Original Investigation

Utility and Validity of Estimated GFR–Based Surrogate Time-to-Event End Points in CKD: A Simulation Study

Tom Greene, PhD,[1] Chia-Chen Teng, MS,[1] Lesley A. Inker, MD, MS,[2] Andrew Redd, PhD,[1] Jian Ying, PhD,[1] Mark Woodward, PhD,[3] Josef Coresh, MD, PhD,[4] and Andrew S. Levey, MD[2]

Background: There is interest in surrogate end points for clinical trials of chronic kidney disease progression because currently established end points—end-stage renal disease (ESRD) and doubling of serum creatinine level—are late events, requiring large clinical trials with long follow-up. Doubling of serum creatinine level is equivalent to a 57% decline in estimated glomerular filtration rate (eGFR). We evaluated type 1 error and required sample size for clinical trials using surrogate end points based on lesser eGFR declines.

Study Design: Simulation study.

Setting & Participants: Simulations evaluating 3,060 scenarios representative of 19 treatment comparisons in 13 chronic kidney disease clinical trials.

Index Tests: Surrogate end points defined as composite end points based on ESRD and either 30% or 40% eGFR declines.

Reference Test: Clinical outcome (ESRD) for type 1 error. Established end point (composite of ESRD and 57% eGFR decline) for required sample size.

Results: Use of the 40% versus 57% eGFR decline end point consistently led to a reduction in sample size $>$ 20% while maintaining risk for type 1 error $<$ 10% in the presence of a small acute effect ($<$1.25 mL/min/1.73 m^2) for: (1) 2-, 3-, or 5-year trials with a high mean baseline eGFR (67.5 mL/min/1.73 m^2), and (2) 2-year trials with an intermediate mean baseline eGFR (42.5 mL/min/1.73 m^2). Use of the 30% versus the 40% eGFR decline end point often led to moderately larger reductions in sample size in the absence of an acute effect, but not in the presence of acute effects.

Limitations: The complexity of eGFR trajectories prevented evaluation of all scenarios for clinical trials.

Conclusions: Use of end points based on 30% or 40% eGFR declines is an appropriate strategy to reduce sample size in certain situations. However, risk for type 1 error is increased in the presence of acute effects, particularly for 30% eGFR declines. The decision to use these end points should be made after thorough evaluation of their expected performance under the conditions of specific clinical trials.

Am J Kidney Dis. 64(6):867-879.

INDEX WORDS: Clinical trials; simulation; kidney end point; renal end point; kidney disease outcome; surrogate end point; chronic kidney disease (CKD); estimated glomerular filtration rate (eGFR) decline; eGFR trajectory; renal function; kidney disease progression; end-stage renal disease (ESRD).

This is the final article in a series of 4 to report the analyses undertaken in conjunction with a 2012 workshop sponsored by the National Kidney Foundation (NKF) and US Food and Drug Administration (FDA) titled "GFR Decline as an Endpoint for Clinical Trials in CKD." The Workshop Report is included in this issue of *AJKD*.[1]

The FDA accepts the composite of doubling of serum creatinine level or kidney failure as an end point for clinical trials of kidney disease progression. However, these are late events in chronic kidney disease (CKD). A doubling of serum creatinine level corresponds approximately to a 57% decline in estimated glomerular filtration rate (eGFR) using the 2009 CKD-EPI (CKD Epidemiology Collaboration) creatinine equation. End points based on lesser eGFR declines provide a greater number of events in a shorter time, raising the possibility of using surrogate end points based on lesser eGFR declines to reduce the follow-up times and/or sample sizes required in clinical trials. The first article in this series showed consistent epidemiologic associations between lesser

From the [1]Division of Epidemiology, Department of Internal Medicine, University of Utah, Salt Lake City, UT; [2]Division of Nephrology, Tufts Medical Center, Boston, MA; [3]Department of Epidemiology, Johns Hopkins University; and [4]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Received April 2, 2014. Accepted in revised form August 27, 2014. Originally published online October 30, 2014.

Because an author of this article is an editor for AJKD, the peer-review and decision-making processes were handled entirely by an Associate Editor (Paul Muntner, PhD, MHS) who served as Acting Editor-in-Chief. Details of the journal's procedures for potential editor conflicts are given in the Information for Authors & Editorial Policies.

Address correspondence to Tom Greene, PhD, Division of Epidemiology, University of Utah, 295 South Chipeta Way, Salt Lake City, UT 84108. E-mail: tom.greene@hsc.utah.edu

0272-6386/$36.00

http://dx.doi.org/10.1053/j.ajkd.2014.08.019

declines in eGFR and subsequent kidney failure and mortality in observational studies.[2] The third article showed consistent epidemiologic associations between end points defined by lesser eGFR declines and subsequent kidney failure in joint analysis of clinical trials across a range of clinical characteristics.[3] However, demonstration of epidemiologic association is a necessary but not sufficient condition for demonstrating that treatment effects on the surrogate end point can be used for predicting treatment effects on the established end point.[4-6] The second article in this series addressed this limitation by evaluating the consistency of treatment effects on the established and alternative end points from a meta-analysis of clinical trials.[7] Unfortunately, the quantity and variety of well-powered clinical trials in CKD was insufficient to establish the consistency of this relationship or determine with certainty the factors affecting it.

Symptomatic kidney failure, as defined by the onset of uremic symptoms signifying a requirement for renal replacement therapy, occurs when eGFR reaches a level of 7 to 15 mL/min/1.73 m^2, a narrow range relative to a normal eGFR of ~120 mL/min/1.73 m^2. Thus, in contrast to many surrogate end points, eGFR has a close mathematical relationship with the established end point of kidney failure. The fact that kidney failure commences when a relatively narrow eGFR range is reached provides the opportunity to model relationships between treatment effects on eGFR-based surrogate end points and the time to kidney failure more precisely than is possible using general criteria for validating surrogate end points. The challenge for such a model-based approach is that patterns of eGFR decline and the relationship of these patterns with time to kidney failure can be complex. To deal with this complexity, we performed a simulation study to characterize the circumstances under which treatment effects on time-to-event end points based on lesser eGFR declines reliably predict treatment effects on time to kidney failure treated by maintenance dialysis or kidney transplantation (end-stage renal disease [ESRD]).The input parameters for the simulations were chosen to represent a wide spectrum of scenarios found in prior randomized controlled trials about CKD.

METHODS

Framework for eGFR Trajectories

In order to investigate the performance of eGFR-based surrogate end points, it first is necessary to provide a framework for describing and simulating eGFR trajectories. We first describe the classic linear random slope and intercept model that has been used most commonly for analyzing longitudinal outcomes such as eGFR[8] and then present extensions to account for deviations from the classic framework and for relationships of eGFR trajectories with death and ESRD.

The Classic Slope-Intercept Model

The classic model is a random-effects linear growth curve model that posits that each patient's eGFR measurements vary randomly about an underlying linear trajectory (Fig 1E in Levey et al[1]) that can be described by an intercept (initial eGFR at baseline) and a single slope (rate of change over time). If the slopes, intercepts, and eGFR deviations are jointly normally distributed, the distribution of eGFR values for patients assigned to a particular treatment can be defined by 6 quantities: (1 and 2) the mean intercept and slope, which determine the average eGFR trajectory; (3 and 4) the standard deviations of the intercepts and slopes, which characterize the variation among individual patients' trajectories; (5) the correlation between the slopes and intercepts; and (6) the residual standard deviation, which defines variation in the eGFR values about their linear trajectory.

Deviations From the Classic Model

Past CKD clinical trials have demonstrated several deviations from the classic framework. The 2 most important deviations pertain to the short- and long-term effects of the treatment. Treatments in CKD trials often produce early changes in mean eGFR that differ from the long term slope[9,10] (Fig 1B in Levey et al[1]). For example, blockade of the renin-angiotensin system by angiotensin-converting enzyme inhibitors or angiotensin receptor blockers usually leads to an initial decline in GFR, and hence eGFR, although these interventions slow the rate of disease progression in the long term. This early effect is referred to as an acute effect and, as in the renin-angiotensin system–blockade example, often is considered to be hemodynamic. We address acute effects by allowing the initial mean slope during the first several months after randomization to differ from the long-term mean slope. Although hemodynamic effects usually occur quickly after interventions are implemented, for most interventions, they are thought to persist over time at a given level of eGFR,[11,12] but may attenuate to 0 as eGFR declines prior to ESRD. Our simulations considered 3 options for attenuation of the acute effect: (1) a linear attenuation of the acute effect to 0 when eGFR reaches 15 mL/min/1.73 m^2, (2) a logarithmic attenuation to 0 when eGFR reaches 15 mL/min/1.73 m^2, and (3) persistence of the initial acute effect as eGFR declines.

The long-term treatment effect also may deviate from the classic model if the treatment slows progression differently depending on the patient's underlying eGFR decline (progression rate). It has been noted that some treatments may have larger effects on eGFR decline among fast progressors (patients with faster rates of eGFR decline), but do not alter eGFR slopes among nonprogressors (patients with stable eGFRs)[13,14] (Fig 1D in Levey et al[1]). Thus, we consider 3 models for the long-term treatment effect: (1) a uniform treatment effect in which the same treatment effect is assumed for fast progressors and nonprogressors, (2) a proportional treatment effect in which the treatment effect is proportional to the rate of eGFR decline, and (3) an intermediate model halfway between the uniform and proportional treatment effect models.

Our simulations also addressed 4 other deviations of eGFR trajectories from the classic model. First, we used the generalized log gamma distribution to simulate the distribution of slopes. The generalized log gamma distribution includes the normal distribution as a limiting case, but also includes negatively skewed distributions to account for a possible subgroup of rapid progressors.[15] Second, we modeled positive correlations between successive eGFR measurements to allow individual patients to have nonlinear trajectories in which contiguous eGFRs may deviate from the underlying linear trajectory for periods of time.[16,17] Third, we used t distributions to examine the implications of outliers in the deviations of individual eGFR measurements from the underlying trajectories. Fourth, in accordance with results of analyses of prior CKD studies (see Item S1, provided as

Table 1. Factors Considered in Simulations

Category and Factor	Values Considered in Simulations
eGFR decline	
1. Mean long-term slope	−2.5, −4.0,[a] or −5.5 mL/min/1.73 m^2
2. SD of long-term slope	3.0, 3.5,[a] or 4.5 mL/min/1.73 m^2 per y
3. Correlation of slope and intercept	0,[a] a negative correlation providing a slope 1 mL/min/1.73 m^2 per y steeper for each 25-mL/min/1.73 m^2 increase in baseline eGFR
4. Slope skewness	Low, moderate,[a] or high, characterized by a generalized log gamma distribution with shape parameter of 1.5, 3,[a] or 5
5. Nonlinearity of individual trajectories	None,[a] moderate, high (described in Item S1)
6. Size of residuals	Residual variance = 0.67 × eGFR value (medium variability)[a] or 1.15 × eGFR value (high variability)
7. Frequency of residual outliers	Low,[a] moderate, or high, represented by normal,[a] *t* with 12 *df*, or *t* with 5 *df* distributions
Acute effect	
8. Mean acute effect	−2.5, −1.25, 0, or +1.25 mL/min/1.73 m^2 at baseline eGFR = 42.5 mL/min/1.73 m^2
9. Attenuation of initial acute effect	Linear to 15, logarithmic to 15,[a] none
10. Variability of acute effect	None,[a] low, or high, characterized by acute-effect SDs of 0,[a] 1, or 3 mL/min/1.73 m^2
Long-term treatment effect	
11. Type of long-term treatment effect	Proportional, uniform, intermediate[a] (see Item S1)
12. Size of long-term treatment effect	0%, 20%, 25%,[a] or 30% reduction in slope for participant with an average long-term slope in absence of treatment
Death and ESRD	
13. Death rate per year	Case 1: 0.03375 − 0.00025[a] × eGFR Case 2: 0.0675 − 0.00050 × eGFR
14. eGFR level associated with onset of ESRD	Uniformly distributed between 6 and 15 mL/min/1.73 m^2[a] or between 10 and 15 mL/min/1.73 m^2
Design and conduct	
15. Accrual period and total F/U	Short trial: 1-y accrual and 1.5-y further F/U Medium trial: 2-y accrual and 2-y further F/U Long trial: 2-y accrual and 4-y further F/U
16. Measurement frequency	3, 6,[a] and 12 mo
17. Mean baseline eGFR	27.5, 42.5, or 67.5 mL/min/1.73 m^2
18. No. of baseline eGFRs	1, 2,[a] or 3
19. Permanent loss to F/U rate	2%[a] or 5% per y
20. Intermittent missing eGFRs	5%, 7.5%,[a] or 10%

Abbreviations: eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; F/U, follow-up; SD, standard deviation.
[a]Input parameters for the base-case scenario. No base-case values were specified for mean baseline eGFR, the accrual period and further follow-up, or for the size of the acute effect.

online supplementary material), we assumed that the variance of the residual eGFR deviations is proportional to eGFR level.

Death and ESRD

The mortality hazard rate was assumed to be related linearly to patients' eGFR levels based on their underlying trajectories at each follow-up time, with higher death rates at lower eGFRs. It was assumed that ESRD occurred when either the eGFR trajectory or a measured eGFR first declined below a patient-specific random threshold between 6 to 15 mL/min/1.73 m^2.

Conduct of Simulations

Choice of Input Parameters

The simulations were defined by a total of 20 input parameters (Table 1, first column). Design parameters included the accrual period and total duration of follow-up, frequency of eGFR measurements, baseline eGFR, number of baseline eGFR measurements, rate of permanent loss to follow-up, and the rate of intermittently missing eGFRs.

The rightmost column of Table 1 presents the values of the 20 input parameters that were considered in the simulations. Values for each input parameter were selected based in part on analyses of 19 treatment comparisons in the 13 largest CKD studies considered in the second article of this series[7] (see Item S2). We considered 2 to 4 values for each parameter to reflect the distribution of values observed in these studies or to reflect the range of biologically plausible values. Of importance, acute effects ranging from −1.25 to +1.25 mL/min/1.73 m^2 were observed frequently in the prior CKD clinical trials, with larger acute effects observed in 2 cases. Our analyses also suggest that mixed long-term treatment effects intermediate between the uniform and proportional treatment effect model may be common in CKD clinical trials.

Evaluation of Utility and Validity

The simulations were constructed to address 2 criteria for the performance of a surrogate end point. First, we characterized circumstances in which the eGFR-based surrogate end points preserved a low risk of type 1 error relative to the clinical outcome of ESRD, which is relevant to the validity of the surrogate. This requirement stipulates that the probability of reporting a statistically significant treatment effect on the surrogate should not substantially exceed the designated type 1 error when the treatment has no effect

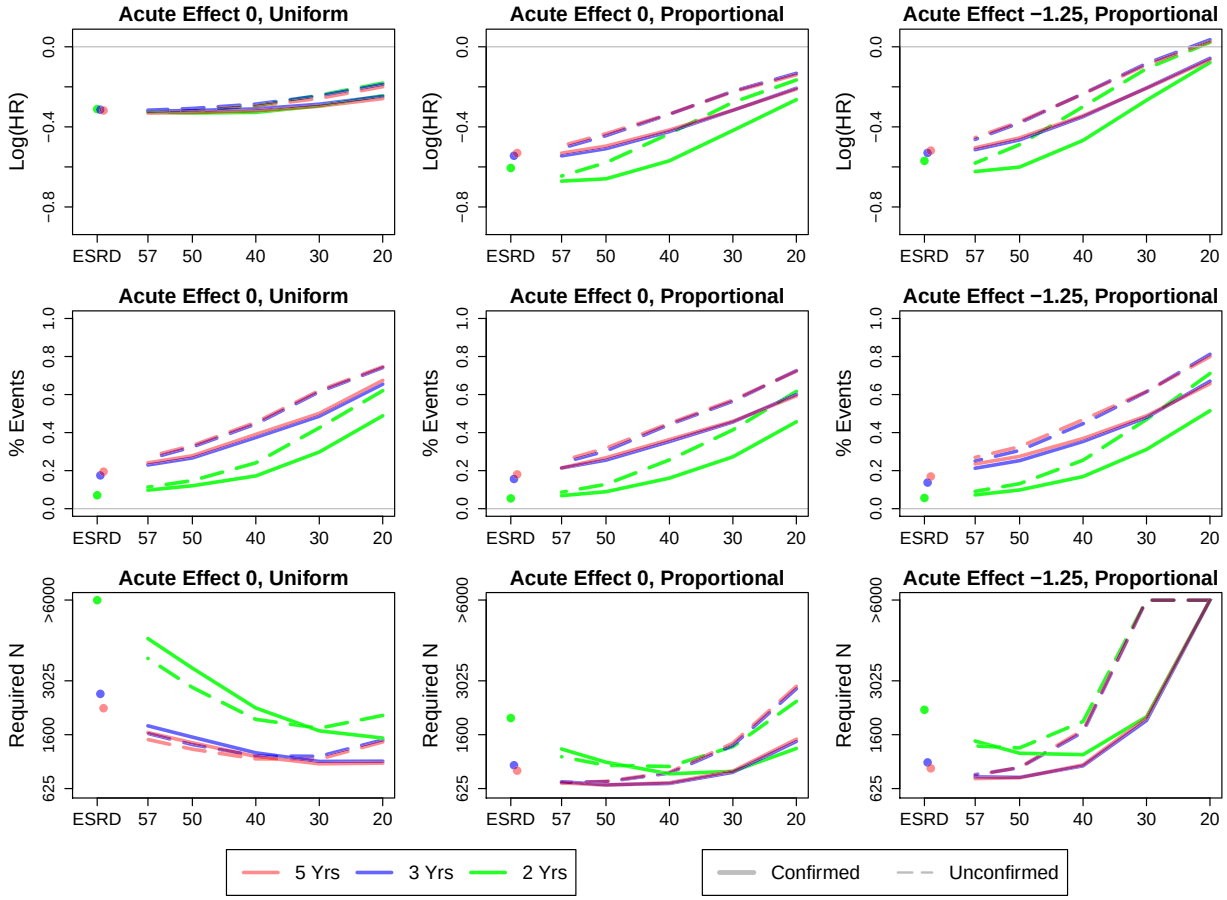

Figure 1. (Top row) Log hazard ratios (HRs), (middle row) percentage of patients with events, and (bottom row) the required N for end points defined by end-stage renal disease (ESRD) alone or composites of designated estimated glomerular filtration rate (eGFR) declines and ESRD (horizontal axis) with and without confirmation of eGFR events (solid and dashed curves) for trials with a short, medium, and long median planned follow-up of 2, 3, or 5 years (green, blue, or red curves) for 3 scenarios for the acute effect and long-term treatment effect: (1) no acute effect and a 25% uniform long term treatment effect (left), (2) no acute effect and a 25% proportional long-term treatment effect, and (3) a negative 1.25 mL/min/1.73 m^2 acute effect of the treatment and a 25% proportional long-term treatment effect. Mean baseline eGFR was 42.5 mL/min/1.73 m^2 for each scenario, with all other input parameters in the simulations set to their base-case value. The required Ns are provided on the square root scale. In some cases, curves fail to display fully due to overlap.

on time to ESRD. Second, we characterized the circumstances in which the surrogate end points decreasing sample size at a fixed statistical power beyond that provided by the established composite outcome when the treatment has a beneficial effect on the time to ESRD, which is relevant to the utility of the surrogate. The targeted 2-sided type 1 error was 5% for each analysis.

The numerous combinations of parameter values (>80 million) precluded examination of all combinations; hence, we evaluated the performance of different eGFR-based surrogate end points for 36 combinations defined by 4 levels of the acute effect, 3 levels of baseline eGFR, and 3 trial durations at fixed values of the remaining input parameters (which we refer to as base-case values), and then varied the remaining parameters one at a time. When evaluating type 1 error, the treatment effect on the long-term slope was fixed at 0 and the acute effect was assumed to attenuate to 0 prior to the time the patient reached ESRD. When evaluating statistical power, we considered proportional, mixed, or uniform treatment effects of 20%, 25%, or 30% on the long-term slope in conjunction with acute effects that either attenuated or persisted as eGFR declined. This led to evaluation of type 1 error under the null hypothesis of no treatment effect on time to ESRD across $36 \times 39 = 1{,}404$ parameter configurations, and evaluation of power in the presence of a treatment effect on time to ESRD across $36 \times 46 = 1{,}656$ parameter configurations, for a total of 3,060 parameter configurations (scenarios).

End Points

For each parameter combination, we examined the occurrence of ESRD alone as well as composite end points defined by either the occurrence of ESRD or eGFR decline of either 20%, 30%, 40%, 50%, or 57%. We evaluated each composite end point both with and without confirmation of eGFR events. For an eGFR event without confirmation, a single measured eGFR had to fall below the designated threshold. For an eGFR event with confirmation, an eGFR value initially falling below the threshold had to be confirmed by a second eGFR at a repeat visit simulated to occur within 1 month after the initial eGFR that triggered the event. Thus, 11 end points were evaluated for each parameter configuration: ESRD alone and 5 eGFR declines with and without confirmation.

Simulation Methods

For each parameter combination, we simulated 800 independent data sets, with each data set consisting of 1,000 patients with equal allocation to the treatment and control groups. For each simulated

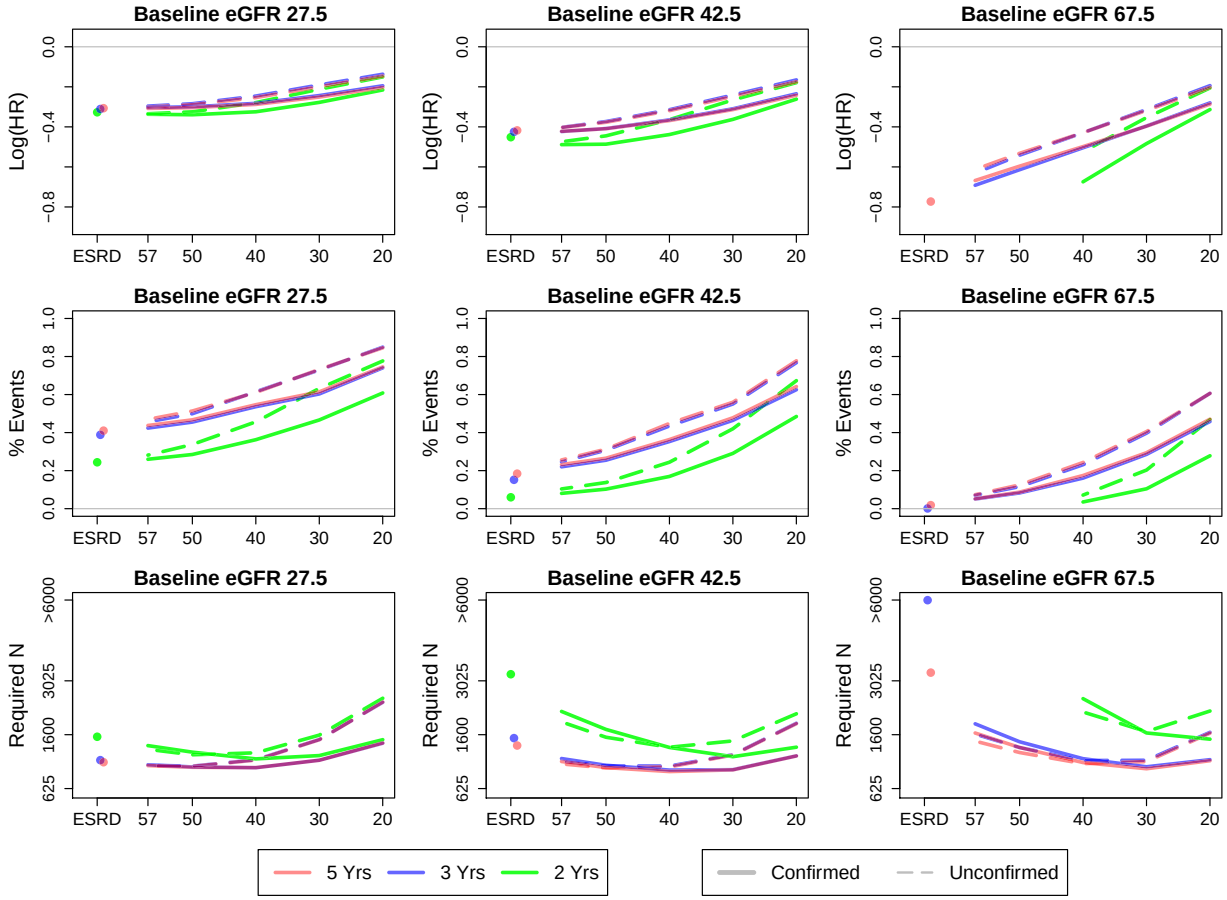

Figure 2. (Top row) Log hazard ratios (HRs), (middle row) percentage of patients with events, and (bottom row) the required N for end points defined by end-stage renal disease (ESRD) alone or composites of designated estimated glomerular filtration rate (eGFR) declines and ESRD (horizontal axis) with and without confirmation of eGFR events (solid and dashed curves) for trials with a short, medium, and long median planned follow-up of 2, 3, or 5 years (green, blue, or red curves) and with a (left) low mean baseline eGFR of 27.5 mL/min/1.73 m^2, (middle) intermediate mean baseline eGFR of 42.5 mL/min/1.73 m^2, or (right) high mean baseline eGFR of 67.5 mL/min/1.73 m^2. All other input parameters in the simulations were set to their base-case value including the assumption of a mixed proportional/additive model for the long-term treatment effect. The required Ns are provided on the square root scale. In some cases, curves fail to display fully due to overlap.

data set, we applied Cox proportional hazards regression to estimate the treatment effect corresponding to each of the 11 end points described previously while censoring mortality. We estimated type 1 error and statistical power based on the mean and estimated standard errors of Cox regression coefficients across the 800 simulated data sets. Assessments of statistical power were based on the sample size required to achieve 90% power with 2-sided $\alpha = 0.05$.

Dependence of Sample Size on Percent of Patients Reaching Events and Estimated Hazard Ratio

The sample size required to achieve a designated statistical power is related inversely to the statistical power that can be achieved with a designated sample size. In time-to-event analyses, statistical power is determined by the product of the square root of the percent of patients with events and the absolute value of the average log hazard ratio (HR):

$$Statistical\ \ Power = f\left(\sqrt{(\%\ \text{Reaching Events})} \times \left|\mathrm{E}\left\{\log\left(\widehat{HR}\right)\right\}\right|\right)$$

The expression $\mathrm{E}\{\log(\widehat{HR})\}$ represents the "expected value" of the estimated log HR under the scenario being evaluated. The absolute value $|\mathrm{E}\{\log(\widehat{HR})\}|$ is equal to 0 when the expected HR is equal to 1 and increases to values above 0 when the expected HR falls either below or above 1, reflecting an effect of the treatment. Reducing the threshold for eGFR events necessarily increases the first term because the percent of patients with lesser eGFR declines always equals or exceeds the percent with larger declines. By contrast, using a lesser eGFR decline may either increase the second term if it yields HRs further from 1 or reduce it if applying a lesser eGFR decline attenuates the HR toward 1. Considering both terms, using lesser declines provides increased power if the proportional increase in events exceeds the effect of any attenuation in the HR. The simulations explore the variation of these terms across the ranges for the input parameters described above.

RESULTS

Figure 1 displays HRs, proportions of patients reaching events, and required N's for the 11 end points for trials with average planned durations of 2, 3, or 5 years for 3 parameter settings for the acute and long-term treatment effects: (1) no acute effect and uniform treatment effect (left), (2) no acute effect and proportional treatment effect (middle), and (3) a

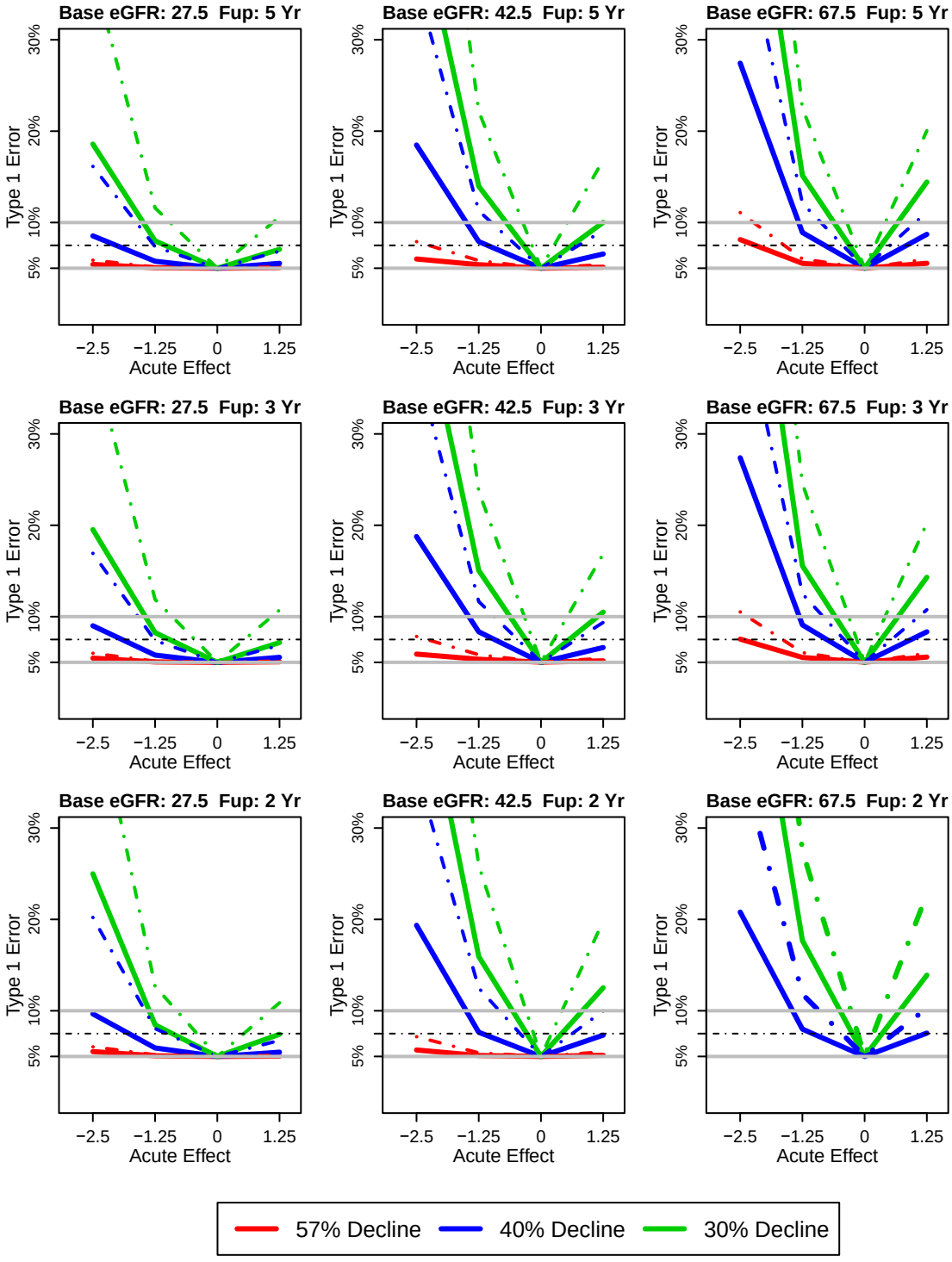

Figure 3. Type 1 error of composite end points based on end-stage renal disease (ESRD) and either a 57% estimated glomerular filtration rate (eGFR) decline (red curve), 40% eGFR decline (blue curve), or 30% eGFR decline (green curve) with or without confirmation (solid or dashed curves) relative to the clinical end point of ESRD when there are varying acute effects of the treatment (horizontal axis), but no treatment effect on time to ESRD. The targeted type 1 error is 5%, so deviations of the curves above 5% represent inflated risk of a false conclusion of treatment harm for negative acute effects or inflated risk of a false conclusion of treatment effect for positive acute effects. The rows reflect different levels of baseline eGFR and the columns represent mean planned trial durations of 2, 3, or 5 years. All other input parameters in the simulations were set to their base-case value. In general, in this surrogate end point setting, we expect that a small inflation of type 1 error relative to the clinical end point may be acceptable in certain circumstances, but that large increases in type 1 error to levels close to 10% or higher are unacceptable. In regulatory settings, the consequences of a false-positive conclusion of treatment benefit may be more severe than for a false-positive conclusion of treatment harm. Abbreviation: Fup, follow-up.

moderate negative acute effect of −1.25 mL/min/1.73 m^2 and a proportional treatment effect (right). HRs were stable across end points with different eGFR declines when there was no acute effect and a uniform long-term effect (top left). The proportional treatment effect, which implies a larger effect on the slopes of faster progressors, strengthened the HRs for ESRD alone and for the end points based on the larger eGFR declines, where events are restricted to the fastest progressors. However, HRs attenuated toward 1 for the lesser eGFR declines. Attenuation of the HR with lesser eGFR declines was increased in the presence of a negative acute effect, reflecting a greater relative impact of an acute eGFR decline for end points based on lesser eGFR declines than for events based on larger eGFR declines. Use of lesser eGFR declines substantially increased the number of events for each scenario (middle panels). Reflecting the combined effects of the HRs and number of events, lesser eGFR declines reduced the required N in the absence of an acute effect when the treatment effect is uniform, but increased the required N in the presence of a small negative acute effect with a proportional long-term treatment effect. Requiring confirmation of eGFR events substantially reduced the required N for end points based on 30% or 40% eGFR declines, but was not a major determinant of the required N with a 57% decline. The figures in Item S3 show that using lesser eGFR declines also leads to attenuation of the HRs for an intermediate long-term treatment effect, but not as much as for a proportional treatment effect.

Figure 2 provides similar summaries for parameter configurations with low (27.5), intermediate (4.25), and high (67.5 mL/min/1.73 m^2) mean baseline eGFRs under the intermediate long-term treatment effect model without an acute effect. The required N is stable across the different eGFR decline end points for 3- or 5-year studies when baseline eGFR is 27.5 mL/min/1.73 m^2, but is reduced with lesser eGFR declines for parameter configurations where events are less frequent (ie, 2-year studies and/or high baseline eGFR).

Figure 3 displays type 1 error for the composite end points relative to ESRD when there is a nonzero acute effect but no treatment effect on ESRD. Type 1 error increases above the nominal 5% for 30% eGFR declines for positive acute effects, which lead to inflated risk of falsely concluding treatment benefit, as well as negative acute effects, which lead to inflated risk of falsely concluding treatment harm. The type 1 error exhibits analogous but smaller increases in the presence of acute effects for 40% eGFR declines and is highly robust across the full range of acute effects for the stricter threshold of 57%. Failure to require confirmation of eGFR events inflates type 1 error in the presence of acute effects with all thresholds, but especially so for the 30% and 40% eGFR declines.

Tables 2 to 4 display the ratios of the required sample sizes for 30% and 40% confirmed eGFR declines versus a confirmed 57% eGFR decline in relation to changes in 8 input parameters that we found to be the most influential for determining power. The 3 tables examine 2-, 3-, and 5-year trials with intermediate baseline eGFR. Smaller ratios indicating greater power with the lesser eGFR declines are obtained when baseline eGFR is higher and the treatment effect is uniform rather than proportional, whereas larger ratios are seen when long-term eGFR slopes or individual eGFR measurements have increased variability. The susceptibility of the 30% and 40% eGFR decline thresholds to the adverse consequences of a negative acute effect of the treatment is increased if mean eGFR slope is less steep (as might be expected, eg, in trials with low levels of baseline proteinuria), if the long-term treatment effect is smaller, if eGFR measurements are obtained more frequently, and if variation of the acute effect between patients is higher, particularly for short studies. Table 5 presents similar results for a trial with high baseline eGFRs with a median planned follow-up of 3 years. End points based on 30% or 40% eGFR declines consistently provide greater savings in required N for studies with high baseline eGFRs compared with studies with intermediate baseline eGFRs. The tables in Item S3 provide expanded summaries of the required sample size ratios considering an expanded set of input parameters and trials with low, intermediate, and high baseline eGFRs. Tables 6 and 7 provide a synopsis of these results.

DISCUSSION

This simulation study provides an extensive investigation of surrogate end points for time-to-event analyses based on alternative eGFR declines across a large number of scenarios chosen to be representative of 13 of the largest clinical trials performed in patients with CKD. The simulations confirm that the established composite end point of doubling of serum creatinine level (corresponding approximately to a 57% eGFR decline) or ESRD preserves a low risk of type 1 error and produces treatment effect estimates in agreement with the clinical end point of ESRD across all parameter configurations that we considered. The simulations also show that composite end points based on a confirmed 30% or 40% eGFR decline or ESRD can substantially reduce the required sample size compared to the established composite end point in certain situations. For example, when no acute effect is present, both end points with lesser eGFR declines reduced the required sample sizes by at least

Table 2. Ratio of Required Sample Size for 30% or 40% eGFR Decline End Points Compared to 57% eGFR Decline End Point: Intermediate Baseline eGFR and Planned Follow-up of 2 Years

	Acute Effect = −1.25 mL/min/1.73 m^2			No Acute Effect		
	N Ratio 30% vs 57%	N Ratio 40% vs 57%	Required N With 57%	N Ratio 30% vs 57%	N Ratio 40% vs 57%	Required N With 57%
Long-term treatment effect						
Proportional	1.45	0.84	1,405	0.66	0.64	1,307
Mixed[a]	1.24	0.82	2,388	0.54	0.60	2,225
Uniform	0.98	0.74	5,305	0.38	0.51	4,315
Mean slope						
−5.50 mL/min/1.73 m^2 per y	1.15	0.83	1,216	0.59	0.64	1,115
−4.00 mL/min/1.73 m^2 per y[a]	1.24	0.82	2,388	0.54	0.60	2,225
−2.50 mL/min/1.73 m^2 per y	1.96	0.91	5,042	0.51	0.57	4,650
Mean initial eGFR						
27.50 mL/min/1.73 m^2	1.31	0.97	1,518	0.84	0.81	1,339
42.50 mL/min/1.73 m^2[a]	1.24	0.82	2,388	0.54	0.60	2,225
67.50 mL/min/1.73 m^2	57% infeasible	57% infeasible	57% infeasible	57% infeasible	57% infeasible	57% infeasible
Size of treatment effect						
20%	1.78	0.98	3,715	0.58	0.63	3,268
25%[a]	1.24	0.82	2,388	0.54	0.60	2,225
30%	1.00	0.77	1,664	0.53	0.59	1,554
Frequency of eGFR determination						
Every 3 mo	1.58	0.84	2,521	0.55	0.58	2,187
Every 6 mo[a]	1.24	0.82	2,388	0.54	0.60	2,225
Every y	1.09	0.83	3,265	0.55	0.63	2,741
Long-term slope SD						
3.00 mL/min/1.73 m^2 per y	1.05	0.73	2,622	0.48	0.56	2,167
3.50 mL/min/1.73 m^2 per y[a]	1.24	0.82	2,388	0.54	0.60	2,225
4.50 mL/min/1.73 m^2 per y	1.64	1.00	2,397	0.72	0.73	2,122
eGFR variability						
0.67 mL/min/1.73 m^2[a]	1.24	0.82	2,388	0.54	0.60	2,225
1.10 mL/min/1.73 m^2	1.96	0.95	3,037	0.67	0.66	2,330
Acute-effect SD						
0[a]	1.24	0.82	2,388	0.54	0.60	2,225
1.00	1.29	0.85	2,352	0.57	0.61	2,143
3.00	2.40	1.08	2,400	0.74	0.72	2,197

Note: Shown are ratios of required sample size (N) for 90% power with a 2-sided $\alpha = 0.05$ for composite end points based on ESRD or 30% or 40% eGFR declines versus a composite based on ESRD or a 57% eGFR decline for a trial with a median planned follow-up of 2 years and mean baseline eGFR of 42.5 mL/min/1.73 m^2. Scenarios with a negative acute effect and no acute effect are presented. Each block of rows shows the effect of varying 1 input parameter, with all remaining input parameters set to their base case in Table 1 (indicated by [a]). Simulations could not be performed for the composite of ESRD or a 57% eGFR decline when baseline eGFR = 67.5 mL/min/1.73 m^2 due to an insufficient number of events.

Abbreviations: eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; SD, standard deviation.

34% across all parameter configurations considered for 2-year trials in patients with an intermediate mean baseline eGFR of 42.5 mL/min/1.73 m^2 and for trials of 2 to 3 years in patients with a high mean baseline eGFR of 67.5 mL/min/1.73 m^2.

Both the utility and validity of the end points based on 30% or 40% eGFR decline are contingent on the absence of moderate to large acute effects (±1.25 mL/min/1.73 m^2 or more). Acute effects resulting from hemodynamic changes in GFR have been common among interventions that previously have been evaluated for CKD progression, including renin-angiotensin system blockade, low blood pressure, low-protein diet, dihydropyridine calcium channel blockers, and cyclosporine.[9,10,18-22] Acute effects on eGFR also can result from changes in non-GFR determinants of the filtration marker, in this case creatinine. Our simulations indicate that use of lesser eGFR declines often reduces statistical power for treatments causing acute declines ≥1.25 mL/min/1.73 m^2 and increases type 1 error, leading to false conclusions of treatment benefit for treatments

Table 3. Ratio of Required Sample Size for 30% or 40% eGFR Decline End Points Compared to the 57% eGFR Decline End Point: Intermediate Baseline eGFR and Planned Follow-up of 3 Years

	Acute Effect = −1.25 mL/min/1.73 m^2			No Acute Effect		
	N Ratio 30% vs 57%	N Ratio 40% vs 57%	Required N With 57%	N Ratio 30% vs 57%	N Ratio 40% vs 57%	Required N With 57%
Long-term treatment effect						
Proportional	2.45	1.25	788	1.16	0.94	719
Mixed[a]	1.64	1.06	1,224	0.87	0.84	1,077
Uniform	1.07	0.87	2,080	0.57	0.68	1,822
Mean slope						
−5.50 mL/min/1.73 m^2 per y	1.51	1.07	669	0.89	0.85	592
−4.00 mL/min/1.73 m^2 per y[a]	1.64	1.06	1,224	0.87	0.84	1,077
−2.50 mL/min/1.73 m^2 per y	2.42	1.21	2,738	0.89	0.85	2,305
Mean initial eGFR						
27.50 mL/min/1.73 m^2	1.55	1.09	1,082	1.04	0.95	983
42.50 mL/min/1.73 m^2[a]	1.64	1.06	1,224	0.87	0.84	1,077
67.50 mL/min/1.73 m^2	1.10	0.86	2,199	0.51	0.59	1,834
Size of treatment effect						
20%	2.16	1.22	1,934	0.88	0.83	1,724
25%[a]	1.64	1.06	1,224	0.87	0.84	1,077
30%	1.47	1.03	826	0.83	0.81	768
Frequency of eGFR determination						
Every 3 mo	2.15	1.18	1,235	0.91	0.84	1,128
Every 6 mo[a]	1.64	1.06	1,224	0.87	0.84	1,077
Every y	1.45	1.04	1,184	0.84	0.83	1,103
Long-term slope SD						
3.00 mL/min/1.73 m^2 per y	1.50	1.01	1,105	0.78	0.78	962
3.50 mL/min/1.73 m^2 per y[a]	1.64	1.06	1,224	0.87	0.84	1,077
4.50 mL/min/1.73 m^2 per y	1.88	1.18	1,579	1.03	0.95	1,309
eGFR variability						
0.67 mL/min/1.73 m^2[a]	1.64	1.06	1,224	0.87	0.84	1,077
1.10 mL/min/1.73 m^2	2.27	1.22	1,420	0.99	0.88	1,156
Acute-effect SD						
0[a]	1.64	1.06	1,224	0.87	0.84	1,077
1 mL/min/1.73 m^2	1.82	1.15	1,216	0.88	0.85	1,085
3 mL/min/1.73 m^2	2.32	1.21	1,334	1.03	0.90	1,170

Note: Shown are ratios of required sample size (N) to achieve 90% power with a 2-sided $\alpha = 0.05$ for composite end points based on ESRD or 30% or 40% eGFR declines versus a composite based on ESRD or a 57% eGFR decline for a randomized trial with a median planned follow-up of 3 years and mean baseline eGFR of 42.5 mL/min/1.73 m^2. Scenarios with a negative acute effect and no acute effect are presented. Each block of rows shows the effect of varying 1 input parameter, with all remaining input parameters set to their base case in Table 1 (indicated by [a]).

Abbreviations: eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; SD, standard deviation.

causing an acute increase in eGFR of a similar magnitude. The adverse consequences of acute effects on type 1 error and statistical power were substantially greater for 30% declines than 40% declines.

The utility of eGFR decline–based end points also can be limited for interventions that cause treatment effects that are either proportional to the patient's underlying rate of eGFR decline or intermediate between uniform and proportional effects. Proportional or intermediate treatment effects produce stronger HRs for the established end point than for end points based on lesser eGFR declines, hence partially or fully negating the benefit of the increased numbers of events based on lesser eGFR declines. Trial duration, baseline eGFR, size of the long-term effect, and variability of the eGFR measurements and slopes of the eGFR trajectories also are important determinants of the performance of end points with lesser eGFR declines. Due to this complexity, designers of future CKD clinical trials may benefit from simulations similar to those presented in this article adapted to the specific circumstances of the trial. Even with such an evaluation, uncertainty usually will remain. Investigators are unlikely to be certain of the size of the

Table 4. Ratio of Required Sample Size for 30% or 40% eGFR Decline End Points Compared to 57% eGFR Decline End Point: Intermediate Baseline eGFR and Planned Follow-up of 5 Years

	Acute Effect = −1.25 mL/min/1.73 m^2			No Acute Effect		
	N Ratio 30% vs 57%	N Ratio 40% vs 57%	Required N With 57%	N Ratio 30% vs 57%	N Ratio 40% vs 57%	Required N With 57%
Long-term treatment effect						
Proportional	2.76	1.39	751	1.33	1.04	676
Mixed[a]	1.79	1.15	1,070	0.97	0.91	95
Additive	1.14	0.93	1,644	0.64	0.72	1,482
Mean slope						
−5.50 mL/min/1.73 m^2 per y	1.58	1.11	605	0.96	0.90	534
−4.00 mL/min/1.73 m^2 per y[a]	1.79	1.15	1,070	0.97	0.91	953
−2.50 mL/min/1.73 m^2 per y	2.59	1.31	2,418	0.97	0.90	2,132
Mean initial eGFR						
27.50 mL/min/1.73 m^2	1.60	1.13	1,034	1.06	0.97	959
42.50 mL/min/1.73 m^2[a]	1.79	1.15	1,070	0.97	0.91	953
67.50 mL/min/1.73 m^2	1.28	0.95	1,643	0.64	0.69	1,360
Size of treatment effect						
20%	2.23	1.26	1,754	0.97	0.90	1,529
25%[a]	1.79	1.15	1,070	0.97	0.91	953
30%	1.52	1.07	756	0.91	0.87	676
Frequency of eGFR determination						
Every 3 mo	2.24	1.23	1,105	1.01	0.92	995
Every 6 mo[a]	1.79	1.15	1,070	0.97	0.91	953
Every y	1.53	1.08	1,095	0.89	0.87	1,020
Long-term slope SD						
3.00 mL/min/1.73 m^2 per y	1.63	1.08	936	0.88	0.84	824
3.50 mL/min/1.73 m^2 per y[a]	1.79	1.15	1,070	0.97	0.91	953
4.50 mL/min/1.73 m^2 per y	1.99	1.25	1,466	1.08	0.98	1,252
eGFR variability						
0.67 mL/min/1.73 m^2[a]	1.79	1.15	1,070	0.97	0.91	953
1.10 mL/min/1.73 m^2	2.42	1.29	1,230	1.08	0.93	1,032
Acute-effect SD						
0[a]	1.79	1.15	1,070	0.97	0.91	953
1.00	1.87	1.18	1,102	0.95	0.91	979
3.00	2.36	1.29	1,183	1.12	0.97	1,032

Note: Shown are ratios of required sample size (N) to achieve 90% power with a 2-sided $\alpha = 0.05$ for composite end points based on ESRD or 30% or 40% eGFR declines versus a composite based on ESRD or a 57% eGFR decline for a randomized trial with a median planned follow-up of 5 years and mean baseline eGFR of 42.5 mL/min/1.73 m^2. Scenarios with a negative acute effect and no acute effect are presented. Each block of rows shows the effect of varying 1 input parameter, with all remaining input parameters set to their base case in Table 1 (indicated by [a]).

Abbreviations: eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; SD, standard deviation.

acute effect or whether the long-term treatment effect is uniform or proportional. Importantly, detection of an acute effect of ±1.25 mL/min/1.73 m^2 typically requires a data set with 600 to 1,200 participants, which may not be available to the investigators prior to the start of the definitive trial. It would be prudent to restrict the use of lesser eGFR declines to situations in which the projected reduction in the required sample size is substantial (ie, ≥15%-20%). If there is doubt about whether an acute effect is present, it also would be prudent to restrict use of these end points to situations in which it can be verified that performance is not seriously compromised by acute effects of up to ±1.25 mL/min/1.73 m^2.

As displayed in Table 6, the composite end point of a confirmed 40% eGFR decline or ESRD provides a substantial reduction in required sample size in the absence of an acute effect while retaining reasonable robustness for acute effects of up to ±1.25 mL/min/1.73 m^2 in magnitude if baseline eGFR is high or baseline eGFR is intermediate and the planned trial duration is short (ie, 2 years). A 40% eGFR decline can be a useful option in certain other situations, but these should be evaluated on a case-by-case basis. A

Table 5. Ratio of Required Sample Size for 30% or 40% eGFR Decline End Points Compared to 57% eGFR Decline End Point: High Baseline eGFR and Planned Follow-up of 3 Years

	Acute Effect = −1.25 mL/min/1.73 m^2			No Acute Effect		
	N Ratio 30% vs 57%	N Ratio 40% vs 57%	Required N With 57%	N Ratio 30% vs 57%	N Ratio 40% vs 57%	Required N With 57%
Long-term treatment effect						
Proportional	1.34	0.85	1,165	0.63	0.64	1,027
Mixed[a]	1.10	0.86	2,199	0.51	0.59	1,834
Uniform	0.66	0.65	6,572	0.32	0.47	4,441
Mean slope						
−5.50 mL/min/1.73 m^2 per y	1.04	0.84	1,077	0.58	0.65	943
−4.00 mL/min/1.73 m^2 per y[a]	1.10	0.86	2,199	0.51	0.59	1,834
−2.50 mL/min/1.73 m^2 per y	1.53	0.90	4,801	0.53	0.61	3,720
Mean initial eGFR						
27.50 mL/min/1.73 m^2	1.55	1.09	1,082	1.04	0.95	983
42.50 mL/min/1.73 m^2	1.64	1.06	1,224	0.87	0.84	1,077
67.50 mL/min/1.73 m^2[a]	1.10	0.86	2,199	0.51	0.59	1,834
Size of treatment effect						
20%	1.48	0.94	3,591	0.53	0.62	2,869
25%[a]	1.10	0.86	2,199	0.51	0.59	1,834
30%	0.91	0.76	1,527	0.49	0.57	1,354
Frequency of eGFR determination						
Every 3 mo	1.38	0.89	2,125	0.55	0.61	1,776
Every 6 mo[a]	1.10	0.86	2,199	0.51	0.59	1,834
Every y	0.96	0.80	2,293	0.51	0.60	1,924
Long-term slope SD						
3.00 mL/min/1.73 m^2 per y	0.91	0.74	2,294	0.40	0.50	2,015
3.50 mL/min/1.73 m^2 per y[a]	1.10	0.86	2,199	0.51	0.59	1,834
4.50 mL/min/1.73 m^2 per y	1.52	1.05	2,143	0.71	0.73	1,731
eGFR variability						
0.67 mL/min/1.73 m^2[a]	1.10	0.86	2,199	0.51	0.59	1,834
1.10 mL/min/1.73 m^2	1.63	0.99	2,318	0.58	0.59	1,931
Acute-effect SD						
0[a]	1.10	0.86	2,199	0.51	0.59	1,834
1.00	1.19	0.86	2,246	0.51	0.59	1,869
3.00	1.83	1.00	2,533	0.68	0.68	2,004

Note: Shown are ratios of required sample size (N) to achieve 90% power with a 2-sided $\alpha = 0.05$ for composite end points based on ESRD or 30% or 40% eGFR declines versus a composite based on ESRD or a 57% eGFR decline for a randomized trial with a median planned follow-up of 3 years and mean baseline eGFR of 67.5 mL/min/1.73 m^2. Scenarios with a negative acute effect and no acute effect are presented. Each block of rows shows the effect of varying 1 input parameter, with all remaining input parameters set to their base case in Table 1 (indicated by [a]).

Abbreviations: eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; SD, standard deviation.

composite end point of a confirmed 30% eGFR decline or ESRD may be appropriate in select circumstances in which there is high confidence that an acute effect is absent (Table 7).

We note 4 limitations to this study. First, while the simulations considered many factors, we did not evaluate long-term effects of treatments on non-GFR determinants of serum creatinine that are not apparent in the first several months after randomization. More generally, characterization of trajectories is highly multifactorial, and our conclusions extend to only the specific input parameters that we considered in the simulations. Second, our analyses of prior CKD clinical trials for determination of input parameters in Item S2 were limited in scope and more extensive analyses of issues such as the nature of long-term treatment effect are warranted for future research. Third, our assessment of type 1 error focused on the setting in which the acute effect attenuates to 0 prior to ESRD. If the initial acute effect persists until ESRD, a positive acute effect would represent a true benefit of the treatment; in this case, using lesser eGFR declines would improve power without incurring inflated type 1 error. Fourth, we considered only a limited collection of time-to-event end points based on designated eGFR declines from baseline. Some of

Table 6. Scenarios of Favorable Performance of 40% eGFR Decline Compared to 57% eGFR Decline When the Acute Effect Is Expected to Be Absent or Smaller Than 1.25 mL/min/1.73 m^2

Baseline eGFR	2-y Trial	3-y Trial	5-y Trial
27.5 mL/min/1.73 m^2	In Select Cases	No Exceptions: Uniform treatment effect, low-slope SD	No Exceptions: Uniform treatment effect
42.5 mL/min/1.73 m^2	Yes Exception: Substantial increase in required N if a negative acute effect does not attenuate	No Exceptions: Uniform treatment effect, low-slope SD	No Exceptions: Uniform treatment effect
67.5 mL/min/1.73 m^2	Yes	Yes	Yes

Yes scenarios require:
1) compared to a 57% eGFR decline, a 40% eGFR decline reduces the required N by at least 20% if there is no acute effect;
2) compared to a 57% eGFR decline, a 40% eGFR decline either reduces the required N or requires an increase in N by no more than 10% when there is a negative acute effect of -1.25 mL/min/1.73 m^2; and
3) type 1 error relative to the clinical end point of end-stage renal disease is $<10\%$ in the presence of a positive acute effect not greater than 1.25 mL/min/1.73 m^2.

Abbreviations: eGFR, estimated glomerular filtration rate; SD, standard deviation.

Table 7. Scenarios of Favorable Performance of 30% eGFR Decline Compared to 57% eGFR Decline When Acute Effect Is Expected to Be Absent

Baseline eGFR	2-y Trial	3-y Trial	5-y Trial
27.5 mL/min/1.73 m^2	No Exceptions: Uniform treatment effect, low-slope SD	No Exception: Uniform treatment effect	No
42.5 mL/min/1.73 m^2	Yes	No Exceptions: Uniform treatment effect, low-slope SD	No
67.5 mL/min/1.73 m^2	Yes	Yes	Yes

Yes scenarios require:
Compared to a 57% eGFR decline, a 30% eGFR decline reduces the required N by at least 20% if there is no acute effect.
Caution:
Use of a 30% decline will lead to large increases in required N compared to both 40% and 57% declines if the treatment unexpectedly produces a negative acute effect with magnitude of at least 1.25 mL/min/1.73 m^2, and to severely inflated type 1 error if the treatment unexpectedly produces a positive acute effect of 1.25 mL/min/1.73 m^2 or greater.

Abbreviations: eGFR, estimated glomerular filtration rate; SD, standard deviation.

the circumstances in which time-to-event end points based on lesser declines perform well, namely the absence of an acute effect and uniform treatment effects, also should be conducive to analyses based on the mean eGFR slope. Analyses using a post-randomization baseline also have been suggested as a means of limiting the impact of the acute effect. In future work, the simulation-based approach developed in this article also may be used to evaluate broader classes of end points.

In conclusion, for treatments with small or no acute effects, use of a 40% eGFR decline in some cases can substantially increase statistical power compared to the current 57% threshold while avoiding a prohibitive inflation of type 1 error relative to the clinical end point of ESRD, thereby allowing smaller sample sizes and/or reduced follow-up times. Use of a 30% eGFR decline can provide moderately increased savings in required sample size compared to a 40% decline in certain situations, but is problematic if a moderate or large acute effect is present. The decision to use these end points should be made after thorough evaluation of their expected performance under the conditions of specific clinical trials.

ACKNOWLEDGEMENTS

The planning committee of the NKF-FDA workshop contributed to the formulation of the research questions and presentation of the concepts at the meeting (members: Andrew S. Levey, MD [chair], Aliza M. Thompson, MD [FDA; co-chair], Josef Coresh, MD, PhD, Kerry Willis, PhD [NKF], Norman Stockbridge, MD, PhD [FDA], Edmund Lewis, MD, Dick de Zeeuw, MD, PhD, and Alfred K. Cheung, MD).

Key results presented in this article were presented in abstract form at the meeting of the American Society of Nephrology, November 5-10, 2013, in Atlanta, GA.

The participating CKD-EPI clinical trials/collaborators are listed in Item S4.

Support: The workshop was supported and facilitated by the NKF. The NKF gratefully acknowledges Abbott, Amgen, ChemoCentryx, Lilly, Mitsubishi Tanabe Pharma, Novartis, Pfizer, Reata, Sanofi, and Takeda, which provided grants to the NKF to support the workshop and the supporting data analysis. This research was also supported by the University of Utah Study Design and Biostatistics Center, with funding in part from the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through grant 8UL1TR000105 (formerly UL1RR025764).

Financial Disclosure: Dr Greene is a consultant for the following companies: Keryx Biopharmaceuticals, Jansen Pharmaceuticals, Pfizer, GenKyoTex S.A. The other authors declare that they have no other relevant financial interests.

Contributions: Research idea and study design: TG, LAI, ASL, JC; data acquisition: LAI, ASL; statistical analysis/interpretation: TG, C-CT, JY, MW; software development and validation: C-CT, JY, AR. Each author contributed important intellectual content during manuscript drafting or revision and accepts accountability for the overall work by ensuring that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved. TG takes responsibility that this study has been reported honestly, accurately, and transparently; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

SUPPLEMENTARY MATERIAL

Item S1: Details of simulations of eGFR trajectories and clinical events.

Item S2: Analyses of prior CKD randomized controlled trials.

Item S3: Detailed summaries of simulation results.

Item S4: List of participating CKD-EPI clinical trials/collaborators.

Note: The supplementary material accompanying this article (http://dx.doi.org/10.1053/j.ajkd.2014.08.019) is available at www.ajkd.org

REFERENCES

1. 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.

2. Coresh J, Turin TC, Matsushita K, et al. Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *JAMA*. 2014;311(24):2518-2531.

3. Lambers Heerspink HJ, Tighiouart H, Sang Y, et al. GFR decline and subsequent risk of established kidney outcomes: a meta-analysis of 37 randomized controlled trials. *Am J Kidney Dis*. 2014;64(6):860-866.

4. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996;125(7):605-613.

5. Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. *Biometrics*. 2009;65(2):530-538.

6. Stoycheff N, Pandya K, Okparavero A, Schiff A, Levey AS, Greene T. Early change in proteinuria as a surrogate outcome in kidney disease progression: a systematic review of previous analyses and creation of a patient-level pooled dataset. *Nephrol Dial Transplant*. 2010;26:848-857.

7. Inker LA, Lambers Heerspink HJ, Mondal H, 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*. 2014;64(6):848-859.

8. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982;38:963-974.

9. Klahr S, Levey AS, Beck GJ, et al. The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. *N Engl J Med*. 1994;330(13):877-884.

10. Wright JT Jr, Bakris G, Greene T, et al; and African American Study of Kidney Disease and Hypertension Study Group. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: results from the AASK trial. *JAMA*. 2002;288(19):2421-2431.

11. Hansen HP, Rossing P, Tarnow L, Nielsen FS, Jensen B, Parving HH. Elevated GFR after withdrawal of chronic antihypertensive treatment in diabetic nephropathy [abstract]. *J Am Soc Nephrol*. 1994;5:378.

12. Apperloo AJ, deZeeuw D, van Essen G, de Jong P. The initial GFR decline on anti-hypertensive treatment in long term interaction studies is reversible after withdrawal of treatment [abstract]. *J Am Soc Nephrol*. 1994;5:324.

13. Greene T. A model for a proportional treatment effect on disease progression. *Biometrics*. 2001;57(2):354-360.

14. Levey AS, Greene T, Beck GJ, et al. Dietary protein restriction and the progression of chronic renal disease: what have all of the results of the MDRD Study shown? *J Am Soc Nephrol*. 1999;10(11):2426-2439.

15. Prentice RL. A log gamma model and its maximum likelihood estimation. *Biometrika*. 1974;61(3):539-544.

16. Li L, Astor BC, Lewis J, et al. Longitudinal progression trajectory of GFR among patients with CKD. *Am J Kidney Dis*. 2012;59(4):504-512.

17. Chi EM, Reinsel GC. Models for longitudinal data with random effects and AR (1) errors. *J Am Stat Assoc*. 1989;84(406):452-459.

18. Bauer JH, Reams GP, Hewett J, Klachko D. A randomized, double-blind, placebo-controlled trial to evaluate the effect of enalapril in patients with clinical diabetic nephropathy. *Am J Kidney Dis*. 1992;20:443-457.

19. Slataper R, Vicknair N, Sadler R, Bakris GL. Comparative effects of different anti-hypertensive treatments of progression of diabetic renal disease. *Arch Intern Med*. 1993;153:973-980.

20. Bakris GL, Barnhill BW, Sadler R. Treatment of arterial hypertension in diabetic man: importance of therapeutic selection. *Kidney Int*. 1992;41:912-919.

21. Bauer JH, Reams GP. Renal protection in essential hypertension: how do angiotensin converting enzyme inhibitors compare with calcium antagonists. *J Am Soc Nephrol*. 1990;1:580-587.

22. Klein IH, Abrahams A, van Ede T, Hené RJ, Koomans HA, Ligtenberg G. Different effects of tacrolimus and cyclosporine on renal hemodynamics and blood pressure in healthy subjects. *Transplantation*. 2002;73(5):732-736.