKD progression rate defined as the control arm chronic slope, in ml/min per 1.73 m^2/yr) and defined strong evidence for modification when 95% posterior credible intervals for interaction terms excluded zero.

Results There was no evidence for modification by disease severity when evaluating 3-year total slope (95% credible intervals for the interaction slope: baseline GFR [−0.05 to 0.03]; baseline UACR [−0.02 to 0.04]; CKD progression rate [−0.07 to 0.02]). There was strong evidence for modification in evaluations of chronic slope (95% credible intervals: baseline GFR [0.02 to 0.11]; baseline UACR [−0.11 to −0.02]; CKD progression rate [0.01 to 0.15]).

Conclusions These analyses indicate consistency of the performance of total slope over 3 years, which provides further evidence for its validity as a surrogate end point in RCTs representing varied CKD populations.
CJASN 18: 183–192, 2023. doi: https://doi.org/10.2215/CJN.0000000000000050

Due to the number of contributing authors, the affiliations are listed at the end of this article.

Correspondence:
Dr. Lesley A. Inker, Division of Nephrology, Tufts Medical Center, 800 Washington St., Box #391, Boston, MA 02111. Email: LInker@ tuftsmedicalcenter.org

Introduction

A critical challenge in the design of randomized controlled trials (RCTs) in CKD is that the established clinical end point, typically a composite of time-to-kidney failure with replacement therapy or doubling of serum creatinine, involves late-stage events and often takes many years to observe for patients with early-stage or slowly progressing disease.[1,2] High-quality surrogate end points are thus crucial to expanding feasibility of RCTs in CKD. In some circumstances, the slope of the GFR may reveal clinically meaningful effects of therapy with shorter follow-up or smaller sample size.[1,2] Meta-analyses of previously conducted CKD RCTs have demonstrated strong potential for GFR slope-based end points to be used to predict treatment effects on clinical end points.[1,2,4]

The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) collection of RCTs used for evaluations of GFR slope includes a spectrum of patient population CKD severities.[1] This heterogeneity raises the question of whether there is a consistent relationship between treatment effects on GFR slope and treatment effects on the clinical end point across trials that differ by CKD severity. Previously, heterogeneity in the behavior of GFR slope as a surrogate was examined informally in a smaller set of studies by performing separate analyses in subgroups of trials defined by GFR or albuminuria thresholds.[1] These analyses suffered from reduced statistical power due to small numbers of trials within subgroups and failed to account for the remaining heterogeneity across trials within subgroups.[5] Analyzing severity measures as continuous in model fitting may increase statistical power and provide a more granular assessment of surrogate end point performance.

For this paper, we performed analyses using an expanded collection of 66 trials[1] to rigorously

investigate whether trial-level associations of treatment effects on the clinical and slope-based end points are modified by three measures of disease severity: baseline GFR, baseline albuminuria, and rate of CKD progression, approximated as the mean chronic slope in the control arm. We consider and contrast analyses using two different GFR slope-based end points: the total slope over a 3-year follow-up period and the chronic slope starting at 3-months follow-up. These analyses are intended to facilitate more targeted interpretation of treatment effects on slope-based end points, which will be beneficial for design and interpretation of future trials.

Methods

Part 1 of the Supplemental Material provides further detail for the analytical methods.

Data Description

A pooled database of RCTs comprising individual patient data was used. We identified trials through a systematic review of the literature. Search terms, inclusion and exclusion criteria, and bias assessment have been previously published.[1,19] After applying inclusion and exclusion criteria, there were 66 randomized treatment comparisons as the main unit of analysis. For analyses using baseline albuminuria, four studies were removed due to lack of albuminuria measurements.

End Points and Estimation of Treatment Effects

As previously mentioned, the clinical end point was defined as the first occurrence of any of the following: initiating dialysis or receiving a kidney transplant, sustained GFR below 15 ml/min per 1.73 m[2], or doubling of serum creatinine.[1,2] Treatment effects on the clinical end point were expressed as log-hazard ratios (HRs) estimated using proportional hazards regression.

We estimated GFR using the 2009 CKD-EPI creatinine equation (eGFR, expressed in ml/min per 1.73 m[2]) and estimated treatment effects on GFR slope as well as standard errors using a shared parameter mixed effects model.[1,6,7] Interventions in CKD can cause a short-term change in GFR that may differ from the long-term effect.[1,2] To allow for a short-term acute treatment effect, the mixed effects model allows for a nonlinear acute GFR response up to 3 months after baseline and allows for a potentially different linear mean GFR slope thereafter.[7] Under this model, the differences between the randomized groups in mean change from baseline to 3 months, mean slopes from 3 months onward, and mean changes from baseline to a given follow-up time of interest are defined as treatment effects on the acute, chronic, and total GFR slopes, respectively (expressed in ml/min per 1.73 m[2]/yr). This model allows estimation of the mean total slope over a given follow-up of interest, which we define as a time-weighted average of the mean acute and chronic slopes. Our prior analyses have demonstrated that treatment effects calculated under this model for total GFR slope over 3 years (herein total slope) and chronic GFR slope (herein chronic slope) can be used to accurately predict treatment effects on the clinical end point.[1] Because there are different strengths and weaknesses of the total and chronic slopes, we evaluate both in separate analyses in this paper.[1–4]

Measures of Disease Severity

We included three measures of severity of kidney disease that are often targeted in a trial design: baseline GFR, baseline albuminuria, and mean chronic slope in the control arm. The measures of albuminuria varied across studies, with most measuring protein excretion rate. Because guidelines recommend use of urine albumin-to-creatinine ratio (UACR), we express UACR using a commonly used conversion of UACR as 60% of protein excretion rate.[8] For model fitting, we log-transformed UACR to account for positive skewness.

Trial-Level Analyses

Our analyses extend previous work using meta-regression models that relate treatment effects on the clinical end point to treatment effects on the GFR slope-based surrogate end point across RCTs.[1] The trial-level analysis consists of two steps.[1,5,9–12] In the first step, we used intent-to-treat analyses to estimate the treatment effects on each end point within each trial. We also estimated standard errors and correlations of sampling errors of the estimated effects on the different end points using robust sandwich estimates. In the second step, we fit Bayesian meta-regression models to the treatment effect estimates while accounting for their standard errors and correlations to assess the strength of the association between treatment effects on the separate end points across trials.[5,9] Four meta-regression parameters are used to interpret surrogate quality: the intercept, slope, meta-regression residual SD, which we refer to as root-mean-squared error (RMSE), and the coefficient of determination, denoted R^2.[1,5,9,13,24] The intercept should be zero, implying a null effect on slope corresponds to a null average effect on the clinical end point. The meta-regression slope should be nonzero, indicating that trials with beneficial (harmful) effects on slope correspond to beneficial (harmful) effects on the clinical end point. Finally, the RMSE should be small relative to the overall SD between studies in the true treatment effects on the clinical end point (corresponding to R^2 close to 1), indicating that GFR slope can be used to accurately predict the clinical effect.

In this study, we extend the second-stage model to allow meta-regression parameters to vary with severity, which is accomplished by adding interaction terms. Our primary question was whether the meta-regression intercept and/or slope may change with disease severity, reflecting variation in how an effect size on GFR slope translates to an effect size on the clinical end point. Thus, our primary model includes interaction terms for the meta-regression intercept and slope. A nonzero interaction for the intercept indicates that the implications of a null effect on GFR slope differ depending on the level of disease severity. A nonzero interaction for the meta-regression slope indicates that the implications of beneficial (harmful) treatment effects on GFR slope differ depending on disease severity. For our analyses, we centered disease severity data using the mean across trials so that the meta-regression intercept could be interpreted for a trial with the average value defining severity. Strong evidence for effect modification was defined as a 95% credible interval for the interaction terms that excluded zero.

We estimated treatment effects, standard errors, and within-study correlations using SAS version 9.4 and fit the trial-level models using R version 4.0.3, and the RStan package (version 2.21.12) for sampling posterior distributions using Markov Chain Monte Carlo algorithms. Example code and assessments of MCMC performance are detailed in Part 1 of the Supplemental Material.

Prediction Summaries

We summarize posterior predictive distributions for predicted effects on the clinical end point for a hypothetical future trial for a range of treatment effects on the slope-based end points (from a null to a large treatment effect) and values defining CKD severity (for the 25th and 75th percentiles of each measure from the data). These distributions properly account for sampling error in the new trial and uncertainty in the estimated model parameters. For illustration, we use a fixed large trial design (corresponding to roughly 1600 patients), similar to previous work.[1,19]

Sensitivity Analyses

We performed sensitivity analyses to assess the robustness of our results, which are summarized in Part 2 of the Supplemental Material. We evaluated results under two models, which relaxed key assumptions made in modeling for our primary analyses: one in which each trial's disease severity measure was regarded as fixed as opposed to being drawn from a common distribution across studies and the second in which the meta-regression RMSE was allowed to vary with disease severity. We also refit models after removing disease subgroups that stood out because of distinctiveness in their population CKD severity (cardiovascular disease trials because of high baseline GFR and glomerular disease trials because of high baseline UACR), and intervention subgroups with unique mechanisms of action or known extrarenal effects on serum creatinine (immunosuppressant and dietary modification trials). We also refit models allowing interaction with log-transformed mean baseline GFR (instead of with mean baseline GFR itself). Finally, we also evaluated the total slope over 2 years rather than 3 years as the surrogate end point.

Results

Among the 66 trials, there were 186,949 patients and 19,533 (10%) experienced the clinical event (Table 1). Across trials, the mean baseline GFR was 55 (SD=23) ml/min per 1.73 m^2, the median baseline UACR was 389 (interquartile range=894) mg/g, and the median control arm GFR slope was −3.0 (interquartile range=2.5) ml/min per 1.73 m^2/yr.

**Table 1. Summary of trials used for analyses**

Characteristics	Trials Overall	Trials with Low Mean Baseline GFR (<40 ml/min per 1.73 m^2)	Trials with Moderate Mean Baseline GFR (40–75 ml/min per 1.73 m^2)	Trials with High Mean Baseline GFR (>75 ml/min per 1.73 m^2)
Patients, N (median)	186,949 (564)	14,073 (289)	51,864 (985)	121,012 (5519)
Clinical events, total (% of patients)	19,533 (10)	4601 (33)	7379 (14)	7553 (6)
Baseline UACR, median (interquartile range)	389 (894)	575 (655)	643 (1041.9)	29 (772)
Control arm chronic slope, median (interquartile range)	−3.0 (2.5)	−3.7 (1.8)	−3.6 (2.1)	−1.7 (1.9)
Diseases, total (% of trials)				
Cardiovascular	7 (11)	0 (0)	2 (9)	5 (23)
CKD not otherwise specified	28 (42)	19 (86)	7 (32)	2 (9)
Diabetes	21 (32)	2 (9)	8 (36)	11 (50)
Glomerular	10 (15)	1 (5)	5 (23)	4 (18)
Interventions, total (% of trials)				
Albuminuria targeted	1 (2)	1 (5)		
Allopurinol	2 (3)	2 (9)		
Antiplatelet	3 (5)	1 (5)		2 (9)
DPP-4 inhibitor	3 (5)		1 (5)	2 (9)
Endothelin receptor antagonist	1 (2)		1 (5)	
GLP-1 agonist	2 (3)			2 (9)
Immunosuppression	9 (14)	1 (5)	5 (23)	3 (14)
Intensive glucose	1 (2)			1 (5)
Low versus usual BP	7 (11)	3 (14)	1 (5)	3 (14)
Low versus usual diet	2 (3)	2 (9)		
Mineralocorticoid receptor antagonist	2 (3)		2 (9)	
Nurse-coordinated care	2 (3)	1 (5)	1 (5)	
RASB versus CCB	4 (6)	1 (5)	2 (9)	1 (5)
RASB versus control	21 (32)	9 (41)	7 (32)	5 (23)
RASB+CCB	1 (2)			1 (5)
SGLT-2 inhibitor	4 (6)		2 (9)	2 (9)
Statin+ezetimibe	1 (2)	1 (5)		

Baseline UACR is represented in mg/g, and the mean control arm chronic slope in ml/min per 1.73 m^2/yr. UACR, urine albumin-to-creatinine ratio; DPP-4 inhibitor, dipeptidyl peptidase-4; GLP-1 agonist, glucagon-like peptide-1 receptor; RASB, renin-angiotensin system blockers; CCB, calcium channel blockers; SGLT-2, sodium-glucose cotransporter-2 inhibitor.

Trials with lower GFR typically had higher UACR and faster progression and had higher proportions of patients experiencing the clinical event (Supplemental Results Figures 1 and 2). Blood pressure modification, RAS blockade, and immunosuppressants were evaluated more frequently than other intervention classes and were evaluated across a broad range of baseline GFR or ACR, and rate of progression (Table 1 and Supplemental Results Tables 3, 4). There was less variation across disease severity measures within certain intervention classes. For example, GLP-1 agonists were only evaluated in trials with high baseline GFR and moderate progression rates.

For total slope, there was no evidence that either the meta-regression intercept or slope varied with baseline GFR, UACR, or control arm chronic slope (Figure 1 and Table 2). The 95% credible intervals for the interaction coefficients were narrow and centered near zero in analyses for each disease severity measure. By contrast, for chronic slope, there was strong evidence for variation in both the meta-regression intercept and slope depending on severity (Figure 2 and Table 2). Specifically, these results indicated that the meta-regression slope is larger in magnitude for trials with higher baseline GFR, lower baseline UACR, or slower control arm progression compared with trials with the opposite. We found that the major implications of our results were not substantially changed in any of the sensitivity analyses considered (Supplemental Results Tables 5–12). The results using the 2-year total slope were similar to those
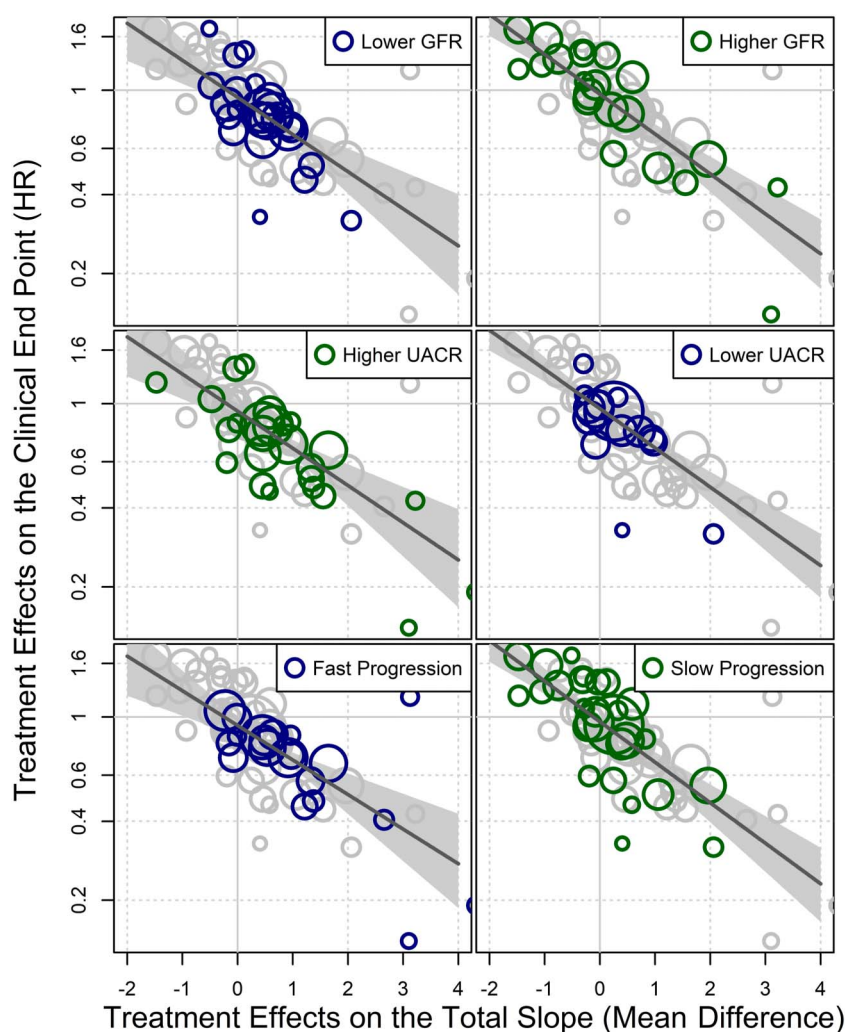
Figure 1. Trial-level associations using total slope. Displayed are trial-level associations between treatment effects on the clinical end point and treatment effects on the total slope. In the left panels, displayed are the meta-regression line and credible bands when mean baseline GFR (top row), mean baseline ACR (middle row), and the mean control arm slope (bottom row) each are fixed at the 25th, 75th, and 25th percentiles for values of those variables observed in the data, respectively (baseline GFR: 40 ml/min per 1.73 m^2; baseline UACR: 1000 mg/g; control arm slope: −4 ml/min per 1.73 m^2/yr). In the right panels, displayed are the meta-regression line and credible bands when baseline GFR, baseline ACR, and the control arm slope are fixed at the 75th, 25th, and 75th percentiles for values observed in the data, respectively (baseline GFR: 75 ml/min per 1.73 m^2; baseline UACR: 71 mg/g; control arm slope: −1.5 ml/min per 1.73 m^2/yr). The colored circles indicate trials that fall within a unit of the value used to define disease severity used to plot the relevant regression line.

Table 2.   Meta-regression posterior summaries for continuous interactions models

Parameter	Baseline GFR (per 10 ml/min per 1.73 m² Higher)	Baseline UACR (per Two-Fold Higher in mg/g)	Control Arm Slope (per 1 ml/min per 1.73 m²/yr Higher)
Total slope			
Meta-regression intercept at mean level for each severity measure	−0.05 (−0.11 to 0.01)	−0.05 (−0.12 to 0.01)	−0.06 (−0.13 to −0.003)
Meta-regression slope at mean level for each severity measure	−0.34 (−0.42 to −0.24)	−0.34 (−0.42 to −0.25)	−0.32 (−0.41 to −0.23)
Change in meta-regression intercept for a designated change in severity (interaction intercept)	0.01 (−0.02 to 0.04)	−0.01 (−0.03 to 0.05)	0.01 (−0.02 to 0.05)
Change in meta-regression slope for a designated change in severity (interaction slope)	−0.01 (−0.05 to 0.03)	0.01 (−0.03 to 0.05)	−0.02 (−0.07 to 0.02)
RMSE	0.06 (0.02 to 0.13)	0.05 (0.01 to 0.12)	0.06 (0.02 to 0.13)
R^2	0.96 (0.81 to 1.00)	0.93 (0.77 to 0.99)	0.94 (0.78 to 0.99)
Chronic slope			
Meta-regression intercept at mean level for each severity measure	−0.03 (−0.13 to 0.06)	−0.09 (−0.21 to 0.02)	−0.08 (−0.20 to 0.05)
Meta-regression slope at mean level for each severity measure	−0.28 (−0.41 to −0.16)	−0.27 (−0.41 to −0.13)	−0.30 (−0.44 to −0.15)
Change in meta-regression intercept for a designated change in severity (interaction intercept)	0.06 (0.02 to 0.11)[a]	−0.09 (−0.16 to −0.03)[a]	0.07 (0.01 to 0.15)[a]
Change in meta-regression slope for a designated change in severity (interaction slope)	−0.08 (−0.16 to −0.003)[a]	0.13 (0.05 to 0.21)[a]	−0.10 (−0.18 to −0.03)[a]
RMSE	0.16 (0.10 to 0.24)	0.14 (0.06 to 0.22)	0.16 (0.09 to 0.24)
R^2	0.70 (0.41 to 0.89)	0.67 (0.40 to 0.85)	0.68 (0.41 to 0.88)

Displayed are the posterior medians and 95% credible intervals. Note that for analyses using UACR as the severity measure, the interaction coefficients were converted such that they should be interpreted as multiplicative changes in HRs as a function of a two-fold higher of UACR in mg/g for a given effect on the total or chronic slope. Whereas for columns corresponding to baseline GFR and the control arm slope, the interaction coefficients should be interpreted as additive changes in log-HRs as a function of designated changes for those corresponding variables. For interpretation of the RMSE, the total SD (as opposed to residual SD) for the true treatment effects on the clinical end point is estimated to have a posterior median and 95% credible interval of 0.26 (0.19 to 0.36). UACR, urine albumin-to-creatinine ratio; RMSE, root-mean-squared error; HR, hazard ratio.
[a]95% credible intervals for interaction terms that exclude zero.

obtained using the 3-year total slope, with no strong evidence of interactions (Supplemental Results Table 5).

Table 3 illustrates the implications for predicted effects on the clinical end point. The predicted HRs for the clinical end point were nearly identical between lower and higher levels of each of the three disease severity measures for prediction using total slope. For example, a treatment effect of 1.5 ml/min per 1.73 m²/yr on the total slope translated to a nearly identical range in predicted HRs (between 0.59 and 0.60) when the mean baseline GFR was 75 to 40 ml/min per 1.73 m². However, there were large changes in the predicted HRs across the same range in severity values when using chronic slope. For example, for a trial with baseline GFR equal to 40 ml/min per 1.73 m², the predicted HR changed from 0.87 if the treatment had no effect on the chronic slope to 0.70 if the treatment had a 1.5 ml/min per 1.73 m²/yr benefit on the chronic slope (a 24% difference). By contrast, for a trial with baseline GFR equal to 75 ml/min per 1.73 m², the predicted HR changed from 1.06 if the treatment had no effect on the chronic slope to 0.61 if the treatment had a 1.5 ml/min per 1.73 m²/yr benefit on the chronic slope (a 74% difference).

Discussion

This article extends our prior work evaluating GFR slope.[1] We assessed whether the relationship between treatment effects on slope and treatment effects on an established clinical end point differs depending on the severity of CKD in the study population. We considered three metrics for severity: baseline GFR, UACR, and control arm GFR slope. We found the association of effects on 3-year total slope with effects on the clinical end point consistent across measures of patient population disease severity. By contrast, we found variation by disease severity in the association when evaluating chronic slope. Our analyses showed that treatment effects on chronic slope translated to progressively larger treatment effects on the clinical end point for patient populations with less as opposed to more severe disease. These results can aid in interpreting treatment effects on slope-based end points in ongoing and future trials.

As we describe in a separate study using the same 66 trials, treatment effects on the 3-year total slope predict treatment effects on the clinical end point with greater accuracy than using chronic slope.[19] The results of this study suggest that this greater accuracy may be explained, in part, by a greater stability in the trial-level association of the total versus chronic slope across measures of CKD severity. The explanation for why severity influences the trial-level association for chronic slope, but not for the total slope, is not completely clear. One possible explanation arises from the relationship between acute effects and treatment effects on chronic slope. Glomerular hyperfiltration is a common pathway underlying CKD progression.
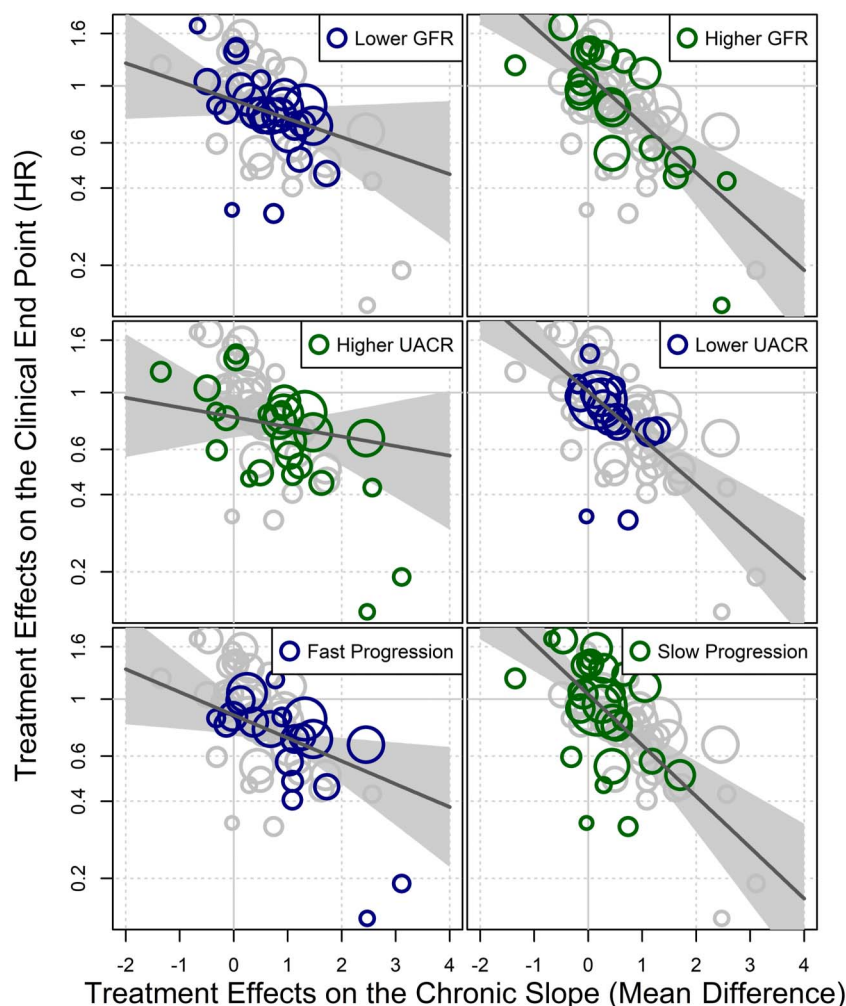
Figure 2. Trial-level associations using chronic slope. Displayed are trial-level associations between treatment effects on the clinical end point and treatment effects on the chronic slope. In the left panels, displayed are the meta-regression line and credible bands when mean baseline GFR (top row), mean baseline ACR (middle row), and the mean control arm slope (bottom row)—each fixed at the 25th, 75th, and 25th percentiles for values of those variables observed in the data, respectively (baseline GFR: 40 ml/min per 1.73 m^2; baseline UACR: 1000 mg/g; control arm slope: −4 ml/min per 1.73 m^2/yr). In the right panels, displayed are the meta-regression line and credible bands when baseline GFR, baseline ACR, and the control arm slope are fixed at the 75th, 25th, and 75th percentiles for values observed in the data, respectively (baseline GFR: 75 ml/min per 1.73 m^2; baseline UACR: 71 mg/g; control arm slope: −1.5 ml/min per 1.73 m^2/yr). The colored circles indicate trials that fall within a unit of the value used to define disease severity used to plot the relevant regression line.

Reducing glomerular hyperfiltration with dietary or pharmacological interventions can result in a negative acute effect on GFR in the short term, but in the long term, such interventions can stabilize GFR, as reflected by the chronic slope.[2,18,20–23] This mechanism may lead to a negative correlation between the effect on the acute GFR slope and the effect on the chronic slope, in which case the clinical benefit of slowing decline in the chronic slope may be partially offset by the clinical implications of greater acute declines in GFR. Some have suggested that the inverse association between acute and chronic effects may be stronger for patients with more advanced disease.[21–23] As such, it is possible that treatment benefits on the chronic slope would be counteracted by the negative acute effects to a greater degree in studies of patients with more advanced disease compared with less advanced disease. Such a pattern could potentially explain the observed

attenuation of the hazard reduction associated with a given effect on the chronic slope with more severe disease. Alternatively, because the treatment effect on total slope incorporates the acute effect, effects on the total slope would remain weaker than effects on the chronic slope for studies with larger negative acute effects, mitigating the tendency for apparent strongly beneficial slope-based effects in studies with severe CKD.

Our finding of an association between CKD severity and how treatment effects on the chronic slope, which are expressed on the absolute scale, relate to treatment effects on the clinical end point, which are expressed on the relative scale, can be seen as analogous to findings reported in other medical domains. For example, across a collection of trials for heart failure, stroke, and other diseases, Kent *et al.* found heterogeneity in treatment effect sizes across quartiles defined by baseline disease risk for effects

Table 3. Predicted treatment effects on the clinical end point as a function of effects on total or chronic slope and disease severity

Disease Severity Levels	Observed Treatment Effect on GFR Slope: 0 Median HR (95% Prediction Interval)	Observed Treatment Effect on GFR Slope: 0.75 Median HR (95% Prediction Interval)	Observed Treatment Effect on GFR Slope: 1.5 Median HR (95% Prediction Interval)	% Difference in HR Between Treatment Effects of 0 and 1.5 ml/min/1.73 m^2/yr
Total slope				
Low baseline GFR	0.93 (0.76 to 1.14)	0.74 (0.60 to 0.90)	0.59 (0.47 to 0.74)	58
High baseline GFR	0.96 (0.78 to 1.19)	0.75 (0.60 to 0.93)	0.59 (0.47 to 0.74)	63
High baseline UACR	0.92 (0.75 to 1.13)	0.75 (0.61 to 0.90)	0.59 (0.47 to 0.72)	56
Low baseline UACR	0.95 (0.77 to 1.16)	0.74 (0.60 to 0.89)	0.59 (0.47 to 0.72)	61
Fast GFR progression	0.91 (0.74 to 1.13)	0.74 (0.60 to 0.90)	0.60 (0.48 to 0.74)	52
Slow GFR progression	0.96 (0.77 to 1.18)	0.75 (0.60 to 0.92)	0.59 (0.46 to 0.73)	63
Chronic slope				
Low baseline GFR	0.87 (0.60 to 1.26)	0.78 (0.54 to 1.11)	0.70 (0.47 to 1.02)	24
High baseline GFR	1.06 (0.71 to 1.60)	0.81 (0.54 to 1.20)	0.61 (0.39 to 0.93)	74
High baseline UACR	0.81 (0.58 to 1.14)	0.75 (0.53 to 1.07)	0.70 (0.49 to 0.94)	16
Low baseline UACR	0.98 (0.69 to 1.41)	0.75 (0.55 to 1.01)	0.58 (0.39 to 0.83)	69
Fast GFR progression	0.84 (0.57 to 1.23)	0.74 (0.52 to 1.05)	0.66 (0.45 to 0.93)	27
Slow GFR progression	1.02 (0.69 to 1.53)	0.78 (0.51 to 1.14)	0.58 (0.37 to 0.88)	76

Observed treatment effects on slope-based end points are expressed as mean differences, expressed in ml/min per 1.73 m^2/yr. Low baseline GFR: 40 ml/min per 1.73 m^2, high baseline GFR: 75 ml/min per 1.73 m^2; low baseline UACR: UACR of 71 mg/g; high baseline UACR: 1000 mg/g; fast GFR progression: −4.0 ml/min per 1.73 m^2/yr; slow GFR progression: −1.5 ml/min per 1.73 m^2/yr. The % differences in the final columns compare the largest predicted median HR in a given row to that row's smallest predicted median HR. A larger value indicates a larger degree of heterogeneity in predicted clinical effects across possible treatment effects on the surrogate. This table highlights that the differences in predicted HRs for treatment effects on the clinical end point can vary considerably more across the range of treatment effects on slope when chronic slope is used than when total slope is used, where variability is dependent on values in disease severity metrics. HR, hazard ratio; UACR, urine albumin-to-creatinine ratio.

represented as absolute risk differences, but less so for effects represented as relative risk differences. These authors noted that the relationship between treatment effects expressed as absolute and relative differences can often be predicted by disease severity as measured by the absolute risk of the outcome.[14,15] There are important reasons why we use absolute (mean differences) as opposed to relative differences (a ratio) for treatment effects on GFR slope. In general, a ratio of means can be an unstable quantity to estimate when the numerator and denominator can be positive or negative, and especially when the denominator (estimated mean control arm slope) can be close to zero.

A major implication of this study is that the relationship between treatment effects on total slope and those of the clinical end point appears consistent across levels of disease severity. These results suggest that the overall meta-regression model for 3-year total slope can be used to predict treatment effects on the clinical end point across the range of patient population disease severity. On the other hand, our results indicate that disease severity should be considered when translating the magnitude of a treatment effect on chronic slope to the magnitude of the projected treatment effect on the clinical end point. Our results did show that the predicted HRs are relatively consistent across levels of disease severity for effects on chronic slope near 0.75 ml/min per 1.73 m^2/yr. However, the validity of a surrogate end point relies on consistent interpretation across the full spectrum of effect sizes, from that of the null hypothesis to a large treatment benefit. Finally, although we found no evidence that the prediction accuracy on the basis of the total slope (expressed as the RMSE)

varies with disease severity, there was a trend to suggest the chronic slope might be used to more accurately predict effects on the clinical end point in trials with more compared with less severe disease.

There are several strengths to the analyses reported. We obtained individual participant data from 66 well-powered RCTs. The abundance of high-quality data enabled us to use more intricate models than typically possible in surrogate evaluations. Use of individual-level data allowed us to appropriately account for multiple sources of variation in the treatment effect estimates and enabled us to evaluate the influence of three measures of disease severity.[9,16,17] Having multiple measures helped us corroborate results from separate analyses. By using Bayesian modeling, we were able to report probability for treatment effects in hypothetical future trials, which improved the interpretability and intuition behind our results

There are also several limitations. First, disease severity is multifaceted and associated with distinct intervention options. We did not attempt to identify the causal mechanism behind the interactions observed. We also did not attempt to evaluate if our results persisted within all subgroups of trials defined by specific disease conditions or specific interventions, as statistical power for evaluating interactions within many subgroups is very low. We did observe that the results of our analyses were not substantially changed after removing subsets of trials whose populations or treatments exhibited unique characteristics relative to the broader collection of trials considered. It is possible that the clinical implications of the disease severity indices may differ between the earlier and later trials

included in our analyses because of improvements in standards of care. We used a guideline-recommended method to convert urine protein to UACR. However, this may vary by disease and other population characteristics. These analyses were also limited by use of clinical events acquired only over the duration of the trial as the reference standard, but this is a common challenge to surrogate end point evaluation using previously conducted trials. Finally, our calculation for the total slope over 3 years requires an extrapolation of the data for trials with <3 years of follow-up. However, treatment effects on 3-year GFR slope defined this way have been found to predict treatment effects on the clinical end point with high accuracy, and this work further corroborates this result.

Overall, the analyses reported help improve understanding of how to interpret treatment effects on GFR slope-based end points in RCTs. Our results indicate the general interpretability and applicability of the trial-level association for the total slope, which is consistent across levels of disease severity in the patient population. However, our results indicate that treatment effects on chronic slope should be interpreted in the context of population disease severity. Future work should continue to refine optimal trial design strategies for trials on the basis of CKD severity.

Funding

Acknowledgments

Author Contributions

Data Sharing Statement

We have been provided data under strict data use agreements from the investigators or sponsors who conducted trials used in our meta-analysis. We are not allowed to share these data publicly. Some datasets were, however, acquired through public sharing platforms like Vivli, CSDR, and NHLBI. These can be requested by interested parties by visiting the respective websites.

Supplemental Material

This article contains the following supplemental material online at http://links.lww.com/CJN/B566.

Supplemental Material Part 1. Additional Detail on Statistical Methods.

Supplemental Material Part 2. Trial Summaries, Results of Secondary and Sensitivity Analyses, and Collaborator Acknowledgments and Funding Sources.

References

1. Inker LA, Heerspink HJL, Tighiouart H, 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.* 2019;30(9):1735-1745. doi:10.1681/ASN.2019010007
2. Levey AS, Inker LA, Matsushita K, 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.* 2014;64(6):821-835. doi:10.1053/j.ajkd.2014.07.030
3. Grams ME, Sang Y, Ballew SH, 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.* 2019;30(9):1746-1755. doi:10.1681/ASN.2019010008
4. Greene T, Ying J, Vonesh EF, et al. Performance of GFR slope as a surrogate endpoint for kidney disease progression in clinical trials: a statistical simulation. *J Am Soc Nephrol.* 2019;30(9):1756-1769. doi:10.1681/ASN.2019010009
5. Papanikos T, Thompson JR, Abrams KR, et al. Bayesian hierarchical meta-analytic methods for modeling surrogate relationships that vary across treatment classes using aggregate data. *Stat Med.* 2020;39(8):1103-1124. doi:10.1002/sim.8465
6. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate [published correction appears in *Ann Intern Med.* 2011;155(6):408]. *Ann Intern Med.* 2009;150(9):604-612. doi:10.7326/0003-4819-150-9-200905050-00006
7. Vonesh E, Tighiouart H, Ying J, et al. Mixed-effects models for slope-based endpoints in clinical trials of chronic kidney disease [published correction appears in *Stat Med.* 2021;40(14):3400-3401]. *Stat Med.* 2019;38(22):4218-4239. doi:10.1002/sim.8282
8. Kidney Disease: Improving Global Outcomes (KDIGO). KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl.* 2013;3(1):1-150.
9. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Stat Med.* 1997;16(17):1965-1982. doi:10.1002/(sici)1097-0258(19970915)16:17<1965::aid-sim630>3.0.co;2-m

10. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996;125(7):605-613. doi:10.7326/0003-4819-125-7-199610010-00011

11. Bujkiewicz S, Thompson JR, Spata E, Abrams KR. Uncertainty in the Bayesian meta-analysis of normally distributed surrogate endpoints. *Stat Methods Med Res.* 2017;26(5):2287-2318. doi:10.1177/0962280215597260

12. Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. *Biometrics.* 2009;65(2):530-538. doi:10.1111/j.1541-0420.2008.01106.x

13. 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 Intern Med.* 2015;175(8):1389-1398. doi:10.1001/jamainternmed.2015.2829

14. van Klaveren D, Varadhan R, Kent DM, et al. The predictive approaches to treatment effect heterogeneity (PATH) statement. *Ann Intern Med.* 2020;172(11):776. doi:10.7326/l20-0427

15. Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG, Hayward RA. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *Int J Epidemiol.* 2016;45(6):2075-2088. doi:10.1093/ije/dyw118

16. Ishak KJ, Platt RW, Joseph L, Hanley JA. Impact of approximating or ignoring within-study covariances in multivariate meta-analyses. *Stat Med.* 2008;27(5):670-686. doi:10.1002/sim.2913

17. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BMJ* 2010;340:c2289. doi:10.1136/bmj.c2289

18. Neuen BL, Tighiouart H, Heerspink HJL, et al. Acute treatment effects on GFR in randomized clinical trials of kidney disease progression. *J Am Soc Nephrol.* 2022;33(2):291-303. doi:10.1681/ASN.2021070948

19. Inker LA, Collier W, Heerspink HJL, et al. GFR slope as a surrogate end point for kidney failure in clinical trials: an updated meta-analysis of treatment effects of randomized controlled trials. In preparation.

20. de Galan BE, Perkovic V, Ninomiya T, et al. Lowering blood pressure reduces renal events in type 2 diabetes. *J Am Soc Nephrol.* 2009;20(4):883-892. doi:10.1681/ASN.2008070667

21. Heerspink HJL, Persson F, Brenner BM, et al. Renal outcomes with aliskiren in patients with type 2 diabetes: a prespecified secondary analysis of the ALTITUDE randomised controlled trial. *Lancet Diabetes Endocrinol.* 2016;4(4):309-317. doi:10.1016/S2213-8587(15)00469-6

22. Apperloo AJ, de Zeeuw D, de Jong PE. A short-term antihypertensive treatment-induced fall in glomerular filtration rate predicts long-term stability of renal function. *Kidney Int.* 1997;51(3):793-797. doi:10.1038/ki.1997.111

23. Holtkamp FA, de Zeeuw D, Thomas MC, et al. An acute fall in estimated glomerular filtration rate during treatment with losartan predicts a slower decrease in long-term renal function. *Kidney Int.* 2011;80(3):282-287. doi:10.1038/ki.2011.79

24. Gelman A, Goodrich B, Gabry J, Vehtari A. R-squared for Bayesian regression models. *Am Stat.* 2019;73(3):307-309. doi:10.1080/00031305.2018.1549100

*Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI):
Raymond O. Estacio, Mark Woodward, Hans-Henry Parving, Pietro Canetta, Barry M. Brenner, Brendan Barrett, Bruce Neal, Vlado Perkovic, Kenneth W. Mahaffey, David Johnson, Meg Jardine, Fernando Fervenza, Maximilian von Eynatten, Eduardo Verde, Ursula Verdalles, David Arroyo, Arlene Chapman, Vicente Torres, Alan Yu, Godela Brosnahan, Thierry Hannedouche, Kai-Ming Chow, Cheuk-Chun Szeto, Chi-Bon Leung, Di Xie, Fan Fan Hou, Jamie Dwyer, Marc Pohl, Itamar Raz, Lawrence Hunsicker, Gavin J. Becker, Ritsuko Katafuchi, An Vanacker, Thomas Malfait, Giuseppe Maschio, Jan van den Brand, Jack F.M. Wetzels, Peter Blankestijn, Arjan van Zuilen, Enyu Imai, Fumiaki Kobayashi, Hirofumi Makino, Juliana C.N. Chan, Patrizia Passerini, Lucia Del Vecchio, Francesco Locatelli, Simeone Andrulli, Claudio Pozzi, Donatella Casartelli, Manuel Praga, Hernando Trujillo, Teresa Cavero, Angel Sevillano, Piero Ruggenenti, Annalisa Perna, Fabiola Carrara, Giulia Gherardi, William F. Keane, Carlo Manno, Richard Haynes, William G. Herrington, Colin Baigent, Martin Landray, Thomas Rauen, Claudia Seikrit, Stefanie Wied, Robert D. Toto, Dick de Zeeuw, Paul E. de Jong, Mauro Saddelli, and Pietro Zucchelli.

AFFILIATIONS

[1]Population Health Sciences, University of Utah School of Medicine, Salt Lake City, Utah

[2]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California

[3]Division of Nephrology, Tufts Medical Center, Boston, Massachusetts

[4]Division of Nephrology, Columbia University Medical Center and the New York Presbyterian Hospital, New York, New York

[5]Renal and Metabolic Division, the George Institute for Global Health, Newtown, New South Wales, Australia

[6]Instituto de Investigación Hospital 12 de octubre (i+12), Madrid, Spain

[7]Division of Nephrology, RWTH Aachen University, Aachen, Germany

[8]Department of Nephrology, Hospital General Universitario Gregorio Marañón, Madrid, Spain

[9]Nakayamadera Imai Clinic, Takarazuka, Japan

[10]Program in Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore

[11]Division of Nephrology, Vanderbilt University, Nashville, Tennessee

[12]Division of Nephrology, Prince of Wales Hospital, Chinese University of Hong Kong, Shatin, Hong Kong

[13]Department of Nephrology, Alessandro Manzoni Hospital (past Director), ASST Lecco, Italy

[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]Division of Nephrology, University Hospital of Würzburg, Würzburg, Germany

[18]Department of Clinical Pharmacy and Pharmacology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands