**Surrogate End Points for Clinical Trials of Kidney Disease Progression**

Lesley A. Stevens,* Tom Greene,† and Andrew S. Levey*

*Division of Nephrology, Tufts-New England Medical Center, Boston, Massachusetts; and †Department of Quantitative Health Sciences, Cleveland Clinic Foundation, Cleveland, Ohio

Copyright © 2006 by the American Society of Nephrology ISSN: 1555-9041/104-0874

Published online ahead of print. Publication date available at www.cjasn.org.

Address correspondence to: Dr. Lesley A. Stevens, Tufts-New England Medical Center, 750 Washington Street, Box 391, Boston, MA 02111. Phone: 617-636-2569; Fax: 617-636-5740; E-mail: lstevens1@tufts-nemc.org

---

Chronic kidney disease (CKD) worsens over time by transitions through a sequence of stages, regardless of the specific cause of kidney damage or rate of progression (Figure 1). Recent guidelines and public health campaigns have focused on early detection and treatment of CKD on the basis of the rationale that treatments that are initiated early in the disease course will have a greater effect on slowing the progression of kidney disease and delaying the onset of kidney failure. At this time, there is a paucity of therapies to slow kidney disease progression, reflecting, in part, the difficulty in selecting appropriate end points for clinical trials to prove the benefits of new interventions.

Kidney failure (GFR <15 ml/min per 1.73 m² or the initiation of dialysis or transplantation) is an accepted clinical end point for clinical trials that evaluate the progression of CKD but is usually impractical because many CKD progress slowly and predominantly affect older individuals (Figure 2). A long duration of follow-up therefore would be required and many patients would die of other complications of CKD, such as cardiovascular disease, before reaching kidney failure. Hence, use of surrogate end points could accelerate testing of new therapies, particularly in earlier stages of CKD. The purpose of this review is to examine the concepts underlying the use of surrogate end points for kidney disease progression and review implications for clinical trial design and interpretation. First, we review the definition and that use of surrogate end points in clinical trials. Next, we discuss the use of end points that are based on GFR decline as surrogate end points for kidney failure in CKD stages 3 to 4 and their relationship to changes in serum creatinine (Sₖₑₐᵢₜᵢₜ). Finally, we examine the use of proteinuria as a surrogate end point for GFR decline in CKD stages 3 to 4 and possibly in earlier stages of CKD.

**Surrogate End Points**

The medical field is evolving rapidly with the growth of biotechnology, creating novel methods to measure and monitor disease. This has led to a great deal of excitement for use of biomarkers in clinical practice and clinical trials. Simultaneously, there is a growing awareness of the importance of rigorous analysis and testing of these markers, leading to more sophisticated nomenclature and biostatistical approaches related to evaluation of biomarkers and surrogate end points. Table 1 shows a glossary of terms for end points in clinical trials (1,2).

A clinical end point is a characteristic that reflects how a patient feels or functions or how long a patient survives, and it is the most definitive end point for a clinical trial. New therapies must have a substantial impact on a clinical end point as a prerequisite for widespread acceptance by the medical community as well as favorable review from regulatory agencies. In contrast, a biomarker is an indicator of normal biologic processes, pathogenic processes, or pharmacologic response to a therapeutic intervention. For example, biomarkers that are relevant to studies of cardiovascular disease prevention include physical signs, such as BP, or biochemical measurements, such as serum lipids. Biomarkers that are relevant to studies of CKD progression include levels of GFR, Sₖₑₐᵢₜᵢₜ, and urine protein.

We use the expression “surrogate end point” to refer to an end point that is intended to substitute for a clinical outcome in the evaluation of therapeutic efficacy in clinical trials. The key empirical criterion for the validity of a surrogate end point is that it is possible to predict the effect of the treatment on the clinical end point on the basis of the effect of the treatment on the surrogate. In most cases, surrogate end points are biomarkers, but they also may be clinical end points (e.g., hospitalizations) that are used in place of a less frequently observed target clinical end point. Surrogate end points are expected to predict clinical benefit or harm (or lack of benefit or harm) on the basis of epidemiologic, therapeutic, pathophysiologic, or other scientific evidence. For example, levels of BP and serum lipids are accepted as surrogate end points for cardiovascular disease because laboratory investigations demonstrate biologic plausibility and because of the strong relationship to cardiovascular disease in observational studies and intervention trials.

Intermediate end points are a subset of biomarkers that are on the causal pathway between the intervention and the clinical end point and can be evaluated as surrogate end points. However, an intermediate end point could fail to be a valid surrogate if an intervention affects the clinical end point by a separate causal pathway that does not include the intermediate end point. For example, ventricular arrhythmias are an intermedi-
ate end point for sudden cardiovascular death. However, sudden cardiovascular death occurs by other pathways. Lack of understanding of this key point led to the initial approval of the antiarrhythmics encainide, flecainide, and moricizine, because of their effect on ventricular arrhythmias, which later were shown to increase mortality (3).

Surrogate end points have several potential roles in the evaluation of a new therapeutic intervention (4) (Table 2): (1) Surrogate end points can help to improve the design of pilot and full-scale studies and potentially even refine the intervention itself. For example, the magnitude and the time course of the effect of a therapeutic intervention on the surrogate end point can help in choosing the dose range and titration steps, as for example in phase II trials. This use is widely accepted as a valid use for surrogate end points. (2) Surrogate end points can improve understanding of effects of the intervention. For example, the baseline level or response of the surrogate to an intervention can help to identify populations that are more or less likely to experience benefit or harm in response to the intervention. This use, too, is widely accepted when the “pop-
Table 1. Definitions of outcomes for clinical trials

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomarker (biological marker)</td>
<td>A characteristic that is measured and evaluated objectively as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.</td>
</tr>
<tr>
<td>Clinical end point</td>
<td>A characteristic or variable that reflects how a patient feels or functions or how long a patient survives.</td>
</tr>
<tr>
<td>Surrogate end point</td>
<td>A biomarker that is intended to substitute for a clinical end point, ( i.e., ) a biomarker that is expected to predict clinical benefit, harm, or lack of benefit or harm.</td>
</tr>
<tr>
<td>Intermediate end point</td>
<td>A characteristic that is intermediate in the causal pathway between an intervention and the clinical end point.</td>
</tr>
</tbody>
</table>

Table 2. Motivation for surrogate end points in clinical trials

- Replace a distal end point with a more proximal one, can be measured earlier
- Can be measured more easily or frequently
- Can be measured with higher precision or less subject to competing risks
- Less affected by other treatment modalities
- Reduced sample size requirements
- Faster decision-making

As children. Similarity of response in a surrogate end point in children to an intervention whose efficacy has been proved in adults can provide evidence for the efficacy in the new population. For example, use of clinical end points may not be possible in clinical trials in small target populations, such as children. Surrogates is problematic when clinical characteristics of the populations differ greatly. (4) Finally, when validated in previous clinical trials, surrogate end points may be used in clinical trials of new agents within established interventions, such as new antihypertensive and lipid-lowering agents. Extension to new therapeutic interventions remains controversial. For all these uses, the better the understanding of the biologic and epidemiologic relationship between the surrogate end point and the clinical end point, the greater the potential for establishing the validity of the surrogate.

Despite the potential promise of surrogate end points for clinical trials of kidney disease progression, we must be attentive to the lessons learned from other domains in which initial beliefs as to the efficacy of treatment on the basis of trials that used surrogate end points subsequently were reversed with further evidence that was based on clinical end points. The lessons of the antiarrhythmic agents are important examples that should not be forgotten. Failure to evaluate all potential pathways may lead to unexpected effects on clinical end points despite seemingly beneficial effects on the surrogate. It is critical that the introduction of any new surrogate marker be put to rigorous test.

GFR Decline as a Surrogate End Point and Its Relationship with Changes in \( S_{cr} \)

Physiology of GFR Decline

GFR is widely accepted to be the best overall marker of kidney function. Normal GFR in young men and women is approximately 130 and 120 ml/min per 1.73 m², respectively, and declines by approximately 1 ml/min per 1.73 m²/yr after age 40 yr (5,6). By definition, GFR must decline for patients to develop kidney failure. Therefore, substantial GFR decline is an intermediate outcome for the clinical end point of kidney failure, irrespective of the study population or therapeutic intervention, and an accepted surrogate end point for clinical trials. A number of clinical trials have used decline in GFR as a surrogate end point (7–10) and have highlighted a number of important issues. First, substantial GFR decline can be appreciated only late in the course of CKD (Figure 2). Second, this outcome is most useful in studies of “fast progressors,” which we define as a GFR decline >4 ml/min per 1.73 m²/yr, based on an anticipated progression from a GFR <60 ml/min per 1.73 m² to kidney failure (GFR < 15 ml/min per 1.73 m²) within 10 yr. This section reviews the physiologic basis for decline in GFR in CKD, measurement and estimation of GFR in CKD, and statistical issues related to design and interpretation of clinical trials that are based on the expression of GFR decline.

The level of GFR is a product of the number of nephrons and the average filtration rates of the nephrons, as expressed in the formula GFR = \( n \times SNGFR \), where SNGFR is the single-nephron GFR and \( n \) is the total number of nephrons in both kidneys, and SNGFR = \( K_f \times \Delta P = \text{Area} \times \text{Permeability} (\Delta P_{tr} - \Delta P_C) \), where \( K_f \) is the ultrafiltration coefficient, defined as the product of glomerular surface area (Area) available for filtration and its hydraulic permeability (Permeability), and \( \Delta P \) is the net filtration pressure, defined as the difference between the transglomerular hydrostatic (\( \Delta P_{tr} \)) and oncotic (\( \Delta P_C \)) pressure gradients.

In experimental animals, progression of kidney disease usually is defined as the loss of functioning nephrons, which
ultimately is reflected by a decrease in GFR. However, it is widely recognized that SNGFR can increase secondary to an increase in ΔPf (glomerular capillary hypertension) or in surface area (glomerular hypertrophy). These increases in SNGFR blunt the decline in GFR after reduction in nephron number, thereby decreasing the sensitivity of changes in GFR to detect the onset and progression of kidney disease.

Conversely, changes in GFR can occur as a result of hemodynamic effects on SNGFR, rather than alteration in nephron number (11). Therefore, changes in GFR might not be specific for the progression and remission of kidney disease. Conditions other than kidney disease that affect GFR include pregnancy, reduced kidney perfusion, marked surfeit or deficit of extracellular fluid volume, and nonsteroidal anti-inflammatory drugs. In addition, some of the interventions that are used to slow the progression of kidney disease may have opposite effects on SNGFR and nephron number. For example, in experimental models of CKD, strict glycemic control in diabetes, dietary protein restriction, strict BP control, angiotensin-converting enzyme inhibition, and angiotensin receptor blockade all cause short-term reductions in SNGFR but slow the loss of nephrons. By contrast, treatment with dihydropyridine calcium channel blockers induces an increase in SNGFR but does not slow the loss of nephrons. In clinical trials, effects on GFR that are apparent shortly after initiation of therapy usually are attributed to hemodynamic effects on SNGFR, whereas long-term changes in GFR usually are attributed to loss of nephrons. However, hemodynamic effects may persist throughout therapy; therefore, the final change in GFR would reflect changes in both SNGFR and nephron number. In humans, neither the number of nephrons nor SNGFR can be measured, making it difficult to interpret small changes in GFR.

Measurement and Estimation of GFR

It is difficult to measure GFR. Urinary or plasma clearance of an exogenous filtration marker is the most accurate method but requires time and expense for preparation of the marker and patient, a skilled technician, specialized laboratory assays, and exposure to radiation for some filtration markers. In clinical practice and in most research studies, creatinine is used as an endogenous filtration marker to estimate the level of GFR. Urinary creatinine clearance (Ccr) approximates GFR, but timed urine collections are inconvenient and fraught with error. The steady-state Scr level is inversely proportional to the GFR. The relationship between the levels of Ccr and Scr to GFR is as follows:

\[
GFR = \frac{C_{cr} - T_{S_{cr}}}{S_{cr}}
\]

\[
GFR = \frac{(G_{cr} - T_{S_{cr}} - E_{cr})}{S_{cr}}
\]

where G is the generation, TS is tubular secretion of creatinine, and E is the extrarenal elimination, expressed as rates (e.g., mg/min). Estimating equations, such as the four-variable Modification of Diet in Renal Disease (MDRD) Study equation (12,13), use demographic variables as indices for creatinine generation:

\[
GFR = 186 \times S_{cr}^{-1.154} \times \text{age}^{-0.203} \times 1.212 \times 0.742 \text{ (if black) } \times 0.742 \text{ (if female)}
\]

where GFR is expressed as ml/min per 1.73 m², Scr is expressed as mg/dl, and age is expressed as years.

In general, changes in Ccr, reciprocal of Scr, and GFR estimated from Scr parallel the changes in GFR, and a number of clinical trials have used changes in Scr as an index for changes in GFR (14–16). Because the terms in estimating equations other than Scr usually are approximately constant over time, the choice of estimating equation typically has little affect when the effects of treatments on longitudinal change are evaluated.

As discussed next, analyses of end points that are defined by creatinine-based GFR estimates are affected by the same limitations as those that affect end points that are defined by measured GFR. In addition, end points that are derived from creatinine-based GFR estimates may be biased if the interventions directly or indirectly affect creatinine generation, secretion, or extrarenal elimination. For example, in the MDRD Study, dietary protein restriction had significant effects to lower creatinine secretion and generation, as well as effects on GFR (17). By the end of the study, the low-protein diet had significantly slowed the mean decline in Ccr and the mean rise in Scr but did not significantly affect the mean decline in GFR. If GFR estimates are based on Ccr or Scr, then the effects of the intervention on creatinine generation, secretion, and extrarenal elimination should be explored before finalization of study design.

Slope-Based End Points in Clinical Trials

One intuitive approach to evaluating effects of treatment interventions would be to compare the average rate of change in GFR over time (slope) between the treatment groups. This approach has been applied both to measured and estimated GFR. When the end point is based on sequential measures or estimates of GFR, the comparison of the mean slope between the treatment groups provides greater statistical power than any alternative method of analysis under two key assumptions: (1) the mean rate of GFR decline is constant during the follow-up interval in each treatment group, and (2) the effect of the treatment effect on the clinical end point is the same, regardless of the patient’s underlying rate of disease progression (18,19). However, both assumptions often are violated.

As discussed earlier, many interventions of interest produce short-term “acute” hemodynamic effects on GFR early after randomization, which differ from their long-term “chronic” effects, violating the first condition. When this occurs, the early hemodynamic effect may lead to an ambiguous pattern in which the comparison of the long-term “chronic” GFR slopes starting several months after randomization differs from the comparison of the total mean GFR decline from baseline to the end of the trial. Consider, for example, the six possible scenarios in Figure 3. Among these, only A and D represent definitive comparisons in favor of one of the treatments being compared (20). This danger of such ambiguity has been indicated aptly by past clinical trials; the primary comparisons of both the usual versus low-protein diet and usual versus low BP goal in the MDRD Study were inconclusive, as in Figure 3C, and the primary comparisons of amlopidine versus metoprolol and amlopidine versus ramipril in the African American Study of Kid-
The statistical power of slope-based analyses can be compromised when the treatment effect is proportional to the underlying rate of disease progression (i.e., not uniform) and the overall mean rate of progression is slow. In this situation, the larger treatment effect in the fast progressors is diluted by the smaller treatment effect in the slow progressors, leading to a smaller difference between mean slopes and less statistical power, especially when the underlying rate of progression is slow. This effect was observed clearly in the MDRD Study, in which the treatment effect of the low-BP intervention was greater in patients with higher levels of proteinuria at baseline (faster progressors) (20). In addition to these fundamental problems, there also are technical challenges with slope-based analyses. Slopes that are computed for patients with short follow-up times tend to be imprecise and highly variable, necessitating the use of specialized statistical approaches such as the use of mixed-effects models to avoid an unacceptable loss of statistical power (18,21). These approaches provide more precise estimates of mean slopes by assigning a greater weight to patients with longer follow-up times, thereby deemphasizing the imprecise slopes for patients with shorter follow-up. A common complication is the presence of informative censoring that results from attrition as a result of kidney failure, death, or loss to follow-up before the scheduled end of the study (22,23). Because patients who reach these events tend to have a shorter follow-up time

**Figure 3.** Possible scenarios for the comparison of two treatment groups. Shown are six alternative scenarios for the effects of a treatment compared with a reference group on the mean change in GFR from baseline to 4 yr using the two-slope model. In all panels, the vertical axis is GFR and the horizontal axis is time. The chronic slope is depicted by the slope from 3 mo (1/4 yr) to 4 yr and the total mean slope as the average rate of change from baseline to 4 yr. (A and D) Definitive scenarios in which the comparisons between treatment groups of the mean chronic and total slopes are in agreement. (B, C, E, and F) Inconclusive scenarios in which the comparisons of the chronic and total means slopes are not in agreement.
than those who complete the study, the increased weighting in common statistical methods of patients with longer follow-up can lead to biased estimates. Statistical models have been developed for analysis of slopes in the presence of informative censoring (1,23–25) but at the cost of requiring more complex and less transparent analytic methods.

As a result of these limitations, we recommend caution in the use and interpretation of slope-based outcomes that are defined from GFR or $S_{eGFR}$. Use of slope-based outcomes should be restricted to settings in which (1) it is known that the interventions do not produce hemodynamic effects and (2) either the hypothesized treatment effect is uniform (i.e., unrelated to the patients’ underlying rates of progression) or the study population consists predominantly of “fast progressors” (e.g., GFR decline $\geq 4$ ml/min per 1.73 m$^2$ per year).

**Time-to-Event End Points**

Time-to-event analysis also is known as survival analysis and refers to the method of measuring the rate of occurrence of discrete outcomes. When sequential measurements or estimates of GFR are obtained, one can analyze the time from baseline until the first recorded decline to a predesignated level. Examples of time-to-event outcomes include time from randomization until kidney failure, a 50% reduction in GFR, a doubling of $S_{eGFR}$, or a decrease in GFR or increase in $S_{eGFR}$ to a designated threshold. In principle, a large decline in GFR would be similar to the occurrence of kidney failure, the clinical end point. However, smaller declines in GFR also can be defined as a surrogate end point. Composite end points that incorporate the occurrence of kidney failure as well as these lesser declines in GFR usually are used to account for patients who develop kidney failure without first recording change in GFR by the amount required to trigger an event.

In contrast to slope-based analyses, the time-to-event analyses are sensitive primarily to the treatment effect in the subset of patients with faster disease progression. This is illustrated in Figure 4, in which GFR versus time curves are drawn for several patients with varying rates of GFR decline. Patients with slow progression usually will reach the end of the study without attaining the threshold for an event and therefore are censored from the analysis. The results of the time-to-event analysis reflect the treatment effect in patients who experienced events, which tend to be the patients with more rapid disease progression. When the treatment effect is hypothesized to be proportional to the underlying progression rate, the time-to-event analysis focuses on these patients with the largest hypothesized effect. This avoids the dilution of the mean treatment effect that is incurred in slope-based analyses by the inclusion of slow progressors in computation of mean slopes.

The focus of the time-to-event analysis on patients who have large changes also reduces the sensitivity of the results to confounding by hemodynamic effects. This is because the hemodynamic effects typically are $<5$ ml/min per 1.73 m$^2$ (7,17) and thus constitute a relatively small fraction of the total GFR decline in patients who exhibit large enough declines to trigger events. The relative contribution of the hemodynamic effect to the total GFR decline is greater in slow progressors, but these patients are censored from the analysis. A further advantage of the time-to-event outcome is that large changes in kidney function (e.g., a 50% decline in GFR, a doubling of $S_{eGFR}$) are conceptually closer approximations to the target clinical end point of the occurrence of kidney failure than is GFR slope.

For all these reasons, time-to-event end points that are based on large GFR declines generally can be expected to be more reliable surrogate end points than slope-based end points. However, there are limitations to time-to-event analyses. First, the treatment effect principally reflects differences in fast progressors, so generalizability of the results to slow progressors may be unclear. Second, a high risk for competing events, such as death as a result of cardiovascular disease, can lead to a high rate of attrition and incur the risk of informative censoring (i.e., patients who would have reached kidney events were not observed to do so because they died first). This difficulty can be addressed by expanding the definition of the composite end point to include death as an event but with the recognition that the composite outcome then may reflect a combination of kidney disease progression and other diseases. Third, the length of time that is required to accrue enough end points to achieve adequate statistical power may be excessively long in study populations with a slow average progression rate. This limitation is especially important in populations with early stages of CKD, in which low event rates usually make time-to-event analyses that are based on GFR declines infeasible.

We suggest consideration of CKD stage 4 as a surrogate end point for kidney failure. The rationale is that CKD stage 4 is more common than kidney failure but has the same importance as a clinical event: It is accompanied by important clinical complications and requires a distinct change in the nature and the degree of care delivered (26). Recent studies show a higher prevalence of all complications and higher risk for morbidity and mortality for CKD stage 4 (27,28). In addition, interventions during this stage may reduce subsequent morbidity and

![Figure 4](image-url)
Proteinuria as a Surrogate End Point
Proteinuria in CKD shares many of the features of accepted surrogate end points in other domains, including biologic plausibility and a strong relationship to the clinical end points. However, unlike GFR decline, an increase in proteinuria is not a necessary intermediate end point on the path to kidney failure and therefore requires rigorous evaluation as to its role as a surrogate end point for trials of kidney disease progression.

Proteinuria is a marker of kidney damage, and some have argued that it is one of the pathogenic processes that lead to the progression of kidney disease independent of cause. Proteinuria also is strongly correlated with GFR decline, and changes in proteinuria after an intervention often occur earlier and by a larger amount than changes in GFR. Therefore, a change in proteinuria is a reasonable candidate as a surrogate end point for kidney disease progression in clinical trials. Establishment of proteinuria as a valid surrogate end point for progression of CKD would have substantial implications for clinical research, potentially shortening the duration of follow-up and reducing the number of patients who are required for clinical trials of CKD stages 3 and 4 and enabling clinical trials in CKD stages 1 to 2, in which use of GFR decline as a surrogate end point is not feasible (Figure 2). In this section, we first outline the biologic rationale for changes in proteinuria as a surrogate marker for kidney disease progression and then outline the statistical approaches that may be used to evaluate the validity of this surrogate.

Physiologic Basis for Proteinuria and Its Relationship to Kidney Disease Progression

Proteinuria as a Marker of Kidney Damage. Under normal circumstances, only low molecular weight substances gain access to the urinary space; large molecules are excluded by an intact glomerular barrier. Glomerular injury leads to increased glomerular permeability to macromolecules, such that albumin and other large serum proteins are able to gain access to the mesangium and tubular fluid. In some morphologic studies, the magnitude of proteinuria or albuminuria is correlated with the magnitude of severity of glomerular and tubulointerstitial damage, as well as GFR decline (32,33). It is well established that proteinuria is part of the natural history of many CKD. Proteinuria also may reflect generalized systemic endothelial injury in addition to kidney disease.

Proteinuria as a Cause of Kidney Disease Progression. There is accumulating experimental evidence linking proteinuria to worsening kidney damage and GFR decline. There are two main hypotheses for the mechanism by which this may occur: Mesangial toxicity and tubular injury. Accumulation of toxic molecules in the mesangium may cause further glomerular injury and glomerulosclerosis. Accumulation in tubular fluid is followed by reabsorption and endocytosis in proximal tubule lysosomes, causing tubular injury and interstitial fibrosis (32,34–36). However, other experimental work suggests that proteinuria is not damaging in the absence of extension of the initial glomerular damage and that tubular damage is confined to damaged nephrons (37).

Empirical Evaluation of Proteinuria as a Surrogate End Point
Next, we outline three general statistical approaches that can be used to evaluate the validity of proteinuria as a surrogate end point for kidney disease progression.

Individual-Level Association. The first approach is to examine the relationship between proteinuria and clinical outcomes in individual patients in observational studies or clinical trials. Initial proteinuria has been found to be an independent predictor of both faster GFR decline and progression to kidney failure in studies of patients with diabetes, specific glomerular diseases, polycystic kidney disease, and other kidney diseases not otherwise specified (9,38–42). In most studies, proteinuria had the strongest association with kidney disease outcomes among all factors tested. In addition, reductions in proteinuria after administration of treatments in randomized trials have been found consistently to be associated with slower GFR decline and lower risk for progression to kidney failure (14,15).

Demonstration of a strong and consistent association between proteinuria and progression of kidney disease provides preliminary support to the validity of proteinuria as a surrogate end point. However, these analyses cannot demonstrate that the effect of a treatment on the change in proteinuria predicts the effect of the same treatment on progression of kidney disease. Even in the presence of a strong association between proteinuria and progression, there are numerous potential mechanisms under which the relationship between treatment effects on proteinuria and treatment effects on progression could break down. Figure 5A depicts a scenario in which proteinuria is on the causal pathway to kidney disease progression and the treatment causes a change in protein; however, the treatment also effects progression of kidney disease by a separate causal pathway separate from proteinuria. Depending on the direction and the size of the effect through this second pathway, the treatment effect on progression may differ markedly from its effect on proteinuria. Alternatively, as in the scenario depicted in Figure 5C, the treatment does affect both proteinuria and progression, but these effects are unrelated to one another. Here, the association between proteinuria and progression is the result of effects of other factors that are unrelated to the treatment effects.

Prentice Criteria. Formal statistical criteria for validation of surrogate end points first were proposed in a landmark paper by Prentice (43). We refer the reader to the original reference and subsequent publications for a detailed description of these criteria. In practice, the Prentice criteria are operationally evaluated by testing whether the treatment effect on the clinical end point is reduced to zero after statistical adjust...
ment for the surrogate. Under certain models for causal inference, establishment of this operational criterion would demonstrate the absence of a causal pathway independent of proteinuria between the treatment and the true clinical end point (Figure 5B) (44). The importance of such alternative pathways is underscored by past failures of surrogate end points in which the adverse effects of the treatment on the clinical outcome eventually were traced to such alternative pathways. These differences may be related, in part, to the different analytic approaches to establishment of the relationships between treatment, proteinuria, and outcomes in the individual studies and likely reflect low precision with which the PTE can be determined within a single study (43).

There are other serious limitations to the application of the Prentice criteria or the PTE for establishing the validity of surrogate end points, including proteinuria. First, if PTE values <1.0 are accepted as supporting the validity of the surrogate, then this means overlooking potential adverse effects through alternative causal pathways. Second, when the surrogate end point has substantial measurement error (as is the case with proteinuria), the standard methods for statistical adjustment that have been used in past studies tend to undercorrect for the surrogate, leading to underestimation of the PTE. Most important, recent work under formal statistical frameworks for causal modeling has clarified that even exact fulfillment of the Prentice criterion (or a PTE of 1.0) fails to establish the validity of the surrogate except under the unlikely assumption that there are no confounding factors that influence both the surrogate and the clinical end point (Figure 5D) (46–48). In the case of proteinuria, it is plausible that genetic or other biologic factors influence both the initial change in proteinuria and the rate of progression to kidney failure. If a confounding factor is associated both with an increase in proteinuria and with faster progression, then failure to control for that factor in the analysis will bias the PTE upward, leading to a false impression that the surrogate accounts for a high proportion of the total treatment effect.

**Figure 5.** Possible scenarios for the relationship among treatment, change in proteinuria, and clinical end points. (A) Treatment affects both change in proteinuria and clinical end points, but there are separate causal pathways. (B) Treatment effect is mediated only through reduction in proteinuria. (C) Treatment affects both proteinuria and progression, but these effects are unrelated to one another. (D) Treatment affects both proteinuria and progression, and there are no confounding factors that influence both.

**Trial-Level Association.** The trial-level approach is a recent development in the statistical approach in validation of surrogate markers, which directly evaluates the association between treatment effects on the surrogate end point and treatment effects on the clinical end point at the trial level (49–51). This approach requires a joint analysis or a meta-analysis of multiple randomized trials. A regression model is developed from previous randomized trials to predict the treatment effect on the clinical end point from the treatment effect on the surrogate, as illustrated in Figure 6. This produces an equation for estimating the treatment effect on the clinical end point from the treatment effect on the surrogate that can be applied to a new study within the same class of drugs in which the clinical end point is not available. The method also gives a confidence interval for the treatment effect on the clinical end point to convey the uncertainty in this estimate. The major advantage of
The trial-level approach is that the estimated treatment effect on both the surrogate and the clinical end points are based on randomized comparisons, thereby reducing the risk for bias from confounding that confronts evaluations of individual-level association and the Prentice criteria.

The trial-level approach has not yet been implemented for the evaluation of proteinuria as a surrogate end point, but a large number of randomized trials in CKD with assessments of proteinuria and progression end points potentially are available. A joint analysis of these trials using the trial-level approach should do much to clarify the extent to which treatment effects on proteinuria have in fact predicted treatment effects on progression in studies to date. Still, the trial-level approach also has limitations. The most important limitation is that the method assumes that the previous studies in which the model is fit are representative of a new study to which the surrogate end point is to be applied. Extrapolation to studies with substantially different features than the original validation studies will entail an additional level of uncertainty beyond that captured by the error terms in the statistical model and must rely primarily on biologic arguments. Because many randomized trials in CKD have evaluated interventions that were designed to block the renin-angiotensin-aldosterone system, the trial-level approach would be most suited for evaluation of proteinuria as a surrogate end point for future trials that test interventions that are hypothesized to act through this particular mechanism. From analyses of data, one is unable to extrapolate to other interventions without explicit testing. The limited number of trials that test interventions that act through other mechanisms could provide a limited assessment of the validity of proteinuria for the more general setting in which alternative mechanisms are hypothesized.

Where to Go from Here?

Until further data are available on the validity of a change in proteinuria as a valid surrogate end point for kidney disease progression, a number of strategies for clinical trials of new interventions could be considered. For example, concurrent trials of the same intervention using end points that are based on change in GFR or Scr in patients with CKD stages 3 to 4 and change in proteinuria in CKD stages 1 to 2. Alternatively, end points that are based on change in GFR or Scr could be used for a large-scale trial with wide entry criteria, and changes in proteinuria could be used as a surrogate for predefined tests for effect modification within subgroups.

Recent revisions to Food and Drug Administration regulations for the approval of new drugs ("Subpart H") allow use of surrogate end points in a clinical trial for initial approval of a drug but require subsequent postmarketing testing to verify and describe both long-term clinical benefits and adverse outcomes (52). To apply this regulation to trials of kidney disease, empirical evaluation of proteinuria as a surrogate end point first must be performed along the steps outlined above. The National Institute of Diabetes and Digestive and Kidney Dis-

---

### Table 3. Proportion treatment effect explained

<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention</th>
<th>Index</th>
<th>Outcome</th>
<th>PTE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIPRD (nondiabetic kidney disease, 11 studies)</td>
<td>Antihypertensive regimens with \textit{versus} without ACE inhibitors</td>
<td>$\Delta$ total protein</td>
<td>Kidney failure alone</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kidney failure or doubling of serum creatinine</td>
<td>26</td>
</tr>
<tr>
<td>AASK (hypertensive nephrosclerosis)</td>
<td>Ramipril \textit{versus} amlodipine</td>
<td>$\Delta$ total protein</td>
<td>Death or kidney failure</td>
<td>41</td>
</tr>
<tr>
<td>RENAAL (type 2 diabetes)</td>
<td>Losartan \textit{versus} placebo</td>
<td>$\Delta$ albumin</td>
<td>Death, kidney failure or doubling of serum creatinine</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kidney failure alone</td>
<td>51</td>
</tr>
<tr>
<td>IDNT (type 2 diabetes)</td>
<td>Irbesartan \textit{versus} placebo</td>
<td>$\Delta$ albumin</td>
<td>Kidney failure</td>
<td>36</td>
</tr>
</tbody>
</table>

*Note: AIPRD, ACE Inhibitors in Progressive Renal Disease; AASK, African American Study of Kidney Disease; ACE, angiotensin-converting enzyme; IDNT, Irbesartan in Diabetic Nephropathy Trial; PTE, proportion of the treatment effect; RENAAL, Reduction of Endpoints in NIDDM with the Angiotensin II Antagonist Losartan.*

---

**Figure 6.** Trial-level approach: Ideal hypothetical example. Dots represent a single study or groups within a study. Thick black line represents the regression line depicting the relationship between change in proteinuria and treatment effect on clinical end points. Vertical gray line represents the point estimate with confidence interval for a treatment effect associated with change in proteinuria. This is for illustration purposes only and does not reflect any real relationships.
eases (NIDDK) is currently funding a research group, Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI), to do so. CKD-EPI will use individual patient meta-analysis and the rigorous three-step approach described above to evaluate the strength of the evidence. Nonetheless, validation of proteinuria as a surrogate end point is likely to be intervention specific, with uncertain extrapolation to other interventions with different hypothesized mechanisms.

Acknowledgment
This study was supported by grant UO1 DK 053869 from NIDDK.

References


52. Food and Drug Administration: *Approval Based on a Surrogate Endpoint or on an Effect on a Clinical Endpoint Other than Survival or Irreversible Morbidity*, Bethesda, Department of Health and Human Services, 2005, pp 125–132