Change in Albuminuria and GFR as End Points for Clinical Trials in Early Stages of CKD: A Scientific Workshop Sponsored by the National Kidney Foundation in Collaboration With the US Food and Drug Administration and European Medicines Agency

Andrew S. Levey, Ron T. Gansevoort, Josef Coresh, Lesley A. Inker, Hiddo L. Heerspink, Morgan E. Grams, Tom Greene, Hocine Tighiouart, Kunihiro Matsushita, Shoshana H. Ballew, Yingying Sang, Edward Vonesh, Jian Ying, Tom Manley, Dick de Zeeuw, Kai-Uwe Eckardt, Adeera Levin, Vlado Perkovic, Luxia Zhang, and Kerry Willis

The US Food and Drug Administration (FDA) and European Medicines Agency (EMA) are currently willing to consider a 30% to 40% glomerular filtration rate (GFR) decline as a surrogate end point for kidney failure for clinical trials of kidney disease progression under appropriate conditions. However, these end points may not be practical for early stages of kidney disease. In March 2018, the National Kidney Foundation sponsored a scientific workshop in collaboration with the FDA and EMA to evaluate changes in albuminuria or GFR as candidate surrogate end points. Three parallel efforts were presented: meta-analyses of observational studies (cohorts), meta-analyses of clinical trials, and simulations of trial design. In cohorts, after accounting for measurement error, relationships between change in urinary albumin-creatinine ratio (UACR) or estimated GFR (eGFR) slope and the clinical outcome of kidney disease progression were strong and consistent. In trials, the posterior median R^2 of treatment effects on the candidate surrogates with the clinical outcome was 0.47 (95% Bayesian credible interval [BCI], 0.02-0.96) for early change in UACR and 0.72 (95% BCI, 0.05-0.99) when restricted to baseline UACR > 30 mg/g, and 0.97 (95% BCI, 0.78-1.00) for total eGFR slope at 3 years and 0.96 (95% BCI, 0.63-1.00) for chronic eGFR slope (ie, the slope excluding the first 3 months from baseline, when there might be acute changes in eGFR). The magnitude of the relationships of changes in the candidate surrogates with risk for clinical outcome was consistent across cohorts and trials: a UACR reduction of 30% or eGFR slope reduction by 0.5 to 1.0 mL/min/1.73 m^2 per year were associated with an HR of ~0.7 for the clinical outcome in cohorts and trials. In simulations, using GFR slope as an end point substantially reduced the required sample size and duration of follow-up compared with the clinical end point when baseline eGFR was high, treatment effects were uniform, and there was no acute effect of the treatment. We conclude that both early change in albuminuria and GFR slope fulfill criteria for surrogacy for use as end points in clinical trials for chronic kidney disease progression under certain conditions, with stronger support for change in GFR than albuminuria. Implementation requires understanding conditions under which each surrogate is likely to perform well and restricting its use to those settings.

Complete author and article information provided before references.

Am J Kidney Dis. 75(1): 84-104. Published online August 28, 2019.

doi: 10.1053/j.ajkd.2019.06.009

© 2019 by the National Kidney Foundation, Inc.

Introduction

Although chronic kidney disease (CKD) is a substantial global public health problem, progression of CKD is usually slow and there are few specific symptoms until kidney failure occurs. There is general agreement that biomarkers will be needed to approve new drugs to slow the progression of CKD.[1-4] The 2 most widely studied biomarkers in CKD are glomerular filtration rate (GFR) and albuminuria, but there is controversy about their validity as surrogate end points for important clinical outcomes (often referred to as clinical end points) in clinical trials, especially in the early stages of CKD.[5-7]

Editorials, p. 4, 6, 9

The 2012 scientific workshop sponsored by the National Kidney Foundation (NKF) and US Food and Drug Administration (FDA) on GFR decline as an end point in CKD clinical trials showed strong relationships between change in GFR and kidney failure and mortality in observational studies, and based on these analyses and analyses from past clinical trials and simulations, participants at the workshop proposed that a confirmed 30% or 40% decline in GFR would be an acceptable surrogate end point in clinical trials in some circumstances.[8] These end points are less applicable at higher baseline GFRs and in the context of agents that cause an "acute effect" on GFR decline (an early treatment effect of the intervention that differs from the later treatment effect), making them less practical for drugs targeted at earlier stages of kidney disease and drugs with potential hemodynamic effects. Surmounting these limitations may involve examining changes in albuminuria (or proteinuria) as an earlier marker of kidney disease progression, assessing the rate of GFR decline (slope), and combined use of both these approaches.

On March 15 to 16, 2018, the NKF sponsored a follow-up workshop, "Change in Albuminuria and GFR as End Points for Clinical Trials in Early Stages of Chronic Kidney Disease," in collaboration with the FDA and European Medicines Agency (EMA).[9] The goal of the workshop was to evaluate surrogate end points for trials of kidney disease progression and improve understanding of change in albuminuria and GFR as measures of kidney disease progression in early stages of CKD (Box 1). The anticipated outcome of the workshop was a determination of whether albuminuria change and GFR slope have sufficiently strong relationships with clinical end points in CKD to be used as surrogate end points in clinical trials, especially in early stages of CKD.

In this article, we summarize the clinical, analytic, and regulatory context for the workshop; the methods, results, and conclusions of the data analyses; the proposals for surrogate end points based on changes in albuminuria and GFR and their potential application; and key points from the discussion. Detail for the data analyses is provided in separate publications.[10-15] Perspectives from the FDA, EMA, and patient representatives are provided in accompanying editorials.[16-18] Additional information about the workshop is included in Item S1. The analyses and the workshop were funded by the NKF; contributors to the NKF participated in the workshop, but the planning committee operated independently to prepare the conference and this report.

Context

Kidney Disease Outcomes and Measures

Albuminuria and GFR are widely accepted as measures of kidney damage and function (Table 1). Increased albuminuria and reduced GFR are criteria for the definition and classification of CKD; they are among the strongest risk predictors for complications of kidney disease, including progression to kidney failure, cardiovascular disease, and mortality; and change in either measure has biological plausibility as an end point for clinical trials.[19]

Both albuminuria and GFR are measures of glomerular function. Albuminuria is primarily a measure of the permeability of the glomerular capillary wall to macromolecules and increased albuminuria occurs earlier in the course of many causes of kidney disease than a decline in GFR.[20] Impaired uptake of proteins from tubular fluid may also cause increased albuminuria, and the presence of macromolecules in tubules may directly cause kidney damage.[21,22] Thus, the relationship between change in albuminuria and kidney disease progression may vary among different causes of kidney disease, and it is possible than an increase in albuminuria may not be on the path to kidney failure for all causes of kidney disease. Biological plausibility is greater for diseases that are characterized by increased albuminuria and for interventions in which reducing albuminuria is hypothesized to be one of the main mechanisms of action, such as agents that act on the renin-angiotensin system.

GFR is determined by the filtration pressure, surface area of the glomerular capillary wall, and its permeability to small solutes and water. GFR reflects the product of the number of nephrons and average single-nephron GFR.[23] GFR is generally considered the most useful overall measure of kidney function in health and disease, and the decline in other kidney functions often mirrors the decline in GFR. A severe reduction in GFR is defined as kidney failure; hence, by definition, GFR decline is on the path of progression to kidney failure for all kidney

Box 1. Goals and Aims for the Workshop

Goals: Evaluate surrogate end points for trials of kidney disease progression and improve understanding of change in albuminuria and GFR as measures of kidney disease progression in early stages of CKD

Review aims

1. Review data on pathophysiologic mechanisms by which albuminuria causes kidney damage and may be an appropriate target of therapy
2. Review methodologic and design issues in evaluating slope of GFR vs time as an outcome for clinical trials
3. Review laboratory issues regarding measurement of albuminuria and proteinuria that could affect interpretation of past trial analyses and future designs

Research aims

1. Examine associations of changes in UACR with subsequent adverse outcomes (ESKD and mortality) and examine consistency of associations across subgroups (level of UACR and GFR, disease, and intervention), as well as implications of measurement error
2. Examine associations of slope of GFR with subsequent adverse outcomes (ESKD and mortality) and examine consistency of associations across subgroups (level of UACR and GFR, disease, and intervention), as well as implications of measurement error
3. Examine associations of treatment effects on early change in UACR with treatment effects on established end points and consistency across subgroups (level of UACR and GFR, disease, and intervention)
4. Examine associations of treatment effects on GFR slope (acute, chronic, and total slope) with treatment effects on established end points and consistency across subgroups (level of UACR and GFR, disease, and intervention)
5. Develop methods to combine early change in UACR and GFR as end points

Abbreviations: CKD, chronic kidney disease; ESKD, end-stage kidney disease; GFR, glomerular filtration rate; UACR, urinary albumin-creatinine ratio.

Table 1. Kidney Disease Measures in CKD

	Albuminuria	GFR
Biological plausibility as an end point for clinical trials	Albuminuria is a marker of kidney damage and in some diseases may cause kidney disease progression	GFR decline is on the path of progression to kidney failure
Pathophysiology	AER = GFR * P * θ −TR Increased albuminuria in CKD generally reflects increased θ (permeability of glomerulus to macromolecules); macromolecules in tubular fluid may be harmful; in some diseases, albuminuria increase occurs before GFR decline	GFR = N * SNGFR Decreased GFR in CKD generally reflects decreased N; SNGFR may be increased due to hemodynamic alterations
Measurement	Standardization of urine albumin in progress; no standardization of urine total protein; AER or PER is the gold standard but complex and not usually performed; UACR or UPCR is simple and controls for urinary concentration and dilution, but introduces error due to variability in AER, PER, and creatinine excretion	GFR cannot be measured directly in humans; mGFR from clearance of an exogenous filtration marker is the gold standard, but complex and rarely performed; eGFR from serum concentration of an endogenous filtration marker is simple, but introduces error due to non-GFR determinants; standardized assays for creatinine and cystatin C exist
Use in clinical practice	Criterion for definition and classification of CKD; key risk predictor; not frequently assessed in clinical practice (but should be)	Criterion for definition and classification of CKD; key risk predictor; frequently assessed in clinical practice
Use as an end point in clinical trials	Available drugs improve albuminuria; remission of NS accepted as a valid surrogate end point in some diseases, but not smaller reductions in albuminuria; albuminuria change may be acute and reversible (reflecting functional changes) vs chronic and persistent (reflecting structural changes); difficult to verify assumptions in the clinical trial setting	Few drugs to improve GFR; preventing a large GFR decline (30%-57%) is accepted as a valid surrogate end point in some circumstances, but large declines are less frequent at higher baseline GFRs than at lower GFRs; GFR slope may have greater statistical power than time to GFR decline under certain conditions; GFR slope may be acute and reversible (reflecting functional changes) vs chronic and persistent (reflecting structural changes); assumed extrapolation of chronic slope to large decline; difficult to verify assumptions in the clinical trial setting

Abbreviations: AER, albumin excretion rate; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; GFR, glomerular filtration rate; mGFR, measured GFR; N, number of nephrons; NS, nephrotic syndrome; P, plasma concentration of albumin; PER, protein excretion rate; SNGFR, single-nephron GFR; θ, sieving coefficient for albumin; TR; rate of tubular reabsorption of albumin; UACR, urinary albumin-creatinine ratio; UPCR, urinary protein-creatinine ratio.

diseases, and it is more strongly related to the development of kidney failure and its complications than increased albuminuria.

Measurement methods have been simplified in recent years, and many recommendations for clinical practice are applicable for clinical trials.[24] For albuminuria, clinical practice guidelines recommend measurement of urinary albumin-creatinine ratio (UACR) in untimed ("spot") urine samples, preferably collected in the morning ("early morning sample"). Albumin is preferred rather than total urine protein because measurement can be standardized across clinical laboratories.[25] Possible limitations of UACR include imprecision due to variability in albumin excretion, failure to detect changes in nonalbumin proteins that may be of clinical importance, and bias due to interventions that affect creatinine excretion. An early morning urine sample may limit the impact of diurnal variation, and repeat measurements at baseline and important intervals during follow-up may improve precision. The importance of nonalbumin proteins is best addressed by specific assays for these proteins. Albumin excretion rate, the reference standard, may be required for interventions that may affect creatinine excretion.

For GFR, clinical practice guidelines recommend estimated GFR (eGFR) from serum concentration of creatinine ($eGFR_{cr}$), using a standardized assay and the CKD Epidemiology Collaboration (CKD-EPI) estimating equation.[26] Possible limitations of $eGFR_{cr}$ in clinical trials are imprecision, particularly at higher GFRs, and bias by interventions that affect non-GFR determinants of serum creatinine level (such as muscle mass). Precision may be improved by repeat determinations of $eGFR_{cr}$ at baseline and important intervals during follow-up. Possibly, bias in $eGFR_{cr}$ may be overcome by estimation of GFR from serum concentration of cystatin C ($eGFR_{cys}$), which has also been standardized in recent years. eGFR based on the combination of creatinine and cystatin C ($eGFR_{cr\text{-}cys}$) is generally more accurate than either alone. Use of any alternative filtration marker in a clinical trial would require evaluation of whether the intervention could affect the non-GFR determinants of the marker. Clearance measurements with exogenous filtration markers remain the reference standard if eGFR is not appropriate.

Previous work has examined the validity of change in albuminuria for use as a surrogate end point in clinical trials, but definitive conclusions have not been reached and ongoing debate has highlighted the controversy.[27-32] At higher GFRs, a trial design to compare mean rate of GFR decline versus time between randomized groups (mean slope analysis) may have greater statistical power and may

be more feasible than comparison of time to a GFR decline (time-to-event analysis).[33] However, analyses of GFR slope defined from a prerandomization baseline can be misleading and may have severe reductions in statistical power when the treatment has a substantial negative acute effect. Design strategies proposed to overcome these limitations include evaluation of the chronic slope (ie, the slope excluding the first 3 months from baseline, when there might be acute changes in eGFR) rather than the total slope from randomization. However, there is no generally accepted method.[34,35] Unlike time-to-event end points based on a large decline in eGFR, which can be easily used as outcomes to assess response at the individual level, change in albuminuria and GFR are evaluated here for use as average measures for comparing treatment groups.

End Points for Clinical Trials

The FDA and EMA use similar end point definitions for clinical trials but offer different regulatory pathways for drug approval (Table 2). A clinical outcome is defined based on how a person feels, functions, or survives. By contrast, a surrogate end point is expected to support accurate prediction of effects of treatment interventions on the clinical outcome and can be classified by the level of clinical validation as a "validated," "reasonably likely," or "candidate" surrogate.[36-38] Use of clinical outcomes or validated surrogates can lead to traditional approval (FDA) or full marketing authorization (EMA). Use of reasonably likely valid surrogates can lead to accelerated approval by the FDA but is limited to defined circumstances and generally requires a postmarketing confirmatory trial.[39,40]

European Union pharmaceutical legislation does not mention "reasonably likely surrogates" and the type of end point for efficacy demonstration per se is not a guiding principle for the type of marketing authorization. Use of reasonably likely validated surrogates can lead to full marketing authorization by the EMA, potentially requiring certain postauthorization commitments.[41] Conditional marketing authorization is limited to defined circumstances, including postauthorization commitments. Both agencies strongly encourage investigators and sponsors to discuss the acceptability of end points on a case-by-case basis before beginning the clinical trial.

Patient Perspective

Early detection and early intervention are essential to improving patient outcomes in many diseases. Unlike many other diseases, CKD may be asymptomatic until late in the course when symptoms of kidney failure arise, and the disease may be irreversible. Thus, candidate surrogate end points, such as changes in albuminuria or GFR, take on more important meaning in CKD than other diseases, which is consistent with their importance in clinical practice, research, and public health. For many patients, especially those who are identified early in their disease, preventing disease progression is critically important, and change in albuminuria and GFR represent the disease itself (Box 2).

Overview of Methods

Criteria for Surrogacy and Framework for Analyses

Criteria for surrogacy include biological plausibility, strength and consistency of relationships in epidemiologic data (observational studies), and prediction of the treatment effects on the clinical outcome in clinical trials (Table 3).[42]

Table 2. End Point Definitions for Clinical Trials and Regulatory Pathways for Drug Approval

Term	Definition (similar for FDA and EMA)	
Clinical outcome	How a person feels, functions, or survives	
Surrogate end point	Expected to predict clinical benefit or harm; characterized by the level of clinical validation as a validated, reasonably likely, or candidate surrogate	
Regulatory Pathway	**FDA Approval Pathways**	**EMA Centralized MA Procedures**
Usual pathway	Traditional approval: approval based on a clinical outcome/clinically meaningful end point or a validated surrogate	Full ("standard") MA: benefit/risk assessment based on comprehensive evidence base; benefit demonstrated by showing a clinically relevant effect, using end point(s) representing clinical or surrogate outcomes
Alternative pathway	Accelerated approval: approval based on an effect on a reasonably likely surrogate; product must be for a serious or life-threatening disease or condition and provide a meaningful advantage over available therapies. Postmarketing confirmatory trials are generally required using a clinical outcome or validated surrogate.	Conditional (CMA): in case benefit of immediate availability outweighs the risk for less comprehensive data than usually required; categories that may be eligible: treatment, prevention, or diagnosis of seriously debilitating or life-threatening diseases; products to be used in emergency situations or in response to public health threats; orphan medicinal products. For CMA, all these criteria have to be met: unmet medical need fulfilled, benefit/risk balance at time of assessment is positive, it is likely that comprehensive data can be provided via postauthorization commitments

Abbreviations: CMA, conditional marketing authorization; EMA, European Medicines Administration; FDA, US Food and Drug Administration; MA, marketing authorization.

Box 2. Patient Perspective From the National Kidney Foundation Kidney Advocacy Committee

"As people who have lived with and progressed through the stages of kidney disease, we often wonder how the course of our disease and lives may have been altered by earlier detection and treatment. Kidney diseases can be identified early but there are very few treatments at the earlier stages, so many clinicians choose not to pay close attention to the diagnosis of CKD at earlier stages. We want to take medications that have proven to be effective at preventing progression of kidney disease. Later stages of kidney disease, even before the development of kidney failure, have consequences for us and for our families."

Abbreviation: CKD, chronic kidney disease.

Classifying the strength of evidence for these criteria into categories can be difficult. Currently, no FDA guidance document provides a detailed description of the evidence needed to establish a "validated surrogate end point" for use in the traditional approval pathway; however, the FDA has stated that the standard is high. We accepted the biological plausibility for albuminuria change and GFR decline as surrogate end points for kidney failure and designed an analytic plan for evaluation for the remaining criteria.

For our analyses, we used data from prospective cohorts to evaluate the strength of the associations of candidate surrogates with clinical end points and data from clinical trials to evaluate the strength of the association of treatment effects of interventions on candidate surrogates with their effects on clinical end points. Because criteria for surrogacy include a consideration of a wide range of clinical circumstances and interventions, we attempted to include all available studies, irrespective of cause and stage of kidney disease and intervention.

For both cohorts and trials, we applied a 2-step approach whereby each study was first analyzed separately, followed by a random-effects meta-analysis of all studies. For GFR-based end points, we also conducted simulations based on data from past trials under alternative study designs and different assumptions about the short- and long-term effects of the treatment. For scenarios with beneficial treatment effects on time to the clinical outcome, we compared the simulated statistical power for candidate and validated surrogate end points and the clinical outcome. For scenarios with null treatment effects on the time to the clinical outcome, we compared the simulated risk for false conclusion of clinical benefit or harm when the analysis is based on the surrogate end point. False conclusions of benefit or harm based on an analysis of the surrogate when there is no effect on the clinical end point represent an extension of the concept of type 1 error to the surrogate end point setting, which we refer to as type 1 errors of the surrogate end point relative to the clinical outcome.

The strengths of analyses of cohorts are the long duration of follow-up and large sample size, enabling more accurate assessment of associations than in trials. The principal limitation of the analyses of cohorts is that it predicts the clinical outcome from the candidate surrogate within individual patients, but not whether treatment effects on the surrogate can predict treatment effects on the clinical outcome. Variability among individuals and measurement error in the surrogate may lead to attenuation of the association with the clinical outcome due to regression dilution, which may be accounted for in the analyses. However, accurate prediction of the clinical outcome may not translate to accurate prediction of treatment effects on the clinical outcome if the association between the surrogate and clinical outcome results from confounding factors extraneous to the treatment and if treatments affect the clinical outcome through pathways distinct from the surrogate. The cohort analyses are also challenged by variations in study design and the possibility that commonly used treatments may not be recorded.

The main strength of the analyses of trials is that the intention-to-treat analysis directly evaluates the accuracy with which treatment effects on the surrogate end point predict treatment effects on the clinical end point, which is the fundamental issue for the application of the surrogate end point in clinical trials. The analyses of trials overcome some of the limitations of the analyses of cohorts by relating estimated treatment effects on the surrogate and clinical end points, which are each based on intent-to-treat comparisons of randomized groups. However, there are also limitations to the analyses of trials, including imprecision of estimates of treatment effects on the surrogate end points and clinical end points due to small sample size in many trials. Limitations in statistical power limit the ability to determine whether relationships demonstrated in the full set of trials apply also for specific treatments or study populations. The prediction of the clinical end point is limited to the follow-up of the individual trials, and our analyses of trials are unable to address whether treatment effects on the surrogate end points accurately predict longer-term effects of the treatment on future clinical events.

The main strength of the simulations is their ability to compare the statistical power provided by the GFR-based surrogate end points and the clinical outcome under a wide range of study designs, population characteristics, and assumptions concerning the short- and long-term effects of the treatment. The simulations thus augment the information provided by overall measures of association from the analyses of cohorts and trials by determining the specific conditions under which different GFR-based surrogates can be used to shorten follow-up or reduce the sample size while maintaining adequate statistical power without inflating risk for false-positive conclusions. The simulations are useful to show the effects of variation in intervention, population characteristics, and study design and can help explain observations in past clinical trials. The

Table 3. Criteria for Evaluation of Surrogacy of Change in Albuminuria and GFR and Framework for Analysis

Criteria and Analysis	Comment
Criterion: biological plausibility	Whether surrogate is on pathophysiologic pathway leading to clinical outcome of interest (causal or necessary intermediate)
Analysis: synthesize data from cellular, animal, and human studies on causal mechanisms linking change in albuminuria and GFR to CKD progression and risk for ESKD	Important but difficult to integrate information from multiple sources; this workshop accepted a summary by experts supporting biological plausibility for both changes in albuminuria and in GFR
Criterion for analysis in cohorts: strength and consistency of epidemiologic data	Supporting relationship between candidate surrogate and clinical outcome of interest
Analysis: Regression of clinical end point vs UACR change or eGFR slope	
Relative risk for ESKD (HR), before and after accounting for measurement error	Consistency across cohorts and subgroups (BL eGFR and UACR)
Absolute risk for ESKD (risk reduction)	Anticipate larger absolute risk reduction with longer F/U interval, higher BL eGFR, and BL UACR
Criterion for analysis in trials: prediction of treatment effects on the clinical outcome of interest from treatment effects on the surrogate	With drugs in the same/related pharmacologic class? With drugs from distinct pharmacologic classes/regardless of the mechanism of the intervention?
Analysis: regression of treatment effect on clinical end point (HR) vs treatment effect on UACR change or eGFR slope	
Significant slope, nonsignificant intercept, high R^2, and low RMSE for regression	Consistency across subgroups (BL eGFR and UACR), anticipate less power to evaluate subgroups by disease and intervention
Point estimate for predicted benefit on end point	Magnitude of point estimate for predicted benefit on clinical end point predicted by an observed effect on candidate surrogate
Threshold for the minimum observed benefit on the surrogate required to provide a high PPV (eg, 97.5%) for a nonzero benefit on the clinical outcome ($PPV_{0.975}$)	Magnitude of point estimate for observed effect on candidate surrogate to be greater than threshold, varies with sample size, contingent on similarity of future RCTs with RCTs in the analysis
Analysis: simulations for eGFR-based surrogate end points (time to eGFR decline and eGFR slope)	
Increased power compared with clinical outcome in scenarios with beneficial treatment effect	Identify patient characteristics and trial conditions that are favorable (and unfavorable) for surrogate end points vs clinical outcome
Preserved type 1 errors for the surrogate end point relative to the clinical outcome in scenarios with null treatment effects	

Note: Changes in albuminuria and GFR are considered here for use as average measures comparing 2+ groups, not at the individual level. Criteria and comments apply to both change in albuminuria and GFR, unless otherwise noted. We transformed measures of albuminuria and proteinuria as UACR and UPCR, expressed as mg/g, using established conversions, and calculated eGFR from Scr level, expressed as mL/min/1.73 m^2 body surface area, using the CKD-EPI equation and standardized Scr level when available or an established conversion for nonstandardized Scr level. Change in UACR or UPCR was expressed on the log scale (GMR, with a value < 1.0 corresponding to an improvement and a value > 1.0 corresponding to worsening). Because albuminuria may change soon after an intervention, our primary interest was an early change. For cohorts, not all participants had frequent measurements of albuminuria, so we evaluated changes over 1, 2, and 3 years and accounted for measurement error using regression dilution from data from selected cohorts with repeated measures. For trials, we evaluated early changes in albuminuria (from randomization to ~6 or 12 months in follow-up). Change in eGFR was expressed as slope in mL/min/1.73 m^2 per year. Differences in slope were expressed on the raw scale rather than the proportional scale because this has been the practice in most prior clinical trials, ratios of means can be unstable when the denominator is small, and statistical distribution theory applies better for differences in means than for ratios of means. For cohorts, we computed slope using linear regression over 1, 2, and 3 years and accounted for measurement error by using mixed models. For clinical trials, we computed acute slope (from randomization to ~3 months in follow-up), chronic slope (from 3 months to end of the trial), and total slope (from randomization to 1, 2, 3, or 4 years) using a simplified linear mixed-effects model based on a single slope starting at 3 months follow-up while adjusting for BL eGFR.[15] Additional methods were used to account for between-participant variability in eGFR trajectories, variability in individual eGFR assessments, informative censoring by ESKD and death, and uniform versus proportional long-term treatment effects (defined as treatment effects that are independent or proportional to the underlying rate of progression, eg, similar for fast and slow progressors or larger in fast progressors than in slow progressors). We also conducted preliminary analyses using combinations of UACR change and eGFR slope as a candidate surrogate end point, but these are not reported here.

The kidney disease clinical outcome in cohorts was ESKD, defined as initiation of kidney replacement therapy (long-term dialysis or kidney transplantation). For some analyses, we also considered mortality because it is an important clinical end point and a competing event for kidney disease events. In trials, there were fewer ESKD events, so the clinical outcome was a composite outcome including ESKD, eGFR < 15 mL/min/1.73 m^2 (the usual definition of kidney failure), or confirmed doubling of Scr level (equivalent to a 57% decline in eGFR), and in some analyses a confirmed 40% decline in eGFR. For simulations, we defined the composite outcome as a 57% reduction in GFR or ESKD, in which ESKD was assumed to occur when GFR crossed a random threshold between 6 and 15 mL/min/1.73 m^2. We examined the same methods for slope and time-to-event analysis in the simulations that were used for the analyses of the real data from the clinical trials. For cohorts and trials, we defined subgroups based on BL eGFR, UACR, cause of disease, intervention (for trials, agents that act on the renin-angiotensin system vs others), and other clinical characteristics if sufficient data were available.

To take into account the precision of the estimate of the treatment effect, we computed $PPV_{0.975}$ for an RCT of infinite, large, or modest size. An infinite sample size would provide the "true" effect. For UACR change, a large RCT was defined as one in which the treatment effect can be estimated to within an SE of 0.05, corresponding to a total sample size (N) of approximately 1,090. A modest-sized RCT was defined as having SE of 0.12 (N ~ 190). For GFR slope, a large RCT was defined as one in which the treatment effect can be estimated to within an SE of 0.25 (N ~ 1,900 for RCTs for which average follow-up accorded with the RCTs in the analyses). A modest-sized RCT was defined as having SE of 0.4 (N ~ 720).

Abbreviations: BL, baseline; CKD, chronic kidney disease; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration; eGFR, estimated glomerular filtration rate; ESKD, end-stage kidney disease; F/U, follow-up; GMR, geometric mean ratio; HR, hazard ratio; RCT, randomized controlled trial; RMSE, root mean squared error; Scr, serum creatinine; SE, standard error; UACR, urinary albumin-creatinine ratio; UPCR, urinary protein-creatinine ratio.

Table 4. Summary of Studies, Participants, and Outcomes

	Studies for Analysis of Albuminuria Change		Studies for Analysis of GFR Slope		
	Cohorts (for 2-y BL period)	Clinical Trials	Cohorts (for 2-y BL period)	Clinical Trials	Simulations
Studies	28 cohorts total: 20 for ESKD outcomes, 25 for mortality outcomes	41 treatment comparisons	14 cohorts total: 14 for ESKD outcomes, 14 for mortality outcomes	47 treatment comparisons	20 input parameters derived from 47 treatment comparisons
Participants	693,816 participants total: 557,583 with DM (80%), 675,904 for ESKD outcomes, 690,513 for mortality outcomes	29,979 participants categorized by 3 causes of CKD • DM and DKD: 21,102 (71%) • Glomerular diseases: 1,325 (4%) • Other and unspecified causes: 7,552 (25%)	3,881,215 participants total: 3,881,215 for ESKD outcomes, 3,881,215 for mortality outcomes	60,620 participants categorized by 3 causes of CKD • DM and DKD: 43,481 (72%) • Glomerular diseases: 1,389 (2%) • Other and unspecified causes: 15,750 (26%)	225 parameter configurations; for each parameter configuration, 800 data sets consisting of 500 participants (250 assigned to treatment and 250 to control)
BL eGFR (mL/min/1.73 m^2)[a]	78 ± 21	58.2 ± 25.0	47 ± 10 in eGFR < 60 stratum 87 ± 19 in eGFR ≥ 60 stratum	61.7 ± 26	Not applicable
BL UACR (mg/g)[b]	11 [5-33]	272 [30-1,134]	Not reported	60 [13-554]	Not applicable
Outcomes	7,461 ESKD events, 75,761 deaths	3,935 composite events (ESKD, eGFR < 15, Scr doubling)	12,635 ESKD events, 564,196 deaths	7,115 composite events (ESKD, eGFR < 15, Scr doubling)	5 outcomes per simulated data set (3-y total slope; chronic slope; composite end point for ESKD [GFR 6-15]; and confirmed 30%, 40%, or 57% GFR decline)

Note: Data from cohorts were collected by the CKD-PC, which consists of more than 70 cohorts across more than 40 countries, with data for Scr level and albuminuria and outcomes.[44] For the analyses of albuminuria change,[10] cohorts that included a repeat measure of albuminuria during an elapsed period of 8 months to 4 years and subsequent ESKD or mortality F/U were invited to participate. Cohorts with urine albumin and urine protein as the measure of albuminuria were analyzed separately. For analyses of GFR slope,[12] cohorts with at least 3 eGFR assessments in an initial 2-year observation period and subsequent longitudinal F/U for ESKD were invited to participate. We stratified analyses by BL eGFR, conducting separate meta-analyses for individuals with eGFRs < 60 and ≥60mL/min/1.73 m^2. Cohorts could contribute to both meta-analyses if they had sufficient numbers of individuals who developed ESKD (>10 events) within the given eGFR category.
Data from clinical trials were collected previously by CKD-EPI and the REASSURE Consortium, and newly collected for this workshop.[27,28,45] For the albuminuria change analyses,[11] key inclusion criteria were biological plausibility of albuminuria change as a surrogate end point for the intervention, quantifiable measurements of albuminuria or proteinuria at BL and within 12 months of F/U, and information on ESKD incidence. For the GFR slope analyses,[13] key inclusion criteria were Scr measurements at BL and at F/U at 12 months or earlier, and at least 12 months' F/U after that second measurement. For both analyses, small studies (n < 100) were pooled if the disease and intervention were the same. For trials that evaluated more than 1 intervention, a separate group for each independent treatment comparison was included, such that some participants were included in more than 1 analytical comparison. Subgroups were defined by average study level of BL UACR (<30 or ≥30 mg/g), eGFR (<60 or ≥60 mL/min/1.73 m^2), cause of disease (DM and DKD, glomerular diseases, or other or unspecified causes of CKD), and intervention.
For simulations,[14] a total of 20 input parameters were modeled, including rates and distributions of GFR declines, magnitudes of acute effects, patterns of long-term treatment effect, types of study design, rates of mortality and missing data, and relationship of GFR to initiation of maintenance dialysis or kidney transplantation. For consistency with an earlier simulation study evaluating 30% and 40% GFR decline, data analyses for determination of most input parameters were based on 14 trials from the CKD-EPI data set (above) with at least 1 year of eGFR F/U in at least 100 participants.[46] Additional analyses in the full set of 47 treatment comparisons were used to update the values considered for key parameters defining the mean and SD of GFR slopes, the acute effect, and nature of the long-term treatment effect.
Abbreviations: BL, baseline; CKD, chronic kidney disease; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration; CKD-PC, Chronic Kidney Disease Prognosis Consortium; DKD, diabetic kidney disease; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate (in mL/min/1.73 m^2); ESKD, end-stage kidney disease; F/U, follow-up; REASSURE, Reducing Albuminuria as Surrogate Endpoint; Scr, serum creatinine; SD, standard deviation; UACR, urinary albumin-creatinine ratio.
[a]Values expressed as mean ± SD.
[b]Values expressed as median [interquartile range].

chief weakness of the simulations results from the multifactorial nature of GFR trajectories and the fact that investigators will usually have substantial uncertainties concerning the mean rate and variability of GFR progression and about the nature of short- and long-term effects of the treatment when designing a specific randomized trial. These uncertainties translate to uncertainties in the assumptions that should be used in any given simulation and thus in evaluation of sample sizes and follow-up times that will be required for the different end points. Because the relationship between UACR trajectories and the clinical outcomes are not as well defined as for GFR, the simulations are more valuable for end points based on GFR than for albuminuria.

Variables and Evidence Synthesis

We considered change in albuminuria and GFR as candidate surrogate end points and various measures of kidney failure and previously validated surrogate end points for kidney failure as the clinical outcomes. Overall, we summarized evidence as shown in Table 3. In cohorts, we evaluated the relative risk for end-stage kidney disease (ESKD) for UACR change or eGFR slope before and after accounting for measurement error and consistency of findings across

Table 5. Summary of Evidence for Albuminuria Change

Analysis	Findings	Comment
Cohorts		
2-y change	Median UACR or UPCR fold increase: 1.12 (IQR, 0.61-2.17)	
Relative risk for ESKD	HRs of 0.83 (0.74-0.94) and (adjusted for measurement error) 0.78 (0.66-0.95) for 30% UACR reduction over 2 y; similar findings for 30% UPCR reduction	Stronger at higher BL UACR, consistent across cohorts and subgroups (BL eGFR, DM status)
Prediction of absolute risk for ESKD	1%-2% absolute risk reduction at 10 y for 30% UACR reduction when BL UACR > 300 mg/g and GFR = 60 mL/min/1.73 m^2	Higher for lower BL eGFRs
Trials		
6-mo change in treatment and control groups	Reduction in UACR GMR: 34% (27%-40%) in treatment group, 16% (8%-24%) in control group	
Treatment effect[a]	Reduction in UACR GMR: 22% (18%-26%)	
Regression of treatment effect on clinical outcome (HR) vs UACR reduction (GMR)	Slope significant, intercept nonsignificant, R^2 = 0.47 (BCI, 0.02-0.96); 30% UACR GMR reduction corresponds to 27% (BCI, 5%-45%) lower average risk for clinical outcome; stronger relationship in participants with BL UACR ≥ 30 g/mg: R^2 = 0.72 (BCI, 0.05-0.99)	Consistency across BL eGFR, cause of disease, and intervention, or for 12-mo change in UACR, but insufficient power for definitive evaluation
Prediction for a new trial restricted to BL UACR ≥ 30 mg/g	21%-27% $PPV_{0.975}$	Threshold is lower for a large trial than for a modest-size trial

Note: Albuminuria change is used as an average for comparing 2 or more groups, not at the individual level. Simulations section not shown because it is not applicable. Values in parentheses for HRs and GMRs are 95% confidence intervals.
Abbreviations: BCI, Bayesian credible interval; BL, baseline; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; ESKD, end-stage kidney disease; GMR, geometric mean ratio; HR, hazard ratio; IQR, interquartile range; $PPV_{0.975}$, threshold UACR reduction to provide ≥97.5% PPV for nonzero benefit on clinical outcome; UACR, urinary albumin-creatinine ratio; UPCR, urinary protein-creatinine ratio.
[a]The percent reduction in GMR in the treatment versus control group.

cohorts and subgroups and used these data to predict absolute risk for ESKD. In trials, we evaluated the regression of treatment effects on the clinical outcome versus treatment effects on UACR or eGFR slope and consistency of findings across cohorts and subgroups and used these data for predictions for future trials, including thresholds (discussed next). We used the designations of low, moderate, and strong trial-level association as defined by $R^2 < 0.49$, 0.49-0.72, and ≥0.72, respectively.[43] In simulations, we determined conditions in which GFR-based surrogates (time to designated declines in GFR or GFR slope) increased statistical power compared to the clinical outcome, allowing shorter follow-up or reduced sample size in scenarios with beneficial treatment effects on time to the clinical outcome, and preserved a low risk of type 1 error for the surrogate end point relative to the clinical outcome in scenarios with null treatment effects on time to ESKD.

Threshold for Prediction of a Beneficial Effect

Because of uncertainty in the relationship of changes in surrogate end points to subsequent clinical outcomes, we used the trial-level metaregression to determine thresholds for the minimum observed benefit on the surrogate required to provide a high (97.5%) probability for a nonzero benefit on clinical outcome; threshold probabilities are denoted as $PPV_{0.975}$. Effects of the intervention on the surrogate larger than $PPV_{0.975}$ provide high confidence that the intervention will have a significant effect on the clinical outcome. Determination of the threshold for the treatment effect on the surrogate end point provides a basis for limiting the use of the surrogate to settings in which there is stronger evidence of validity.

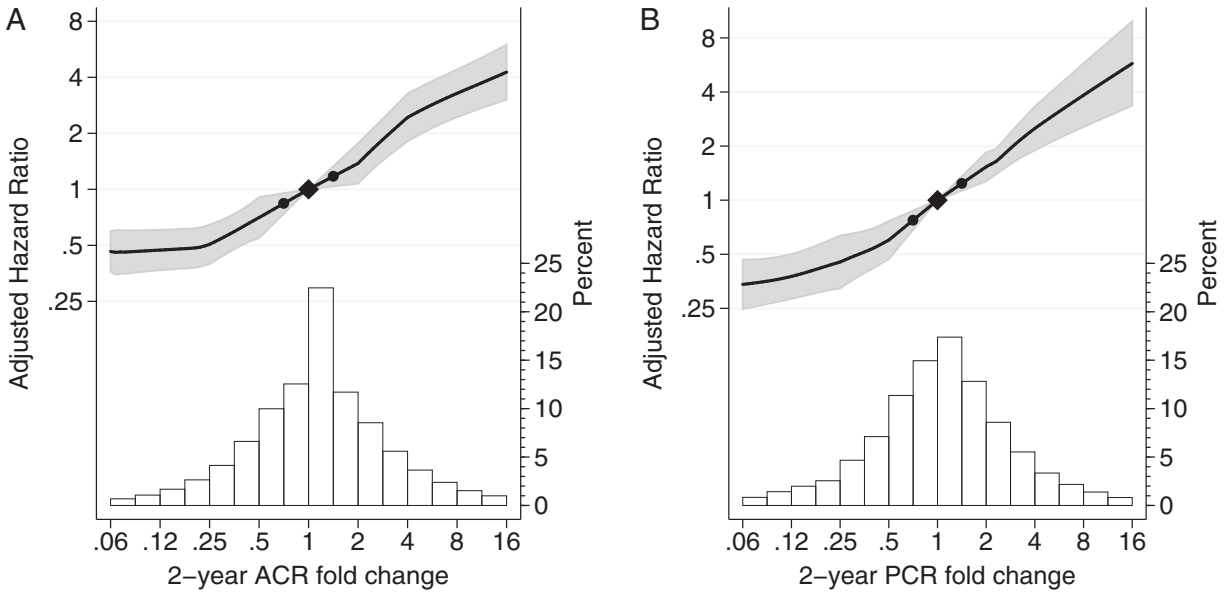
Sources of Data

Data from cohorts on albuminuria change[10] and GFR slope[12] in cohorts were collected by the CKD Prognosis Consortium (Table 4).[44] Data from clinical trials on albuminuria change[11] and GFR slope[13] were collected previously by CKD-EPI and the Reducing Albuminuria as Surrogate Endpoint (REASSURE) Consortium and newly collected for this workshop.[27,28,45] Data from clinical trials were used in simulations.[14,46]

Results and Interpretation

Albuminuria Change

Results from analyses of albuminuria change are summarized in Table 5. In cohorts,[10] a 30% UACR reduction over a 2-year interval was associated with a multivariable adjusted hazard ratio (HR) for subsequent ESKD of 0.83 (95%



C

ESKD Risk Associations With Albuminuria Change Over 1, 2, and 3-y Baseline Periods: ESKD HR for 30% ACR or PCR Decrease		
	Empirical*	Adjusted for Regression Dilution (Median Reliability)**
ACR		
1-year	0.82 (0.74-0.91)	0.75 (0.64-0.87)
2-year	0.83 (0.74-0.94)	0.78 (0.66-0.92)
3-year	0.80 (0.71-0.90)	0.76 (0.65-0.87)
PCR		
1-year	0.86 (0.76-0.97)	0.80 (0.67-0.95)
2-year	0.77 (0.68-0.87)	0.69 (0.58-0.83)
3-year	0.74 (0.61-0.89)	0.68 (0.54-0.86)

Figure 1. Analysis in cohorts: associations between population distribution of change in albuminuria and end-stage kidney disease (A and B). Adjusted hazard ratio (HR) for end-stage kidney disease and population distribution of change in albuminuria over a 2-year baseline period measured by (A) urinary albumin-creatinine ratio (ACR) and (B) urinary protein-creatinine ratio (PCR). Central diamond is the reference point (HR = 1, change in albuminuria = 1) and circles denote −30% and +43% change in albuminuria. Lines indicate point estimates and shaded areas 95% confidence intervals. (C) ESKD risk associations with albuminuria change over 1, 2, and 3-year baseline period.*Adjusted for age, sex, race/ethnicity (blacks vs nonblacks), systolic blood pressure, total cholesterol level, diabetes, history of cardiovascular disease, current smoking, former smoking, and first estimated glomerular filtration rate and albuminuria. **Based on estimates for ACR and PCR in 19 studies. Median (interquartile range; 25th to 75th percentile) reliability estimates (λ) for 1-, 2-, and 3-year change were 0.677 (0.533-0.770), 0.721 (0.650-0.808), and 0.789 (0.713-0.852). The same reliability estimates were used for ACR and PCR. Adapted from Coresh et al[10] with permission of Elsevier; original material © 2019 Elsevier.

confidence interval [CI], 0.74-0.94), which became stronger after adjustment for regression dilution for measurement error (0.78 [95% CI, 0.66-0.92]; Fig 1]. Similar results were observed for 30% UACR reductions over 1 and 3 years and for 30% UPCR reductions. The results were stronger at higher baseline UACRs (P interaction < 0.05) and null at UACRs < 30 mg/g and consistent across cohorts and across subgroups stratified by baseline eGFR and diabetes status. The absolute risk reduction depends on level of UACR, eGFR, and length of follow-up. For persons with UACR of 300 mg/g or UPCR of 500 mg/g and baseline eGFR of 60 mL/min/1.73 m^2, a true reduction in UACR of 30% over 2 years was estimated to confer a 1% absolute reduction in 10-year ESKD risk (from ∼5% to 4%).

In trials,[11] the association of treatment effect on UACR over 6 months with the treatment effect on the clinical end point had a significant regression slope and nonsignificant intercept and a posterior median squared correlation (R^2) of 0.47 (95% Bayesian credible interval [BCI], 0.02-0.96; Fig 2, left panel). Each 30% larger treatment effect on the geometric mean ratio for UACR in the treatment group compared to the control group was associated with an

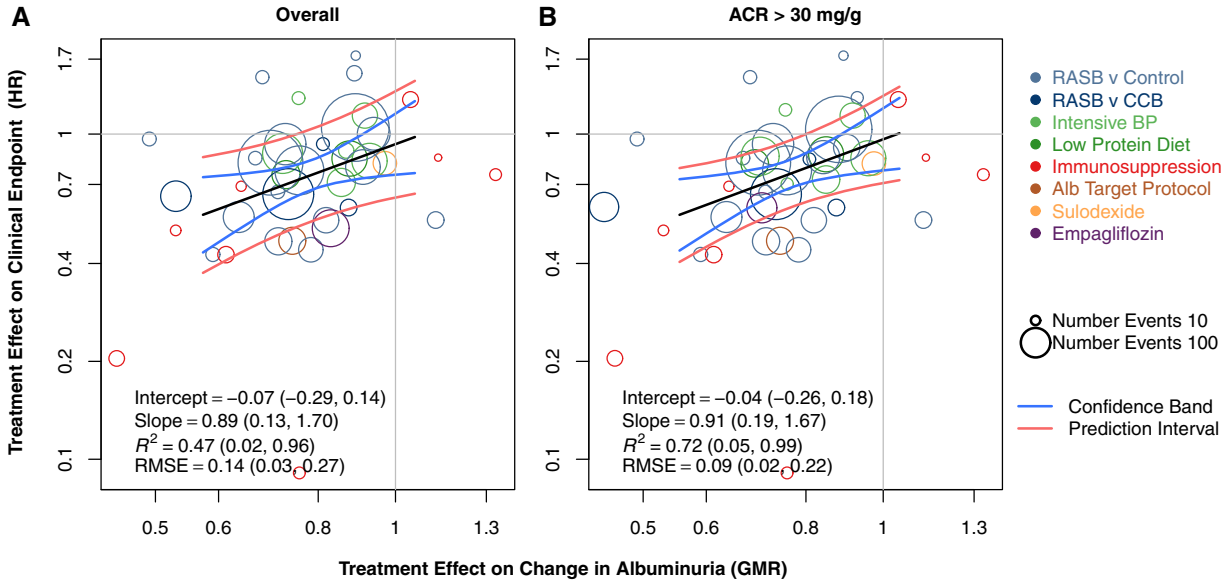


Figure 2. Trial-level analyses: association between treatment effects on change in albuminuria and treatment effects on the clinical end point for (A) the pooled population and (B) participants with baseline urinary albumin-creatinine ratio (ACR) > 30 mg/g. The vertical axes are the estimated treatment effects on the clinical end point (hazard ratio [HR]) and the horizontal axes are the estimated treatment effects on 6-month change in albuminuria (geometric mean ratio [GMR] of log-transformed ACR). The composite clinical end point was end-stage kidney disease, doubling of serum creatinine concentration, or estimated glomerular filtration rate < 15 mL/min/1.73 m^2. The different colored circles indicate intervention types, and each circle is a separate intervention with the size of the circle proportional to the number of events. The black line is the line of regression through the studies. The blue line is the confidence band. The pink lines are the prediction bands computed from the model. (A) The posterior median squared correlation (R^2) of 0.47 (95% Bayesian credible interval [BCI], 0.02-0.96) corresponds to Bayesian probabilities of 0.53, 0.28, and 0.19 for the R^2 falling into low, moderate, or high ranges. (B) The R^2 = 0.72 (95% BCI, 0.05-0.99) corresponds to Bayesian probabilities of 0.27, 0.24, and 0.49 for the R^2 falling into low, moderate, or high ranges. Abbreviations: Alb, albuminuria; BP, blood pressure; CCB, calcium channel blocker; RASB, renin–angiotensin system blocker; RMSE, root mean squared error. Adapted from Heerspink et al[11] with permission of Elsevier; original figure © 2019 Elsevier.

average 27% (95% BCI, 5%-45%) lower hazard for the clinical end point. The association appeared to be stronger when analyses were restricted to patients with baseline UACR > 30 mg/g (median R^2 = 0.72 [95% BCI, 0.05-0.99]; Fig 2, right panel). No clear differences were observed for subgroups defined by baseline eGFR, cause of disease, intervention, or when 12-month changes in UACR were analyzed, but power was limited to detect such differences. For future trials of participants restricted to baseline UACR ≥ 30 mg/g, the $PPV_{0.975}$ threshold treatment effect on geometric mean UACR required to establish a 0.975 probability of a nonzero benefit on the clinical end point was a 21% reduction for a large trial and 27% reduction for a modest-size trial. Somewhat larger hypothesized effect sizes would be required for the future trial to provide favorable statistical power for these thresholds to be achieved.

Overall, these 2 very different analyses and populations provided results that align both qualitatively and quantitatively (Table 5). Together they extend previous epidemiologic and trial-level analyses and support the role of early change in albuminuria as surrogate end point in randomized controlled trials (RCTs) in patients with higher baseline albuminuria. A relatively large treatment effect on UACR (a 20%-30% observed difference in geometric mean) is necessary to provide high confidence in a nonzero benefit of the treatment on the clinical outcome.

GFR Slope

Results from analyses of GFR slope are summarized in Table 6. In cohorts,[12] a reduction in eGFR slope by 0.75 mL/min/1.73 m^2 over 2 years was protective for ESKD (multivariable-adjusted HRs of 0.79 [95% CI, 0.77-0.81] and 0.84 [95% CI, 0.82-0.87], respectively; Fig 3). Associations were stronger when adjusted for measurement error (adjusted HRs of 0.71 [95% CI, 0.69-0.74] and 0.70 [95% CI, 0.68-0.72], respectively). Associations were weaker when observed over 1 year and stronger when observed over 3 years. Results were consistent across cohorts and across subgroups stratified by baseline UACR or diabetes status. The reduction in absolute risk for ESKD associated with a reduction in eGFR slope of 0.75 mL/min/1.73 m^2 per year was larger with longer follow-up and with faster expected eGFR decline. For a hypothetical population with a mean eGFR of 75 mL/min/1.73 m^2 and mean eGFR slope of −5 ± 4 (standard deviation) mL/min/1.73 m^2 per year, an intervention that reduced eGFR slope by 0.75 mL/min/1.73 m^2 per year would be

Table 6. Summary of Evidence for GFR Slope

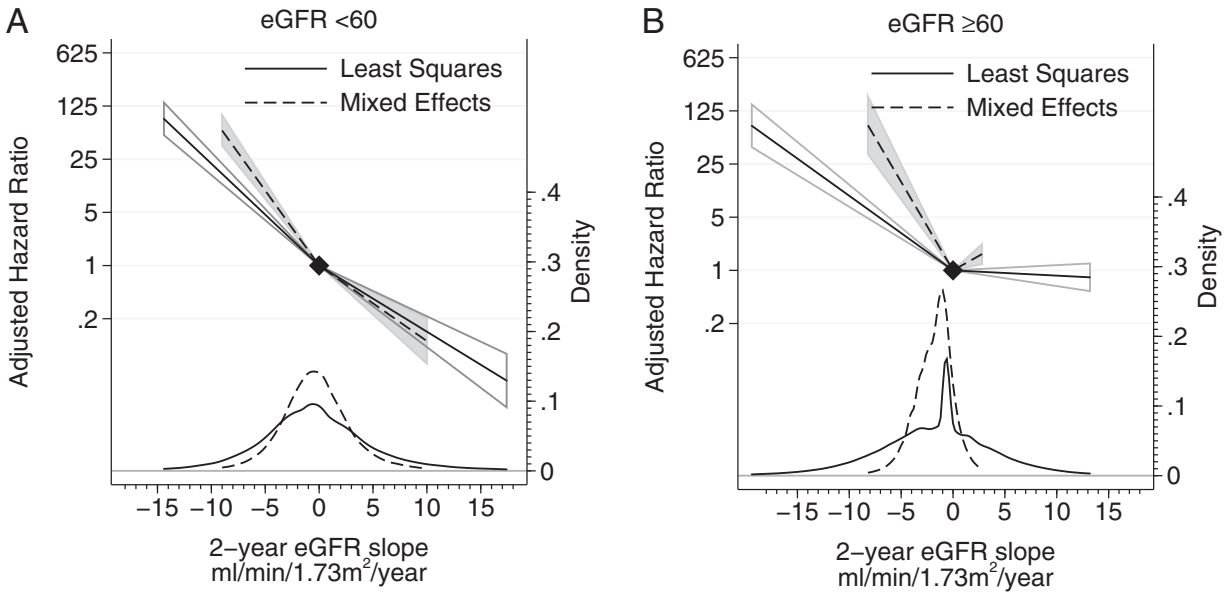
Analysis	Findings	Comment
Cohorts		
2-y median eGFR slope in median cohort	−0.68 mL/min/1.73 m^2/y for eGFR < 60 mL/min/1.73 m^2 stratum; −2.07 mL/min/1.73 m^2/y for eGFR ≥ 60 mL/min/1.73 m^2 stratum	
Relative risk for ESKD	HRs of 0.71 (GFR < 60 mL/min/1.73 m^2) and 0.70 (GFR ≥ 60 mL/min/1.73 m^2) for 0.75-mL/min/1.73 m^2/y eGFR slope reduction (slope estimated over 2 y, adjusted for measurement error)	Stronger at longer BL interval; consistent across cohorts and subgroups (BL eGFR, BL UACR, DM status)
Absolute risk for ESKD	1.6% absolute risk reduction at 5 y for 0.75-mL/min/1.73 m^2/y eGFR slope reduction (eGFR slope of −5 mL/min/1.73 m^2/y in control group and BL GFR = 75 mL/min/1.73 m^2)	Higher risk for lower BL eGFR
Trials		
Mean (CI) eGFR slope in treatment and control groups	Total slope at 3 y: −2.94 (−3.45 to −2.43) mL/min/1.73 m^2/y in the treatment group, −3.49 (−4.04 to −2.93) mL/min/1.73 m^2/y in the control group; chronic slope: −3.03 (−3.49 to −2.57) mL/min/1.73 m^2/y in the treatment group, −3.55 (−4.07 to −3.02) mL/min/1.73 m^2/y in the control group	
Treatment effect (mean [CI] for eGFR slope reduction)	Total slope at 3 y, 0.45 (0.19-0.72) mL/min/1.73 m^2/y; chronic slope, 0.53 (0.32-0.74) mL/min/1.73 m^2/y	
Regression of treatment effect on clinical outcomes (HR) vs slope reduction	3-y total slope significant, intercept nonsignificant, R^2 = 0.97 (BCI, 0.78-1.00); 0.75 mL/min/1.73 m^2/y eGFR slope reduction corresponds to 27% (BCI, 20%-34%) lower average risk for clinical outcome; similar relationship for chronic slope: R^2 = 0.96 (BCI, 0.63-1.00)	Acute effects are common; regression with total slope is less strong over shorter F/U; consistency across BL eGFR and UACR; no apparent differences by disease or intervention, but insufficient power for definitive evaluation
Prediction for a new trial	For 3-y total slope, 0.48-0.74 mL/min/1.73 m^2/y $PPV_{0.975}$; for chronic slope, 0.62-0.85 mL/min/1.73 m^2/y $PPV_{0.975}$	Threshold is lower for a large trial than for a modest-size trial; for total slope, predictions are weaker for shorter trials
Simulations		
Power compared to clinical outcome in scenarios with beneficial treatment effect	In absence of an acute effect, slope analysis has greater power than time-to-event analysis at higher vs lower BL GFRs, shorter vs longer follow-up, and uniform vs proportional treatment effects	For total slope, F/U < 2 y generally leads to substantial increase in required sample size and greater susceptibility to bias and reduced power from acute effects; optimum eGFR-based end point depends on the rate of eGFR decline, type of treatment effect, and study design
Type 1 errors for surrogate end point relative to clinical outcome in scenarios with null treatment effects	In the presence of an acute effect, errors are increased for chronic slope analysis	

Note: GFR slope is used as an average for comparing 2 or more groups, not at the individual level.
Abbreviations: BCI, Bayesian credible interval; BL, baseline; CI, confidence interval; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; ESKD, end-stage kidney disease; F/U, follow-up; HR, hazard ratio; $PPV_{0.975}$, threshold slope reduction to provide ≥97.5% PPV for nonzero benefit on clinical outcome; UACR, urinary albumin-creatinine ratio.

expected to reduce the 5-year ESKD risk by 1.6% (from 8.3% to 6.7%).

In trials,[13] the association of treatment effect on total slope over 3 years with the treatment effect on the clinical end point had a significant regression slope and nonsignificant intercept and posterior median squared correlation (R^2) of 0.97 (95% BCI, 0.78-1.00; Fig 4, left panel). Each 0.75–mL/min/1.73 m^2 per year larger treatment effect on the total slope was associated with an average 27% (95% BCI, 20%-34%) lower hazard for the clinical end point. Results were weaker when the total slope was calculated over shorter intervals. Similar results were shown for the chronic slope (R^2 of 0.96 [95% BCI, 0.63-1.00]; Fig 4, right panel). No clear differences were observed for subgroups defined by baseline UACR, eGFR, cause of disease, or intervention, but power was limited to detect such differences. For future trials, the $PPV_{0.975}$ threshold treatment effect on the total GFR slope at 3 years required to provide a 0.975 probability of a nonzero benefit on the clinical outcome are reductions of 0.48 and 0.74 mL/min/1.73 m^2 per year for large and modest-size trials, respectively. Predictions were weaker for shorter trials. For chronic slope, the corresponding $PPV_{0.975}$ thresholds required were 0.62 and 0.85 mL/min/1.73 m^2 per year, respectively.

In simulations,[14] GFR-based surrogate end points substantially improved the efficiency (reduced the required N) compared with the clinical outcome for a trial with 2 or



C

ESKD Risk Associations With eGFR Slope: ESKD HR for 0.75 mL/min/1.73 m^2/y Difference		
	Least Square Mean Regression*	Linear Mixed Models Regression**
Baseline eGFR < 60		
1-year	0.88 (0.86, 0.91)	0.79 (0.76, 0.83)
2-year	0.79 (0.77, 0.81)	0.71 (0.69, 0.74)
3-year	0.71 (0.68, 0.73)	0.63 (0.60, 0.67)
Baseline eGFR ≥ 60		
1-year	0.93 (0.92, 0.94)	0.74 (0.69, 0.80)
2-year	0.84 (0.82, 0.87)	0.70 (0.68, 0.72)
3-year	0.77 (0.74, 0.80)	0.66 (0.64, 0.68)

Figure 3. Analysis in cohorts: associations between population distribution of change in glomerular filtration rate (GFR) slope and end-stage kidney disease (A and B). Meta-analyzed adjusted hazard ratios for association between end-stage kidney disease and 2-year eGFR decline for patients with (A) eGFR <60 mL/min/1.73 m^2 and (B) eGFR ≥ 60 mL/min/1.73 m^2. Mixed effects indicates the best linear unbiased prediction from linear mixed models; the least squares is the β coefficient from linear regression of eGFR on time. The distribution of slopes is shown in the kernel density plot in the bottom half of the graph, demonstrating the substantial shrinkage, particularly in the higher eGFR group. (C) ESKD risk associations with eGFR slope. *Empirical: β coefficient from linear regression of eGFR on time. **Best linear unbiased prediction from linear mixed models. All eGFR values within a given observation period (1, 2, and 3 years ± 30%) were used to estimate slope coefficient. Adapted from Grams et al[12] with permission of the American Society of Nephrology (ASN); original figure © 2019 ASN.

more years of follow-up when baseline GFR was high and there was no acute effect. For total slope, a follow-up period less than 2 years generally led to a substantial decrease in efficiency and greater susceptibility to bias and reduced power from acute effects. For treatments without an acute effect, analyses based on the total slope were able to attain the same power as analyses of the clinical end point or on 30% or 40% GFR declines while reducing both sample size and follow-up time, with greater efficiency with higher versus lower mean baseline GFR, shorter versus longer follow-up, and uniform versus proportional treatment effects (Table 7). As an example, in the absence of an acute effect and with an intermediate or fast mean rate of progression, using the total slope instead of the clinical end point allows investigators to reduce follow-up from 4-6 years to 2 years while also improving efficiency by 17% to 64% (corresponding to sample size savings of 14%-39%) across the scenarios considered, including a 29% reduction for the intermediate case with baseline GFR of 42.5 mL/min/1.73 m^2 and GFR slope of −3.25 mL/min/1.73 m^2 per year. The presence of even a small negative acute effect, which was common in the trials in our database, attenuated the statistical power advantages of the total slope compared to the clinical end point, particularly for a slower versus

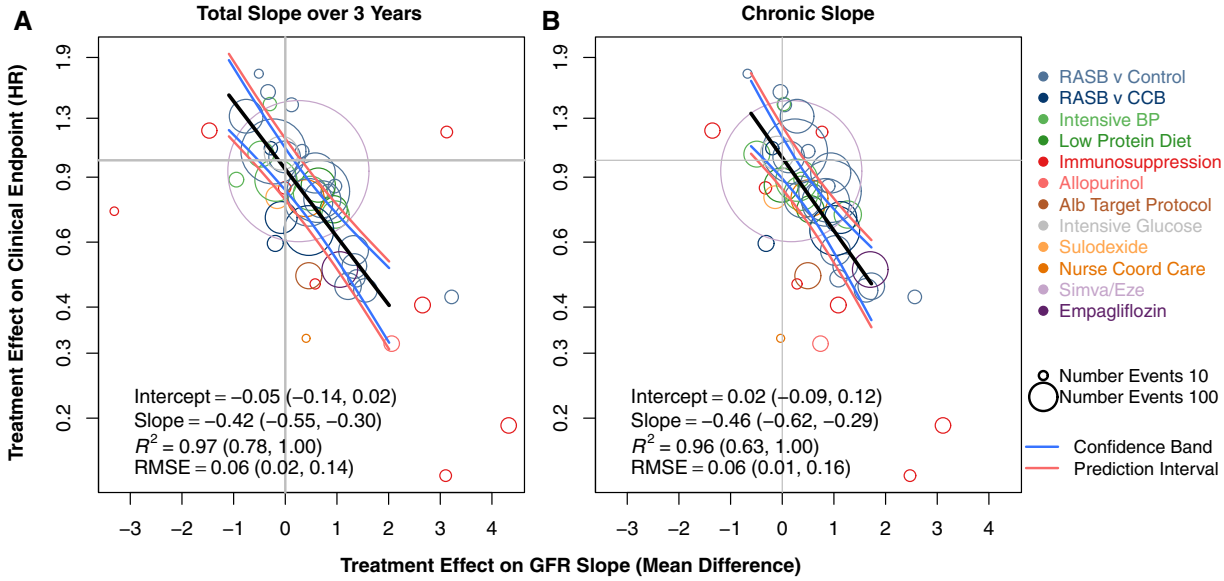


Figure 4. Trial level analyses: association between treatment effects on glomerular filtration rate (GFR) slope and treatment effects on the clinical end point. (A) Total slope at 3 years. (B) Chronic slope. Shown is the relationship between estimated treatment effects on the clinical end point (end-stage kidney disease [ESKD]; GFR, 15 mL/min/1.73 m^2, or doubling of serum creatinine) on the vertical axis and estimated treatment effects on the GFR slope on the horizontal axis. Treatment effects on GFR slope are expressed as mean difference in treatment minus control and are expressed in mL/min/1.73 m^2/y. The clinical end point is defined as treated kidney failure, doubling of creatinine level, or GFR of 15 mL/min/1.73 m^2. Treatment effect on the clinical end point is expressed as hazard ratio (HR). The colors indicate intervention type. Each circle is a separate intervention with the size of the circle proportional to the number of events. The black line is the line of regression through the studies. The blue line is the confidence band. The pink lines are the prediction bands computed from the model. Abbreviations: Alb, albuminuria; BP, blood pressure; CCB, calcium channel blocker; RASB, renin-angiotensin system blocker; RMSE, root mean squared error. Reproduced from Inker et al[13] with permission of the American Society of Nephrology (ASN); original figure © 2019 ASN.

faster average rate of progression or a shorter versus longer follow-up period. In the presence of a negative acute effect, the chronic slope had substantially greater power than either the clinical end point or other time-to-event end points, but higher risk for type 1 errors compared to the clinical outcome. The optimum GFR-based end point depended on the rate of GFR decline, the nature of the short- and long-term treatment effects, and study design. These topics are explored in detail in a separate publication and on an interactive spreadsheet available on the CKD-EPI website.[14,45]

The results of GFR slope from cohorts and trials agree both qualitatively and quantitatively and extend previous epidemiologic and trial-level analyses (Table 6). Of note, the results of the trial-level analyses for GFR slope are stronger than for albuminuria change, compare favorably with widely used surrogate end points in other fields,[47-50] and provide strong support for the role of both total and chronic GFR slopes as surrogate end points in RCTs in patients with early and late stages of CKD. In addition, simulations provide guidance when to use slope versus time-to-event analysis for GFR decline and when to use the chronic versus total slope. The $PPV_{0.975}$ threshold treatment effects on GFR slope of 0.5 to 1.0 mL/min/1.73 m^2 per year required to achieve a high probability of nonzero clinical benefit may appear small compared to the mean baseline GFR of trial participants in our database but represent approximately a 15% to 30% reduction in the mean rate of decline in the control group at 3 years.

Recommendations

The following recommendations represent our synthesis of all the information available to us, including evidence presented at the workshop (reviews of topics, data analyses addressing research aims, breakout group reports, and discussions), data in the published articles, and discussions with representatives of the FDA and EMA after the workshop. We anticipate that the recommendations may serve as a general guide for implementation of these candidate surrogate end points; they are not specific enough to cover all the important details required for design of the clinical trial, and for clinical trials for drug approval, regulatory agencies consider whether particular surrogate end points are appropriate in the context of a specific development program. In principle, using standardized measurement methods for determination of albuminuria and GFR and using multiple measures will improve precision of the estimates of change. Logistical considerations may also affect decisions about measurement methods and number of measurements; some of these questions might be addressed from past trials.

Table 7. Simulations: Gain in Efficiency for Total Slope and Chronic Slope Compared to the Clinical Outcome When Long-term Treatment Effect Is Intermediate Between Uniform and Proportional

		Total Slope: No Acute Effect			Total Slope: Acute Effect of −1.25 mL/min/1.73 m²			Chronic Slope: Acute Effect of −1.25 mL/min/1.73 m²		
		Relative Efficiency			Relative Efficiency			Relative Efficiency		
Mean BL GFR, mL/min/1.73 m²	Mean GFR Slope, mL/min/1.73 m²/y	Total Slope in 2 y RCT vs Clinical Outcome in 4-6 y RCT	Total Slope in 4-6 y RCT- vs Clinical Outcome in 4-6-y RCT	Required N for Clinical Outcome in 4-6-y RCT	Total Slope in 2 y RCT vs Clinical Outcome in 4-6-y RCT	Total Slope in 4-6-y RCT vs Clinical Outcome in 4-6-y RCT	Required N for Clinical Outcome in 4-6-y RCT	Chronic Slope in 2-y RCT vs Clinical Outcome in 4-6-y RCT	Chronic Slope in 4-6-y RCT vs Clinical Outcome in 4-6-y RCT	Required N for Clinical Outcome in 4-6-y RCT
27.5	−1.5	1.14	1.07	4,980	0.37	0.41	7,140	1.27	0.76	7,140
	−3.25	1.51	1.58	2,170	0.82	1.49	2,190	1.29	1.81	2,190
	−5.0	1.24	1.40	870	1.13	1.53	960	1.32	1.78	960
42.5	−1.5	0.71	1.11	4,130	0.28	0.39	5,010	0.71	1.68	5,010
	−3.25	1.40	1.70	1,750	0.26	1.23	1,940	1.13	2.32	1,940
	−5.0	1.17	1.61	830	0.64	1.60	930	1.26	2.34	930
67.5	−1.5	1.06	1.83	6,090	0.46	0.46	8,240	1.18	2.82	8,240
	−3.25	1.42	2.16	2,480	0.16	0.69	2,940	1.42	3.65	2,940
	−5.0	1.64	2.43	1,310	0.09	1.25	1,260	1.36	3.29	1,260

Note: All calculations assume a 25% intermediate long-term effect. Relative efficiencies are given by the ratio of sample size (N) for the clinical end point over 4 to 6 years versus the slope analysis over the indicated follow-up period. Relative efficiencies >1 indicate that a smaller sample size is required to achieve the same statistical power with the slope outcome over the indicated follow-up period compared to the clinical end point over 4 to 6 years. Increases in required N for the clinical end point with BL GFR of 27.5 compared to 42.5 mL/min/1.73 m² are due to assumptions that the acute effect is smaller at lower GFRs, a higher proportion of slow progressors reaching clinical events with smaller effect sizes under the intermediate long-term treatment effect model, and larger effects of random error in GFR measurements on end points requiring smaller GFR change. The 2-year total slope has very low relative efficiencies when the acute effect is −1.25 mL/min/1.73 m² and progression rate is slow because the treatment effect in the chronic phase is too small to reverse the acute effect.
Abbreviations: BL, baseline; GFR, glomerular filtration rate; RCT, randomized controlled trial.

Box 3. Recommendations for Albuminuria Change as a Surrogate End Point

Conclusions
- Early albuminuria change can be a reasonably likely or valid surrogate end point in phase 3 RCTs of participants with moderate to severe albuminuria, depending on the context
- Appropriateness varies by disease and by intervention; it is appropriate for diseases characterized by elevated albuminuria and for interventions in which reducing albuminuria is hypothesized to be one of the main mechanisms of action
- A large treatment effect on the surrogate is required to reliably predict a treatment effect on the clinical end point
- Larger studies with longer follow-up may be necessary to evaluate change in GFR or safety

Supportive Results From Our Analysis
- Strength of association consistent in cohorts and trials
- Stronger effects at higher baseline UACRs
- No significant differences by baseline GFR or cause of disease
- Treatment effect can be detected within 6 mo
- Threshold UACR reduction of 21%-27% (geometric mean) provides 97.5% PPV for a nonzero benefit on clinical outcome in a new trial in participants with UACR ≥ 30 mg/g (threshold varies with sample size)

Application
- Populations at high risk for GFR decline, such as DKD with baseline UACR > 300 mg/g
- Albuminuria preferred over total protein, but can measure both
- Multiple measurements at beginning and end to improve precision
- Measurement of posttreatment effect is desirable but sustained efficacy in absence of the drug might not be required for determination of efficacy (analogous to BP, cholesterol level, glycemia)
- Longer studies would be required to evaluate change in GFR or safety

Circumstances in Which Albuminuria Change May Not be Applicable as a Surrogate End Point
- Diseases *not* characterized by albuminuria or interventions in which reducing albuminuria is *not* hypothesized to be one of the main mechanisms of the intervention; in such cases, potential solutions include:
  - ◇ GFR decline as an end point (time to event or slope)
  - ◇ Other markers of kidney damage as an end point, specific for disease and intervention (not currently approved by regulatory agencies)
    - ▪ Total kidney volume in early PKD
    - ▪ Markers of tubular injury in tubulointerstitial diseases
- Interventions that do not lower albuminuria enough; in such cases, potential solutions include:
  - ◇ GFR decline as an end point
  - ◇ Combined end point with GFR decline

Note: Albuminuria change is used as an average for comparing 2 or more groups, not for determining response at the individual level.
Abbreviations: BP, blood pressure; DKD, diabetic kidney disease; GFR, glomerular filtration rate; PKD, polycystic kidney disease; RCT, randomized controlled trial; UACR, urinary albumin-creatinine ratio.

Albuminuria Change

We conclude that an early change in albuminuria can be a reasonably likely or valid surrogate end point for kidney disease progression in phase 3 RCTs, depending on the context (Box 3). Its appropriateness may vary by disease and by intervention; it appears more appropriate for diseases characterized by moderate to severe albuminuria and for interventions in which reducing albuminuria is hypothesized to be one of the main mechanisms of action. A large treatment effect (20%-30% reduction in geometric mean UACR) is likely to be necessary to ensure a significant treatment effect on the clinical outcome. Furthermore, requiring a threshold for the treatment effect on albuminuria change strengthens its validity as a surrogate end point.

Our results have implications for trial design. Our observation of larger changes in albuminuria at higher baseline UACRs and stronger relationships of albuminuria change with the clinical outcome at higher baseline UACRs suggest that entry criteria should include a minimum UACR. Our observation that the beneficial treatment effect on albuminuria may be observed in a 6-month follow-up period suggests that depending on the hypothesized mechanism of action of the intervention, the time frame of the trial can be shortened, allowing more efficient and likely less-expensive designs. Such trials would have more limited follow-up than established kidney disease end point trials but would provide a basis for firm, precise conclusions on UACR reduction. Postapproval studies with longer follow-up may be necessary to evaluate efficacy on GFR decline, as well as safety.

Other aspects of trial design are noteworthy. In principle, measurement of albumin would be preferred over total protein because it can be standardized, and a central laboratory would be preferred over local laboratories to

minimize measurement variability. It can be expected that increasing the number of albuminuria measurements for each participant at each time point (eg, at baseline and 6 months) would increase precision of the treatment effect and decrease the required threshold for albuminuria change, especially when the sample size is modest, but we were not able to evaluate this in our analyses.

In our view, these results would be applicable for diabetic kidney disease with higher albuminuria levels at baseline. There is a high prevalence of disease and a high risk for kidney disease progression, even in early stages of disease, but few effective therapies and therefore a large unmet clinical need.[51] The threshold of 20% to 30% reduction in geometric mean UACR could be used to evaluate results of phase 2 trials to determine which interventions have greatest promise to bring into phase 3 trials, and in phase 3 trials, the threshold could be used as a requirement for efficacy. Measurement of posttreatment effect is desirable to understand the mechanism of effect, but sustained efficacy in the absence of the drug might not be required for determination of efficacy (analogous to blood pressure, cholesterol level, and glycemia). A longer study would be required to ensure superiority for GFR decline and noninferiority for cardiovascular disease, mortality, and other potential adverse outcomes specific to the intervention.

These results complement ongoing efforts to define clinical end points for RCTs in membranous nephropathy and focal and segmental glomerulosclerosis with nephrotic syndrome or in immunoglobulin A nephropathy.[52-55] We do not propose that the current results replace these initiatives. In less common diseases, the large sample size needed to reliably assess treatment effects on albuminuria may not be available. Even so, investigators designing studies in rare diseases could potentially use the PPV results presented here and consider applying less stringent criteria for a minimal treatment effect on UACR change. Our findings may also be applicable to RCTs evaluating interventions in patients with CKD of other or unspecified cause with higher levels of albuminuria, in whom absolute risk is high.

There are a number of circumstances in which the proposal may not be applicable, including diseases not characterized by albuminuria or interventions in which reducing albuminuria is not hypothesized to be one of the main mechanisms of action, or interventions in which the treatment effect on albuminuria is less than the threshold. In these circumstances, other end points would be preferable.

GFR Slope

We conclude that GFR slope can be a valid surrogate end point in phase 3 RCTs (Box 4). Because GFR decline is on the path to progression to kidney failure, it may be applicable to many causes of CKD. A treatment effect of 0.5 to 1.0 mL/min/1.73 m^2 per year on the total slope or on the chronic slope is likely to be necessary to ensure a high probability of a nonzero benefit of the treatment effect on the clinical outcome. Requiring a threshold for treatment effect on the GFR slope strengthens the validity of GFR slope as a surrogate end point.

Our results also have implications for trial design. Our observation of stronger treatment effects with longer duration of follow-up suggests that a follow-up of at least 2 years is likely to be required to reliably predict a treatment effect on the clinical end point. A difference in slopes observed during a longer follow-up is more likely to predict a benefit on the clinical end point, but longer follow-up complicates trial design and conduct. We did not observe differences based on level of GFR, albuminuria, cause of disease, or demographic factors, although we had low power to detect differences in performance among subgroups. Our results suggest that GFR slope may be more useful relative to time-to-event end points for study populations with higher versus lower baseline GFRs and shorter versus longer follow-up, and for interventions with more uniform versus proportional treatment effects, and without acute effects. Many factors influence the decision to use total versus chronic slope; key among them is the potential for an acute effect, which was common in our analyses and which may complicate interpretation of the treatment effect on both the chronic and total slopes. In particular, a negative acute effect can attenuate or reverse the statistical power advantages of the total slope compared to the clinical end point and can increase the risk that use of the chronic slope as a surrogate end point could lead to a type 1 error relative to the clinical end point.

Other aspects of trial design are noteworthy. In principle, multiple measurements may improve the precision of determining treatment effects and shorten the required duration of follow-up, although we had few trials with sufficiently frequent eGFR assessments to evaluate this hypothesis empirically. However, a short follow-up period would not be sufficient to ensure noninferiority for detrimental effect on cardiovascular disease, mortality, and other outcomes specific to the intervention. Ascertainment of acute effects in phase 2 trials can aid in the design of phase 3 trials, although a large sample size may be necessary to detect small acute effects. If acute effects are suspected, it is advisable that the phase 3 trial assess acute, chronic, and total slope and if possible, the reversibility of potential acute effects after withdrawal of treatment at the beginning or end of the trial.

In our view, these results are most applicable to populations at high risk for progressive kidney disease because the likelihood of achieving the threshold treatment effect of 0.5 to 1.0 mL/min/1.73 m^2 per year is greater if the underlying rate of GFR decline is faster. If an acute effect is detected, modification of the phase 3 trial design may be necessary, with approval from the regulatory agency. Other circumstances that may require modification of trial design include effects of interventions on non-GFR determinants of serum creatinine level or proportional rather than uniform treatment effects.

Box 4. Recommendations for GFR Slope as a Surrogate End Point

Conclusion
- GFR slope can be a valid surrogate end point in phase 3 RCTs
- A difference in slopes observed during a longer follow-up is more likely to predict a benefit on the clinical end point
- A 2- to 3-year follow-up period may be required to estimate slopes, which reliably predict a treatment effect on longer term clinical outcomes
- GFR slope is more useful than time-to-event end points
  - ◇ For study populations with higher vs lower baseline GFRs
  - ◇ For shorter vs longer durations of follow-up
  - ◇ For interventions with more uniform vs proportional treatment effects
- Acute effects may complicate interpretation of the treatment effect; acute, chronic, and total slope should be assessed
- Assessment of pre- and posttreatment reversibility of acute effects may be helpful in interpretation

Supportive Results From Our Analysis
- Strength of association consistent in cohorts and trials
- No significant differences by baseline GFR, UACR, or cause of disease
- Treatment effect can be detected within 2 y
- Threshold GFR slope reduction of 0.5-1.0 mL/min/1.73 m^2 per year (difference in mean slope) provides 97.5% PPV for a nonzero benefit on clinical outcome in a new trial (threshold varies with sample size)
- Acute effects are common
- Many factors influence the decision to use total vs chronic slope

Application
- Populations at high risk for progressive kidney disease, no significant differences among subgroups
- Multiple measurements at important time points to improve precision
- Measurements at beginning, at early time point (≈3 mo), and at end to assess for acute effect
- What to do if there is an acute effect?
  - ◇ Extend follow-up long enough so the chronic slope can overcome risk for type 1 error of the surrogate end point relative to the clinical outcome due to initial negative acute effect in the analysis of the total slope
  - ◇ Require a larger treatment effect on chronic slope to overcome risk for type 1 error of the surrogate end point relative to the clinical outcome due to initial negative acute effect
  - ◇ Obtain a before-treatment eGFR and final eGFR after withdrawing treatment (evaluate total "off-treatment slope")
  - ◇ Use time-to-event analysis based on 40%-57% eGFR decline or ESKD instead of slope

Circumstances Other Than Acute Effects in Which GFR Slope May Not be Applicable as a Surrogate End Point
- Effects of the interventions on non-GFR determinants of serum creatinine level; in such cases, potential solutions include:
  - ◇ Measure other filtration markers (cystatin C, etc)
  - ◇ Measure GFR
- Proportional treatment effects; in such cases, a potential solution is time-to-event end points

Note: GFR slope is used as an average for comparing 2 or more groups, not for determining response at the individual level.
Abbreviations: eGFR, estimated glomerular filtration rate; ESKD, end-stage kidney disease; GFR, glomerular filtration rate; RCT, randomized controlled trial; UACR, urinary albumin-creatinine ratio.

Combinations

Preliminary analyses, not reported here, showed stronger evidence for surrogacy for UACR change plus GFR decline than for either alone in some settings. More work is required in this area. Novel designs can take advantage of the shorter period of follow-up required to assess the treatment effect on albuminuria compared to the longer period required to assess the treatment effect on GFR decline (eg, sequential or adaptive designs) while appropriately accounting for multiple hypothesis tests.

Hierarchy of End Points

Based on the strength of the clinical trial analyses and other considerations, the analytic team proposed a hierarchy of end points for clinical trials of CKD progression (Table 8). The strength of evidence represents a compilation of recommendations from the 2012 NKF-FDA Workshop and the 2018 NKF-FDA-EMA Workshop. Many factors affect which end point is most appropriate for the disease, type of intervention, study population, phase 3 trial design, and approval process of the FDA and EMA.

Strengths and Limitations of Our Analyses

Our approach is based on individual participant data, uses multiple sources of evidence, follows a consistent analytical approach, and gives consistent results across cohorts and trials and subgroups based on level of GFR, cause of kidney disease, and demographic data.

However, a key weakness is that there were a limited number of clinical trials available for analysis, leading to

Table 8. Hierarchy of End Points for Kidney Disease Progression for Phase 3 RCTs

End Point	Strength of Evidence
Kidney failure	Clinical outcome
Doubling of Scr (confirmed) (57% eGFR decline)	Valid surrogate end point
GFR decline > 40% (confirmed) GFR slope reduction (mean) > 0.5-1.0 mL/min/1.73 m^2/y	Valid surrogate end point
GFR decline > 30% (confirmed) UACR reduction (mean) > 30%	Reasonably likely surrogate end point in many trials and valid surrogate end point in some trials

Note: Hierarchy may vary depending on study population and trial design. Doubling of Scr and 30% to 40% decline in eGFR are used to determine response at the individual level; UACR reduction and GFR slope reduction are used as averages for comparing 2 or more groups.
Abbreviations: eGFR, estimated glomerular filtration rate; RCT, randomized controlled trial; Scr, serum creatinine; UACR, urinary albumin-creatinine ratio.

a limited variety of interventions and causes of kidney disease. Thus, the predictions for future trials reflect not only the implications of varying treatment effects on UACR change and GFR slope, but also the specific results in the previously conducted studies. Consequently, achieving a treatment effect on UACR change or GFR slope large enough to provide a high PPV in a future trial may not guarantee a low risk for falsely concluding that an ineffective treatment has a clinical benefit. Standardizing definitions and voluntarily sharing clinical trial data would facilitate using additional trials to update these analyses, as would allow for validation in separate studies.

As another limitation, we note that explicit criteria for acceptance of a new surrogate were not defined in advance. Also, some results showed heterogeneity among studies, perhaps on account of differences in study populations, assays (for serum creatinine, urine albumin, and protein), and outcome definitions. Further, because of the small number of clinical trials including children, we excluded children from the analyses. Finally, we evaluated only kidney disease progression and mortality as CKD outcomes.

Conclusion

As described in this report, we have analyzed a large number of prospective cohorts, clinical trials, and simulations; considered the strengths and limitations of the proposed candidate surrogates; and described settings in which they may or may not be applicable. Our results support important roles for both early change in albuminuria and GFR slope as surrogate end points for kidney disease progression in clinical trials. The evidence supporting the validity of these surrogate end points is stronger for GFR slopes than for albuminuria change. In addition, use of albuminuria change is appropriate only for diseases characterized by albuminuria and interventions in which reducing albuminuria is hypothesized to be one of the main mechanisms of action, whereas use of GFR slope is more generally appropriate but requires attention to acute effects and requires a longer duration of follow-up to ascertain the treatment effect than for albuminuria change. The proposed thresholds for treatment effects on albuminuria reduction (geometric mean of 30% within 6 months) or on GFR slope (0.5 to 1.0 mL/min/1.73 m^2 over 2 to 3 years) are reliably associated with significant treatment effects on the clinical end point of kidney disease progression under some conditions. Implementation requires understanding conditions under which the surrogate is likely to perform well and restricting its use to those settings. Appropriate implementation of surrogates could facilitate clinical trials in earlier stages of CKD. The optimal end point based on change in albuminuria or GFR depends on many factors that must be considered when designing a trial for a specific disease and a specific drug. We encourage careful consideration of these proposed surrogate end points in the design of future clinical trials.

Supplementary Material

Supplementary File (PDF)

Item S1: Supplementary information.

Article Information

Authors' Full Names and Academic Degrees: Andrew S. Levey, MD*, Ron T. Gansevoort, MD, PhD*, Josef Coresh, MD, PhD, Lesley A. Inker, MD, MS, Hiddo L. Heerspink, PharmD, PhD, Morgan E. Grams, MD, MHS, PhD, Tom Greene, PhD, Hocine Tighiouart, MS, Kunihiro Matsushita, MD, PhD, Shoshana H. Ballew, PhD, Yingying Sang, MS, Edward Vonesh, PhD, Jian Ying, PhD, Tom Manley, RN, Dick de Zeeuw, MD, PhD, Kai-Uwe Eckardt, MD, Adeera Levin, MD, Vlado Perkovic, MBBS, PhD, Luxia Zhang, MD, PhD, and Kerry Willis, PhD. *Conference Co-Chairs.

Authors' Affiliations: Division of Nephrology, Tufts Medical Center, Boston, MA (ASL, LAI); Department of Nephrology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands (RTG); Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (JC, KM, SHB, YS); Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, Groningen, the Netherlands (HLH, DdZ); Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD (MEG); Division of Epidemiology, Department of Internal Medicine, University of Utah, Salt Lake City, UT (TG, JY); The Institute for Clinical Research and Health Policy Studies, Tufts Medical Center (HT); Tufts Clinical and Translational Science Institute, Tufts University, Boston, MA (HT); Division of Biostatistics, Department of Preventive Medicine, Northwestern University, Chicago, IL (EV); National Kidney Foundation, New York, NY (TM, KW); Department of Nephrology and Medical Intensive Care, Charité-Universitätsmedizin Berlin, Berlin, Germany (K-UE); Division of Nephrology, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada (AL); George Institute for Global Health, UNSW, Sydney, Australia (VP); Peking University Institute of

Nephrology (LZ); and Division of Nephrology, Peking University First Hospital, Beijing, China (LZ).

Address for Correspondence: Andrew S. Levey, MD, Division of Nephrology, Tufts Medical Center, 800 Washington St, Box #391, Boston, MA 02111. E-mail: alevey@tuftsmedicalcenter.org

Support: The following companies provided a grant to the NKF to support the data analysis and workshop: Alexion, Astellas, AstraZeneca, Aurinia, Bayer, Boehringer-Ingelheim, Calliditas, Gilead, Ironwood, Janssen, Kyowa Hakko Kirin, Merck, Novartis, Novo-Nordisk, Reata, Retrophin, Tricida, and Vifor Fresenius Medical Care. The workshop sponsors had no role in the development of the workshop agenda or objectives. The sponsors were restricted from viewing any part of the workshop report manuscript until it was accepted for publication and therefore had no role in the content developed for this report.

Financial Disclosure: Dr Levey reports grants from the National Institutes of Health (NIH) and the NKF during the conduct of the study, and funding from Siemens outside the submitted work. Dr Coresh has grants from the NIH and the NKF related and unrelated to this research. Dr Inker reports funding to Tufts Medical Center for research and contracts with the NIH, NKF, Retrophin, Omeros, and Reata Pharmaceuticals and consulting agreements with Tricida Inc. and Omeros Corp. Drs Inker, Levey, and Coresh have a patent "Precise estimation of glomerular filtration rate from multiple biomarkers" pending to Coresh, Inker, and Levey; 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 payments from Abbvie, AstraZeneca, Boehringer Ingelheim, and Janssen; and other payments from Astellas, Fresenius, Gilead, and Merck (all outside the submitted work). Dr Greene reports grants from the NKF during the conduct of the study; and personal fees from DURECT Corp, Janssen Pharmaceuticals, and Pfizer Inc outside the submitted work. Dr Matsushita reports research funding and personal fee outside of the work from Kyowa Hakko Kirin and personal fee outside of the work from Akebia. Dr Vonesh is currently under contract as a consultant to Prometic, and Tricida Inc (related to current article) and Epividian Inc (unrelated to current article) and reports prior funding from the NKF for related work. Dr Perkovic reports receiving personal fees for Advisory Boards or Scientific Presentations from Retrophin, Janssen, Merck, and Servier; was a member of the SONAR Steering Committee; has served on Steering Committees for trials funded by Abbvie, Boehringer Ingelheim, GSK, Janssen, Novo Nordisk, Retrophin, and Tricida; and participated in Scientific Presentations/Advisory boards with Abbvie, Astellas, Astra Zeneca, Bayer, Baxter, BMS, Boehringer Ingelheim, Dimerix, Durect, Eli Lilly, Gilead, GSK, Janssen, Merck, Mitsubishi Tanabe, Novartis, Novo Nordisk, Pfizer, Pharmalink, Relypsa, Retrophin, Sanofi, Servier, Vifor, and Tricida. Dr Gansevoort received grant support and consultancy fees from Abbvie, Bayer, Ipsen, Sanofi-Genzyme, and Otsuka; all money was paid to the employing institution. Dr de Zeeuw was on the advisory boards and/or speaker for Bayer, Boehringer Ingelheim, Fresenius, Mundipharma, and Mitsubishi-Tanabe; Steering Committees and/or speaker for AbbVie and Janssen; and Data Safety and Monitoring Committees for Bayer. Mr Manley and Dr Willis are employed by NKF. All other authors declare that they have no relevant financial interests.

Acknowledgements: The authors are grateful to Juhi Chaudhari, MPH, for assistance with manuscript preparation.

Peer Review: Received April 1, 2019, in response to an invitation from the journal. Evaluated by 3 external peer reviewers, with direct editorial input from an Associate Editor and a Deputy Editor. Accepted in revised form June 13, 2019.

References

1. Perkovic V, Craig JC, Chailimpamontree W, et al. Action plan for optimizing the design of clinical trials in chronic kidney disease. *Kidney Int Suppl*. 2017;2:138-144.
2. Baigent C, Herrington WG, Coresh J, et al. Challenges in conducting clinical trials in nephrology: conclusions from a Kidney Disease—Improving Global Outcomes (KDIGO) Controversies Conference. *Kidney Int*. 2017;92(2):297-305.
3. Inrig JK, Califf RM, Tasneem A, et al. The landscape of clinical trials in nephrology: a systematic review of Clinicaltrials.gov. *Am J Kidney Dis*. 2014;63(5):771-780.
4. Palmer SC, Sciancalepore M, Strippoli GF. Trial quality in nephrology: how are we measuring up? *Am J Kidney Dis*. 2011;58(3):335-337.
5. Jun M, Turin TC, Woodward M, et al. Assessing the validity of surrogate outcomes for ESRD: a meta-analysis. *J Am Soc Nephrol*. 2015;26(9):2289-2302.
6. Palmer SC, Ruospo M, Teixeira-Pinto A, Craig JC, Macaskill P, Strippoli GFM. The validity of drug effects on proteinuria, albuminuria, serum creatinine, and estimated GFR as surrogate end points for ESKD: a systematic review. *Am J Kidney Dis*. 2018;72(6):779-789.
7. Thompson A. Proteinuria as a surrogate end point–more data are needed. *Nat Rev Nephrol*. 2012;8(5):306-309.
8. 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.
9. National Kidney Foundation. NKF-FDA-EMA Workshop. https://www.kidney.org/CKDEndpoints. Accessed January 14, 2019.
10. Coresh J, Heerspink HJL, Sang Y, et al. Change in albuminuria and subsequent risk of end-stage kidney disease: an individual participant-level consortium meta-analysis of observational studies. *Lancet Diabetes Endocrinol*. 2019;7(2):115-127.
11. Heerspink HJL, Greene T, Tighiouart H, et al. Change in albuminuria as a surrogate endpoint for progression of kidney disease: a meta-analysis of treatment effects in randomised clinical trials. *Lancet Diabetes Endocrinol*. 2019;7(2):128-139.
12. Grams ME, Sang Y, Ballew SH, et al. Evaluating glomerular filtration rate slope as a surrogate end point for ESKD in clinical trials: an individual participant meta-analysis of observational data. *J Am Soc Nephrol*. 2019, ASN-2019010008.
13. Inker LA, Heerspink HJL, Tighiouart H, et al. GFR slope as a surrogate end point for kidney disease progression in clinical trials: a meta-analysis of treatment effects of randomized controlled trials. *J Am Soc Nephrol*. 2019, ASN-2019010007.
14. Greene T, Ying J, Vonesh EF, et al. Performance of GFR slope as a surrogate endpoint for kidney disease progression in clinical trials: a statistical simulation. *J Am Soc Nephrol*. 2019, ASN-2019010009.
15. Vonesh EF, Tighiouart H, Ying J, et al. Mixed-effects models for slope-based endpoints in clinical trials of chronic kidney disease. *Stat Med*. 2019;38(22):4218-4239.
16. Thompson A, Smith K, Lawrence J. Change in estimated GFR and albuminuria as end points in clinical trials: a viewpoint from the FDA. *Am J Kidney Dis*. 2020;75(1):4-5.

17. Holtkamp F, Gudmundsdottir H, Maciulaitis R, Benda N, Thomson A, Vetter T. Change in albuminuria and estimated GFR as end points for clinical trials in early stages of CKD: a perspective from European regulators. *Am J Kidney Dis*. 2020;75(1):6-8.
18. Baliker M, Bressler K, Forfang D, Fowler KJ, Grandinetti A. Change in estimated GFR and albuminuria as end points in clinical trials: a perspective of people living with kidney disease. *Am J Kidney Dis*. 2020;75(1):9-10.
19. Levey AS, de Jong PE, Coresh J, et al. The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. *Kidney Int*. 2011;80(1):17-28.
20. Moeller MJ, Tenten V. Renal albumin filtration: alternative models to the standard physical barriers. *Nat Rev Nephrol*. 2013;9(5):266-277.
21. Comper WD. The limited role of the glomerular endothelial cell glycocalyx as a barrier to transglomerular albumin transport. *Connective Tissue Res*. 2014;55(1):2-7.
22. Remuzzi G, Bertani T. Is glomerulosclerosis a consequence of altered glomerular permeability to macromolecules? *Kidney Int*. 1990;38(3):384-394.
23. Hostetter TH, Olson JL, Rennke HG, Venkatachalam MA, Brenner BM. Hyperfiltration in remnant nephrons: a potentially adverse response to renal ablation. *Am J Physiol*. 1981;241(1):F85-F93.
24. Kidney Disease: Improving Global Outcomes (KDIGO). KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl*. 2013;1:1-150.
25. Miller WG, Bruns DE, Hortin GL, et al. Current issues in measurement and reporting of urinary albumin excretion. *Clin Chem*. 2009;55(1):24-38.
26. Earley A, Miskulin D, Lamb EJ, Levey AS, Uhlig K. Estimating equations for glomerular filtration rate in the era of creatinine standardization: a systematic review. *Ann Intern Med*. 2012;156(11):785-795.
27. Inker LA, Levey AS, Pandya K, Stoycheff N, Okparavero A, Greene T. Early change in proteinuria as a surrogate end point for kidney disease progression: an individual patient meta-analysis. *Am J Kidney Dis*. 2014;64(1):74-85.
28. Lambers-Heerspink HJ, Kropelin TF, Hoekman J, de Zeeuw D. Drug-induced reduction in albuminuria is associated with subsequent renoprotection: a meta-analysis. *J Am Soc Nephrol*. 2015;26(8):2055-2064.
29. Lambers Heerspink HJ, Gansevoort RT. Albuminuria is an appropriate therapeutic target in patients with CKD: the pro view. *Clin J Am Soc Nephrol*. 2015;10(6):1079-1088.
30. Fried LF, Lewis J. Albuminuria is not an appropriate therapeutic target in patients with CKD: the con view. *Clin J Am Soc Nephrol*. 2015;10(6):1089-1093.
31. Fried LF, Lewis J. Rebuttal of the pro view: albuminuria is an appropriate therapeutic target in patients with CKD. *Clin J Am Soc Nephrol*. 2015;10(6):1095-1098.
32. Lambers Heerspink HJ, Gansevoort RT. Rebuttal of the con view: albuminuria is an appropriate therapeutic target in patients with CKD. *Clin J Am Soc Nephrol*. 2015;10(6):1099.
33. Greene T. A model for a proportional treatment effect on disease progression. *Biometrics*. 2001;57:354-360.
34. Short-term effects of protein intake, blood pressure, and antihypertensive therapy on glomerular filtration rate in the Modification of Diet in Renal Disease Study. *J Am Soc Nephrol*. 1996;7(10):2097-2109.
35. Gassman JJ, Greene T, Wright JT Jr, et al. Design and statistical aspects of the African American Study of Kidney Disease and Hypertension (AASK). *J Am Soc Nephrol*. 2003;14(7)(suppl 2):S154-S165.
36. Leptak C, Menetski JP, Wagner JA, et al. What evidence do we need for biomarker qualification? *Sce Transl Med*. 2017;9(417), eaal4599.
37. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001;69(3):89-95.
38. Robb MA, McInnes PM, Califf RM. Biomarkers and surrogate endpoints: developing common terminology and definitions. *JAMA*. 2016;315(11):1107-1108.
39. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Guidance for industry expedited programs for serious conditions – drugs and biologics. 2014. https://www.fda.gov/downloads/Drugs/Guidances/UCM358301.pdf. Accessed January 16, 2019.
40. United States Code. 21 CFR Part 314, Subpart H, Accelerated Approval of New Drugs for Serious or Life-Threatening Illnesses. https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=2ec0bc1571c06727e63ffa2c72b2e08d&mc=true&n=pt21.5.314&r=PART&ty=HTML#sp21.5.314.h. Accessed January 16, 2019.
41. European Medicines Agency. Guideline on the clinical investigation of medicinal products to prevent development/slow progression of chronic renal insufficiency. 2017. https://www.ema.europa.eu/documents/scientific-guideline/draft-guideline-clinical-investigation-medicinal-products-prevent-development/slow-progression-chronic-renal-insufficiency_en.pdf. Accessed January 14, 2019.
42. Desai M, Stockbridge N, Temple R. Blood pressure as an example of a biomarker that functions as a surrogate. *AAPS J*. 2006;8(1):E146-E152.
43. Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. *JAMA Intern Med*. 2015;175(8):1389-1398.
44. Johns Hopkins University School of Public Health. Chronic Kidney Disease Prognosis Consortium (CKD-PC). https://www.jhsph.edu/research/centers-and-institutes/chronic-kidney-disease-prognosis-consortium/index.html. Accessed January 14, 2019.
45. Chronic Kidney Disease Epidemiology Collaboration. 2016. http://ckdepi.org. Accessed January 31, 2019.
46. Greene T, Teng CC, Inker LA, et al. Utility and validity of estimated GFR-based surrogate time-to-event end points in CKD: a simulation study. *Am J Kidney Dis*. 2014;64(6):867-879.
47. Baigent C, Keech A, Kearney PM, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet*. 2005;366(9493):1267-1278.
48. Turnbull F, Neal B, Pfeffer M, et al. Blood pressure-dependent and independent effects of agents that inhibit the renin-angiotensin system. *J Hypertens*. 2007;25(5):951-958.
49. Delahoy PJ, Magliano DJ, Webb K, Grobler M, Liew D. The relationship between reduction in low-density lipoprotein cholesterol by statins and reduction in risk of cardiovascular outcomes: an updated meta-analysis. *Clin Ther*. 2009;31(2):236-244.

50. Lassere MN, Johnson KR, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? An analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the Biomarker-Surrogacy (BioSurrogate) Evaluation Schema (BSES). *BMC Med Res Methodol*. 2012;12:27.
51. Tuttle KR. Diabetic kidney disease (c. 2018). *Adv Chronic Kidney Dis*. 2018;25(2):119-120.
52. Thompson A, Cattran DC, Blank M, Nachman PH. Complete and partial remission as surrogate end points in membranous nephropathy. *J Am Soc Nephrol*. 2015;26(12):2930-2937.
53. Inker LA, Mondal H, Greene T, et al. Early change in urine protein as a surrogate end point in studies of IgA nephropathy: an individual-patient meta-analysis. *Am J Kidney Dis*. 2016;68(3):392-401.
54. Troost JP, Trachtman H, Nachman PH, et al. An outcomes-based definition of proteinuria remission in focal segmental glomerulosclerosis. *Clin J Am Soc Nephrol*. 2018;13(3):414-421.
55. Thompson A, Carroll K, Inker LA, et al. Proteinuria reduction as a surrogate end point in trials of IgA nephropathy. *Clin J Am Soc Nephrol*. 2019;14(3):469-481.