GFR Slope as a Surrogate End Point for Kidney Disease Progression in Clinical Trials: A Meta-Analysis of Treatment Effects of Randomized Controlled Trials

Lesley A. Inker,¹ Hiddo J. L. Heerspink,² Hocine Tighiouart,³,⁴ Andrew S. Levey,¹ Josef Coresh,⁵ Ron T. Gansevoort,⁶ Andrew L. Simon,¹ Jian Ying,⁷ Gerald J. Beck,⁸ Christoph Wanner,⁹ Jürgen Floege,¹⁰ Philip Kam-Tao Li,¹¹ Vlado Perkovic,¹² Edward F. Vonesh,¹³ and Tom Greene⁷

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

ABSTRACT

Background Surrogate end points are needed to assess whether treatments are effective in the early stages of CKD. GFR decline leads to kidney failure, but regulators have not approved using differences in the change in GFR from the beginning to the end of a randomized, controlled trial as an end point in CKD because it is not clear whether small changes in the GFR slope will translate to clinical benefits.

Methods To assess the use of GFR slope as a surrogate end point for CKD progression, we performed a meta-analysis of 47 RCTs that tested 12 interventions in 60,620 subjects. We estimated treatment effects on GFR slope (mean difference in GFR slope between the randomized groups), for the total slope starting at baseline, chronic slope starting at 3 months after randomization, and on the clinical end point (doubling of serum creatinine, GFR < 15 ml/min per 1.73 m², or ESKD) for each study. We used Bayesian mixed-effects analyses to describe the association of treatment effects on GFR slope with the clinical end point and to test how well the GFR slope predicts a treatment’s effect on the clinical end point.

Results Across all studies, the treatment effect on 3-year total GFR slope (median $R^2 = 0.97$; 95% Bayesian credible interval [BCI], 0.78 to 1.00) and on the chronic slope ($R^2 = 0.96$; 95% BCI, 0.63 to 1.00) accurately predicted treatment effects on the clinical end point. With a sufficient sample size, a treatment effect of 0.75 ml/min per 1.73 m²/yr or greater on total slope over 3 years or chronic slope predicts a clinical benefit on CKD progress with at least 96% probability.

Conclusions With large enough sample sizes, GFR slope may be a viable surrogate for clinical end points in CKD RCTs.

JASN 30: 1735–1745, 2019. doi: https://doi.org/10.1681/ASN.2019010007

CKD is common and harmful, causing kidney failure in its late stages, but with few therapies.¹ One of the challenges in development and evaluation of therapies for CKD is that randomized, controlled trials (RCTs) to assess efficacy and safety of novel therapies traditionally use kidney failure and doubling of serum creatinine as clinical end points,² which are late events in the progression of CKD. In order to obtain sufficient end points, RCTs in CKD often require substantial follow-up periods or are restricted to patients with rapidly progressive or late-stage disease, yet some interventions may have a greater effect when applied earlier versus later in the disease course.³ Alternative end points are thus needed to perform RCTs more efficiently, especially in earlier stages of CKD.
Decline in GFR is on the causal pathway to kidney failure, providing strong biologic plausibility for GFR decline as a surrogate end point for CKD progression in RCTs. There are also strong epidemiologic associations of both GFR level and GFR decline with subsequent kidney failure.4–9 However, concerns that relatively small treatment effects on average GFR slope may not translate to treatment effects on clinical end points have complicated regulatory approval of GFR slope as an end point for clinical trials of CKD. Validation of surrogate end points also requires evidence on the basis of randomized comparisons from RCTs that treatment effects on the surrogate end point predict treatment effects on the clinical end point. On the basis of these sorts of analyses, we previously demonstrated that 30% or 40% declines in GFR are valid surrogate end points for CKD progression, but these end points are not appropriate for all populations or interventions, or at early stages of disease. End points on the basis of the mean GFR slope could overcome some of these limitations. However, relatively small differences in mean GFR are generally feasible in follow-up periods for most therapies. Concerns that these seemingly small effects on mean GFR level may not translate to effects on clinical end points have contributed to the reluctance to use GFR slope as an end point in CKD clinical trials. Moreover, patterns of change in GFR after intervention are often nonlinear, with possibly differing direction and rates of changes in early follow-up (herein called acute slope) and longer-term follow-up (herein called chronic slope). The total decline from beginning to the end of the study incorporates both elements (herein called total slope). The frequent occurrence and uncertain implications of acute effects for CKD therapies have also limited the use of GFR slope. Hence, empirical validation of the total and chronic GFR slopes as surrogate end points is necessary before these end points can play a significant role in trials of kidney disease progression. In March of 2018, the National Kidney Foundation (NKF), Food and Drug Administration (FDA), and European Medicines Agency (EMA) cosponsored a scientific workshop, “Change in Albuminuria and GFR as End Points for Clinical Trials in Early Stages of CKD,” to evaluate surrogate end points for trials of kidney disease progression and to improve understanding of changes in albuminuria and GFR as measures of kidney disease progression in early stages of CKD. For this workshop, we performed an individual patient meta-analysis of 47 RCTs accounting for a total of 60,620 subjects across 12 interventions to provide a comprehensive assessment of total or chronic GFR slope as a surrogate end point for trials of CKD progression. We used Bayesian analyses to examine the agreement between treatment effects on GFR slope and treatment effects on the clinical end point and to inform the use of GFR slope as a surrogate end point in future RCTs.

**METHODS**

A more detailed description of the methods is available in the Supplemental Materials.

**Significance Statement**

Surrogate end points are needed to assess whether treatments are effective in the earlier stages of CKD. Measuring the effects of treatments on GFR decline, which leads to kidney failure, might be one way to identify early benefits of CKD treatments. So far regulators have not approved the use of GFR slope, the difference in the change in GFR between treatment groups over time, as an end point in CKD randomized, controlled trials because they are concerned that small treatment effects on GFR may not translate into meaningful clinical benefits. Using a Bayesian individual patient meta-analysis of 47 studies including 60,620 participants, the authors found, that for sufficiently large studies, treatment effects on GFR slope from baseline and from 3-month follow-up of 0.5–1.0 ml/min per 1.73 m²/yr strongly predict benefits on clinical end points such as doubling of serum creatinine, GFR < 15 ml/min per 1.73 m², or ESKD. GFR slope can play a useful role as a surrogate end point for CKD progression in clinical trials.

**Datasets and Analytic Groups**

For our prior work investigating surrogate end points, we developed a pooled database of RCTs using a systematic search (see Supplemental Table 1 for search terms and Supplemental Table 2 for complete list of inclusion criteria).14 In December of 2016, we updated this search and identified additional RCTs and requested individual patient data. After eliminating studies that did not have sufficient data, we had a total of 49 RCTs. Risks of bias for each study were assessed using the risk-of-bias tool of the Cochrane collaboration11 (Supplemental Figure 1). For RCTs that evaluated more than one intervention, we included a separate randomized treatment comparison for each independent treatment versus control comparison reported, such that some participants were included in more than one analytic unit.12–16 We then pooled small RCTs that had <100 participants if the disease and intervention were the same and thus had 49 randomized treatment comparisons as the main unit of analysis (herein called studies) (Supplemental Figure 2, Supplemental Table 3).17–29 Tufts Medical Center Institutional Review Board approved this study.

**Clinical End Points**

The clinical end point was defined as a composite of any of the following events over the full study duration: ESKD (initiation of chronic treatment with dialysis or kidney transplantation), GFR < 15 ml/min per 1.73 m², or sustained doubling of serum creatinine. Of the 49 studies, 47 had sufficient end points for estimation of treatment effects on the clinical end point and were used for the primary analysis.

**GFR**

GFR was estimated using the Chronic Kidney Disease Epidemiology Collaboration 2009 creatinine equation.30 Creatinine was standardized to isotope dilution mass spectrometry traceable reference methods using direct comparison or was reduced by 5% as has previously been described (Supplemental Table 4).31
Statistical Analyses

Objectives
Our first goal was to evaluate the validity of the chronic and total slopes as surrogate end points by assessing the association between treatment effects on each GFR slope end point and the treatment effects on the clinical end point across studies. Our second goal was to use these results to estimate the probability of clinical benefit associated with treatment effects on GFR slope for application to future studies.

Analyses of the Total and Chronic GFR Slopes (Surrogate End Points)
We used a simplified linear mixed-effects model on the basis of a single slope starting at 3 months postrandomization adjusted for baseline GFR. The Supplemental Material describes in detail how the model accounted for various sources of variation in GFR slopes between and within subjects and treatment arms. Under this model, the differences between the randomized groups in the mean intercepts at 3-month follow-up; the mean slopes after 3 months; and the estimated mean changes from baseline to either 1-, 2-, 3-, or 4-year follow-up factored by the follow-up duration represented the treatment effects on the acute, chronic, and total slopes.

Trial-Level Analysis
The trial-level analysis requires two steps: intent-to-treat estimation of the treatment effects on the surrogate and clinical end points within each RCT and a meta-regression to relate the treatment effects on the surrogate and clinical end points across RCTs. In the first step, treatment effects on GFR slopes were estimated using the shared parametric mixed-effects models described above and were expressed as mean differences in the GFR slopes between the treatment versus control groups, in units of ml/min per 1.73 m²/yr. Treatment effects on the clinical end point were estimated by performing separate Cox proportional hazard regressions to estimate log hazard ratios (HRs) for the treatment in each trial. Summary estimates of treatment effects were obtained by use of random-effects models. In the second step, a Bayesian mixed-effects meta-regression related the estimated treatment effects on the clinical end point to the estimated treatment effects on GFR slope with study as the unit of analysis (details in the Supplemental Material). The model relates the treatment effects on the two end points after accounting for random errors in the estimated effects in each RCT. The meta-regression supports validity of GFR slope as a surrogate end point if (1) the slope of the meta-regression line is statistically significant as defined by 95% Bayesian credible intervals (95% BCI) that do not cross 0, with a large magnitude; (2) the intercept is close to 0, implying absence of an average effect on the clinical end point when the treatment does not affect GFR slope; (3) the R² is high, so that treatment effects on GFR slope account for most of the variation in treatment effects on the clinical end point; and (4) the root mean square error (RMSE) is low, assuring low variation in the clinical end point given a fixed treatment effect on GFR slope. We used the designations of low, moderate, and strong trial-level association as defined by $R^2<0.49$, 0.49–0.71, and ≥0.72, respectively.

Positive Predictive Value
We used positive predictive values (PPVs) to describe the uncertainty in predicting the treatment effect on the clinical end point from the treatment effect on the GFR slope. From the trial-level meta-regression, we computed 95% Bayesian prediction intervals and estimated the probabilities of clinical benefit (defined as HR<1) for an infinite, large, or modest-sized RCT. A large RCT was defined as one in which the treatment effect on GFR slope can be estimated to within an SEM of 0.25, corresponding to a total sample size (N) of about 1900 for RCTs whose average follow-up accorded with the RCTs in the analysis. A modest RCT was defined as having an SEM of 0.4 (N roughly 720). We computed the threshold associated with the smallest observed treatment effect on either the chronic or total slope that would assure a high probability of benefit of the treatment on the clinical end point, which we defined as the treatment effect on the GFR slope end point providing a PPV of 97.5%.

Subgroup and Sensitivity Analyses
We performed the trial-level analysis for the primary analytic dataset overall and by subgroups defined by average study level of baseline albumin-to-creatinine ratio (< or ≥30 mg/g; < or ≥3.4 mg/mmol), GFR (< or ≥60 ml/min per 1.73 m²), cause (diabetes and diabetic kidney disease, glomerular diseases, or other causes of CKD), and intervention. Because of differences in the ranges of treatment effects, accuracy in predicting the treatment effect on the clinical end point is best compared between subgroups using the RMSE.

Analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC) and R 3.16.1 (R Project for Statistical Computing, www.r-project.org).

RESULTS
Table 1 summarizes aggregate characteristics of the included studies stratified by disease. The study characteristics of each individual study are reported in Supplemental Tables 5 and 6. Average baseline mean (SD) GFR and median (25th, 75th percentiles) albumin-to-creatinine ratio were 61.7 (26) ml/min per 1.73 m² and 60 mg/g (13, 554) in the pooled dataset, respectively.

For most studies, the mean total slope at 1, 2, 3, and 4 years and chronic slope were slightly attenuated in the treatment arm compared with the control arm (Supplemental Table 7). For example, the pooled mean total slope at 3 years was −3.49 (95% confidence intervals −4.04, −2.93) ml/min per 1.73 m²/yr in the control arm and −2.94 (−3.45, −2.43) ml/min per 1.73 m²/yr in the treatment arm. The mean treatment effect on the total slope at 3 years (0.45 [0.19, 0.72] ml/min per 1.73 m²/yr) was similar
to that of the chronic slope (0.53 [0.32, 0.74]) ml/min per 1.73 m²/yr) with apparent variation by intervention (Figure 3, Supplemental Figure 3, A–E, Supplemental Table 8).

A total of 7115 patients reached the composite clinical end point across the 47 studies (Supplemental Table 9). Across all interventions, the active treatment led to a reduction in risk for the clinical end point (HR, 0.76; 95% confidence interval, 0.69 to 0.84), with similar results across subgroups (Figure 1, right panel; Supplemental Figure 4).

There was strong agreement between the treatment effects on the total slope at 3 years and those of the clinical end point (Figure 2, Supplemental Table 10). The slope of the meta-regression line was –0.42 (95% BCI, –0.55 to –0.30 per ml/min per 1.73 m²/yr), which indicates that each 0.75 ml/min per 1.73 m²/yr greater treatment effect on the total GFR slope was associated with an average 27% lower hazard for the clinical end point (95% BCI, 20% to 34%). The intercept of the regression line was 0.05 (95% BCI, 0.14 to 0.02), indicating that when the treatment had no effect on the total GFR slope at 3 years, there was a low probability of having a substantial treatment effect on the clinical end point.

The median estimate for $R^2$ was 0.97 (95% BCI, 0.78 to 1.00), with Bayesian probabilities of 0.1%, 0.9%, and 99% for the $R^2$ values falling into low, moderate, or high ranges for the strength of a surrogate end point. Similar results are shown for the chronic slope ($R^2$ 0.96), with Bayesian probabilities of 0.6%, 5%, and 94% for the $R^2$ values falling into low, moderate, or high ranges for the strength of a surrogate end point.

Results were weaker when the total slope was computed over shorter durations (Supplemental Figure 5, Supplemental Table 10).

No clear evidence of significant differences in RMSE was found in comparisons of summary results obtained from subgroups stratified by GFR or cause of disease, but credible intervals were wide for some groups (Supplemental Figure 6, Supplemental Table 11, Table 2).

For application of GFR slope as a surrogate end point in future RCTs, Table 3 shows the predicted HRs and 95% prediction intervals for the treatment effects on the clinical end point. For example, for future large trials, the model predicts that a mean difference between the treatment and control groups of 0.54 or 0.48 ml/min per 1.73 m²/yr in total GFR slope over either 2 or 3 years, respectively, and 0.62 ml/min per 1.73 m²/yr for chronic slope, confers a 97.5% probability of a nonzero clinical benefit. For modest-sized trials, a study would be required to have observed treatment effects of 0.72 or 0.74 ml/min per 1.73 m²/yr mean difference in total GFR slope over either 2 or 3 years, respectively, and 0.85 ml/min per 1.73 m²/yr mean difference for chronic GFR slope, to confer 97.5% probability of clinical benefit. The predicted probabilities for a clinical benefit for total slope over 1 year are lower.

**DISCUSSION**

There is strong biologic plausibility and epidemiologic support for GFR slope as a measure of kidney disease progression.

---

**Table 1.** Clinical characteristics of the population stratified by disease cause.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Patients, N</th>
<th>Age, Mean (SD)</th>
<th>Female, N (%)</th>
<th>Black, N (%)</th>
<th>Diabetes, N (%)</th>
<th>GFR, Mean (SD)</th>
<th>ACR, Median (25, 75th)</th>
<th>Clinical End Points, N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>47</td>
<td>60.620</td>
<td>61.4 (11.3)</td>
<td>22,607 (37.3)</td>
<td>4601 (7.6)</td>
<td>61.7 (26.4)</td>
<td>60 (13, 554)</td>
<td>7115 (11.7)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>12</td>
<td>64.0 (8.5)</td>
<td>21,073 (34.7)</td>
<td>1484 (3.4)</td>
<td>73.6 (30.2)</td>
<td>5 (0.4)</td>
<td>7.2 (2.3)</td>
<td>2431 (5.6)</td>
</tr>
<tr>
<td>Glomerular disease</td>
<td>9</td>
<td>40.8 (12.8)</td>
<td>19 (1.4)</td>
<td>1317 (83.8)</td>
<td>1871 (119)</td>
<td>38.2 (22.1)</td>
<td>150 (36.72)</td>
<td>188 (13.5)</td>
</tr>
<tr>
<td>Other CKD</td>
<td>26</td>
<td>56.2 (13.9)</td>
<td>6137 (39.0)</td>
<td>3098 (19.7)</td>
<td>63 (13, 554)</td>
<td>3851 (119)</td>
<td>150 (36.72)</td>
<td>4496 (28.5)</td>
</tr>
</tbody>
</table>

Additional notes:
- Other CKD refers to causes of CKD other than glomerular disease or diabetes or cause not specified.
- Clinical end point defined as the composite of chronic dialysis or kidney transplantation.
- GFR, glomerular filtration rate. ACR, albumin-to-creatinine ratio.
- RASB, renin-angiotensin system blocker.
- CCB, calcium channel blocker.
- Alb protocol, albumin-targeted protocol.
This report provides the addition of trial-level analyses to support GFR slope as a surrogate end point performed for a scientific workshop cosponsored by the NKF, FDA, and EMA. We found that with computation of GFR slope using a robust method, treatment effects on both the total slope at 3 years and the chronic slope have strong associations with the treatment effect on clinical end points, with similar results across key subgroups, including patients with higher baseline GFR, with weaker associations observed for total slope of shorter duration, especially at 1 year. We provide thresholds for minimum effects on change in GFR slope that provide high confidence for significant treatment effects on the clinical end point, providing guidance as to how to interpret treatment effects on GFR slope in future RCTs. This, together with the two companion papers, supports the validity and utility of GFR slope as a surrogate end point in RCTs of CKD progression. The results provide strong general support for the validity of both total slope over 3 years and chronic slope as surrogate end points in CKD RCTs. The treatment effects on both the chronic and total slopes over 3 years accounted for at least an estimated 96% of the variation between studies in treatment effects on the clinical end point, with Bayesian probabilities of at least 90% that the $R^2$'s exceed 0.72, a threshold suggested for a strong surrogate end point. The strength of this trial-level association compares favorably with widely used surrogate end points in other fields. Our analyses imply that, although an effective treatment may reduce mean GFR decline by what might appear to be a small magnitude over the typical duration of RCTs, treatment effects in the range of 0.5–1.00 ml/min per 1.73 m²/yr can have high predictive values of >98% for benefit on the clinical end point. The companion meta-analysis of observational studies that examined the association between GFR slope and subsequent ESKD incidence in more than 1 million individuals demonstrated results consistent with our results.

The frequently encountered presence of acute effects in therapies for CKD progression has limited the use of GFR slope, and, when used, there are often questions as to how GFR slope should be computed—should we use the total slope, which incorporates both the acute and chronic periods,
There has been reluctance to use the chronic slope as a primary outcome because it is defined by change in GFR from a post-baseline time point at which the GFR has already been modified by the treatment, incurring risk of bias due to attenuation of the acute effect or early discontinuation of the study medication. Future work should guide us on how to minimize bias with the use of chronic slope, such as innovative designs employing off-treatment GFR measurements and application of different prerandomization baseline measurements for the treatment and control arms after introduction of the treatment in a run-in phase, as seen in the recent studies evaluating tolvaptan in polycystic kidney disease.37

There are several implications of these results. First, in phase 3 studies, GFR slope could be considered as a candidate primary end point. However, as shown in the simulations companion paper, total or chronic GFR slope confers the advantage of statistical power only under select circumstances.35 For example, the advantage of total slope over time to event time points is clearly demonstrated when there is no acute effect. The effect of an acute effect on the performance of the total slope is greater when the acute effect is large, the time interval for calculating total slope is short, or the control group progression rate is slow. Hence, the time interval required for good performance of the total slope in specific RCTs is likely to depend on each of these factors. The simulations also show that chronic slope has substantially greater power than the clinical end point when the acute effect is negative, but there is a risk of false conclusions for benefit. Thus, sponsors or investigators who consider GFR slope as an end point should do so after consideration of the entire set of design parameters. Second, trials that utilize clinical end points will be most sensitive to fast progressors. GFR slope can be used to demonstrate whether there is similar or different benefit among slow progressors. GFR slope can be used to explore heterogeneous effects among subgroups for trials that are powered for the clinical end point. Similarly, GFR slope might be an appropriate end point for confirmatory studies in a subsequent study after initial RCT showed benefit on the clinical end point but there is interest in demonstrating benefit of a drug in a population or with a study design where the clinical end point is not practical. Third, GFR slope is likely to have great value in phase 2 studies, which are shorter and smaller than phase 3 studies and cannot be powered for benefit on the clinical end point. Thus, overall, use of GFR slope can be considered
Table 2. Trial-level analysis by subgroup

<table>
<thead>
<tr>
<th>Group</th>
<th>N Patients (N Events)</th>
<th>N Studies (N Interv)</th>
<th>Meta-Regression Slope</th>
<th>RMSE</th>
<th>Intercept</th>
<th>Slope</th>
<th>RMSE</th>
<th>90% CI for Intercepts</th>
<th>90% CI for Slopes</th>
<th>90% CI for Intercept</th>
<th>90% CI for Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>1745</td>
<td>30</td>
<td>-0.92</td>
<td>0.22</td>
<td>0.30</td>
<td>0.20</td>
<td>0.41</td>
<td>-0.08 to 0.02</td>
<td>-0.07 to 0.03</td>
<td>0.06 to 0.16</td>
<td>0.03 to 0.13</td>
</tr>
<tr>
<td>Diabetic</td>
<td>13 (6)</td>
<td>34 (6)</td>
<td>0.50</td>
<td>0.20</td>
<td>0.41</td>
<td>0.18</td>
<td>0.26</td>
<td>-0.03 to 0.16</td>
<td>-0.03 to 0.15</td>
<td>0.05 to 0.15</td>
<td>0.03 to 0.16</td>
</tr>
<tr>
<td>Glomerular</td>
<td>9 (2)</td>
<td>12 (6)</td>
<td>-0.06</td>
<td>0.20</td>
<td>0.35</td>
<td>0.18</td>
<td>0.26</td>
<td>-0.02 to 0.07</td>
<td>-0.02 to 0.05</td>
<td>0.05 to 0.11</td>
<td>0.03 to 0.13</td>
</tr>
<tr>
<td>Other CKD</td>
<td>26 (8)</td>
<td>24 (6)</td>
<td>0.39</td>
<td>0.20</td>
<td>0.38</td>
<td>0.18</td>
<td>0.26</td>
<td>0.04 to 0.16</td>
<td>0.03 to 0.15</td>
<td>0.06 to 0.15</td>
<td>0.04 to 0.16</td>
</tr>
</tbody>
</table>

Unit of GFR is ml/min per 1.73 m². Units of ACR are mg/g. Results for slope, intercept, R², and RMSE are presented as median and 2.5–97.5 Bayesian credible intervals.

www.jasn.org

A potential end point in the design and implementation of multiple phases within a drug development pathways.

The main drawback to using GFR slope as an end point in a pivotal trial is that it does not directly indicate the magnitude of the treatment benefit on the clinical end point. Some clinicians might question whether a small difference in mean GFR slope between treatment arms is sufficient evidence to adopt the intervention. Our results as well as those of our companion paper linking changes in GFR slope over a 1-, 2-, or 3-year period provide quantifiable data on the implications of such differences in GFR slope for longer term outcome.34

Strengths of this study include a systematic literature search to include all available studies, resulting in a large and diverse collection of RCTs, and a rigorous evaluation using individual patient data. Because we analyzed patient-level data, we were able to characterize agreement between the GFR slope and clinical end points after adjusting for spurious correlations in sampling error that resulted from inclusion of the same GFR measurements in the GFR slope and clinical end points. Our application of a robust method for analysis of GFR slope that accounted 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 RCTs. Our use of a Bayesian meta-regression model with diffuse prior distributions allowed us to rigorously account for multiple sources of uncertainty and to translate treatment effects on the surrogate end points to probabilities of benefit on the clinical end point.

There are several limitations. First, because ascertainment of clinical end points was limited to the follow-up period of each trial, we were able to evaluate only the association between the treatment effects on the surrogate and clinical end points during the RCTs, and could not determine whether treatment effects on the surrogate end points predicted the longer term effects of the treatment on future clinical end points. The first companion paper evaluates epidemiologic associations over a longer period of follow-up.34,38 Second, because we used the same slope model for each RCT, somewhat different results might be obtained if the model for slope were tailored to each RCT, including trial-specific strategies for informative censoring and designating the timing of the acute effect.39 Third, because most included trials were not designed as short trials we cannot be certain about the effect of lesser follow-up time on the results, nor could we consider the effect of increased measurement frequencies on such shorter trials. Fourth, our results are dependent on the specific RCTs available to us. Hence, application of these results to future trials with different characteristics to those included here must be done with caution, particularly in trials with larger magnitude of acute effects, or lower rates of GFR decline. Fifth, our analyses do not address the risk that slope-based analyses would lead to false positive conclusions under the null hypothesis of no effect of the treatment on the clinical end point, nor do they evaluate the specific conditions in which analyses of GFR slope outcomes provide superior statistical power than analyses of

JASN 30: 1735–1745, 2019 

GFR Slope as a Surrogate End Point 1741
Table 3. Application of GFR slope as surrogate end point in new RCT: predicted treatment effect on clinical end point and PPV

<table>
<thead>
<tr>
<th>GFR Slope</th>
<th>Observed Treatment Effect on Change in GFR Slope</th>
<th>Infinite Sample Size in New RCT</th>
<th>Large RCT</th>
<th>Modest RCT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median HR and 95% Prediction Interval</td>
<td>PPV</td>
<td>Median HR and 95% Prediction Interval</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total slope over 4 yr</td>
<td>0.5 0.78 (0.69 to 0.87)</td>
<td>1.00</td>
<td>0.78 (0.59 to 1.01)</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>0.75 0.69 (0.61 to 0.78)</td>
<td>1.00</td>
<td>0.69 (0.52 to 0.89)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1.0 0.61 (0.53 to 0.7)</td>
<td>1.00</td>
<td>0.61 (0.46 to 0.8)</td>
<td>1.00</td>
</tr>
<tr>
<td>Threshold for treatment effect on GFR slope to assure PPV ≥ 97.5%</td>
<td>0.20 0.52</td>
<td>0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total slope over 3 yr</td>
<td>0.5 0.77 (0.64 to 0.90)</td>
<td>1.00</td>
<td>0.77 (0.59 to 0.99)</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>0.75 0.69 (0.58 to 0.81)</td>
<td>1.00</td>
<td>0.69 (0.52 to 0.89)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1.0 0.62 (0.52 to 0.74)</td>
<td>1.00</td>
<td>0.62 (0.47 to 0.80)</td>
<td>1.00</td>
</tr>
<tr>
<td>Threshold for treatment effect on GFR slope to assure PPV ≥ 97.5%</td>
<td>0.24 0.48</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total slope over 2 yr</td>
<td>0.5 0.75 (0.56 to 0.98)</td>
<td>0.98</td>
<td>0.75 (0.54 to 1.01)</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>0.75 0.7 (0.52 to 0.91)</td>
<td>0.99</td>
<td>0.7 (0.5 to 0.94)</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>1.0 0.65 (0.48 to 0.85)</td>
<td>1.00</td>
<td>0.65 (0.46 to 0.87)</td>
<td>1.00</td>
</tr>
<tr>
<td>Threshold for treatment effect on GFR slope to assure PPV ≥ 97.5%</td>
<td>0.42 0.54</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total slope over 1 yr</td>
<td>0.5 0.74 (0.49 to 1.1)</td>
<td>0.94</td>
<td>0.74 (0.49 to 1.1)</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>0.75 0.72 (0.47 to 1.06)</td>
<td>0.96</td>
<td>0.72 (0.47 to 1.07)</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>1.0 0.69 (0.46 to 1.03)</td>
<td>0.97</td>
<td>0.69 (0.45 to 1.04)</td>
<td>0.97</td>
</tr>
<tr>
<td>Threshold for treatment effect on GFR slope to assure PPV ≥ 97.5%</td>
<td>1.26 1.32</td>
<td>1.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic slope</td>
<td>0.5 0.8 (0.66 to 0.95)</td>
<td>0.99</td>
<td>0.8 (0.6 to 1.05)</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>0.75 0.72 (0.59 to 0.86)</td>
<td>1.00</td>
<td>0.72 (0.54 to 0.94)</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>1.0 0.65 (0.53 to 0.78)</td>
<td>1.00</td>
<td>0.65 (0.48 to 0.85)</td>
<td>1.00</td>
</tr>
<tr>
<td>Threshold for treatment effect on GFR slope to assure PPV ≥ 97.5%</td>
<td>0.37 0.62</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Units of GFR are ml/min per 1.73 m². Treatment effect on GFR slope is expressed as mean difference and in units of ml/min per 1.73 m²/yr. Treatment effect on the clinical end point is expressed as HR. PPVs are defined as the 97.5% probabilities for clinical benefit, defined as HR<1 for an infinite, large, or modest-sized RCT. A large RCT was defined as one in which the treatment effect on GFR slope can be estimated to within an SEM of 0.25, corresponding to a total sample size (N) of about 1900 for RCTs whose average follow-up accorded with the RCTs in the analysis. A modest RCT was defined as having SEM of 0.4 (N roughly 720).
the clinical end point. Our companion paper uses simulations to address the latter three questions. In summary, the results presented here, together with our companion papers, suggest that total and chronic GFR slope are strong surrogate end points and may be used as end points for RCTs of kidney disease progression in certain circumstances in both early and late CKD. Future work is required to optimize the design of RCTs to deploy slope-based end points to increase efficiency while preserving a sufficiently low risk of false positive conclusions.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the help of Juhi Chaudhari in preparation of the manuscript and Shiyuan Miao in preparation of the figures.

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication. The corresponding author had full access to all of the data in the study and final responsibility for the decision to submit for publication.

Dr. Inker, Dr. Greene, Dr. Heerspink, Dr. Coresh, Dr. Levey, and Dr. Gansevoort conceived of the study concept and design. The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) investigators/collaborators listed below acquired the data. Dr. Tighiouart, Dr. Simon, Dr. Greene, and Dr. Inker analyzed the data. All authors took part in the interpretation of the data. Dr. Inker and Dr. Greene drafted the manuscript, and all authors provided critical revisions of the manuscript for important intellectual content. All collaborators shared data and were given the opportunity to comment on the manuscript. Dr. Inker, Dr. Greene, Dr. Heerspink, and Dr. Levey obtained funding for the CKD-EPI and individual cohort and collaborator support is listed in Supplemental Appendix 2.

CKD-EPI investigators/collaborators (study acronyms/abbreviations are listed in Supplemental Appendix 1 with other abbreviations):


The planning and operations committee of the National Kidney Foundation, Food and Drug Administration, and European Medicines Agency Scientific Workshop on Change in Albuminuria and GFR as End Points for Clinical Trials in Early Stages of CKD contributed to the design and critical review of these analyses. Planning and Operations Committees: Dr. Levey (Chair), Dr. Gansevoort, Dr. Coresh, Dick de Zeeuw, Kai-Uwe Eckardt, Hrefna Gudmundsdottir, Adeera Levin, Romaldas Maculaelaitis, Tom Manley, Dr. Perkovic, Kimberly Smith, Norman Stockbridge, Aliza Thompson, Thorsten Vetter, Kerry Willis, and Luzia Zhang.

FUNDING

The study was funded by the National Kidney Foundation. A variety of sources have supported the randomized, controlled trials included in the Chronic Kidney Disease Epidemiology Collaboration. These funding sources include government agencies such as the National Institutes of Health and medical research councils as well as foundations and industry sponsors listed in Supplemental Appendix 2.

DISCLOSURES

Dr. Inker reports funding to Tufts Medical Center for research and contracts with the National Institutes of Health (NIH), National Kidney Foundation (NKF), Retrophin, Omeros, Reata Pharmaceuticals, and Dialysis Clinic, Inc. She has consulting agreements with Tricida Inc. and Omeros Corp. Tufts Medical Center, John Hopkins University and Metabolon Inc. have a collaboration agreement to develop a product to estimate GFR from a panel of markers. Dr. Levey reports grants from the NIH and the NKF during the conduct of the study, and funding from Siemens outside of the submitted work. Dr. Coresh has grants from the NIH and the NKF related and unrelated to this research. Dr. Inker, Dr. Levey, and Dr. Coresh have a patent "Precise estimation of GFR from multiple biomarkers pending to Dr. Coresh, Dr. Inker, and Dr. Levey; and Tufts Medical Center, John Hopkins University, and Metabolon Inc. have a collaboration agreement to develop a product to estimate GFR from a panel of markers. Dr. Heerspink reports grants and other from Abbvie, other from Astellas, grants and other from AstraZeneca, grants and other from Boehringer Ingelheim, grants and other from Janssen, other from Fresenius, other from Gilead, and other from Merck, outside of the submitted work. Dr. Wanner reports personal fees from Boehringer Ingelheim during the conduct of the study, and personal fees from Lilly, personal fees from AstraZeneca, and personal fees from MSD outside of the submitted work. Dr. Floege has received consultancy honoraria and/or speaker fees from Alnylam, Amgen, Bayer, Cal militia, Chugai, Fresenius, Omeros Corp., and
Vifor. Dr. Perkovic reports personal fees for Advisory Boards or Scientific Presentations from Retrophin, Janssen, Merck, and Servier. He has served on Steering Committees for trials funded by Abbvie, Boehringer Ingelheim, GlaxoSmithKline, Janssen, and Pfizer; and participated in Scientific Presentations/Advisory boards with Abbvie, Astellas, Astra Zeneca, Bayer, Baxter, Bristol-Myers Squibb, Boehringer Ingelheim, Dimeptrix, Duract, Eli Lilly, Gilead, GlaxoSmithKline, Novartis, Novo Nordisk, Pfizer, Pharmalink, Relypsy, Sanofi, Tricida, and Vitae, with fees paid to his institution. Dr. Vonesh served as a paid biostatistics consultant for the NKF for the expressed purpose of developing statistical models for use in the estimation and comparison of GFR slopes as a surrogate end point in CKD randomized, controlled trials. He is also serving as a biostatistics consultant to Prometic and Tricida, Inc., in which some of the work entails consulting on the design and analysis of clinical trials in patients with CKD. Dr. Greene reports grants from the NKF during the conduct of the study, and personal fees from DURECT Corporation, Janssen Pharmaceuticals, and Pfizer Inc., outside of the submitted work. Dr. Beck, Dr. Gansevoort, Dr. Ying, Dr. Tighiouart, Dr. Li, and Dr. Simon have no conflicts to report.

SUPPLEMENTAL MATERIAL

This article contains the following supplemental material online at http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2019010007/-/DCSupplemental.
Supplemental Appendix 1: Abbreviations, units, and terms.
Supplemental Appendix 2: Study funding sources.
Protocol.
Background and rationale.
Dataset development.
Datasets and analytic groups.
Data management.
Clinical end points.
Estimated GFR.
GFR slope.
Analyses.
Trial-level model for relating treatment effects on the clinical end point to treatment effects on GFR slope.
Prediction intervals and positive predictive value.
Supplemental Table 1: Search terms.
Supplemental Table 2: Study inclusion criteria.
Supplemental Table 3: Studies pooled by intervention.
Supplemental Table 4: Description of studies.
Supplemental Table 5: Patient characteristics by study.
Supplemental Table 6: Distribution of the maximum visit time for each person by duration.
Supplemental Table 7: Slopes (95% confidence intervals) by treatment arm for each intervention.
Supplemental Table 8: Treatment effects by intervention.
Supplemental Table 9: End points used by study.
Supplemental Table 10: Trial-level analysis for GFR slope overall and by different duration.
Supplemental Table 11: Summary of trial-level analyses for GFR slope by subgroup.
Supplemental Figure 1: Evaluation of bias.
Supplemental Figure 2: Flowchart.
Supplemental Figure 3: Treatment effect on GFR slope.
Supplemental Figure 3A: Chronic slope.
Supplemental Figure 3C: Total slope at 2 year.
Supplemental Figure 3D: Total slope at 3 year.
Supplemental Figure 3E: Total slope at 4 year.
Supplemental Figure 4: Forest plot for clinical end point.
Legend for Supplemental Figures 5 and 6.
Supplemental Figure 5: Trial-level analyses for the association between treatment effects on total GFR slope by varying duration and treatment effect on the clinical end point.
Supplemental Figure 6: Trial-level analyses for the association between treatment effects on GFR slope and treatment effects on the clinical end point by level of eGFR.
Supplemental Figure 6A: Total GFR slope over 3 Year.
Supplemental Figure 6B: Chronic GFR slope.

REFERENCES

drug class on progression of hypertensive kidney disease: Results from the AASK trial. JAMA 288: 2421–2431, 2002


**AFFILIATIONS**

1Division of Nephrology and 2The Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, Massachusetts; Departments of 3Clinical Pharmacy and Pharmacology and 4Nephrology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 5Tufts Clinical and Translational Science Institute, Tufts University, Boston, Massachusetts; 6Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; 7Division of Epidemiology, Department of Internal Medicine, University of Utah, Salt Lake City, Utah; 8Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio; 9Division of Nephrology, University Hospital of Würzburg, Würzburg, Germany; 10Division of Nephrology, RWTH Aachen University, Aachen, Germany; 11Division of Nephrology, Prince of Wales Hospital, Chinese University of Hong Kong, Shatin, Hong Kong; 12George Institute for Global Health, University of New South Wales, Sydney, Australia; and 13Department of Preventive Medicine, Division of Biostatistics, Northwestern University, Chicago, Illinois