



## 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

Andrew S. Levey, MD,<sup>1</sup> Lesley A. Inker, MD, MS,<sup>1</sup> Kunihiro Matsushita, MD, PhD,<sup>2</sup>  
Tom Greene, PhD,<sup>3</sup> Kerry Willis, PhD,<sup>4</sup> Edmund Lewis, MD,<sup>5</sup>  
Dick de Zeeuw, MD, PhD,<sup>6</sup> Alfred K. Cheung, MD,<sup>7</sup> and Josef Coresh, MD, PhD<sup>2</sup>

The US Food and Drug Administration currently accepts halving of glomerular filtration rate (GFR), assessed as doubling of serum creatinine level, as a surrogate end point for the development of kidney failure in clinical trials of kidney disease progression. A doubling of serum creatinine level generally is a late event in chronic kidney disease (CKD); thus, there is great interest in considering alternative end points for clinical trials to shorten their duration, reduce sample size, and extend their conduct to patients with earlier stages of CKD. However, the relationship between lesser declines in GFR and the subsequent development of kidney failure has not been well characterized. The National Kidney Foundation and Food and Drug Administration sponsored a scientific workshop to critically examine available data to determine whether alternative GFR-based end points have sufficiently strong relationships with important clinical outcomes of CKD to be used in clinical trials. Based on a series of meta-analyses of cohorts and clinical trials and simulations of trial designs and analytic methods, the workshop concluded that a confirmed decline in estimated GFR of 30% over 2 to 3 years may be an acceptable surrogate end point in some circumstances, but the pattern of treatment effects on GFR must be examined, specifically acute effects on estimated GFR. An estimated GFR decline of 40% may be more broadly acceptable than a 30% decline across a wider range of baseline GFRs and patterns of treatment effects on GFR. However, there are other circumstances in which these end points could lead to a reduction in statistical power or erroneous conclusions regarding benefits or harms of interventions. We encourage careful consideration of these alternative end points in the design of future clinical trials.

*Am J Kidney Dis.* 64(6):821-835. © 2014 by the National Kidney Foundation, Inc.

**INDEX WORDS:** Kidney end point; renal end point; kidney disease outcome; surrogate end point; biomarker; chronic kidney disease (CKD); estimated glomerular filtration rate (eGFR) decline; eGFR trajectory; renal function; serum creatinine; kidney disease progression; end-stage renal disease (ESRD).

Chronic kidney disease (CKD) is a significant public health problem in the United States and around the world, but the progression of CKD often is slow and there are few specific symptoms until the stage of kidney failure has been reached. How does one practically develop drugs when the beneficial effects of treatment of direct interest to patients are not expected to manifest for many years? In some settings, a change in a biomarker level is considered a reliable predictor of later clinical outcomes. In the setting of CKD, a sufficiently large change in glomerular filtration rate (GFR) has been considered such a surrogate end point. Accordingly, the US Food and Drug Administration (FDA) accepts halving of GFR, assessed as doubling of serum creatinine level, as an end point for clinical trials of kidney disease progression because it represents a marked loss of kidney function and is expected to be highly predictive of the development of kidney failure. However, a doubling of serum creatinine level also is a late event in CKD, requiring long durations of follow-up and large sample sizes in clinical trials. Thus, there is great interest in alternative GFR-based end points to shorten the duration of clinical trials, reduce sample sizes, and extend their conduct to patients with earlier stages of CKD. However, there is uncertainty about the

associations of lesser declines in GFR with the subsequent development of kidney failure.

On December 2 to 3, 2012, the National Kidney Foundation (NKF) and FDA cosponsored a scientific

*From the* <sup>1</sup>*Division of Nephrology, Tufts Medical Center, Boston, MA;* <sup>2</sup>*Johns Hopkins Bloomberg School of Public Health, Baltimore, MD;* <sup>3</sup>*Division of Epidemiology, University of Utah, Salt Lake City, UT;* <sup>4</sup>*National Kidney Foundation, New York, NY;* <sup>5</sup>*Rush University Medical Center, Chicago, IL;* <sup>6</sup>*Department of Clinical Pharmacy & Pharmacology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands;* and <sup>7</sup>*Division of Nephrology & Hypertension, University of Utah, Salt Lake City, UT.*

*Received May 23, 2014. Accepted in revised form July 29, 2014. Originally published online October 15, 2014.*

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

*Address correspondence to Andrew S. Levey, MD, William B. Schwartz Division of Nephrology, Tufts Medical Center Box 391, 800 Washington St, Boston, MA 02129. E-mail: [alevey@tuftsmedicalcenter.org](mailto:alevey@tuftsmedicalcenter.org)*

© 2014 by the National Kidney Foundation, Inc.

0272-6386/\$36.00

<http://dx.doi.org/10.1053/j.ajkd.2014.07.030>

workshop to determine whether alternative definitions of GFR decline have sufficiently strong relationships with important clinical outcomes of CKD to be used as end points in clinical trials of CKD-related therapies.<sup>1</sup> In preparation for the workshop, the NKF and FDA appointed a planning committee and analytic group to formulate research questions; design and conduct analyses using data from observational studies (cohorts), randomized clinical trials, and simulation studies (Box 1); invite participants to the workshop from academia (including investigators from cohorts and trials that contributed data), industry, and government with expertise in clinical trials of CKD-related therapies; and lead the workshop and disseminate the results. The anticipated outcome of the workshop was the identification of alternative magnitudes of GFR decline having a sufficiently strong relationship with important clinical outcomes of CKD that they can be used as end points in CKD clinical trials.

The planning committee announced the workshop on the NKF website and issued an open invitation to pharmaceutical companies to attend the workshop or a concurrent web-based broadcast. Approximately 1 month prior to the workshop, the planning committee provided introductory material, including a draft analytic plan, to the workshop attendees and invited them to comment. The conference agenda and list of breakout group topics and workshop attendees are included as [Item S1](#) (available as online supplementary material). During the first plenary session, members of the planning committee and analytic group gave introductory presentations, described the analysis of data, and reported their conclusions and proposal for alternative GFR-based end points. Conference attendees then met in breakout groups to discuss analytic issues, outcomes of interest, and implications for drug development programs. All groups were asked to comment on their level of agreement with results from the data analysis and the proposal. During the second plenary session, group leaders summarized discussions in their groups and reported general agreement with the

proposal. The conference concluded with further general discussion and presentations by members of the planning committee.

The purpose of this article is to summarize the clinical, analytic, and regulatory context for the workshop; methods, results, and conclusions of the data analysis; the proposal for an alternative end point based on estimated GFR (eGFR) decline and its potential application; and key points from the discussion. Detailed descriptions of data analysis are reported in separate publications.<sup>2-5</sup>

## CONTEXT

### Kidney Disease Outcomes and Measures

The prevalence of CKD is increasing in the United States and worldwide, with high cost and poor outcomes, especially for patients with kidney failure.<sup>6</sup> There are few proven therapies to slow the progression of CKD. However, despite the availability of simple laboratory tests to identify people with earlier stages of CKD, fewer clinical trials have been performed for kidney disease than for other common diseases.<sup>7,8</sup> It is widely acknowledged that the lack of specific symptoms prior to the stage of kidney failure makes it difficult to define suitable end points for clinical trials in earlier stages of CKD.

[Table 1](#) reviews the definitions of clinical and surrogate end points and biological markers (biomarkers).<sup>9</sup> From a regulatory perspective, clinical end points reflect how a patient feels, functions, or survives. Kidney failure meets the criteria for a clinical end point because it is accompanied by symptoms and a high burden of complications causing functional impairment and shortened survival. In addition, it represents loss of functional organs. However, the operational definition of kidney failure may vary among trials. Clinical practice guidelines define chronic kidney failure as  $\text{GFR} < 15 \text{ mL/min/1.73 m}^2$  (CKD GFR category 5) for 3 or more months or initiation of treatment with maintenance dialysis or kidney transplantation, thus including patients regardless of whether they receive kidney replacement therapy.<sup>10-12</sup> End-stage renal disease (ESRD) is an administrative term in the United States referring to treatment with maintenance dialysis or kidney transplantation and therefore refers to only treated patients. ESRD is easy to ascertain and clinically meaningful; however, GFR at the initiation of treatment varies, as does the decision of whether to initiate treatment. Other important outcomes of CKD include death, cardiovascular disease, metabolic and endocrine disorders, infections, cognitive impairment, and frailty.<sup>10,13</sup> These outcomes occur at high frequency in patients with kidney failure and also in patients with  $\text{GFR}$  of  $15\text{-}29 \text{ mL/min/1.73 m}^2$  (CKD GFR category 4). Acute kidney injury also

#### Box 1. Research Questions

1. What magnitude of decline in estimated GFR (eGFR, based on serum creatinine) is sufficiently strongly related to kidney failure in observational studies and clinical trials to be a candidate surrogate end point for these events?
2. Is the consistency of effects of treatments for various magnitudes of eGFR declines and kidney failure within clinical trials sufficiently high to allow the use of the proposed decline in eGFR as a surrogate end point?
3. Based on a synthesis of all available data and simulation models of different trial designs and analytic methods, what sizes of decline in eGFR can be used as a surrogate end point in new clinical trials?

**Table 1.** Definitions and Applications to Clinical Trials in CKD

	Definition <sup>a</sup>	Application to CKD
Clinical end point	A characteristic or variable that reflects how a patient feels, functions, or survives	Kidney failure (defined as GFR < 15 mL/min/1.73 m <sup>2</sup> ) or ESRD (defined as treatment with maintenance dialysis or kidney transplantation)
Surrogate end point	A biomarker that is intended to substitute for a clinical end point. A surrogate end point is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence	Established surrogate end point: doubling of serum creatinine (equivalent to 57% decline in eGFR <sub>cr</sub> ); alternative surrogate end points: lesser (<57%) declines in eGFR
Biological marker (biomarker)	A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention	eGFR decline; albuminuria increase

Abbreviations: CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; eGFR<sub>cr</sub>, creatinine-based estimated glomerular filtration rate; ESRD, end-stage renal disease.

<sup>a</sup>According to the Biomarkers Definitions Working Group.<sup>9</sup>

occurs commonly in CKD and is associated with high morbidity and mortality.<sup>14,15</sup>

A recent international clinical practice guideline recommends staging the severity of kidney disease according to level of GFR and albuminuria.<sup>10</sup> GFR decline and albuminuria do not meet the definition of a clinical end point, but nonetheless are important measures of kidney disease. The workshop described here focused on GFR decline as an end point in clinical trials of CKD. The data supporting albuminuria as a surrogate end point were reviewed at a prior NKF-FDA conference.<sup>16-18</sup>

### GFR Decline in CKD

By definition, GFR decline is on the pathway to kidney failure. Numerous studies show that the relationship of low eGFR with the subsequent development of kidney failure is very strong, graded, independent, and consistent across populations irrespective of age, sex, race, presence or absence of hypertension and diabetes, level of albuminuria, and cause of kidney disease.<sup>19-27</sup>

Glomerular filtration is the first step in the formation of urine by the nephron. It is the physiologic process of ultrafiltration of plasma across the glomerular capillary wall, and in principle, the level of GFR is the number of nephrons multiplied by the mean single-nephron GFR. Single-nephron GFR reflects dynamic and structural characteristics within the glomerulus and can vary according to physiologic and clinical conditions, including variation in dietary protein, use of antihypertensive agents, and surfeit or deficit of extracellular fluid. In animals with certain experimentally induced kidney diseases, single-nephron GFR in the remaining nephrons often is elevated due to glomerular hyperfiltration and hypertrophy, and as

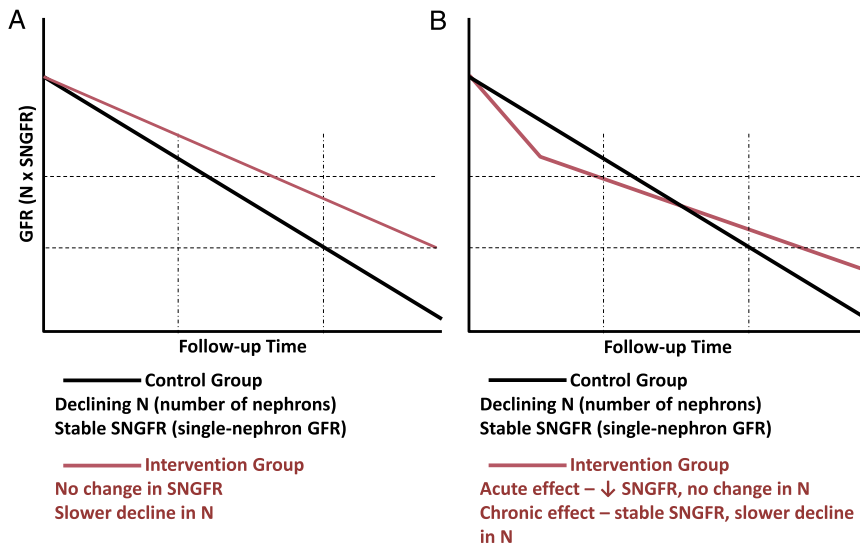
kidney disease progresses, the decline in GFR over time represents the irreversible loss of nephrons. Neither single-nephron GFR nor nephron number can be measured in vivo in humans. Instead, level of GFR is accepted as the best overall index of the level of kidney function. The time until kidney failure depends on the current level of GFR and the subsequent rate of GFR decline. Thus, the rate of decline in GFR over long intervals is accepted as a measure of kidney disease progression.

### Effects of Interventions on GFR

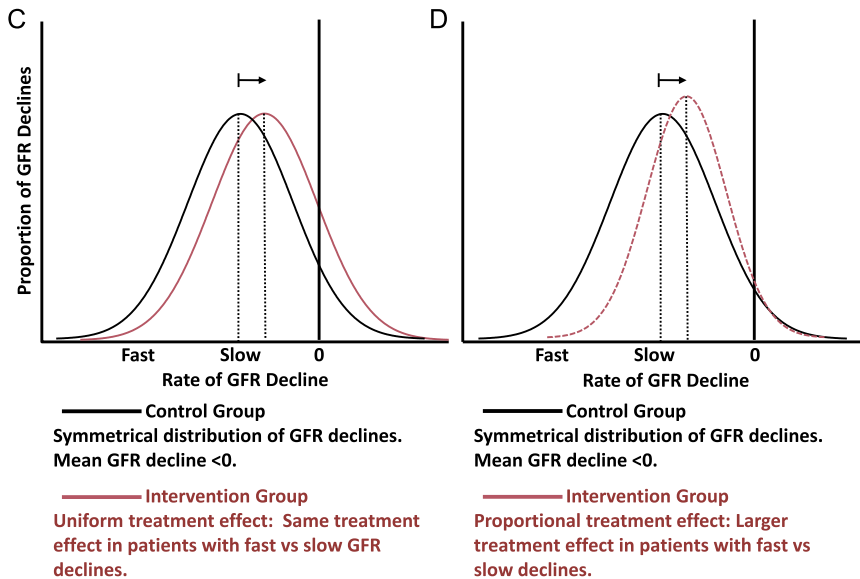
The goal of therapy is to slow the irreversible loss of functional nephrons, thereby preserving GFR level and delaying the onset of kidney failure (Fig 1A). However, the pattern of GFR decline in response to interventions has important implications for the design of clinical trials.

First, interventions may lead to an acute effect on GFR, defined as an early change in GFR different in direction or magnitude to the later change in GFR. For example, a low-protein diet, a low blood pressure goal, angiotensin-converting enzyme inhibitor use, and angiotensin receptor blocker use cause early reversible changes in single-nephron GFR opposite in direction to their hypothesized beneficial effect of slowing the irreversible loss of nephrons; thus, an early faster decline in GFR is followed by a later slower decline (Fig 1B).<sup>28,29</sup> In this setting, the change in GFR in the treatment arm may not be indicative of the irreversible loss of nephrons. Overall, acute effects on GFR complicate the design and interpretation of GFR decline as an end point in clinical trials because it is difficult to determine whether early changes in GFR reflect the acute effects of drugs or the underlying disease process and whether these early changes are reversible or irreversible.

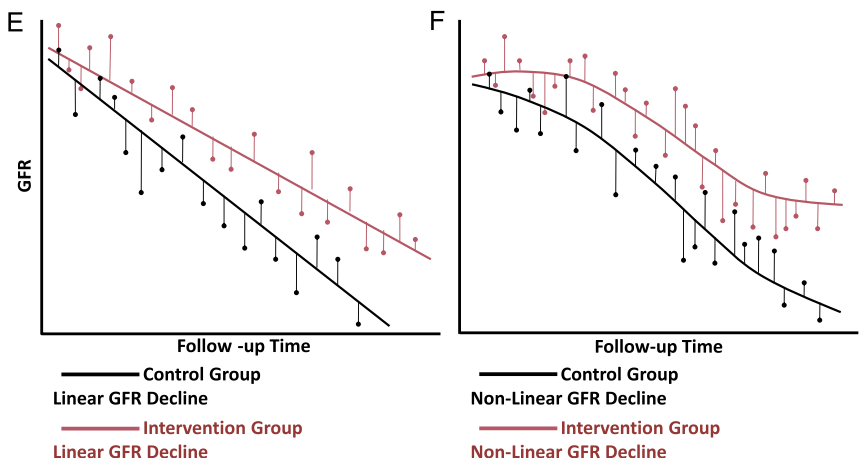
Interventions without vs with an Acute Effect on GFR



Treatment Interventions with Uniform vs Proportional Effects



Linear vs. Non-Linear GFR Declines



**Figure 1.** Glomerular filtration rate (GFR) decline in chronic kidney disease. Treatment interventions without versus with acute effects. (A, B) Hypothetical GFR decline in a patient assigned to the intervention (red lines) or control (black lines) group. Effect of the intervention can be detected using analysis of mean slopes or times to reach specific GFR declines. At baseline, GFR is reduced, with reduction in nephron number (N) but increase in single-nephron GFR (SNGFR). Intervention (A) does not and (B) does cause an “acute effect” on GFR. Acute effect is defined as early change in GFR different in direction or magnitude to the later change in GFR. Acute effects can cause bias in analysis using mean slopes or times to events. Treatment interventions with uniform versus proportional effects. (C, D) Hypothetical distribution of GFR decline in patients assigned to the intervention (red lines) versus control (black lines) groups. (C) Intervention has a uniform effect on GFR decline (uniform shift of the distribution to slower GFR declines across the distribution of GFR declines). Uniform treatment effect may be detected better by a comparison of mean slopes than by a comparison of times to percent decline. (D) Intervention has a proportional effect on GFR decline (a larger shift of faster declines than slower declines). Proportional treatment effect may be detected better by comparison of times to percent GFR decline than comparison of mean slopes. Linear versus nonlinear GFR declines. (E, F) Hypothetical values for measured or estimated GFR in a patient assigned to the intervention (pink dots) or control (black dots) group. The pattern of GFR decline is represented best by (E) a straight line and (F) a curved line.



Second, the hypothesized beneficial effect of the intervention may be uniform or proportional to the rate of GFR decline in the absence of treatment. A uniform treatment effect is characterized by a uniform improvement across the distribution of GFR declines (Fig 1C), whereas a proportional treatment effect is characterized by a larger absolute improvement in participants with faster GFR declines (Fig 1D). The pattern of treatment effect has implications for the analysis of the comparison of treatment groups in a clinical trial. For example, a uniform treatment effect may be detected better by a comparison of mean slopes, whereas a proportional treatment effect may be detected better by a comparison of times elapsed to a percent GFR decline.<sup>30</sup> Proportional treatment effects have been noted in other fields in medicine,<sup>31,32</sup> and analysis of past trials for this workshop suggests that a number of interventions lead to a mixture of uniform and proportional treatment effects.<sup>5</sup>

### Measurement and Estimation of GFR and GFR Decline

GFR cannot be measured directly; instead, it is measured indirectly as the clearance of exogenous filtration markers (referred to as measured GFR [mGFR]) or estimated from the serum level of endogenous filtration markers, such as creatinine or cystatin C (referred to as eGFR).<sup>33</sup> GFR estimating equations use the serum level of an endogenous filtration marker and demographic and clinical variables to estimate the level of mGFR. Demographic and clinical variables serve as measures for non-GFR determinants of the serum level; for example, age, sex, and race are related to creatinine generation by muscle and diet. GFR estimating equations have been derived in cross-sectional studies; in this setting, the use of multiple data elements provides more accurate GFR estimates than using the serum level of the filtration marker alone. In longitudinal studies, changes in the serum level of the filtration marker alone may be as accurate as changes in eGFR if there are no changes in demographic and clinical variables. In both cross-sectional and longitudinal studies, use of eGFR rather than serum level of the filtration marker alone enables inferences about the level of GFR and its change on the GFR scale. GFR estimating equations now are widely used in clinical practice and clinical trials.

Decline in mGFR or eGFR can be expressed as a continuous or categorical variable, and randomized groups can be compared by computing the mean rates of decline or times to event. It can be difficult to assess the rate of decline in GFR because of imprecision in the measures and the possibility of nonlinearity in GFR decline (Fig 1E and F).<sup>34,35</sup> For a number of reasons, many past trials have compared the time to reach a specified degree of GFR decline

rather than the mean GFR declines (slopes) between randomized groups.

In general, the pattern of decline in eGFR after an intervention is expected to mirror the pattern of decline in mGFR; however, interventions can affect the non-GFR determinants of the endogenous filtration markers used to estimate GFR, as well as the level of GFR. For example, dietary protein restriction reduces creatinine generation, leading to an increase in eGFR but a reduction in mGFR.<sup>36</sup> In principle, acute effects, proportional effects, and nonlinear effects of interventions on eGFR could be due to their effects on non-GFR determinants of endogenous filtration markers in addition to their effects on mGFR.

Historically, a doubling of serum creatinine level has been used as an end point in clinical trials of kidney disease progression. Using the CKD-EPI (CKD Epidemiology Collaboration) 2009 creatinine equation, a doubling of serum creatinine level approximately corresponds to a 57% decline in eGFR based on serum creatinine level.<sup>37</sup> Using the CKD-EPI 2012 cystatin C equation, a 57% decline in eGFR based on cystatin C level approximately corresponds to a 1.9-fold increase in serum cystatin C level.<sup>38</sup> Table 2 shows examples of baseline and final eGFRs for a 57% eGFR decline and lesser percent eGFR declines.

### Evaluation of Candidate Surrogate End Points

Before a biological marker is accepted as a surrogate end point, its validity and utility as a surrogate end point should be demonstrated. Candidate surrogate end points that do not meet these criteria may lead to

**Table 2.** Alternative Surrogate End Points for Time-to-Event Analyses

Initial eGFR	Final eGFR			
	57% eGFR Decline	40% eGFR Decline	30% eGFR Decline	20% eGFR Decline
67.5	29.0	40.5	47.3	54.0
42.5	18.3	25.5	29.8	34.0
27.5	11.8	16.5	19.3	22.0

*Note:* Using the CKD-EPI 2009 creatinine equation, a 57% decline in eGFR corresponds to a doubling of serum creatinine level from a baseline of  $\geq 0.9$  mg/dL in men or  $\geq 0.7$  mg/dL in women.<sup>37</sup> Using the CKD-EPI 2012 cystatin C equation, a 57% decline in eGFR corresponds to a 1.9-fold increase in serum cystatin C level from a baseline of  $\geq 0.8$  mg/L in men and women.<sup>38</sup> A 40%, 30%, and 20% decline in eGFR corresponds to a 1.5-fold, 1.3-fold, and 1.2-fold increase in serum creatinine or cystatin C level, respectively.

Abbreviations: CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration; eGFR, estimated glomerular filtration rate (in mL/min/1.73 m<sup>2</sup>).

**Table 3.** Criteria to Be Considered in the Evaluation of Candidate Surrogate End Points for Clinical Trials and Application to GFR Decline as a Surrogate for Kidney Failure

Criteria <sup>a</sup>		Application to GFR Decline
Biological plausibility	Sometimes intuitive, sometimes supported by animal data or by favorable responses in extreme cases	Strong, because it is on the pathway to kidney failure; a sufficient GFR decline defines kidney failure
Epidemiologic data (observational studies)	Increases (or decreases) in the putative surrogate are correlated with unfavorable (or favorable) clinical outcomes	Evaluate associations <ul style="list-style-type: none"> <li>• Goal: Explore associations between established and alternative surrogates with clinical end points</li> <li>• Strengths: Long duration of follow-up, large sample size; ability to assess relative and absolute risk</li> <li>• Limitations: Potential for bias (confounding)</li> </ul>
Clinical trials	Changes in the putative surrogate resulting from at least 1 type of intervention, and preferably many types, working by different mechanisms, affect clinical outcomes in a predictable manner that is substantially attributable to the effect on the surrogate	Evaluate treatment effects <ul style="list-style-type: none"> <li>• Goal: Compare treatment effects on surrogates vs treatment effects on clinical end points</li> <li>• “Case studies” using past clinical trials <ul style="list-style-type: none"> <li>– Strengths: Real world applications</li> <li>– Limitations: Few trials in which treatment effect on clinical outcomes is known with certainty; many sources of variation</li> </ul> </li> <li>• Simulations in which treatment effect is known <ul style="list-style-type: none"> <li>– Strengths: Ability to compare type 1 error and statistical power; ability to assess effects of variation in CKD parameters, analysis methods</li> <li>– Limitations: No direct demonstration of validity or utility</li> </ul> </li> </ul>

Abbreviations: CKD, chronic kidney disease; GFR, glomerular filtration rate.

<sup>a</sup>Based on Desai et al<sup>39</sup> and Biomarkers Definitions Working Group.<sup>9</sup>

underestimation of benefit, leading to rejection of effective therapies, or overestimation of benefit, leading to the adoption of ineffective therapies and exposure to harm. Table 3 shows criteria that often are considered when evaluating a candidate surrogate and how our analyses relate to these criteria.<sup>9,39</sup> There is no single criterion for surrogacy. Acceptance of a surrogate end point for use in clinical trials requires a synthesis of evidence from numerous sources. In principle, the surrogate should be easy to measure and occur earlier than the clinical end point. The association with the clinical end point should be supported by strong biological plausibility and empirical evidence in observational studies. When used in a clinical trial, the treatment effect on the surrogate should be consistent with the treatment effect on the clinical outcome, the risk of type 1 error with the surrogate should be low when there is no effect on the clinical outcome, and the statistical power for the treatment effect for the surrogate should be higher than that for the clinical outcome.

## METHODS, RESULTS, AND INTERPRETATION

In this section, we summarize methods, results, and interpretation of analyses reported elsewhere in more detail.<sup>2-5</sup> We begin with a definition of established and alternative surrogate end points, then discuss the general framework for analysis of observational studies

(cohorts), clinical trials, and simulations, including strengths and limitations of each source of data. Next, we discuss the sources of data, then the main results and interpretation for each analysis.

## Definitions of Established and Alternative Surrogate End Points

For this workshop, we considered kidney failure, defined as either  $GFR < 15 \text{ mL/min/1.73 m}^2$  or ESRD, as the “clinical end points” of interest related to kidney disease progression. For some analyses, we also considered mortality because it is an important clinical end point and a competing event for kidney failure. Because a doubling of serum creatinine level is approximately equivalent to a 57% eGFR decline, we considered a doubling of serum creatinine level or a 57% decline in eGFR to be the established surrogate end point and considered lesser declines in eGFR as potential alternative surrogate end points. We expressed these alternative end points as lesser declines in eGFR rather than lesser increases in serum creatinine level because the former more directly reflects the physiologic process of interest, although the interpretation of the results would be equivalent with either description. We conducted most analyses using percent eGFR decline (Table 2) and comparison of randomized groups using time-to-event analysis. In addition, we also considered the time course of the

**Table 4.** Synthesis of Findings and Summary of Results for eGFR Decline > 30% and >40% Over 2-3 Years

Analysis	Criterion	Results	Comment
<b>Cohorts</b>			
Relative risk for ESRD	Validity	Very strong (HR > 5)	Consistent across cohorts, demographic and clinical characteristics, including baseline eGFR and albuminuria
Excess risk for ESRD	Validity	Substantial (42% excess 10-y risk for baseline eGFR = 35)	Based on average baseline risk and meta-analyzed relative risk; varies by baseline eGFR, F/U interval, and cohort
Relative risk for mortality	Validity	Strong (HR, 1.6-1.8)	Consistent across cohorts, demographic and clinical characteristics, including baseline eGFR and albuminuria
Excess risk for mortality	Validity	Moderate (7% excess 5-y risk for baseline eGFR = 50)	Based on average baseline risk and meta-analyzed relative risk; varies by baseline eGFR, F/U interval, and cohort
<b>Trials</b>			
Relative risk for established end point (ESRD, GFR < 15, or doubling of Scr)	Validity	Very strong (HR > 9)	Consistent among trials, demographic and clinical characteristics, including type of kidney disease and intervention
Treatment effect precision	Utility	More precise (more frequent) than for established end point	More frequent with longer duration of follow-up, greater with vs without confirmation with repeat Scr measurement
Treatment effect magnitude	Utility	Generally consistent HR compared to established end point, but HR attenuated in some comparisons	Supportive, but limited due to low event rate in most trials
<b>Simulations</b>			
Type 1 errors in simulations with null treatment effects	Validity	Acceptable (type 1 error $\approx \leq 10\%$ )	Substantial savings for shorter trials and high baseline GFR and no acute effect; no substantial gain in power at low GFR; inflated type I error with even small acute effects (less with 40% eGFR decline)
Power in simulations with positive treatment effects	Utility	Power stronger than 57% eGFR decline (smaller samples size or shorter F/U)	

Note: (e)GFRs given in mL/min/1.73 m<sup>2</sup>.

Abbreviations: eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; F/U, follow-up; GFR, glomerular filtration rate; HR, hazard ratio; Scr, serum creatinine.

eGFR decline and whether the decline was confirmed after repeat measurement.

### Framework for Analysis

For our analysis, we accepted the biological plausibility for eGFR decline as a valid surrogate for kidney failure because GFR decline is a necessary intermediate on the pathway to kidney failure, and a 57% decline as the established surrogate. We used data from cohorts and clinical trials to evaluate alternative surrogate end points and compare them with the clinical end point and established surrogate with respect to frequency, the strength of their associations with clinical end points, and the consistency of treatment effects of interventions using the alternative and established end points (Table 4).

The strengths of analysis of cohorts are the long duration of follow-up and large sample size, enabling more accurate assessment of associations. The limitation of analysis in cohorts is the potential for bias from imbalance of potential confounding factors, variation in study design, the possibility that treatment may not be

recorded, and the possibility of bias. Clinical trials overcome some of these limitations by randomization. We used past randomized clinical trials as “case studies” with substantial relevance. Analysis of associations of alternative surrogate end points with established surrogate or clinical end points in clinical trials allows evaluation of consistency with cohorts and permits evaluation across treatment interventions and types of kidney disease. We also compared the magnitude and precision of treatment effects on alternative end points with treatment effects on established end points in clinical trials. However, there are many limitations to these analyses, including that treatment effects on established surrogate and clinical outcomes often were estimated imprecisely in past clinical trials and that there were many sources of variation in interventions, study populations, and conduct and analysis among trials that are not reflected in the pooled results. Thus, we also conducted simulations based on data from past trials and assumptions about the treatment effects (Table 4). For simulations with null treatment effects on the clinical outcome, we compared the risk of type 1 errors for established and alternative

**Table 5.** Summary of Study Populations and Outcomes

	Cohorts	Clinical Trials	Simulations
Studies	22 cohorts for ESRD outcomes, 35 cohorts for mortality outcomes	37 trials, 43 intervention comparisons	20 input parameters derived from 14 CKD RCTs
Participants	1,530,614 participants for ESRD outcomes, 1,597,807 participants for mortality outcomes	9,488 participants categorized by 5 causes of CKD: DM (n = 4,008), HTN (n = 1,094), IgAN (n = 888), lupus nephritis (n = 228), MN (n = 321), unspecified/other (n = 2,949) 12,821 participants categorized by interventions: RAS blockade vs control (n = 5,748), RAS blockade vs CCB (n = 2,295), intensive BP control (n = 2,655), low-protein diet (n = 839), IS therapy (n = 1284)	3,060 total parameter configurations (1,404 to evaluate type 1 error and 1,656 to evaluate power); for each parameter configuration, 800 data sets consisting of 1,000 participants
Outcomes	12,345 ESRD events, 159,558 deaths	3,070 composite events, 2,029 ESRD events, 1,151 GFR < 15 mL/min/1.73 m <sup>2</sup> events, 2,086 serum creatinine doublings	11 outcomes per simulated data set (ESRD alone or a composite end point including ESRD or eGFR decline of either 20%, 30%, 40%, 50%, or 57%, with and without confirmation of eGFR decline)

Abbreviations: BP, blood pressure; CCB, calcium channel blocker; CKD, chronic kidney disease; DM, diabetes mellitus; ESRD, end-stage renal disease; GFR, glomerular filtration rate; HTN, hypertension; IgAN, immunoglobulin A nephropathy; IS, immunosuppressive; MN, membranous nephropathy; RAS, renin-angiotensin system; RCT, randomized controlled trial.

surrogate end points (false-positive results lead to erroneous conclusions for benefit and false-negative results lead to erroneous conclusions for harm). For simulations with beneficial treatment effects on the clinical outcome, we compared the magnitude of treatment effects and statistical power for established and alternative surrogate end points. Although simulations cannot substitute for direct demonstration of validity and utility, they are especially useful to show the effects of variation in interventions, population characteristics, and analysis methods on these comparisons and can help explain observations in past clinical trials.

Overall, we synthesized the findings based on the potential validity and utility of lesser versus 57% eGFR declines or doubling of serum creatinine level (Table 4). For validity of alternative surrogates, we considered biological plausibility, strength of associations with the clinical end point in cohorts and trials, and preservation of the low risk of type 1 error in simulations. For utility of alternative surrogates, we considered ease of measurement, increase in frequency of end point events in cohorts and trials, increase in precision and preservation of the magnitude of the treatment effect in clinical trials, and increase in statistical power in simulations.

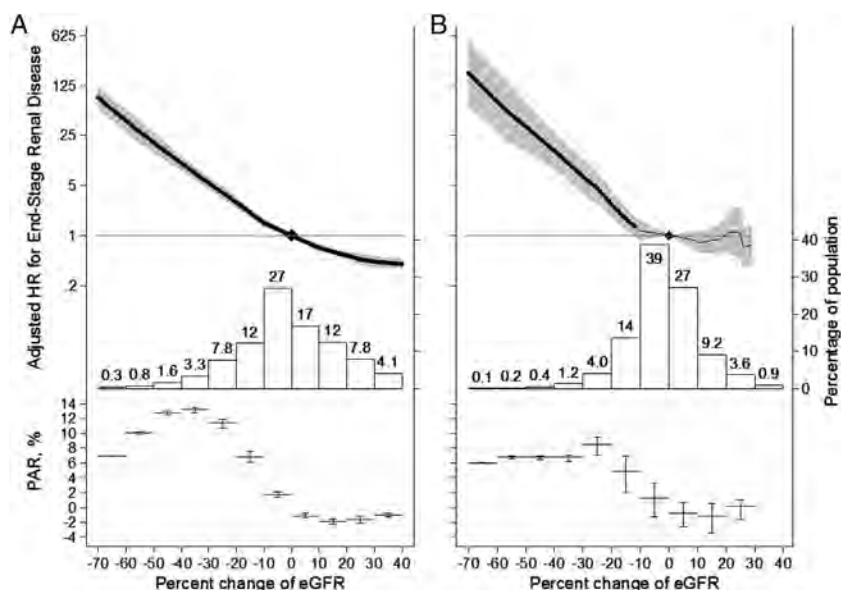
### Sources of Data

Table 5 shows the number of studies, participants, and outcomes for analysis of cohorts and clinical trials and number of parameter configurations for simulations.

Data from cohorts were collected previously by the CKD Prognosis Consortium (CKD-PC). Briefly, the CKD-PC consists of cohorts from the general population, populations with high cardiovascular risk, or populations with CKD, with data for serum creatinine and albuminuria and 50 or more events of outcomes of interest (either mortality or kidney outcome).<sup>19-22,40</sup> General population cohorts were derived from a systematic review of the literature conducted in 2009. Cohorts with high cardiovascular risk and CKD were identified based on consortium members' knowledge of published and unpublished data. General population and high-risk cohorts were required to have at least 1,000 participants. For these analyses, we included cohorts with a repeat measure of serum creatinine during an interval of 0.5 to 3.5 years to determine change in eGFR during a "baseline period" of 1 to 3 years and with data for clinical events following this baseline period. Confirmation of changes in serum creatinine level during the baseline period was not required. For analyses of ESRD as an outcome, we included 22 cohorts that were composed of 4 general population cohorts, 5 cohorts with high cardiovascular risk, and 13 cohorts predominantly containing people with CKD. For analysis of mortality as an outcome, we included 35 cohorts that were composed of 15 general population cohorts, 7 high-risk cohorts, and 13 CKD cohorts. We performed analyses within each cohort and meta-analyses across cohorts. Each meta-analysis was restricted to cohorts with at least 10 ESRD events or



**Figure 2.** Results in cohorts: adjusted hazard ratio (HR) of end-stage renal disease (ESRD), prevalence, and approximate percent population attributable risk (PAR) for percent change in estimated glomerular filtration rate (eGFR) during a 2-year baseline period for (A) baseline eGFR < 60 and (B)  $\geq 60$  mL/min/1.73 m<sup>2</sup>. Adjusted HR of ESRD associated with percent change in eGFR during a 2-year baseline period in eGFR < 60 and  $\geq 60$  mL/min/1.73 m<sup>2</sup> and a histogram of percent change in eGFR and approximate percent PAR proportion of ESRD. Values trimmed at less than -70% change in eGFR (0.22% and 0.05% of the population with eGFR < 60 and  $\geq 60$  mL/min/1.73 m<sup>2</sup>, respectively) and >40% change in eGFR (5.9% and 0.51% of the population with eGFR < 60 and  $\geq 60$  mL/min/1.73 m<sup>2</sup>, respectively). Reproduced with permission from Coresh et al<sup>2</sup> with permission of the American Medical Association.



deaths and participants 18 years or older. Although the high-risk and CKD cohorts were not derived from a systematic search of the literature, prior studies have shown similar relationships between exposures and outcomes in these cohorts as in the general population cohorts.

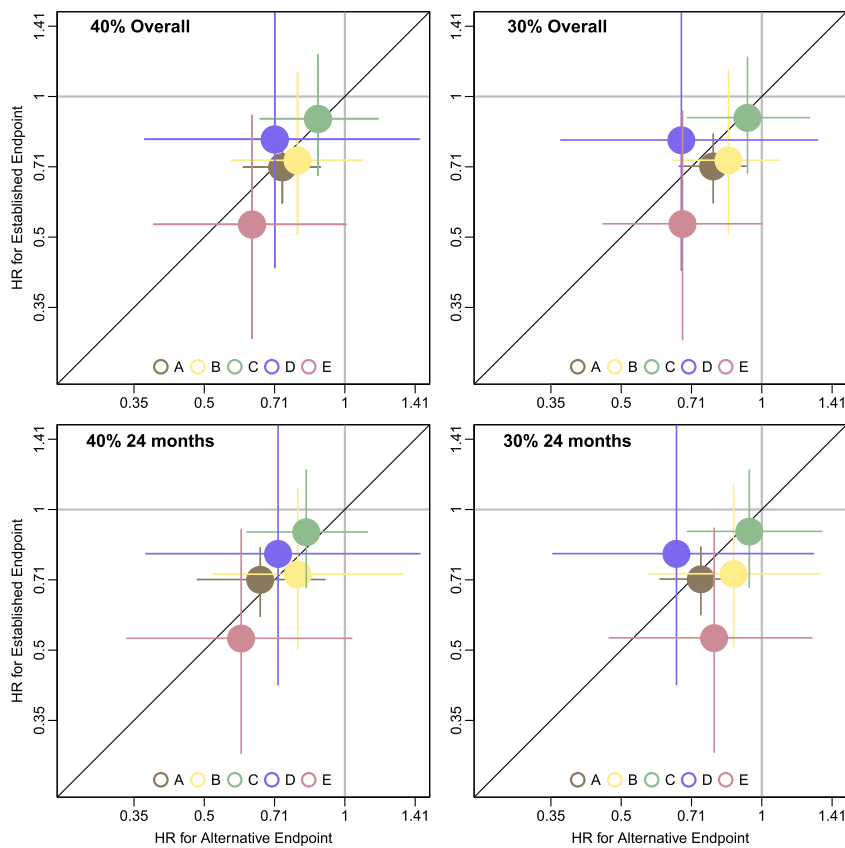
Data from clinical trials were collected previously by the CKD-EPI.<sup>17,41</sup> Briefly, systematic reviews of the literature were performed for kidney disease randomized controlled trials for evaluation of proteinuria as a surrogate end point in CKD in 2007 and for immunoglobulin A nephropathy in 2012. All trials had data for serum creatinine and proteinuria and at least one outcome of interest (either doubling of serum creatinine or ESRD). A total of 37 trials of 5 intervention types were included (renin-angiotensin system [RAS] blockade vs control, RAS blockade vs calcium channel blocker, intensive vs usual blood pressure control, low-protein vs usual-protein diet, and immunosuppressive vs other therapy). Causes of CKD were categorized as diabetes, hypertension, lupus nephritis, membranous nephropathy, and unspecified or other. For trials that evaluated more than one intervention, we included a separate group for each independent treatment comparison, such that some participants were included more than once. Overall, we had 43 analytical comparisons; we performed analyses within each trial and meta-analyses across comparisons. Although the database does not include all recent trials, we thought it contained a sufficient number of representative large and small trials for this purpose.

For simulations, a total of 20 input parameters were modeled, including rates and distributions of eGFR declines, magnitudes of acute effects, patterns of long-term treatment effect, types of study design,

rates of mortality and missing data, and relationship of eGFR to initiation of maintenance dialysis therapy or kidney transplantation. Data analysis for determination of input parameters was performed for 14 trials from the CKD-EPI data set (above) with at least 1 year of eGFR follow-up in at least 100 participants. We considered a total of 3,060 parameter configurations: 1,404 parameter configurations to evaluate type 1 error (assuming a null treatment effect on the clinical outcome) and 1,656 parameter configurations to evaluate power (assuming a beneficial treatment effect on the clinical outcome). For each parameter configuration, we simulated 800 independent data sets, with each data set consisting of 1,000 patients, with 500 assigned to the treatment group and 500 assigned to the control group. For each simulated data set, we considered 11 outcomes, including ESRD alone and composite end points including varying percent eGFR declines, with or without confirmation. For each simulated data set, we applied Cox proportional hazards regression to estimate the treatment effect corresponding to each outcome while censoring mortality. We estimated the standard errors of the Cox regression coefficients both empirically, based on the variation in the estimated coefficients across the 800 simulations, and as the root mean square of the model-based standard errors.

### Results of Analysis in Cohorts and Interpretation

The main results are the comparison of the number (prevalence) of end points for lesser versus 57% eGFR declines during a 1-, 2-, or 3-year baseline period and the strength of their association (hazard ratios [HRs]) with subsequent outcomes (Table 4). We anticipated a reciprocal relationship between the prevalence of



**Figure 3.** Results in clinical trials: ratios of established to alternative confirmed end points by intervention. Relationship between estimated treatment effects on the clinical outcome (end-stage renal disease [ESRD], glomerular filtration rate [GFR] < 15 mL/min/1.73 m<sup>2</sup>, or doubling of serum creatinine level) on the vertical axis to estimated treatment effects on the alternative end points (on the horizontal axis) aggregated for the 5 interventions. Treatment effects are expressed as hazard ratios (HRs). Diagonal line is the line of identity. Horizontal and vertical lines around each circle indicate the Bayesian credible intervals for the treatment effect on the established and alternative end points. The colors indicate intervention type: brown, renin-angiotensin system blockade versus control (A); yellow, renin-angiotensin system blockade versus calcium channel blocker (B); green, intensive blood pressure control (C); blue, low-protein diet (D); and pink, immunosuppressive therapies (E). Estimated GFR (eGFR) decline of (top left) 40% in the overall study duration, (top right) 30% in the overall study duration, (bottom left) 40% at 24 months, and (bottom right) 30% at 24 months. Reproduced from Inker et al<sup>3</sup> with permission of the National Kidney Foundation.

eGFR declines during the baseline period and the HRs of subsequent outcomes; therefore, we used a population-attributable risk (PAR) method to combine both metrics. Figure 2 shows the adjusted HR for ESRD (top panel), prevalence during the baseline period (middle panel), and PAR (lower panel) for percent decline in eGFR during the preceding 2-year baseline period in separate meta-analyses of participants with first baseline eGFRs < 60 and > 60 mL/min/1.73 m<sup>2</sup> (left and right panels, respectively). In the lower-eGFR group, a 57% eGFR decline in 2 years was associated with a very high HR (31.4), but a cumulative prevalence of this outcome of only 0.7% of participants in this interval. By contrast, a 30% decline was associated with a lower although still high HR (5.3) and occurred in 10 times as many people (cumulative prevalence of 6.5%). The 57% eGFR decline accounted for 11% of ESRD events (cumulative PAR), whereas a 30% eGFR decline accounted for 44% of ESRD events. Similar results were obtained in the higher-eGFR group and for 1- and 3-year baseline periods in both the lower- and higher-eGFR groups. Results were qualitatively consistent across studies, although heterogeneity was statistically significant. Metaregression showed no significant variation by age, diabetes status, baseline eGFR, or albuminuria. Absolute risk showed a similar pattern, but was influenced

strongly by baseline eGFR and duration of follow-up and varied more among studies than variation in HRs (Table 4). Based on these analyses, we concluded that a 30% decline over 1, 2, or 3 years is associated sufficiently strongly and consistently with ESRD to support its use as a surrogate end point. Analyses using mortality as the end point showed a similar pattern, but with lower HRs, lower PAR, and lower absolute risk than for ESRD (Table 4).

### Results of Analysis in Clinical Trials and Interpretation

The main results are the comparisons of the number of end points for lesser eGFR declines versus the established end point during the full duration of follow-up and during shorter intervals of 12, 18, and 24 months and the treatment effects (HRs for intervention vs control) using these end points (Table 4). Because there were fewer ESRD end points in the clinical trials than in the cohorts, we used a composite outcome of ESRD, GFR < 15 mL/min/1.73 m<sup>2</sup>, or doubling of serum creatinine level as the established end point for these analyses.

In principle, the significance of the treatment effect in a clinical trial reflects the magnitude of the HR and the precision with which the treatment effect is estimated. Precision is related to the number of end points. For lesser eGFR declines versus the established end

**Figure 4.** Results in simulations: summary of performance of alternative time-to-event end points for trials with 2- to 3-year durations of follow-up. Assumptions: glomerular filtration rate (GFR) based on estimated GFR from serum creatinine level, treatment effect is mixed proportional/uniform; acute effect attenuates to 0 at end-stage renal disease; follow-up duration 3 years for high GFR and 2 years for medium and low GFR. Key: acceptable type 1 error  $\approx \leq 10\%$ ; improved power =  $>15\%$  smaller samples size for lesser versus 57% GFR decline in same trial duration. Baseline GFR (mL/min/1.73 m<sup>2</sup>): high, 67.5; medium, 42.5; low 27.5. Acute effect (mL/min/1.73 m<sup>2</sup>): moderate-to-large  $>\pm 1.25$ ; small,  $>0$  to  $\pm 0.5-1.25$ ; none,  $\approx 0$  (note: moderate-to-large positive acute effects are extrapolated from results from simulations).

Acute effects	Baseline GFR high (Stage 2)		Baseline GFR medium (Stage 3)		Baseline GFR low (Stage 4)	
	40% decline	30% decline	40% decline	30% decline	40% decline	30% decline
Moderate-to-large negative						
Small negative						
None						
Small positive						
Moderate-to-large positive						
Key	Type 1 error not acceptable (false positive for benefit)		Type 1 error not acceptable (false positive for harm)		Type 1 error acceptable but power not improved	
					Type 1 error acceptable and power improved	

point, we would anticipate a larger number of end points, which would lead to improved precision. If the HR for the treatment effect were maintained for lesser eGFR declines versus the established end point, we would expect that improved precision would translate into a more significant result. However, attenuation of the HR would lead to less significant results while augmentation of the HR would lead to an even more significant result.

As expected, for any follow-up interval during the clinical trial, more participants reached a lesser eGFR decline than the established end point, and more participants reached an end point during longer than shorter follow-up intervals. Figure 3 shows the pooled ratios of HRs for eGFR declines of 40% and 30% versus the established end point for each intervention during the full duration of follow-up during the trials and during a 24-month follow-up interval. In general, the pooled ratios of the HRs were near 1.0, indicating some support for the consistency of treatment effects on lesser eGFR declines versus the established end point. However, due to the small number of high-powered trials (trials with a large number of established end points), Bayesian credible intervals (analogous to confidence intervals) often were wide, particularly in trials with participants with high baseline eGFRs, precluding definitive conclusions. For 4 of the 5 interventions (RAS blockade vs control, RAS blockade vs calcium channel blocker, intensive vs usual blood pressure control, and immunosuppressive vs other therapy), point estimates for the pooled ratios for the 30% eGFR decline versus the established end point during the same or shorter follow-up intervals were greater than 1.0, indicating an attenuation of the treatment effect for the 30% eGFR decline (HR closer

to 1). Possible causes of attenuation of treatment effects for the lesser eGFR declines for these interventions include: (1) acute effects of the intervention or the control in the direction opposite to the chronic effect, (2) proportional effects of the interventions causing larger absolute differences between treatment groups for larger GFR declines, or (3) random error in eGFR obscuring effects of interventions on smaller eGFR declines. However, for one intervention (the low- vs usual-protein diet), the point estimate for the pooled ratio for the 30% versus 57% eGFR decline was less than 1.0, indicating augmentation of the treatment effect for the lesser eGFR declines (HR farther from 1). The cause of augmentation of the treatment effect for lesser eGFR declines for this intervention is an acute effect of the low-protein diet on creatinine generation, leading to an increase in eGFR and thus a larger effect on eGFR than on mGFR. Point estimates for ratios of HRs for 40% eGFR decline versus the established end point generally were closer to 1.0, indicating greater consistency. Use of the non-confirmed end points for lesser eGFR declines resulted in a 10% to 50% increase in the number of events over that of the confirmed end points, but resulted in greater attenuation of the HRs.

In addition, the association of percent eGFR decline during a 1-year baseline period with risk of subsequent outcomes was examined using a similar analysis as for the cohorts (above). Adjusted HRs for a 30% and 40% eGFR with the established end point were 9.6 (7.3 to 12.6) and 20.3 (14.1 to 29.2), respectively. Meta-regression showed no significant variation by intervention or cause of CKD (Table 4).

Altogether, these results provide some support for the use of lesser eGFR declines as a surrogate end

**Table 6.** Circumstances in Which the Proposed Alternative Surrogates May Not Be Applicable and Potential Solutions and Unanswered Questions

Circumstances	Potential Solutions and Unanswered Questions
Effects of the interventions on non-GFR determinants of serum creatinine	<ul style="list-style-type: none"> <li>• Measure GFR</li> <li>• Measure other filtration markers (cystatin C, others)</li> </ul>
Acute effects of the intervention on GFR	<ul style="list-style-type: none"> <li>• Rely on the clinical end point (kidney failure) or require larger eGFR decline as a surrogate; both will require longer follow-up</li> <li>• Modifications to trial design on a case-by-case basis</li> </ul>
Insufficient power, requiring long follow-up intervals or large sample sizes	<ul style="list-style-type: none"> <li>• Consider non-GFR kidney disease end points, such as markers of kidney damage, specific for disease and intervention</li> </ul>
<ul style="list-style-type: none"> <li>• Slow GFR decline</li> <li>• High GFR (early stages of kidney disease)</li> <li>• Uncommon diseases</li> </ul>	
High competing risk	<ul style="list-style-type: none"> <li>• Consider composite end points including other kidney outcomes, such as mortality or other important clinical outcomes</li> </ul>
<ul style="list-style-type: none"> <li>• Slower GFR decline</li> <li>• Older population</li> <li>• Comorbid conditions</li> </ul>	

Abbreviation: eGFR, estimated glomerular filtration rate.

point, with stronger support for the 40% than 30% decline. Results suggest that the decision to use a lesser eGFR decline as an end point depends on knowledge of the effect of the treatment on the pattern of eGFR decline and on the non-GFR determinants of the endogenous filtration marker used to estimate GFR.

### Results of Simulations and Interpretation

The main results are the comparisons of type 1 error relative to the clinical outcome and power for simulated trials using lesser versus 57% eGFR declines as the end point across parameter configurations (Table 4). Because there is a consistent increase in end point events in simulations using lesser versus 57% eGFR declines, variation in the HR is the key determinant in whether type 1 error is preserved and power is improved for lesser versus 57% eGFR declines. For simulations in which the interventions had no effect on the clinical outcome (simulated HR of 1), a positive acute effect (an increase in eGFR) can lead to a type 1 error in favor of treatment (HR < 1) and false conclusion of benefit, whereas a negative acute effect (a decline in eGFR) can lead to a type 1 error against treatment (HR > 1) and a false conclusion of harm. Some increase in type 1 error of a surrogate relative to the clinical outcome beyond the targeted 5% is unavoidable, reflecting the inherent uncertainty in the use of surrogate end points, but excessive increases signify poor validity. For simplicity, we label a type 1 error of the surrogate as “acceptable” if it remains <10%, recognizing that this will vary depending on context and may differ between false conclusions of treatment benefit and of treatment harm. The type 1 error rate was higher with larger versus smaller acute effects. The type 1 error rate remained <10% for a wide range of acute effects for the 57% eGFR decline, but it was higher for moderate to large acute effects (>1.25 mL/

min/1.73 m<sup>2</sup>) for a 40% eGFR decline and for small acute effects (<1.25 mL/min/1.73 m<sup>2</sup>) for a 30% eGFR decline. For simulations in which the interventions had a beneficial effect on the clinical outcome (HR < 1), preservation of HR and improvement in power were better for lesser versus 57% eGFR declines for smaller versus larger acute effects, for higher versus lower baseline eGFRs, and for shorter versus longer trials. Figure 4 shows a synthesis of results regarding type 1 error and power for short trials using 30% and 40% versus 57% eGFR declines, according to the magnitude of acute effects and baseline eGFR. Type 1 error rate is acceptable and power is improved for both the 30% and 40% eGFR declines in the absence of acute effects (green shading). However, even small acute effects, either in the opposite direction to the hypothesized beneficial effect of the intervention (yellow shading) or in the same direction (red shading), can lead to erroneous conclusions of benefit or harm. The 40% eGFR decline is acceptable across a wider range of acute effects than the 30% eGFR decline and in many simulations provides almost as much improvement in statistical power as the 30% eGFR decline.

## RECOMMENDATIONS

### The Proposal

Based on these results, the planning committee and analytic group proposed and the workshop participants agreed that under some circumstances, a GFR decline of 30% could be a valid and useful surrogate end point for progression to kidney failure in clinical trials of CKD (Table 4). Evidence was stronger for a GFR decline of 40% as the end point, which represents a more cautious approach and is likely to be more widely applicable (Fig 4). Using the CKD-EPI 2009 creatinine



equation, a 30% and 40% decline in eGFR correspond to a 1.3 and 1.5-fold increase in serum creatinine level, respectively (Table 2). For both end points, we recommended performing a second measurement of serum creatinine at baseline and after reaching the end point to confirm the eGFR decline. For both end points, we recommended a follow-up during the trial of at least 2 to 3 years to allow a thorough evaluation of benefits and harms. However, sample sizes for 90% power often will need to be large ( $n > 1,000$ ) and the duration of follow-up often will be long ( $>3$  years), especially if baseline eGFR is high, and there are many circumstances in which these alternative end points could lead to reduced statistical power or erroneous conclusions regarding benefits or harms of interventions compared to the clinical end point of kidney failure or the established surrogate of doubling of serum creatinine level.

### Circumstances in Which the Proposal May Not Be Applicable

Effects of interventions on the non-GFR determinants of endogenous filtration markers can cause bias in GFR estimates based on that marker (Table 6). Examples for serum creatinine include interventions that affect creatinine generation from muscle (eg, low-protein diets or drugs causing muscle wasting), tubular secretion of creatinine (eg, cimetidine), or extrarenal elimination of creatinine (gastrointestinal bacterial overgrowth). Effects of interventions on non-GFR determinants of serum creatinine level should be considered during the development phase. If non-GFR determinants are detected, GFR can be measured using clearance methods or alternative filtration markers could be used.

Acute effects of the intervention on GFR pose important problems due to type 1 errors (Table 6) and all interventions should be evaluated for potential acute effects. Even a small acute effect ( $<1.25$  mL/min/ $1.73$  m<sup>2</sup>) can cause an unacceptable increase in the rate of type 1 errors for eGFR declines of 30% to 40% (Fig 4). Detection of small acute effects requires a large study population and may not be possible in phase 2 studies. Interventions with acute effects generally will require modifications to clinical trial design on a case-by-case basis. Potential solutions include using the clinical end point (kidney failure) or a larger eGFR decline as the end point; the accepted surrogate of a doubling of serum creatinine level (a 57% eGFR decline) is relatively robust to moderate acute effects, but may require a longer follow-up. Other potential solutions require further study, such as evaluation for acute effects during the trial and a prespecified adaptation of the trial design if an acute effect is detected.

Power may be insufficient for detecting a beneficial effect on eGFR declines of 30% or 40% if GFR decline

is slow, baseline GFR is high, or the disease is uncommon (Table 6). In such circumstances, other end points may need to be considered. For example, markers of kidney damage specific for the disease and intervention, such as change in albuminuria for some glomerular diseases and change in cyst volume for polycystic kidney disease, have been proposed.<sup>18,41</sup> A high competing risk from mortality also may complicate design in diseases with slow GFR decline or in older populations or populations with comorbid conditions (Table 6). If the intervention is hypothesized to reduce mortality or other adverse clinical events, it may make sense to use a composite end point that includes these events. Composite end points including other kidney outcomes, such as occurrence of CKD GFR category 4 or acute kidney injury, require more study.

### Strengths and Limitations of Our Analysis

Strengths are that the approach is based on data, using multiple sources of evidence, and a consistent analytic approach, with consistent results across subgroups based on age, sex, cause of kidney disease, and level of GFR and proteinuria, when available. The major weakness is the limited number of clinical trials available for analysis, especially with high baseline GFRs, representing a limited spectrum of interventions and limited representation of kidney diseases. Standardization of definitions and voluntary sharing of data by clinical trial groups would facilitate an update of the analyses presented here, as well as a validation in separate studies. Other limitations include the fact that explicit criteria for acceptance of a new surrogate were not defined in advance; heterogeneity among studies in some results, possibly due to variation in study populations, assays for serum creatinine, and outcome definitions; exclusion of children from the analyses due to the small number of clinical trials in children; and evaluation of only kidney failure and mortality as outcomes of CKD.

### CONCLUSION

In summary, our results support the use of alternative eGFR-based end points as a surrogate for kidney failure in clinical trials. We have analyzed a large number of cohorts and clinical trials and developed a tool to simulate outcomes for alternative eGFR-based end points based on participant clinical characteristics and trial design. We have proposed eGFR decline of 30% as an alternative surrogate end point in trials of CKD, with stronger evidence for a 40% eGFR decline. We have considered the strengths and limitations of these alternative end points and described settings in which these alternative end points may be applicable and other settings in which these alternative end points may lead to reduction in statistical power or erroneous conclusions regarding

benefits or harms of interventions. We encourage careful consideration of these alternative end points in the design of future clinical trials.

### ACKNOWLEDGEMENTS

The workshop planning committee comprised Andrew S. Levey, MD (chair), Aliza M. Thompson, MD (FDA; co-chair), Josef Coresh, MD, PhD, Kerry Willis, PhD (NKF), Norman Stockbridge, MD, PhD (FDA), Edmund Lewis, MD, Dick de Zeeuw, MD, PhD, and Alfred K. Cheung, MD. The analytical group was chaired by Josef Coresh, MD, PhD, and comprised the observational studies subgroup (Kunihiro Matsushita, MD, PhD [lead], Josef Coresh, MD, PhD, Mark Woodward, PhD, Morgan Grams, MD, MS, Yingying Sang, MS, and Shoshana Ballew, PhD), the randomized trials subgroup (Lesley A. Inker, MD, MS [lead], Christopher H. Schmid, PhD, Andrew S. Levey, MD, Hocine Tighiouart, MS, Hasi Mondal, MPH, Tonya Logvinenko, PhD, Farzad Noubary, PhD, Cassandra Becker, BS, Neal Shah, MD, Hiddo Lambers-Heerspink, PharmD, PhD, and Tom Greene, PhD), and the clinical trial simulations subgroup (Tom Greene, PhD [lead], Chia-Chen Teng, MS, Jian Ying, PhD, Andrew Redd, PhD, Mark Woodward, PhD, Lesley A. Inker, MD, MS, Josef Coresh, MD, PhD, and Andrew S. Levey, MD). John Lawrence, PhD, served as FDA representative. Breakout group discussion leaders comprised Glenn Chertow, MD, MPH (Stanford), Kai-Uwe Eckardt, MD (University of Erlangen-Nürnberg), Michael Flessner, MD, PhD (National Institute of Diabetes and Digestive and Kidney Diseases), Susan Furth, MD, PhD (Children's Hospital of Philadelphia), Ron Gansevoort, MD, PhD (University Hospital Groningen), Brenda Hemmelgarn, MD, PhD (Calgary Foothills Medical Center), Tazeen Jafar, MD, MPH (Aga Khan University), Bert Kasiske, MD (Hennepin County Medical Center), Adeera Levin, MD (St. Paul's Hospital/UBC), Julia Lewis, MD (Vanderbilt University), Vlado Perkovic, MD, PhD (The George Institute), Ron Perrone, MD (Tufts Medical Center), Michael Shlipak, MD, MPH (San Francisco Veterans Affairs Medical Center), Ravi Thadhani, MD, MPH (Massachusetts General Hospital & Harvard University), Marcello Tonelli, MD (University of Alberta), and Christoph Wanner, MD (University Hospital of Würzburg).

The participating CKD-PC cohort investigators/collaborators are listed in [Item S2](#). The participating CKD-EPI clinical trials/collaborators are listed in [Item S3](#).

We thank Tom Manley for assisting in the facilitation of the workshop and Aghogho Okparavero, MD, MPH, for assisting in preparation of the manuscript.

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

**Support:** The workshop was supported and facilitated by the NKF. NKF gratefully acknowledges Abbott, Amgen, Chemo-Centrix, Lilly, Mitsubishi Tanabe Pharma, Novartis, Pfizer, Reata, Sanofi, and Takeda, which provided grants to the NKF to support the workshop and the related publications.

**Financial Disclosure:** Dr Cheung reports receiving consulting fees from Baxter and Amgen, speaking fees from Merck, and royalties from contributions to Up-to-Date. Dr Coresh reports receiving a research grant from Amgen during the past 3 years. Dr de Zeeuw reports receiving funds to his institution from consultant agreements with Astra Zeneca, Amgen, Abbott, MSD, BMS, Novartis, VITAE, Takeda, Hemocuem, J&J, REATA, Abbott, Astellas, Abbvie, and Chemocentrix. Dr Greene reports receiving research grants from Pharmalink AB, Jansen Pharmaceuticals, Keryx Biopharmaceuticals, and Genkyotex SA. Dr Inker reports receiving research grants from Pharmalink AB and Gilead Sciences and a consulting agreement with

Otsuka. Dr Levey reports funding to Tufts Medical Center for research and contracts with the National Institutes of Health, NKF, Amgen, Pharmalink AB, and Gilead Sciences. Dr Matsushita reports receiving an honorarium from Mitsubishi Tanabe Pharm. Drs Lewis and Willis report that they have no relevant financial interests.

### SUPPLEMENTARY MATERIAL

Item S1: Conference agenda, list of breakout group topics, and workshop attendees.

Item S2: List of participating CKD Prognosis Consortium investigators/collaborators.

Item S3: List of participating CKD-EPI clinical trials/collaborators.

Note: The supplementary material accompanying this article (<http://dx.doi.org/10.1053/j.ajkd.2014.07.030>) is available at [www.ajkd.org](http://www.ajkd.org)

### REFERENCES

1. National Kidney Foundation. GFR decline as an endpoint in clinical trials for CKD. 2012. [http://www.kidney.org/professionals/research/research\\_info.cfm](http://www.kidney.org/professionals/research/research_info.cfm). Accessed November 26, 2013.
2. Coresh J, Turin TC, Matsushita K, et al. Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *JAMA*. 2014;311(24):2518-2531.
3. Inker LA, Lambers Heerspink HJ, Mondal H, et al. GFR decline as an alternative end point to kidney failure in clinical trials: a meta-analysis of treatment effects from 37 randomized trials. *Am J Kidney Dis*. 2014;64(6):848-859.
4. Lambers Heerspink HJ, Tighiouart H, Sang Y, et al. GFR decline and subsequent risk of established kidney outcomes: a meta-analysis of 37 randomized controlled trials. *Am J Kidney Dis*. 2014;64(6):860-866.
5. Greene T, Teng C-C, 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.
6. Collins AJ, Foley RN, Herzog C, et al. US Renal Data System 2012 annual data report. *Am J Kidney Dis*. 2013;61(1)(suppl 1):e1-e480.
7. Strippoli GF, Craig JC, Schena FP. The number, quality, and coverage of randomized controlled trials in nephrology. *J Am Soc Nephrol*. 2004;15(2):411-419.
8. Palmer SC, Sciancalepore M, Strippoli GF. Trial quality in nephrology: how are we measuring up? *Am J Kidney Dis*. 2011;58(3):335-337.
9. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001;69(3):89-95.
10. Kidney Disease: Improving Global Outcomes (KDIGO). KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl*. 2013;3(1):1-150.
11. Hemmelgarn BR, James MT, Manns BJ, et al. Rates of treated and untreated kidney failure in older vs younger adults. *JAMA*. 2012;307(23):2507-2515.
12. Thomas MC, Cooper ME, Shahinfar S, Brenner BM. Dialysis delayed is death prevented: a clinical perspective on the RENAAL Study. *Kidney Int*. 2003;63(4):1577-1579.
13. Uhlig K, Levey AS. Developing guidelines for chronic kidney disease: we should include all of the outcomes. *Ann Intern Med*. 2012;156(8):599-601.
14. Bedford M, Farmer C, Levin A, Ali T, Stevens P. Acute kidney injury and CKD: chicken or egg? *Am J Kidney Dis*. 2012;59(4):485-491.

15. Bellomo R, Kellum JA, Ronco C. Acute kidney injury. *Lancet*. 2012;380(9843):756-766.
16. Levey AS, Cattran D, Friedman A, et al. Proteinuria as a surrogate outcome in CKD: report of a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *Am J Kidney Dis*. 2009;54(2):205-226.
17. Stoycheff N, Pandya K, Okparavero A, et al. Early change in proteinuria as a surrogate outcome in kidney disease progression: a systematic review of previous analyses and creation of a patient-level pooled dataset. *Nephrol Dial Transplant*. 2011;26(3):848-857.
18. 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.
19. Matsushita K, van der Velde M, Astor BC, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. *Lancet*. 2010;375(9731):2073-2081.
20. van der Velde M, Matsushita K, Coresh J, et al. Lower estimated glomerular filtration rate and higher albuminuria are associated with all-cause and cardiovascular mortality. A collaborative meta-analysis of high-risk population cohorts. *Kidney Int*. 2011;79(12):1341-1352.
21. Gansevoort RT, Matsushita K, van der Velde M, et al. Lower estimated GFR and higher albuminuria are associated with adverse kidney outcomes. A collaborative meta-analysis of general and high-risk population cohorts. *Kidney Int*. 2011;80(1):93-104.
22. Astor BC, Matsushita K, Gansevoort RT, et al. Lower estimated glomerular filtration rate and higher albuminuria are associated with mortality and end-stage renal disease. A collaborative meta-analysis of kidney disease population cohorts. *Kidney Int*. 2011;79(12):1331-1340.
23. Mahmoodi BK, Matsushita K, Woodward M, et al. Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without hypertension: a meta-analysis. *Lancet*. 2012;380(9854):1649-1661.
24. Fox CS, Matsushita K, Woodward M, et al. Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without diabetes: a meta-analysis. *Lancet*. 2012;380(9854):1662-1673.
25. Hallan SI, Matsushita K, Sang Y, et al. Age and association of kidney measures with mortality and end-stage renal disease. *JAMA*. 2012;308(22):2349-2360.
26. Nitsch D, Grams M, Sang Y, et al. Associations of estimated glomerular filtration rate and albuminuria with mortality and renal failure by sex: a meta-analysis. *BMJ*. 2013;346:f324.
27. Wen CP, Matsushita K, Coresh J, et al. Relative risks of chronic kidney disease for mortality and end-stage renal disease across races are similar. *Kidney Int*. 2014;86:819-827.
28. Modification of Diet in Renal Disease Study Group; Levey AS, Beck GJ, Bosch JP, et al. 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.
29. Agodoa LY, Appel L, Bakris GL, et al. Effect of ramipril vs amlodipine on renal outcomes in hypertensive nephrosclerosis: a randomized controlled trial. *JAMA*. 2001;285(21):2719-2728.
30. 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.
31. Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Stat Med*. 1998;17(17):1923-1942.
32. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA*. 2007;298(10):1209-1212.
33. Levey AS, Inker LA, Coresh J. GFR estimation: from physiology to public health. *Am J Kidney Dis*. 2014;63(5):820-834.
34. Li L, Astor BC, Lewis J, et al. Longitudinal progression trajectory of GFR among patients with CKD. *Am J Kidney Dis*. 2012;59(4):504-512.
35. Heerspink HJ, Liang L, Smink P, Greene T, Parving HH, De Zeeuw D. Longitudinal eGFR trajectories in patients with type 2 diabetes and nephropathy [ASN abstract FR-PO155]. *J Am Soc Nephrol*. 2012;23:404A.
36. Modification of Diet in Renal Disease Study Group; Levey AS, Bosch JP, Coggins CH, et al. Effects of diet and antihypertensive therapy on creatinine clearance and serum creatinine concentration in the Modification of Diet in Renal Disease Study. *J Am Soc Nephrol*. 1996;7(4):556-566.
37. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150(9):604-612.
38. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-29.
39. 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.
40. Matsushita K, Mahmoodi BK, Woodward M, et al. Comparison of risk prediction using the CKD-EPI equation and the MDRD Study equation for estimated glomerular filtration rate. *JAMA*. 2012;307(18):1941-1951.
41. Chapman AB, Bost JE, Torres VE, et al. Kidney volume and functional outcomes in autosomal dominant polycystic kidney disease. *Clin J Am Soc Nephrol*. 2012;7(3):479-486.