Change in albuminuria and GFR slope as joint surrogate endpoints for kidney failure - Implications for phase 2 clinical trials in chronic kidney disease

Contents

Database management

For our prior work investigating surrogate endpoints, we had performed a systematic search of the Medline database from January 1, 1946 to May 15 2007[1]. To update this dataset for the current analysis, we repeated our systematic search beginning May 16 2007 when the initial search had been completed and ending in December 15, 2016 (see Figure A). Key inclusion criteria were quantifiable measurements of albuminuria or proteinuria at baseline and within 12 months of follow-up and information on ESKD incidence. Risks of bias for each RCT were assessed using the risk-of-bias tool of the Cochrane collaboration, as described in our previous publication[2].

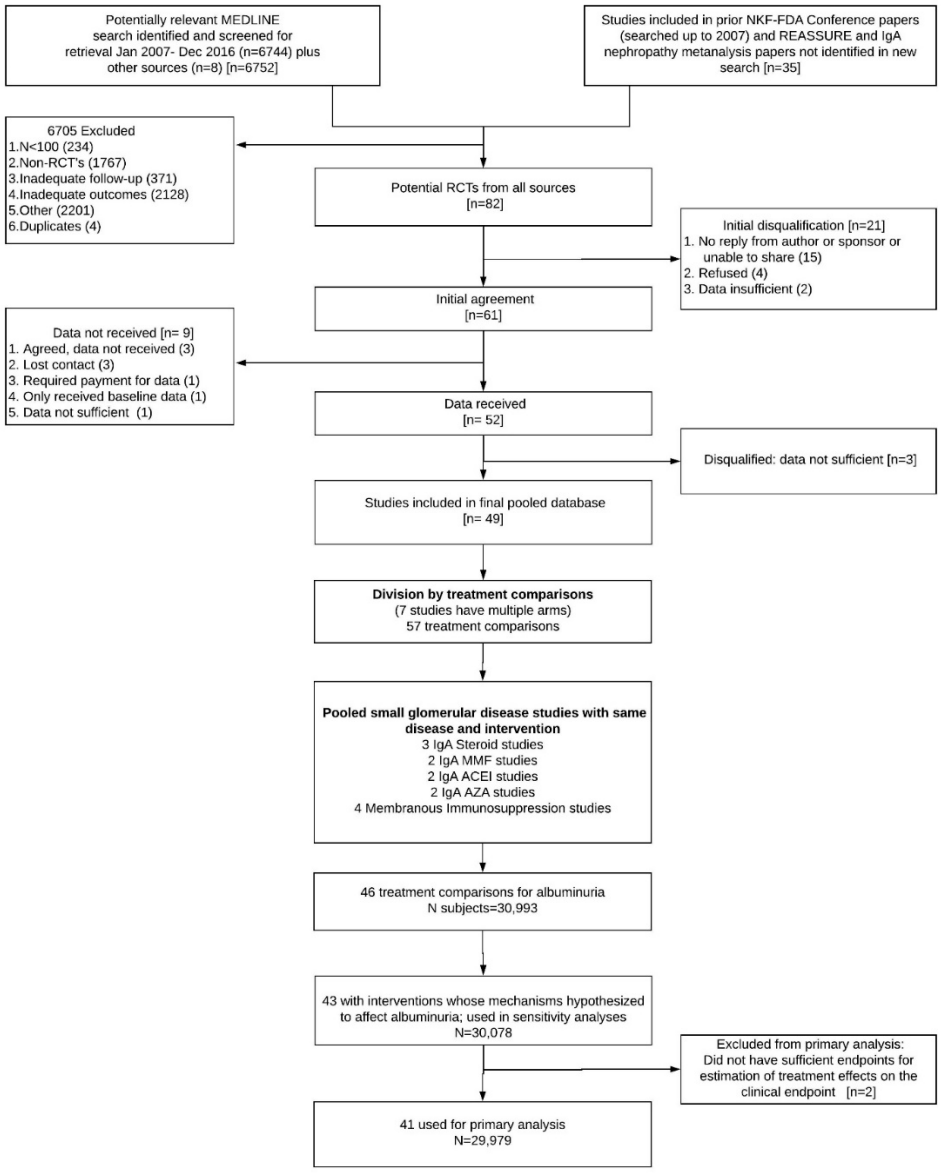


Figure A. Data acquisition flow chart

For each study, we defined the active treatment as the treatment hypothesized to produce the greater reduction in the risk of the clinical endpoint. We categorized the studies by intervention type: renin angiotensin system blockade (RASB) vs. control, RASB vs. calcium channel blocker (CCB), intensive blood pressure control, low protein diet; immunosuppressive therapy (including steroid, azathioprine, tacrolimus, fish oil, plasmapheresis). We categorized disease as diabetes

(studies of people with diabetes not restricted to CKD, and studies of diabetic kidney disease), glomerular disease and other CKD (other causes or cause not specified).

As previously described, if the study defined censoring dates were not available we approximated them as the time from randomization to the final recorded visit date in the data provided plus 6 months plus the study-specific 90th percentile of the average interval between visits with serum creatinine measurements[3-18]. The purpose of adding 6 months to the estimated right censoring date is to retain a higher proportion of clinical outcome events which occurred following the patient's final study visit. We included events event time occurred prior to 1 month following administrative censoring time. Patients who had events but no visits were included if event occurred before 12 months.

Statistical Methods

The first part of this statistical methods supplement provides the rationale for adopting a Bayesian framework for evaluation and application of surrogate endpoints in clinical trials, and provides additional detail on the analyses of GFR slope that were performed within each trial.
The second part of this supplement describes the analytic details for the six steps in the approach that were outlined in the statistical methods section of the manuscript. The six steps, which are summarized schematically in Figure 1 of the main manuscript, begin with Bayesian trial-level meta-regression analyses to characterize the relationship between the three endpoints – UACR, GFR slope, and the clinical endpoint – in previously conducted RCTs, and finish with the design and interpretation of results of a newly conducted Phase 2 clinical trial.

Part 1A: Rationale for Bayesian Approach.

The trial-level meta-regression approach for characterizing the performance of surrogate endpoints for predicting the treatment effect on the clinical endpoint has been implemented under both frequentist[22, 23] and Bayesian frameworks[24, 25]. We use the Bayesian framework for this research primarily for three reasons:
a) The Bayesian framework regards the true hazard ratio for the treatment effect on the clinical endpoint in a newly conducted trial as random variable, thereby allowing intuitive and interpretable statements concerning the posterior probability that the true hazard ratio falls in designated ranges of clinical interest.
b) The Bayesian approach provides a natural and principled way of combining prior information, available from the results of previous studies, with the observed data on the surrogate endpoints in a newly conducted study to perform statistical inferences concerning the effect of the treatment in the new study.
c) The availability of Monte-Carlo Markov Chain (MCMC) sampling under the Bayesian framework solves several numerical and analytic challenges that complicate maximum likelihood inference under a frequentist framework. For example, methods for frequentist inference for the multilevel mixed effects models used in this work require assumptions of large sample size which are not always satisfied, particularly in subgroup analyses with relatively small numbers of trials. In certain analyses, the maximum likelihood estimates for one of the variance components in the model can be equal to 0, which complicates statistical inference under the frequentist approach. This challenge is avoided in the Bayesian approach by using diffuse prior distributions which assign a probability of 0 to a variance component equal to 0. Finally, MCMC sampling provides straightforward statistical inference on any function of the model parameters.

Part 1B: Details on Analyses of GFR Slope Within Each Trial.

As described in the primary manuscript and elaborated below, our analyses depend on estimates of the effects of the treatment on GFR slope within each trial. This section of the supplement provides additional details on the analysis of GFR slope.

We used a simplified linear mixed effects model based on a single slope starting at three months post randomization adjusted for baseline GFR. Under this model, the trajectories of the mean GFR level are characterized for each treatment group by the mean intercepts at 3 months which indicate the acute changes in the mean GFR level over the first 3 months of follow-up, and by the mean slopes after 3 months, which we refer to as the mean chronic slopes in the two treatment groups. The mean rate of GFR change over a designated follow-up period which includes *both* the first 3 months and later follow-up after 3 months is referred to as the mean total slope over the designated period and is computed using a weighted average of the mean acute and chronic slopes. The treatment effects on the acute, chronic and total slopes are estimated as the differences in the respective mean slopes between the randomized treatment groups. In this manuscript, we focus on the mean treatment effects on the chronic slope, beginning at 3 months follow-up.

Our shared parameter mixed effects model accounted for between-subject variability in GFR trajectories by inclusion of random slopes and intercepts, for greater variation in individual GFR measurements at higher GFR using a power of the mean (POM) model, and for non-uniform treatment effects in which treatments slowed progression by a greater extent among patients with faster GFR decline than for patients with slower GFR decline by allowing for different between-patient slope variances in the treatment and control groups[19]. In studies in which at least 15 subjects died or reached ESRD, we accounted for informative censoring by these events by nesting the mixed model for the GFR measurements within a shared parameter model in which the risk of ESRD or death was assumed to be related to the random slopes and intercepts of the GFR part of the model[20, 21]. Simplified models were used in cases where convergence could not be obtained with the full model. The full shared parameter mixed effects models were fit using the SAS (version 9.4) nonlinear mixed-effects regression procedure, NLMIXED.

Part 2: Six steps to applying a trial-level meta-regression of previous trials to the design and analysis of a new randomized trial

We provide below the analytic details for the six steps of our analyses that were documented in the methods section and overviewed in Figure 1 of the primary manuscript. Step 1 consists of the Bayesian meta-regressions which characterize the joint distribution of the treatment effects on the surrogate and clinical endpoints. In Step 2, the results of the meta-regressions are used to construct a prior distribution for the relationship between the treatment effects on the three endpoints for application in a newly conducted phase 2 trial. Standard errors for the estimated treatment effects on the two surrogate endpoints are obtained for candidate designs for the phase 2 trial in Step 3. Step 4 uses a Bayesian framework to apply the prior distribution from Step 2 and the standard errors from Step 3 to compute the posterior probabilities of a benefit on the clinical endpoint (which we refer to as the trial-level positive predictive value, denoted PPV_{trial}) across a grid of hypothetical values for the observed treatment effects on the two surrogate endpoints. Step 5 presents frequentist characteristics of PPV_{trial} under each candidate phase 2 design under consideration. In this step, PPV_{trial} is interpreted as a sample statistic whose frequentist characteristics can be evaluated based on its sampling distribution conditional on the true treatment effects on the surrogate endpoints. Finally, in Step 6 the posterior probability of clinical

benefit is calculated from the actual results of the phase 2 trial under the prior distribution from Step 2.

Step 1. Conduct meta-regressions of previous trials.

Bayesian meta-regression models: We used a pair of Bayesian models to characterize the relationship among treatment effects on three endpoints UACR, GFR slope, and the clinical endpoint – across the previously conducted RCTs which are indexed by $i = 1,2, \ldots, 41$. We express the treatment effects on the clinical endpoint as log hazard ratios, denoted θ_i. We express treatment effects on GFR slope as a mean difference in slopes, expressed in ml/min/1.73m^2/yr and denoted as γ_{1i}, and express treatment effects on proteinuria as a log transformed ratio geometric means, denoted by γ_{2i}.

As described in the statistical methods, the estimated treatment effects on these three endpoints and their associated standard errors were obtained by performing separate intent-to-treat (ITT) analyses in each trial, using Cox regression for the clinical endpoint, the shared parameter mixed effects model described above for GFR slope, and analysis of covariance for UACR which was analyzed on the log scale. In these analyses we also obtained robust sandwich estimates of the correlations in the sampling errors between the three endpoints in addition to their standard errors. We use the notation $\hat{\theta}_i, \hat{\gamma}_{1i}$, and $\hat{\gamma}_{2i}$ to represented the estimated treatment effects on the three endpoints for the i^{th}trial, with θ_i, γ_{1i}, and γ_{2i} representing the corresponding true effects. Using the estimated treatment effects, their associated standard errors, and the correlations between the sampling errors across the three endpoints as inputs, we then used a 2-stage mixed effects model to characterize the relationships between the treatment effects on the three endpoints across the previously conducted RCTs. The first stage model stipulates the estimated treatment effects on the three endpoints follow a multivariate normal distribution conditional on the true effects within each trial, as expressed in equation A1 below.

$$\begin{bmatrix} \hat{\theta}_i \\ \hat{\gamma}_{1i} \\ \hat{\gamma}_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} \theta_i \\ \gamma_{1i} \\ \gamma_{2i} \end{bmatrix}, \begin{bmatrix} \sigma_i^2 & r_{\theta 1i}\sigma_i\delta_{1i} & r_{\theta 2i}\sigma_i\delta_{2i} \\ r_{\theta 1i}\sigma_i\delta_{1i} & \delta_{1i}^2 & r_{12i}\delta_{1i}\delta_{2i} \\ r_{\theta 2i}\sigma_i\delta_{2i} & r_{12i}\delta_{1i}\delta_{2i} & \delta_{2i}^2 \end{bmatrix} \right). \qquad (EQ\ 1)$$

The assumption of multivariate normality for the conditional distribution of the estimated treatment effects given the true treatment effects is justified to a close approximation by the central limit theorem. The variance terms in the covariance matrix are the squared standard errors of the estimated treatment effects. For simplicity, the analyses of this report ignore the sampling error in the covariance matrix for the estimated treatment effects within each trial.

The second stage of our mixed model stipulates a multivariate normal model for the true effects on the three endpoints:

$$\begin{bmatrix} \theta_i \\ \gamma_{1i} \\ \gamma_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_\theta \\ \mu_{\gamma 1} \\ \mu_{\gamma 2} \end{bmatrix}, \begin{bmatrix} \sigma_\theta^2 & R_{\theta\gamma 1}\sigma_\theta\sigma_{\gamma 1} & R_{\theta\gamma 2}\sigma_\theta\sigma_{\gamma 2} \\ R_{\theta\gamma 1}\sigma_\theta\sigma_{\gamma 1} & \sigma_{\gamma 1}^2 & R_{\gamma 1\gamma 2}\sigma_{\gamma 1}\sigma_{\gamma 2} \\ R_{\theta\gamma 2}\sigma_\theta\sigma_{\gamma 2} & R_{\gamma 1\gamma 2}\sigma_{\gamma 1}\sigma_{\gamma 2} & \sigma_{\gamma 2}^2 \end{bmatrix} \right), i = 1,2, \ldots., 41. \qquad (EQ\ 2)$$

In this case, the assumption of multivariate normality is a mathematical convenience, and cannot be justified by the central limit theorem. Under the multivariate normal framework, the Stage 2 model contains three unknown mean parameters and six unknown covariance parameters. This Stage 2 model can be re-parameterized in terms of two simultaneous meta-regressions as

$$\mathrm{E}(\theta_i|\gamma_{1i}, \gamma_{2i}) = \alpha_\theta + \beta_1\gamma_{1i} + \beta_2\gamma_{2i} \qquad (EQ\ 3A)$$

$$\mathrm{Var}(\theta_i|\gamma_{1i}, \gamma_{2i}) = \lambda_\theta^2 \qquad (EQ\ 3B)$$

$$\mathrm{E}(\gamma_{1i}|\gamma_{2i}) = \alpha_{\gamma_1} + \omega\gamma_{2i} \qquad (EQ\ 4A)$$

$\mathrm{Var}(\gamma_{1i}|\gamma_{2i})=\lambda_{\gamma 1}^2$ $(EQ\ 4B)$

$\mathrm{E}(\gamma_{2i}) = \mu_{\gamma 2}$ $(EQ\ 5A)$

$\mathrm{Var}(\gamma_{2i}) = \sigma_{\gamma 2}^2$ $(EQ\ 5B)$.

In this reparameterization, the linearity of the conditional expectations follows from the assumption of multivariate normality for the second stage of the mixed effects model. In this formulation, $EQ\ 3A$ and $EQ\ 3B$ describe the first meta-regression which relates the treatment effects on the clinical endpoint on the treatment effects on UACR and on GFR slope, with the term λ_θ^2 accounting for deviations of the true treatment effects on the clinical endpoint from the linear regression on the treatment effects on the two surrogates. In this meta-regression, the parameters α_θ, β_1 and β_2 represent the intercept and the slopes for the treatment effects on GFR slope and on UACR, respectively. Further, $EQ\ 4A$ and $EQ\ 4B$ define the second meta-regression which relates the true treatment effects on GFR slope to the treatment effects on UACR, with the term $\lambda_{\gamma 1}^2$ accounting for deviations of the true treatment effects on GFR slope from the linear regression on UACR. Finally, $\mu_{\gamma 2}$ and $\sigma_{\gamma 2}^2$ in $EQ\ 5A$ and $EQ\ 5B$ represent the mean and variance of the treatment effects on UACR in the previous RCTs.

The first meta-regression which jointly relates the treatment effect on the clinical endpoint to the treatment effects on GFR slope and ACR is viewed as supporting the validity of the two surrogate endpoints in combination if the following conditions are satisfied:

a) The meta-regression accurately predicts the true treatment effects on the clinical endpoint as reflected by a low value of the residual standard deviation of the conditional distribution of the true treatment effects on the clinical endpoint given the true treatment effects on the surrogates. We refer to this residual standard deviation as the root mean square error (RMSE) for the meta-regression model. The RMSE is represented by the parameter $\lambda_\theta$in $EQ\ 2B$.
b) The meta-regression predicts a high proportion of the variation in the true treatment effects on the clinical endpoint, as reflected in the R^2 for the meta-regression model.
c) At least one of the slopes, β_1 and β_2, for the treatment effects on the two surrogate endpoints must be non-zero, and
d) The intercept α_θ for the meta-regression model should be close to 0, indicating that the treatment effect on the clinical endpoint is close to 0 when there is no treatment effect on either of the two surrogates.

The trial-level meta-regression approach outlined above is closely related to frameworks for trial-level analyses for validating surrogate endpoints which have been developed by other authors[22-26].

Diffuse Prior Distributions for Model Parameters. The full designation of the Bayesian model requires the specification of prior distribution for the 9 model parameters appearing in $EQ\ 3A$ through $EQ\ 5B$. We use diffuse priors for mean treatment effect on UACR and for the coefficients of the two meta-regressions:

$\mu_{\gamma 2} \sim N(0,10000)$

$\alpha_\theta \sim N(0,10000)$

$\alpha_{\gamma_1} \sim N(0,10000)$

$\beta_1 \sim N(0,10000)$

$\beta_2 \sim N(0,10000)$

$\omega \sim N(0,10000)$.

We consider two alternative sets of diffuse priors for the variance terms $\sigma^2_{\gamma 1}$, $\sigma^2_{\gamma 2}$, and λ^2_θ. For the first set of diffuse priors, we use:
$\lambda^2_{\gamma 1}$~Inverse Gamma with shape 0.261 and scale 0.005.
$\sigma^2_{\gamma 2}$~Inverse Gamma with shape 0.261 and scale 0.000408
λ^2_θ~Inverse Gamma with shape 0.261 and scale 0.000408
For $\sigma^2_{\gamma 2}$ and λ^2_θ this choice of shape and scale parameters for the inverse gamma distributions assigns a 1/3 probability to low heterogeneity (defined by an SD for the log HR or log geometric mean ratio < 0.05, 1/3 to medium heterogeneity given by SDs between 0.05 and 0.20, and 1/3 to high heterogeneity given by SDs > 0.20). For slope, the prior for the variance of the treatment effects assigns 1/3 probabilities each to slope SDs < 0.175 ml/min/1.73m^2/yr, 0.175 to 0.70 ml/min/1.73m^2/yr, and > 0.70 ml/min/1.73m^2/yr. For the second set of diffuse priors, we used the inverse gamma distributions with shape and scale both equal to 0.001 for each of the variance terms $\sigma^2_{\gamma 1}$, $\sigma^2_{\gamma 2}$, and λ^2_θ. The inverse gamma distribution with shape and scale equal to 0.001 is a commonly used diffuse prior for variances in Bayesian analysis.
Estimation of posterior distributions for the meta-regression parameters. We obtained the posterior distribution for the model parameters $\xi = \left(\mu_{\gamma 2}, \sigma^2_{\gamma 2}, \alpha_{\gamma_1}, \omega, \lambda^2_{\gamma 1}, \beta_1, \beta_2, \alpha_\theta, \lambda^2_\theta\right)$ by applying a Markov Chain Monte-Carlo(MCMC) algorithm in the STAN package. The inputs to the MCMC algorithm are the estimated treatment effects on the respective endpoints, their standard errors, and the correlations between the sampling errors of the estimated treatment effects across the previous 41 RCTs. The output is a sequence of 50,000 random draws of the model parameters, which we denote as $\xi(j) = \left(\mu_{\gamma 2}(j), \sigma^2_{\gamma 2}(j), \alpha_{\gamma_1}(j), \omega(j), \lambda^2_{\gamma 1}(j), \beta_1(j), \beta_2(j), \alpha_\theta(j), \lambda^2_\theta(j)\right)'$,
$j = 1, 2, \ldots 50{,}000$. The different $\xi(j)$ are not independent, but we used a thinning process within the MCMC algorithm to assure they are not too highly correlated. The posterior distribution the meta-regression parameters are estimated from the distribution of the $\xi(j), j = 1,2, \ldots 50{,}000$. We similarly obtained the posterior distributions of functionals $f(\xi)$ of the model parameters, including the trial-level R^2 for the first meta-regression that relates the true treatment effects on the clinical endpoint to the true treatment effects on the two surrogates.
We verified convergence of the MCMC algorithm by verifying that all effective sample sizes exceeded 1,000 and by evaluating the Gelman-Rubin convergence diagnostic, and further by visual inspection of trace plots for the MCMC chains. We provide additional details in the online supplement to Inker et al (2020)[27].
Step 2. Construct the prior distribution for the treatment effects in a new phase 2 trials.
Let $(\theta_0, \gamma_{10}, \gamma_{20})'$ denote the true treatment effects on the clinical endpoint, GFR slope, and log ACR in a new phase II trial. Bayesian inference can be performed for the effect of the treatment on the clinical endpoint in the phase 2 trial by computing the posterior distribution for θ_0 given estimates of the treatment effects $\hat{\gamma}_{10}$ and $\hat{\gamma}_{20}$ on the two surrogates under a suitable prior distribution for $(\theta_0, \gamma_{10}, \gamma_{20})'$. A common approach to this problem is to assume full exchangeability between the new trial and the previously conducted trials, and define the *prior distribution* for the treatment effects $(\theta_0, \gamma_{10}, \gamma_{20})'$ in the new phase 2 trial as equal to the joint distribution of $(\theta_i, \gamma_{1i}, \gamma_{2i})'$ for a randomly selected trial under the posterior distribution of ξ from the previous RCTs that was obtained in step 1. The joint probability distribution function (pdf) for this distribution is

$$f(\theta_0, \gamma_{10}, \gamma_{20}) = \int f(\theta_0, \gamma_{10}, \gamma_{20}|\xi) f(\xi) d\xi. \qquad (EQ\ 6)$$

where $f(\theta_0, \gamma_{10}, \gamma_{20}|\xi)$ is the conditional pdf of $(\theta_0, \gamma_{10}, \gamma_{20})$' given ξ, and $f(\xi)$ is the posterior pdf of the model parameters given the prior RCT data. In this expression $f(\theta_0, \gamma_{10}, \gamma_{20}|\xi)$ is the multivariate normal pdf defined by EQ 2, evaluated at θ_0, γ_{10} and γ_{20}.

The term $f(\theta_0, \gamma_{10}, \gamma_{20}|\xi)$ factors into the product of three terms $f(\theta_0|\gamma_{10}, \gamma_{20}, \xi) \times f(\gamma_{10}|\gamma_{20}, \xi) \times f(\gamma_{20}|\xi)$, where $f(\theta_0|\gamma_{10}, \gamma_{20}, \xi)$ and $f(\gamma_{10}|\gamma_{20}, \xi)$ are conditional pdf's that correspond to the meta-regressions of the treatment effects on the clinical endpoint on the treatment effects on the two surrogate endpoints, and of the treatment effects on GFR slope on treatment effects on UACR, respectively. Thus, these two terms characterize the relationships of the treatment effects on the clinical and surrogate endpoints with each other. On the other hand, the term $f(\gamma_{20}|\xi)$represents the marginal pdf for the univariate distribution of treatment effects on UACR. This inclusion of $f(\gamma_{20}|\xi)$ thus involves building into the prior distribution an assumption that the treatment effect on UACR in the newly conducted phase 2 trial is consistent with the distribution of treatment effects on UACR in the previously conducted RCTs. This does not mesh well with the goals of the surrogate endpoint framework. Regulatory agencies are likely to require that beneficial effects of new treatments must be demonstrated anew, without presuming a high likelihood that future treatments are beneficial just because past treatments have been beneficial. To address this issue, we replace $f(\gamma_{20}|\xi)$, which is based on the posterior distribution for treatment effects on UACR from previous trials, with a diffuse prior distribution $\pi(\gamma_{20}|\xi)$ which avoids assuming any connection between the treatment effect on UACR in the new trial and the treatment effects on UACR in the previously conducted RCTs. Specially, we take $\pi(\gamma_{20}|\xi)$ to be the pdf of a normal distribution with mean μ_{g2} and variance 100; because treatment effects on UACR are expressed on a log scale, this variance is sufficiently large as to negate any influence of the size and direction of treatment effects on UACR in the previously conducted RCTs when making inferences about the treatment effects in the new phase 2 trial.

Based on this logic, we use the prior distribution with pdf

$$f'(\theta_0, \gamma_{10}, \gamma_{20}) = \int f(\theta_0|\gamma_{10}, \gamma_{20}, \xi) \times f(\gamma_{10}|\gamma_{20}, \xi) \times \pi(\gamma_{20}|\xi) f(\xi) d\xi \qquad (EQ\ 7)$$

to make inferences concerning clinical benefit in the new phase 2 trial. Using this prior invokes the assumption that the joint conditional distribution of the treatment effects on the clinical endpoint and GFR slope is exchangeable between the new phase 2 trial and the previously conducted trials, but avoids assuming exchangeability in the treatment effect on UACR between the new phase 2 trial and the previous trials.

Step 3. Estimate standard errors (SEs) for treatment effects on UACR and GFR slope for candidate designs for the new phase 2 trial

Inferences for the implications of treatment effects on the two surrogate endpoints for the treatment effect on clinical benefit in the new RCT depends on the standard errors (SEs) of the estimated treatment effects on the two surrogate endpoints in the new RCT. To estimate the SEs for a new trial which is being designed, we assume that the treatment effect on UACR in the phase 2 trial will be estimated using an analysis of covariance (ANCOVA) to relate the change in log UACR to the randomized treatment assignment with baseline UACR included as a covariate, and that the treatment effect on GFR slope will be estimated using a mixed effects model that uses a linear spline term to distinguish between the acute and chronic slopes. Appropriate estimates of the standard errors depends on the specific circumstances of the trial, and are a key step in conventional power calculations when designing phase 2 trials with UACR or GFR slope as the primary endpoint.

In this report we estimate the standard error for the estimated treatment effect on log transformed UACR as $\sqrt{2} \times \frac{0.725}{N}$, where 0.725 is a typical residual root mean square error in a regression of the 6-month log ACR on the baseline log ACR in previously conducted trials, and N is the sample size in

each treatment group. We estimated the standard error for the mean chronic slope (starting at 3 months follow-up) from model-based standard errors under the standard 2-slope linear mixed effects model:

$Y_{ij} = \beta_{0c} + \min(t_{ij} - 3,0)\,\beta_{1c} + \min(t_{ij} - 3,0)\,\beta_{2c} + Z_i\beta_{1c} + Z_i \min(t_{ij} - 3,0)\,\beta_{1t} +$
$Z_i \min(t_{ij} - 3,0)\,\beta_{2t} + b_{i0} + \min(t_{ij} - 3,0)\,b_{i1} + \max(t_{ij} - 3,0)\,b_{i2} + \epsilon_{ij}$, where
t_{ij} = time in months of the i^{th} subject's j^{th} GFR measurement,
$Y_{ij} = i^{th}$ subject's GFR measurement at time t_{ij}
Z_i = 0-1 indicator variable for assignment to the active treatment group
$\beta_{0c}, \beta_{1c}, \beta_{2c}$ = fixed effects for the mean intercept, acute and chronic slopes in the control arm
$\beta_{0t}, \beta_{1t}, \beta_{2t}$ = fixed effects for the mean intercept, acute and chronic slopes in the active treatment arm
b_{i0}, b_{i1}, b_{i2} = normally distributed random intercepts, acute and chronic slopes for the i^{th}subject, and
ϵ_{ij} = normally distributed residual for the i^{th} subject's GFR measurement at time t_{ij}.
We stipulated an unstructured covariance model for $(b_{i0}, b_{i1}, b_{i2})'$.

The standard errors for the estimated treatment effects on the mean chronic and total slopes depend on the following inputs, which we have also selected to be typical of previous CKD RCTs:

i) The mean baseline GFR is 40 ml/min/1.73m^2,
ii) The standard deviation of the residuals ϵ_{ij} is $\sqrt{(0.67 \times 40)}$ = 5.16 ml/min/1.73m^2
iii) The random acute and chronic slopes b_{i1} and b_{i2} are taken to be uncorrelated each with a standard deviation of 4 ml/min/1.73m^2/yr.
iv) The loss-to-follow-up rate is 5% per year.

We consider 5 different Phase 2 designs defined by the combinations of following design parameters:
Design A: 60 per group, 1.25 years of follow-up, quarterly GFRs
Design B: 120 patients per group, 1.25 years of follow-up, quarterly GFRs
Design C: 120 patients per group, 1.25 years of follow-up, monthly GFRs
Design D: 120 patients per group, 2 years of follow-up, quarterly GFRs
Design E: 240 patients per group, 2 years of follow-up, quarterly GFRs

We also consider a single Phase 3 design, with design parameters:
Design F: 600 patients per group, 2 years of follow-up, quarterly GFRs.

Under each design, we also assumed that a second GFR measurement was obtained at baseline and at the 3-month assessment, which designates the start of the chronic phase.
Finally, we assumed a correlation in the sampling errors between the treatment effects on the two surrogate endpoints of -0.2, which corresponds roughly to the mean correlation observed in the previous CKD trials.

Step 4. Compare phase 2 designs based on posterior probabilities of clinical benefit (denoted PPV_{trial}) for hypothetical observed treatment effects on UACR & GFR slope

Using the standard errors from Step 3, we can determine the approximate sampling distribution of the estimated treatment effects on the GFR slope and UACR from the phase 2 trial, denoted $(\hat{\gamma}_{10}, \hat{\gamma}_{20})'$, conditional on the true treatment effects $(\gamma_{10}, \gamma_{20})'$ for each candidate design under consideration. Specifically, by invoking the central limit theorem, the sampling distribution of $(\hat{\gamma}_{10}, \hat{\gamma}_{20})'$ conditional on $(\gamma_{10}, \gamma_{20})'$ for a particular design is approximately bivariate normally distributed with mean $(\gamma_{10}, \gamma_{20})'$ and covariance matrix

$$\Sigma_{samp} = Var\left(\begin{bmatrix}\hat{\gamma}_{01}\\ \hat{\gamma}_{02}\end{bmatrix}\right) = \begin{bmatrix} SE^2(\hat{\gamma}_{10}) & R_{\hat{g}_1\hat{\gamma}_2}SE(\hat{\gamma}_{10})SE(\hat{\gamma}_{20}) \\ R_{\hat{g}_1\hat{\gamma}_2}SE(\hat{\gamma}_{10})SE(\hat{\gamma}_{20}) & SE^2(\hat{\gamma}_{20}) \end{bmatrix} \quad (EQ\ 8)$$

where $SE(\hat{\gamma}_{10})$, $SE(\hat{\gamma}_{20})$ and $R_{\hat{g}_1\hat{g}_2}$ are determined for each candidate design as described in Step 3. Let t denote a threshold for the hazard ratio on the clinical endpoint which signifies clinical benefit. In this report, we consider three alternative definitions for clinical benefit given by t = 0.8 0.9, or 1.0. For each candidate design, the posterior probability of clinical benefit, $\Pr(\theta_0 < t|\hat{\gamma}_{10}, \hat{\gamma}_{20})$, can be approximated using the following process:

- For each ξ from the 50,000 MCMC draws from the prior distribution derived in Step 2,

a) Sample γ_{10} and γ_{20} from the bivariate normal distribution $N(\mu'_{post}, \Sigma'_{post})$, where

$$\Sigma'_{post} = \left[\begin{bmatrix} 100 \times \omega^2 + \lambda_{\gamma 1}^2 & 100 \times \omega \\ 100 \times \omega & 100 \end{bmatrix}^{-1} + \Sigma_{samp}^{-1}\right]^{-1}$$

and

$$\mu'_{post} = \Sigma'_{post}\,\Sigma_{samp}^{-1}\begin{bmatrix}\hat{\gamma}_{01}\\ \hat{\gamma}_{02}\end{bmatrix} + \Sigma'_{post}\begin{bmatrix} 100 \times \omega^2 + \lambda_{\gamma 1}^2 & 100 \times \omega \\ 100 \times \omega & 100 \end{bmatrix}^{-1}\begin{bmatrix}\alpha_{\gamma_1} + \omega\mu_{\gamma_2} \\ \mu_{\gamma_2}\end{bmatrix},$$

b) Then sample θ_0 from a random draw from a normal distribution with mean $\alpha_\theta + \beta_1 g_{10} + \beta_2 g_{20}$ and variance λ_θ^2,

- Compute the fraction of random draws of θ_0 which fall below the threshold t.

Σ_{samp} is the covariance matrix for $(\hat{\gamma}_{10}, \hat{\gamma}_{20})'$ given by EQ 8, and $\begin{bmatrix} 100 \times \omega^2 + \lambda_{\gamma 1}^2 & 100 \times \omega \\ 100 \times \omega & 100 \end{bmatrix}$ is the covariance matrix for $(\gamma_{10}, \gamma_{20})$ conditional on ξ under the prior distribution defined in Step 2. Using an analogy with diagnostic testing, we refer to $\Pr(\theta_0 < t|\hat{\gamma}_{10}, \hat{\gamma}_{20})$ as the trial-level positive predictive value for clinical benefit which we denote as PPV_{trial}. In this analogy, a true treatment benefit in the new phase 2 trial substitutes for the presence of true disease in the patient, the observed treatment effects on the surrogate endpoints substitutes for the diagnostic test results, and the prior distribution which relates the treatment effects on the clinical and surrogate endpoints from Step 2 substitutes for disease prevalence.

To assist in the selection of the most appropriate phase 2 design, the values of PPV_{trial} can be provided across a grid of possible values for the estimated treatment effects on UACR and GFR slope for each candidate phase 2 trial design under consideration.

Step 5. Compare phase 2 designs after accounting for precision of the estimated values for PPV_{trial}

From the standpoint of study design, a limitation of Step 4 is that the values of PPV_{trial} condition on estimated effects of the treatment $(\hat{\gamma}_{10}, \hat{\gamma}_{20})'$ on UACR and slope which themselves contain random sampling error, and which are not actually available until after the phase 2 trial is completed. Step 5 addresses this limitation by providing two different frequentist quantities which are based on PPV_{trial} and which are computed at hypothesized values for the true treatment effects on UACR and slope rather than at estimated values for these treatment effects. Note that for evaluating its frequentist properties, PPV_{trial} is interpreted as a statistic which depends on the estimated treatment effects on UACR and slope. Evaluation of frequentist properties of Bayesian methods is often used to assist in evaluation of the study design for Bayesian trials[26].

Given a specific combination of hypothesized true treatment effects on UACR and slope, we define the average estimated posterior probability of clinical benefit as

$$E_{(\gamma_{10},\gamma_{20})}((\Pr(\theta_0 < \log(t)|\hat{\gamma}_{10}, \hat{\gamma}_{20})), \quad (EQ\ 9)$$

where the expectation averages $PPV_{trial} = \Pr(\theta_0 < \log(t)|\hat{\gamma}_{10}, \hat{\gamma}_{20})$ over the sampling distribution of $(\hat{\gamma}_{10}, \hat{\gamma}_{20})$ given the hypothesized true treatment effects $(\gamma_{10}, \gamma_{20})'$ on the two surrogates, and t denotes the threshold for the hazard ratio of the clinical endpoint which designates clinical benefit. Step 5 performs this calculation for a grid of different combinations of hypothesized true treatment effects on UACR and GFR slope which matches the grid of observed effects considered in Step 4. A good phase 2 design from the surrogate endpoint perspective will have sufficient sample size, duration, and frequency of GFR measurements to demonstrate both a low average estimated PPV_{trial} under the null hypothesis of no effect on either UACR or slope, and a high average estimated PPV_{trial} under plausible research hypotheses for treatment benefit on the two surrogates.

Second, given a specific combination of hypothesized true effects $(\gamma_{10}, \gamma_{20})'$ on UACR and slope, we can compute the probability that the estimated PPV_{trial} exceeds a target level, such as 0.85, which is deemed sufficient to proceed to a phase 3 trial. Note that whereas the projected average estimated PPV_{trial} is defined by averaging PPV_{trial} over the sampling distribution of estimated treatment effects on the two surrogates given the hypothesized true effects, the probability that the estimated PPV_{trial} exceeds a designated target level is defined as the probability that PPV_{trial} exceeds the designated target across the same sampling distribution. We can view the probability that the estimated PPV_{trial} exceeds the target threshold as an analogue of Type 1 error for the surrogate endpoint setting under the null hypothesis of no effects on either surrogate, and as an analogue of statistical power for the surrogate endpoint setting when favorable treatment effects are hypothesized for the two surrogates.

Step 6. Use Bayesian computations to estimate the posterior probability of clinical benefit given the observed treatment effects on the two surrogate endpoints in the phase 2 trial

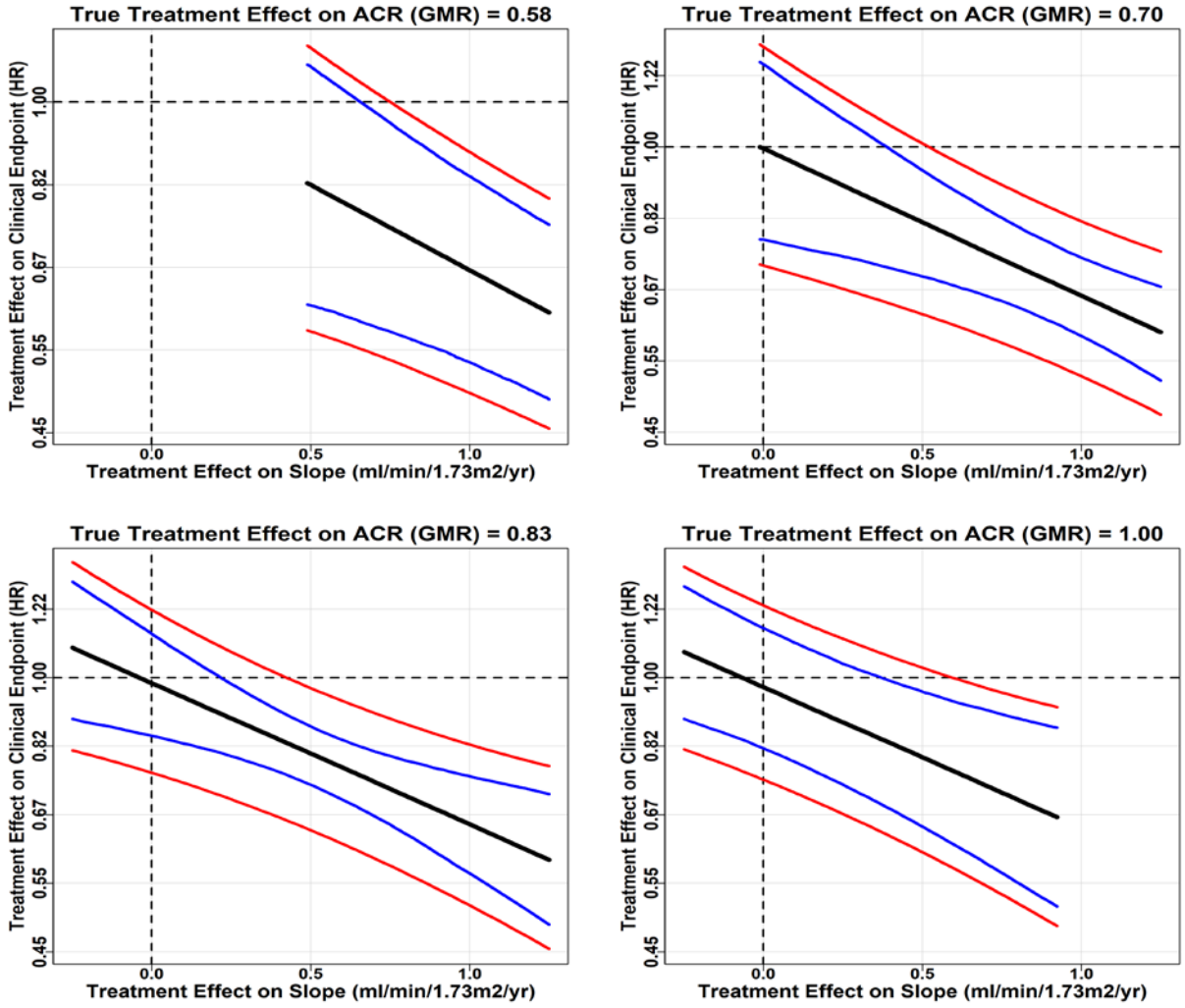
After the phase 2 trial is complete, the posterior probability of clinical benefit, denoted $\Pr(\theta_0 < \log(t)|\hat{\gamma}_{10}, \hat{\gamma}_{20})$ in EQ 9, can be approximated as described in Step 4 based on the values of $(\hat{\gamma}_{10}, \hat{\gamma}_{20})'$ which are actually observed in the phase 2 trial. This computation can be used to assess the joint implications for the estimated treatment effects on GFR slope and UACR for the likelihood that the treatment also benefits the clinical endpoint.

Supplement Table 1: Clinical characteristics of the population stratified by disease etiology

Disease	N Studies	N	Age mean (SD)	Female N (%)	Black N (%)	Diabetes N (%)	eGFR mean (SD)	ACR median (25,75th)	Clinical Endpoints N (%)	Interventions
Overall	41	29979	58.2 (12.6)	9951 (33.2)	3833 (12.8)	21206 (70.7)	58.2 (25.0)	272 (30, 1134)	3935 (13.1)	
Diabetes	10	21102	62.2 (9.9)	6527 (30.9)	1335 (6.3)	21102 (100.0)	61.4 (23.3)	270 (26, 1126)	2103 (10.0)	RASB v CCB Low v Usual BP RASB vs Control Sulodexide Empagliflozin
Glomerular disease	9	1325	40.8 (12.9)	467 (35.2)	18 (1.4)	5 (0.4)	74.2 (29.7)	1311 (838, 2335)	174 (13.1)	Immunosuppression RASB vs Control
Other CKD	22	7552	50.1 (12.9)	2957 (39.2)	2480 (32.8)	99 (1.3)	46.6 (24.5)	126 (30, 838)	1658 (22.0)	RASB vs Control RASB v CCB Low v Usual BP Albuminuria Targeted Protocol Low v Usual Diet

Other CKD refers to causes of CKD other than glomerular disease or diabetes or cause not specified. Clinical end point defined as the composite of chronic dialysis or kidney transplantation, eGFR<15 ml/min/1.73m^2 or confirmed doubling of serum creatinine. CKD, chronic kidney disease; ACR, albumin to creatinine ratio; Age is measured in years. FU time in months; RASB, renin angiotensin system blockers; CCB, calcium channel blocker; BP, blood pressure. Race was defined as Black vs non Black for use in categorization of race in computing eGFR using the CKD-EPI creatinine equation.

Supplement Figure 1: Confidence and prediction interval for true treatment effect on established renal endpoint given true treatment effects on GFR slope and UACR



Shown meta-regression lines (black), 95% Bayesian confidence intervals (blue), and 95% Bayesian prediction intervals (red) for the treatment effect on the clinical endpoint as a function of the treatment effect on the chronic slope (horizontal axis) and the treatment effect on UACR (designated by the 4 panels). The confidence intervals express the uncertainty in the true meta-regression line given the true treatment effects on the two surrogate endpoints, and the prediction intervals express the uncertainty in the true treatment effect on the clinical endpoint given the true treatment effects on the two surrogate endpoints.

Supplement Figure 2: Trial-level positive predictive value (PPV_{trial}) based on observed treatment effects on UACR and Slope to infer clinical benefit defined as HR < 0.9

Design A: 60 per group, 1.25 years of follow-up, quarterly GFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.39	0.46	0.53	0.60	0.68	0.76	0.82	0.88	0.92
0.2	0.39	0.47	0.54	0.62	0.70	0.77	0.83	0.89	0.92
0.4	0.41	0.49	0.56	0.64	0.71	0.78	0.84	0.90	0.93
0.6	0.43	0.49	0.57	0.65	0.73	0.80	0.85	0.90	0.94
0.8	0.45	0.51	0.58	0.67	0.74	0.81	0.86	0.90	0.94
1	0.46	0.53	0.60	0.67	0.75	0.82	0.87	0.91	0.94

Design B: 120 patients per group, 1.25 years of follow-up, quarterly eGFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.37	0.44	0.51	0.59	0.67	0.75	0.82	0.87	0.91
0.2	0.39	0.47	0.54	0.62	0.70	0.77	0.84	0.88	0.92
0.4	0.42	0.49	0.57	0.65	0.73	0.80	0.85	0.90	0.93
0.6	0.44	0.52	0.59	0.67	0.75	0.82	0.87	0.91	0.94
0.8	0.48	0.54	0.63	0.69	0.77	0.84	0.89	0.92	0.95
1	0.50	0.57	0.64	0.73	0.79	0.85	0.90	0.93	0.95

Design C: 120 patients per group, 1.25 years of follow-up, monthly eGFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.37	0.43	0.5	0.57	0.64	0.72	0.77	0.84	0.88
0.2	0.40	0.46	0.54	0.61	0.68	0.75	0.81	0.87	0.90
0.4	0.44	0.51	0.59	0.66	0.73	0.79	0.84	0.89	0.92
0.6	0.49	0.55	0.62	0.69	0.76	0.82	0.87	0.91	0.93
0.8	0.53	0.59	0.67	0.74	0.80	0.85	0.90	0.93	0.95
1	0.58	0.64	0.70	0.77	0.83	0.87	0.91	0.94	0.96

Design D: 120 patients per group, 2 years of follow-up, quarterly eGFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.36	0.42	0.49	0.55	0.61	0.69	0.75	0.81	0.86
0.2	0.41	0.47	0.54	0.61	0.67	0.74	0.80	0.84	0.88
0.4	0.47	0.53	0.59	0.65	0.72	0.79	0.84	0.87	0.91
0.6	0.52	0.58	0.65	0.71	0.77	0.82	0.87	0.90	0.93
0.8	0.58	0.63	0.70	0.75	0.82	0.86	0.90	0.93	0.94
1	0.63	0.69	0.74	0.81	0.85	0.89	0.92	0.94	0.96

Design E: 240 patients per group, 2 years of follow-up, quarterly eGFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.34	0.38	0.43	0.5	0.56	0.63	0.67	0.73	0.78
0.2	0.42	0.47	0.53	0.59	0.65	0.71	0.76	0.80	0.84
0.4	0.51	0.56	0.62	0.67	0.74	0.78	0.83	0.86	0.89
0.6	0.6	0.65	0.7	0.75	0.81	0.85	0.88	0.91	0.92
0.8	0.68	0.72	0.77	0.83	0.87	0.90	0.92	0.94	0.95
1	0.75	0.80	0.83	0.88	0.91	0.93	0.95	0.96	0.97

Supplement Figure 2 Legend.
Shown are the trial-level positive predictive values (PPV_{trial}) for inferring clinical benefit defined as a HR for the clinical endpoint which is less than 0.9 based on combinations of observed treatment effects on slope and ACR in the phase 2 trial. The unshaded region designates values for $PPV_{trial} < 0.80$. The light shaded region indicates values of PPV_{trial} between 0.80 and 0.90. The dark shaded region indicates $PPV_{trial} \geq 0.90$.

If PPV_{trial} is estimated based on the treatment effect on UACR alone, under Design 1 the values of PPV_{trial} are 0.39, 0.46, 0.54, 0.63, 0.72, 0.79, 0.86, 0.91, and 0.94 for observed GMRs ranging from 1.00 to 0.60, respectively. Under Designs 2-4 the corresponding values are 0.38, 0.45, 0.54, 0.64, 0.73, 0.81, 0.87, 0.92, 0.95, respectively. Under Design 5, the corresponding values for PPV_{trial} are 0.37, 0.45, 0.54, 0.64, 0.74, 0.82, 0.88, 0.93, and 0.95.

Abbreviations: UACR urine albumin to creatinine ratio; GMR geometric mean ratio; GFR glomerular filtration rate; PPV_{trial} trial-level positive predictive value

Supplement Figure 3: Trial-level positive predictive value (PPV_{trial}) based on observed treatment effects on UACR and Slope to infer clinical benefit defined as HR < 0.8

Design A: 60 per group, 1.25 years of follow-up, quarterly GFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.23	0.28	0.34	0.42	0.51	0.60	0.68	0.76	0.83
0.2	0.25	0.29	0.36	0.42	0.52	0.61	0.70	0.78	0.84
0.4	0.25	0.31	0.36	0.45	0.53	0.62	0.71	0.79	0.84
0.6	0.27	0.33	0.38	0.46	0.55	0.63	0.72	0.80	0.86
0.8	0.28	0.34	0.40	0.49	0.56	0.65	0.74	0.81	0.86
1	0.29	0.35	0.41	0.50	0.58	0.67	0.74	0.82	0.87

Design B: 120 patients per group, 1.25 years of follow-up, quarterly eGFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.20	0.25	0.31	0.39	0.47	0.57	0.65	0.73	0.80
0.2	0.22	0.27	0.33	0.41	0.50	0.59	0.68	0.76	0.83
0.4	0.24	0.29	0.37	0.43	0.53	0.62	0.70	0.78	0.84
0.6	0.27	0.32	0.38	0.47	0.55	0.64	0.73	0.81	0.86
0.8	0.29	0.35	0.41	0.50	0.58	0.67	0.76	0.82	0.87
1	0.31	0.37	0.45	0.53	0.61	0.71	0.77	0.84	0.89

Design C: 120 patients per group, 1.25 years of follow-up, monthly eGFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.19	0.23	0.29	0.34	0.42	0.51	0.59	0.68	0.76
0.2	0.22	0.27	0.32	0.39	0.47	0.55	0.64	0.72	0.79
0.4	0.26	0.31	0.37	0.44	0.51	0.59	0.68	0.76	0.81
0.6	0.30	0.34	0.41	0.48	0.56	0.64	0.72	0.80	0.84
0.8	0.33	0.39	0.46	0.52	0.60	0.68	0.76	0.82	0.87
1	0.37	0.43	0.49	0.57	0.65	0.72	0.79	0.85	0.89

Design D: 120 patients per group, 2 years of follow-up, quarterly eGFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.18	0.22	0.27	0.33	0.40	0.47	0.56	0.63	0.71
0.2	0.23	0.27	0.32	0.38	0.44	0.52	0.61	0.68	0.75
0.4	0.26	0.32	0.36	0.44	0.51	0.58	0.67	0.74	0.79
0.6	0.31	0.35	0.41	0.50	0.56	0.64	0.71	0.77	0.82
0.8	0.36	0.41	0.48	0.55	0.62	0.69	0.76	0.82	0.87
1	0.41	0.47	0.53	0.60	0.68	0.74	0.80	0.85	0.90

Design E: 240 patients per group, 2 years of follow-up, quarterly eGFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.15	0.18	0.21	0.24	0.30	0.38	0.44	0.52	0.58
0.2	0.21	0.24	0.27	0.33	0.38	0.46	0.53	0.60	0.66
0.4	0.27	0.31	0.35	0.41	0.48	0.55	0.62	0.68	0.74
0.6	0.35	0.39	0.45	0.51	0.57	0.64	0.70	0.76	0.80
0.8	0.44	0.49	0.55	0.60	0.67	0.73	0.78	0.82	0.86
1	0.52	0.58	0.64	0.69	0.76	0.80	0.85	0.87	0.90

Supplement Figure 3 Legend.

Shown are the trial-level positive predictive values (PPV_{trial}) for inferring clinical benefit defined as a HR for the clinical endpoint which is less than 0.8 based on combinations of observed treatment effects on slope and ACR. The unshaded region designates values for $PPV_{trial} < 0.80$. The light shaded region indicates values of PPV_{trial} between 0.80 and 0.90. The dark shaded region indicates $PPV_{trial} \geq 0.90$.

If PPV_{trial} is estimated based on the treatment effect on UACR alone, under Design 1 the values of PPV_{trial} are 0.24, 0.29, 0.36, 0.45, 0.55, 0.64, 0.73, 0.82, and 0.87 for observed GMRs ranging from 1.00 to 0.60, respectively. Under Designs 2-4 the corresponding values are 0.22, 0.28, 0.35, 0.44, 0.54, 0.65, 0.76, 0.83, and 0.89, respectively. Under Design 5, the corresponding values for PPV_{trial} are 0.21, 0.26, 0.34, 0.44, 0.55, 0.66, 0.76, 0.84, and 0.90.

Abbreviations: UACR urine albumin to creatinine ratio; GMR geometric mean ratio; GFR glomerular filtration rate; PPV_{trial} trial-level positive predictive value

Supplement Figure 4: Trial level positive predictive value (PPV_{trial}) based on observed treatment effects on UACR and Slope to infer clinical benefit defined as HR < 1 for an example phase 3 trial

Design F: 600 per group, 2 years of follow-up, quarterly GFRs

Est. effect on chronic slope ml/min/1.73m²/yr	Estimated Treatment Effect on UACR (GMR)								
	1.00	**0.95**	**0.90**	**0.85**	**0.80**	**0.75**	**0.70**	**0.65**	**0.60**
0	0.57	0.60	0.64	0.68	0.72	0.74	0.77	0.78	0.80
0.2	0.70	0.74	0.77	0.81	0.83	0.85	0.86	0.87	0.88
0.4	0.81	0.84	0.87	0.90	0.92	0.93	0.93	0.94	0.93
0.6	0.90	0.92	0.94	0.95	0.96	0.97	0.97	0.97	0.97
0.8	0.94	0.96	0.97	0.98	0.98	0.99	0.99	0.99	0.98
1	0.97	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Supplement Figure 4 Legend.
Shown the trial-level positive predictive values (PPV_{trial}) for inferring clinical benefit defined as a HR for the clinical endpoint which is less than 1 based on combinations of observed treatment effects on slope and ACR for an example phase 3 design. The unshaded region designates values for $PPV_{trial} < 0.80$. The light shaded region indicates values of PPV_{trial} between 0.80 and 0.90. The dark shaded region indicates $PPV_{trial} \geq 0.90$.

Abbreviations: UACR urine albumin to creatinine ratio; GMR geometric mean ratio; GFR glomerular filtration rate; PPV_{trial} trial-level positive predictive value

Supplement Figure 5: Average estimated PPV_{trial} based on hypothesized treatment effects on UACR and GFR slope to infer clinical benefit defined as HR < 0.9

Design A: 60 per group, 1.25 years of follow-up, quarterly GFRs

Hypothesized effect on chronic slope ml/min/1.73m²/yr	Hypothesized Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.40	0.46	0.52	0.59	0.65	0.73	0.79	0.84	0.89
0.2	0.42	0.48	0.54	0.61	0.68	0.74	0.79	0.85	0.89
0.4	0.42	0.48	0.56	0.61	0.68	0.74	0.80	0.86	0.90
0.6	0.44	0.49	0.56	0.63	0.70	0.75	0.82	0.87	0.90
0.8	0.46	0.52	0.58	0.64	0.70	0.77	0.83	0.88	0.91
1	0.47	0.53	0.59	0.65	0.72	0.78	0.83	0.88	0.92

Design B: 120 patients per group, 1.25 years of follow-up, quarterly eGFRs

Hypothesized effect on chronic slope ml/min/1.73m²/yr	Hypothesized Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.39	0.45	0.52	0.58	0.65	0.72	0.79	0.84	0.89
0.2	0.41	0.47	0.54	0.61	0.68	0.74	0.80	0.85	0.90
0.4	0.44	0.5	0.56	0.63	0.69	0.76	0.82	0.87	0.91
0.6	0.45	0.52	0.59	0.65	0.72	0.79	0.84	0.88	0.92
0.8	0.48	0.54	0.60	0.68	0.74	0.80	0.86	0.89	0.92
1	0.51	0.56	0.63	0.69	0.76	0.82	0.87	0.91	0.94

Design C: 120 patients per group, 1.25 years of follow-up, monthly eGFRs

Hypothesized effect on chronic slope ml/min/1.73m²/yr	Hypothesized Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.39	0.44	0.49	0.57	0.62	0.69	0.74	0.80	0.85
0.2	0.42	0.47	0.54	0.6	0.66	0.72	0.78	0.83	0.87
0.4	0.46	0.52	0.57	0.63	0.69	0.75	0.80	0.86	0.89
0.6	0.50	0.55	0.61	0.67	0.73	0.78	0.83	0.87	0.91
0.8	0.54	0.59	0.64	0.70	0.76	0.81	0.86	0.89	0.92
1	0.57	0.61	0.68	0.73	0.79	0.83	0.88	0.91	0.93

Design D: 120 patients per group, 2 years of follow-up, quarterly eGFRs

Hypothesized effect on chronic slope ml/min/1.73m²/yr	Hypothesized Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.39	0.43	0.48	0.54	0.6	0.66	0.73	0.77	0.81
0.2	0.44	0.48	0.53	0.58	0.65	0.71	0.75	0.81	0.85
0.4	0.48	0.53	0.57	0.64	0.7	0.74	0.80	0.84	0.87
0.6	0.52	0.57	0.62	0.69	0.73	0.78	0.83	0.87	0.90
0.8	0.56	0.62	0.67	0.72	0.77	0.82	0.86	0.89	0.92
1	0.62	0.66	0.72	0.76	0.80	0.85	0.89	0.91	0.93

Design E: 240 patients per group, 2 years of follow-up, quarterly eGFRs

Hypothesized effect on chronic slope ml/min/1.73m²/yr	Hypothesized Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.36	0.40	0.45	0.51	0.55	0.61	0.66	0.71	0.75
0.2	0.43	0.48	0.53	0.57	0.63	0.69	0.72	0.77	0.81
0.4	0.51	0.56	0.60	0.66	0.70	0.74	0.79	0.83	0.85
0.6	0.59	0.63	0.67	0.72	0.76	0.81	0.84	0.87	0.89
0.8	0.64	0.70	0.74	0.78	0.82	0.86	0.89	0.91	0.92
1	0.71	0.76	0.79	0.84	0.87	0.90	0.92	0.93	0.95

Supplement Figure 5 Legend:

Shown are the average estimated values of PPV_{trial} for inferring clinical benefit defined as a HR for the clinical endpoint which is less than 0.9 based on combinations of hypothesized true treatment effects on slope and UACR. The estimated values for PPV_{trial} are averaged over the projected sampling distributions for the estimated treatment effects on slope and ACR given the true hypothesized treatment effects on these endpoints. The unshaded region designates average estimated $PPV_{trial} < 0.80$. The light shaded region indicates average estimated PPV_{trial} between 0.80 and 0.90. The dark shaded region indicates average estimated $PPV_{trial} \geq 0.90$.

If the average estimated PPV_{trial} is computed based on the treatment effect on UACR alone, under Design 1 the average estimated PPV_{trial} values are 0.41, 0.48, 0.54, 0.61, 0.69, 0.77, 0.83, 0.88, and 0.93 for true GMRs ranging from 1.00 to 0.60, respectively. Under Designs 2-4 the corresponding values are 0.39, 0.46, 0.54, 0.63, 0.72, 0.79, 0.86, 0.90, and 0.94 respectively. Under Design 5, the corresponding average estimated PPV_{trial} values are 0.37, 0.45, 0.54, 0.64, 0.73, 0.80, 0.88, 0.92, 0.95.

Abbreviations: UACR urine albumin to creatinine ratio; GMR geometric mean ratio; GFR glomerular filtration rate; PPV_{trial} trial-level positive predictive value

Supplement Figure 6: Average estimated PPV_{trial} based on hypothesized treatment effects on UACR and GFR slope to infer clinical benefit defined as HR < 0.8

Design A: 60 per group, 1.25 years of follow-up, quarterly GFRs

Hypothesized effect on chronic slope ml/min/1.73m²/yr	Hypothesized Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.26	0.32	0.38	0.43	0.51	0.57	0.66	0.72	0.79
0.2	0.27	0.32	0.38	0.44	0.52	0.59	0.67	0.74	0.81
0.4	0.28	0.34	0.39	0.46	0.54	0.60	0.69	0.74	0.82
0.6	0.30	0.35	0.42	0.48	0.54	0.62	0.70	0.76	0.83
0.8	0.30	0.36	0.42	0.48	0.56	0.63	0.71	0.77	0.83
1	0.32	0.37	0.43	0.50	0.57	0.64	0.71	0.78	0.83

Design B: 120 patients per group, 1.25 years of follow-up, quarterly eGFRs

Hypothesized effect on chronic slope ml/min/1.73m²/yr	Hypothesized Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.24	0.28	0.34	0.40	0.48	0.55	0.63	0.72	0.78
0.2	0.25	0.29	0.36	0.44	0.51	0.58	0.65	0.73	0.80
0.4	0.28	0.32	0.38	0.45	0.52	0.60	0.68	0.76	0.81
0.6	0.29	0.34	0.40	0.47	0.55	0.64	0.71	0.78	0.83
0.8	0.32	0.37	0.43	0.50	0.57	0.66	0.73	0.79	0.84
1	0.33	0.39	0.45	0.52	0.60	0.67	0.75	0.81	0.86

Design C: 120 patients per group, 1.25 years of follow-up, monthly eGFRs

Hypothesized effect on chronic slope ml/min/1.73m²/yr	Hypothesized Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.23	0.27	0.31	0.37	0.43	0.51	0.58	0.65	0.72
0.2	0.26	0.29	0.35	0.41	0.47	0.55	0.62	0.70	0.75
0.4	0.29	0.33	0.39	0.45	0.51	0.58	0.65	0.72	0.79
0.6	0.32	0.36	0.42	0.48	0.55	0.62	0.70	0.75	0.81
0.8	0.35	0.40	0.47	0.53	0.59	0.66	0.72	0.78	0.84
1	0.40	0.44	0.49	0.56	0.63	0.69	0.75	0.82	0.85

Design D: 120 patients per group, 2 years of follow-up, quarterly eGFRs

Hypothesized effect on chronic slope ml/min/1.73m²/yr	Hypothesized Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.22	0.27	0.31	0.35	0.41	0.47	0.54	0.61	0.68
0.2	0.26	0.30	0.34	0.4	0.46	0.52	0.59	0.66	0.72
0.4	0.29	0.34	0.38	0.44	0.51	0.57	0.63	0.70	0.76
0.6	0.34	0.39	0.44	0.50	0.56	0.62	0.68	0.74	0.79
0.8	0.38	0.43	0.49	0.54	0.60	0.67	0.73	0.78	0.82
1	0.43	0.48	0.53	0.59	0.65	0.71	0.76	0.81	0.86

Design E: 240 patients per group, 2 years of follow-up, quarterly eGFRs

Hypothesized effect on chronic slope ml/min/1.73m²/yr	Hypothesized Treatment Effect on UACR (GMR)								
	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
0	0.19	0.21	0.25	0.29	0.33	0.40	0.45	0.51	0.57
0.2	0.24	0.28	0.31	0.35	0.40	0.47	0.53	0.58	0.65
0.4	0.30	0.34	0.38	0.43	0.48	0.54	0.60	0.66	0.70
0.6	0.37	0.42	0.46	0.52	0.57	0.62	0.67	0.73	0.77
0.8	0.45	0.48	0.53	0.59	0.64	0.69	0.74	0.79	0.83
1	0.52	0.56	0.61	0.65	0.71	0.76	0.80	0.84	0.87

Supplement Figure 6 Legend:

Shown are the average estimated values of PPV_{trial} for inferring clinical benefit defined as a HR for the clinical endpoint which is less than 0.8 based on combinations of hypothesized true treatment effects on slope and UACR. The estimated values for PPV_{trial} are averaged over the projected sampling distributions for the estimated treatment effects on slope and ACR given the true hypothesized treatment effects on these endpoints. The unshaded region designates average estimated $PPV_{trial} < 0.80$. The light shaded region indicates average estimated PPV_{trial} between 0.80 and 0.90. The dark shaded region indicates average estimated $PPV_{trial} \geq 0.90$.

If the average estimated PPV_{trial} is computed based on the treatment effect on UACR alone, under Design 1 the average estimated PPV_{trial} values are 0.27, 0.33, 0.38, 0.45, 0.54, 0.62, 0.72, 0.79, and 0.85 for true GMRs ranging from 1.00 to 0.60, respectively. Under Designs 2-4 the corresponding values are 0.23, 0.29, 0.37, 0.45, 0.54, 0.64, 0.74, 0.81, and 0.88 respectively. Under Design 5, the corresponding average estimated PPV_{trial} values are 0.22, 0.28, 0.35, 0.45, 0.55, 0.65, 0.75, 0.84, and 0.89.

Abbreviations: UACR urine albumin to creatinine ratio; GMR geometric mean ratio; GFR glomerular filtration rate; PPV_{trial} trial-level positive predictive value

Appendix 1: Study funding sources

Study Name	Funding
AASK	Supported by grants to each clinical center and the coordinating center from the National Institute of Diabetes and Digestive and Kidney Diseases. In addition, AASK was supported by the Office of Research in Minority Health (now the National Center on Minority Health and Health Disparities, NCMHD) and the following institutional grants from the National Institutes of Health: M01 RR-00080, M01 RR-00071, M0100032, P20-RR11145, M01 RR00827, M01 RR00052, 2P20 RR11104, RR029887, and DK 2818-02. King Pharmaceuticals provided monetary support and antihypertensive medications to each clinical center. Pfizer Inc., AstraZeneca Pharmaceuticals, Glaxo Smith Kline, Forest Laboratories, Pharmacia and Upjohn also donated antihypertensive medications.
ABCD	Supported by Bayer and the National Institute of Diabetes, Digestive, and Kidney Diseases (DK50298-02)
ADVANCE	ADVANCE was funded by grants from Servier and the National Health and Medical Research Council of Australia
ALTITUDE	Supported by Novartis
Brenner	Supported by Merck & Co.
Donadio 2001	Supported by research grants from Pronova Biocare a.s. (Oslo, Norway) and Mayo Foundation (Rochester, MN)
EMPA-REG OUTCOME	Supported by Boehringer Ingelheim (BI) and Eli Lilly
HALT-PKD A and B	Supported by grants from the National Institute of Diabetes and Digestive and Kidney Diseases (DK62410 to Dr. Torres, DK62408 to Dr. Chapman, DK62402 to Dr. Schrier, DK082230 to Dr. Moore, DK62411 to Dr. Perrone, and DK62401 to Washington University at St. Louis) and the National Center for Research Resources General Clinical Research Centers (RR000039 to Emory University, RR000585 to the Mayo Clinic, RR000054 to Tufts Medical Center, RR000051 to the University of Colorado, RR023940 to the University of Kansas Medical Center, and RR001032 to Beth Israel Deaconess Medical Center), National Center for Advancing Translational Sciences Clinical and Translational Science Awards (RR025008 and TR000454 to Emory University, RR024150 and TR00135 to the Mayo Clinic, RR025752 and TR001064 to Tufts University, RR025780 and TR001082 to the University of Colorado, RR025758 and TR001102 to Beth Israel Deaconess Medical Center, RR033179 and TR000001 to the University of Kansas Medical Center, and RR024989 and TR000439 to Cleveland Clinic), by funding from the Zell Family Foundation (to the University of Colorado), and by a grant from the PKD Foundation.
Hannedouche	Supported by Merck Sharp & Dohme
HKVIN	Supported by Novartis Pharmaceuticals (Hong Kong) Ltd by providing the study medication and placebo
Hou	Supported by a National Nature and Sciences Grant for Major Projects (30330300) and a People's Liberation Army Grant for Major Clinical Research (to Dr. Hou) and in part by Novartis
IDNT	Supported by the Bristol-Myers Squibb Institute for Medical Research and Sanofi–Synthelabo
Ihle/Kincaid	Supported in part by Merck & Co, Inc., West Point, PA
Kamper	Supported by Merck Sharp & Dohme
Lewis 1992	Supported by grants (R01-AM-27769 and R01-AM-27770) from the Public Health Service
Lewis 1993	Supported by grants from the Public Health Service (5 R01-DK 39908, 5 R01-DK 39826, MO1-RR00030, MO1-RR00034, MO1-RR00036, MO1-RR00051, MO1-RR00058, MO1-RR00059, and

	MO1-RR00425) and by the Bristol-Myers Squibb Pharmaceutical Research Institute (Princeton, N.J.).
Maes	The study medication was kindly provided by Hoffmann-LaRoche, Basel, Switzerland
Maschio	Supported by funding from Ciba Geigy-Novartis
MDRD Study A and B	Supported by the National Institute of Diabetes, Digestive, and Kidney Diseases (NIDDK UO1 DK35073 and K23 DK67303, K23 DK02904). Funding for the MDRD Study included the formerly named Health Care and Financing Administration (HCFA); now the Center for Medicare and Medicaid Services.
ORIENT	Supported by a research grant from Daiichi Sankyo
Ponticelli 1989	Supported in part by a grant (82.01308.04) from the Consiglio Nazionale delle Ricerche.
Ponticelli 1998	Supported in part by a grant from Ospedabc Maggiore di Milano
Ponticelli 2006	This was a spontaneous clinical trial sponsored by the grant “Project Glomerulonephritis”
Pozzi 2004	The authors did not receive any financial support
Pozzi 2010	The authors did not receive any financial support
Pozzi 2012	The authors did not receive any financial support
REIN	Supported by a grant from Hoechst Marion Roussel (HMR) Clinical Research Institute, Frankfurt am Main, Germany
RENAAL	Supported by Merck & Co.
ROAD	Supported by a National Nature and Sciences Grant for Major Projects (30330300), a People's Liberation Army Grant for Major Clinical Research (2000), and National 11th Five-Years Plan Foundation (to F.F.H.)
Schena	Supported in part by a grant of University of Bari
STOP- IgAN	Supported by a grant (GFVT01044604) from the German Federal Ministry of Education and Research.
SUN- MACRO	Sponsored by Keryx Biopharmaceuticals, Inc.
Toto	By grant RO1 DK53869A from the U.S. National Institute of Diabetes and Digestive and Kidney Diseases (Dr. Levey); grant RO1 HS 10064 from the Agency for Healthcare Research and Quality (Dr. Schmid); a grant from Dialysis Clinic, Inc., Paul Teschan Research Fund 1097-5 (Dr. Jafar); New England Medical Center St. Elizabeth’s Hospital Clinical Research Fellowship, Boston, Massachusetts (Dr. Jafar); and an unrestricted grant from Merck Research Laboratories, West Point, Pennsylvania (Dr. Levey).
Van Essen	Supported by Merck Sharp & Dohme, Haarlem, The Netherlands

Supplement References

1. Stoycheff N, Pandya K, Okparavero A, Schiff A, Levey AS, Greene T, 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. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association,* 26**:** 848-857, 2011 10.1093/ndt/gfq525
2. Heerspink HJL, Greene T, Tighiouart H, Gansevoort RT, Coresh J, Simon AL, et al.: Change in albuminuria as a surrogate endpoint for progression of kidney disease: a meta-analysis of treatment effects in randomised clinical trials. *The lancet Diabetes & endocrinology,* 7**:** 128-139, 2019 10.1016/S2213-8587(18)30314-0
3. Lewis EJ, Hunsicker LG, Bain RP, Rohde RD: The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. The Collaborative Study Group. *The New England journal of medicine,* 329**:** 1456-1462, 1993 10.1056/NEJM199311113292004
4. Praga M, Barrio V, Juarez GF, Luno J: Tacrolimus monotherapy in membranous nephropathy: a randomized controlled trial. *Kidney international,* 71**:** 924-930, 2007
5. Chan TM, Tse KC, Tang CS, Mok MY, Li FK: Long-term study of mycophenolate mofetil as continuous induction and maintenance treatment for diffuse proliferative lupus nephritis. *J Am Soc Nephrol,* 16**:** 1076-1084, 2005
6. Li PK, Leung CB, Chow KM, Cheng YL, Fung SK, Mak SK, et al.: Hong Kong study using valsartan in IgA nephropathy (HKVIN): a double-blind, randomized, placebo-controlled study. *Am J Kidney Dis,* 47**:** 751-760, 2006 10.1053/j.ajkd.2006.01.017
7. Ponticelli C, Passerini P, Salvadori M, Manno C, Viola BF, Pasquali S, et al.: A randomized pilot trial comparing methylprednisolone plus a cytotoxic agent versus synthetic adrenocorticotropic hormone in idiopathic membranous nephropathy. *Am J Kidney Dis,* 47**:** 233-240, 2006
8. Ponticelli C, Zucchelli P, Passerini P, Cagnoli L, Cesana B, Pozzi C, et al.: A randomized trial of methylprednisolone and chlorambucil in idiopathic membranous nephropathy. *The New England journal of medicine,* 320**:** 8-13, 1989
9. Maes BD, Oyen R, Claes K, Evenepoel P, Kuypers D, Vanwalleghem J, et al.: Mycophenolate mofetil in IgA nephropathy: results of a 3-year prospective placebo-controlled randomized study. *Kidney international,* 65**:** 1842-1849, 2004 10.1111/j.1523-1755.2004.00588.x
10. Pozzi C, Andrulli S, Del Vecchio L, Melis P, Fogazzi GB, Altieri P, et al.: Corticosteroid effectiveness in IgA nephropathy: long-term results of a randomized, controlled trial. *J Am Soc Nephrol,* 15**:** 157-163, 2004
11. Pozzi C, Andrulli S, Pani A, Scaini P, Del Vecchio L, Fogazzi G, et al.: Addition of azathioprine to corticosteroids does not benefit patients with IgA nephropathy. *J Am Soc Nephrol,* 21**:** 1783-1790, 2010 10.1681/ASN.2010010117
12. Pozzi C, Andrulli S, Pani A, Scaini P, Roccatello D, Fogazzi G, et al.: IgA nephropathy with severe chronic renal failure: a randomized controlled trial of corticosteroids and azathioprine. *Journal of nephrology,* 26**:** 86-93, 2013 10.5301/jn.5000110
13. Katafuchi R, Ikeda K, Mizumasa T, Tanaka H, Ando T, Yanase T, et al.: Controlled, prospective trial of steroid treatment in IgA nephropathy: a limitation of low-dose prednisolone therapy. *Am J Kidney Dis,* 41**:** 972-983, 2003 S027263860300194X [pii]
14. Manno C, Torres DD, Rossini M, Pesce F, Schena FP: Randomized controlled clinical trial of corticosteroids plus ACE-inhibitors with long-term follow-up in proteinuric IgA nephropathy. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association,* 24**:** 3694-3701, 2009 10.1093/ndt/gfp356

15. Hou FF, Xie D, Zhang X, Chen PY, Zhang WR, Liang M, et al.: Renoprotection of Optimal Antiproteinuric Doses (ROAD) Study: a randomized controlled study of benazepril and losartan in chronic renal insufficiency. *J Am Soc Nephrol,* 18**:** 1889-1898, 2007 ASN.2006121372 [pii]
10.1681/ASN.2006121372
16. Packham DK, Wolfe R, Reutens AT, Berl T, Heerspink HL, Rohde R, et al.: Sulodexide fails to demonstrate renoprotection in overt type 2 diabetic nephropathy. *J Am Soc Nephrol,* 23**:** 123-130, 2012 10.1681/ASN.2011040378
17. Parving HH, Brenner BM, McMurray JJ, de Zeeuw D, Haffner SM, Solomon SD, et al.: Cardiorenal end points in a trial of aliskiren for type 2 diabetes. *The New England journal of medicine,* 367**:** 2204-2213, 2012 10.1056/NEJMoa1208799
18. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, et al.: Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *The New England journal of medicine,* 373**:** 2117-2128, 2015 10.1056/NEJMoa1504720
19. Vonesh E, Tighiouart H, Ying J, Heerspink HL, Lewis J, Staplin N, et al.: Mixed-effects models for slope-based endpoints in clinical trials of chronic kidney disease. *Statistics in medicine,* 38**:** 4218-4239, 2019 10.1002/sim.8282
20. Vonesh EF, Greene T, Schluchter MD: Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in medicine,* 25**:** 143-163, 2006 10.1002/sim.2249
21. Rizopoulos D: *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*, 1st Ed., New York, Chapman and Hall/CRC, 2012
22. Gail MH, Pfeiffer R, Van Houwelingen HC, Carroll RJ: On meta-analytic assessment of surrogate outcomes. *Biostatistics,* 1**:** 231-246, 2000 10.1093/biostatistics/1.3.231
23. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H: The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics,* 1**:** 49-67, 2000 10.1093/biostatistics/1.1.49
24. Daniels MJ, Hughes MD: Meta-analysis for the evaluation of potential surrogate markers. *Statistics in medicine,* 16**:** 1965-1982, 1997 10.1002/(sici)1097-0258(19970915)16:17<1965::aid-sim630>3.0.co;2-m
25. Bujkiewicz S, Thompson JR, Riley RD, Abrams KR: Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process. *Statistics in medicine,* 35**:** 1063-1089, 2016 10.1002/sim.6776
26. Burzykowski T, Molenberghs G, Buyse M: (Eds.) *The Evaluation of Surrogate Endpoints,* New York, Springer, 2005
27. Inker LA, Heerspink HJL, Tighiouart H, Chaudhari J, Miao S, Diva U, et al.: Association of Treatment Effects on Early Change in Urine Protein and Treatment Effects on GFR Slope in IgA Nephropathy: An Individual Participant Meta-analysis. *Am J Kidney Dis*, 2021 10.1053/j.ajkd.2021.03.007