

ARTICLE TYPE

Mixed-effects models for slope-based endpoints in clinical trials of chronic kidney disease

Edward Vonesh^{*1} | Hocine Tighiouart² | Jian Ying³ | Hiddo L Heerspink⁴ | Julia Lewis⁵ | Natalie Staplin⁶ | Lesley Inker⁷ | Tom Greene⁸

¹Department of Preventive Medicine, Division of Biostatistics, Northwestern University, Chicago, IL, USA

²The Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

³Department of Internal Medicine, School of Medicine, University of Utah, Salt Lake City, Utah, USA

⁴Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, Groningen, the Netherlands

⁵Department of Medicine, Division of Nephrology, Vanderbilt University, Nashville, TN, USA

⁶Medical Research Council Population Health Research Unit, Nuffield Dept. of Population Health, University of Oxford, Oxford, UK

⁷Department of Nephrology, Tufts Medical Center, Boston, MA, USA

⁸Department of Internal Medicine, Division of Epidemiology, University of Utah, Salt Lake City, Utah, USA

Correspondence

*Edward F Vonesh, Department of Preventive Medicine, Division of Biostatistics, Northwestern University, 680 N. Lake Shore Drive, Suite 1400, Chicago, IL 60611
email: e-vonesh@northwestern.edu

Funding information

US National Kidney Foundation

Abstract

In March of 2018, the National Kidney Foundation, in collaboration with the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA), sponsored a workshop in which surrogate endpoints other than currently established event-time endpoints for clinical trials in chronic kidney disease (CKD) were presented and discussed. One such endpoint is a slope-based parameter describing the rate of decline in the estimated glomerular filtration rate (eGFR) over time. There are a number of challenges that can complicate such slope-based analyses in CKD trials. These include the possibility of an early but short-term acute treatment effect on the slope, both within-subject and between-subject heteroscedasticity, and informative censoring resulting from patient dropout due to death or onset of end-stage kidney disease (ESKD). To address these issues, we first consider a class of mixed-effects models for eGFR that are linear in the parameters describing the mean eGFR trajectory but which are intrinsically nonlinear when a power-of-mean (POM) variance structure is used to model within-subject heteroscedasticity. We then combine the model for eGFR with a model for time to dropout to form a class of shared parameter (SP) models which, under the right specification of shared random effects, can minimize bias due to informative censoring. The models and methods of analysis are described and illustrated using data from two CKD studies one of which was one of 56 studies made available to the workshop analytical team. Lastly, methodology and accompanying software for prospectively determining sample size/power estimates are presented.

KEYWORDS:

Linear spline mixed-effects models; power-of-mean; acute and chronic slopes; informative censoring; shared parameter models.

1 | INTRODUCTION

A two-day workshop sponsored by the National Kidney Foundation, in collaboration with the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA), was held on March 15-16, 2018 in Silver Springs, MD USA to examine and discuss the use of changes over time in albuminuria and glomerular filtration rate (GFR) as endpoints for clinical trials in

early stages of chronic kidney disease (CKD). Currently, the accepted endpoints for clinical trials of CKD are the event-time outcomes of end-stage kidney disease (ESKD) and/or a designated decline in estimated glomerular filtration rate (eGFR in mL/min/1.73m²) such as occurs with a doubling of serum creatinine which is equivalent to 57% decline in eGFR. An analysis of these endpoints is carried out using standard event-time survival models like the Cox proportional hazards model. As such, clinical trials are powered on the basis of the number of clinical events required to detect some pre-specified reduction in the hazard ratio (HR) for the designated events below the null hypothesis of 1.00 between treated and control subjects. In many cases, this may result in a prohibitively long and expensive trial particularly when targeting patients who are in the earlier stages of CKD. To illustrate, under a 1:1 sampling ratio of treated to control patients and assuming a two-sided test with type I error of 0.05, the total number of events required to detect a 20% reduction in the HR (treated vs. control) compared to a null HR of 1.00 would be $(Z_\beta + Z_{\alpha/2})^2 / [\ln(0.80)]^2 = 631$ events with 80% power and 845 events with 90% power¹. At earlier stages of disease, these events often do not occur until after 10 or more years. It is not surprising, then, that there would be so much interest on the part of both industry and the regulatory agencies in exploring the use of alternative endpoints for use in clinical trials of CKD. Since a decline in eGFR is a necessary intermediate in the progression to kidney failure, a reasonable alternative endpoint would be the rate of decline in eGFR as measured by slope-based parameters from a linear mixed-effects model. Because such slope-based endpoints can be assessed over a shorter time frame, clinical trials could be conducted among patients in the early stages of CKD. From a patient's perspective, this could lead to interventions that allow for treatment earlier in the disease course.

In this paper, we present various mixed-effects modeling strategies that were developed for and presented during the workshop for the purpose of characterizing and comparing longitudinal trends in the rate of decline in eGFR among individuals with CKD. The models considered are extensions of previous models used to describe and compare rates of decline in eGFR based on both randomized controlled clinical trials and simulation studies.²⁻⁵ The eGFR models considered range in complexity from linear spline mixed-effects models that allow for both acute (short-term) and chronic (long-term) slope effects per treatment group to simple linear mixed-effects models with a single slope parameter per treatment group. Within this class of models, consideration must also be given to different variance-covariance structures required to explain observed patterns of within-subject and between-subject heteroscedasticity. One set of assumptions would be to assume homogeneity for both within-subject and between-subject variability. A second set of assumptions would allow for within-subject heteroscedasticity by assuming an intra-subject power-of-mean (POM) variance structure and between-subject homogeneity assuming a uniform treatment effect on the random slope effects. A third set of assumptions would allow for within-subject homogeneity assuming a constant variance and between-subject heteroscedasticity assuming some form of a treatment-dependent effect on the random slope effects that would induce different random-effects covariance structures between treatment groups. Finally, a fourth set of assumptions would allow for both within-subject heteroscedasticity via a POM variance structure and between-subject heteroscedasticity via a treatment-dependent effect on the random slope effects.

Under the assumptions that the underlying mixed-effects model is correctly specified and that missing values of eGFR are missing at random (MAR), valid statistical inference based on all observed data is possible using likelihood-based methods. However, the assumption that missing data are MAR is often untenable in CKD trials requiring 2 or more years of follow-up. This is especially true for trials involving patients in the later stages of CKD (stages 3a, 3b and 4) as opposed to early stages of CKD (stages 1 and 2) as there is a greater likelihood that some of these patients will die while others will transition into end-stage kidney disease (ESKD) requiring dialysis or transplantation. This leads to the possibility of informative censoring (i.e., non-ignorable dropout) which, in turn, renders standard likelihood-based inference invalid. As has been noted by some authors, patient withdrawal from a trial due to terminal events like death or dialysis renders many of the concepts and methods associated with standard missing data terminology questionable.⁶⁻⁸ For example, standard missing data terminology would portray each patient as having a vector of complete-data that would occur in the absence of missing data due to dropout. However, this is not plausible in late stage CKD trials where many patients who dropout do so because they develop ESKD requiring dialysis. Once on dialysis, eGFR is no longer an accurate reflection of a patient's GFR as it also reflects dialysis clearance which affects remaining renal function.⁹⁻¹⁰ A better conceptual framework for CKD trials is to define the underlying model for eGFR as referring to change in eGFR from the time of randomization until time-to-dropout due to death or ESKD. In this case, the period over which changes in eGFR occur, i.e., the time-to-dropout, is itself a patient-specific random variable. If time-to-dropout is related to the true but unobserved patient-specific eGFR slope or to some other unobserved characteristic of eGFR prior to dropout, then standard likelihood inference that ignores time-to-dropout will suffer from the same set of biases as would occur under the standard missing data formulation. Consequently, one needs to develop an analysis strategy that addresses concerns over possible informative censoring. One such strategy would be to adopt the use of so-called shared-parameter (SP) models.^{6-8,11}

In the CKD setting, shared parameter (SP) models involve jointly modeling trends (i.e., slopes) in eGFR over time together with potentially non-ignorable censoring events (e.g., death, dialysis or transplantation). This consists of jointly fitting a mixed-effects model for eGFR and an event-time model for the censoring events using a set of shared random effects. Estimated parameters associated with the shared random effects determine to what extent, if any, changes in eGFR over time are associated with potentially non-ignorable censoring events. When the association is significant (i.e., informative), a SP model can reduce bias in the estimated eGFR slopes resulting from informative censoring, at least to the extent that the event-time portion of the SP model approximates the true underlying dropout mechanism. As such, several SP models may be fit in a series of sensitivity analyses as a means for assessing what impact potential informative censoring may or may not have on the estimation and comparison of rates of decline in eGFR. This, in turn, will help when evaluating the agreement of treatment effects on endpoints based on eGFR slopes (acute, chronic, and total) with established clinical endpoints across a broad range of CKD clinical trials.

We start in Section 2 by presenting a class of linear spline mixed-effects models for eGFR that allow for 1) the presence of both an acute (short-term) and chronic (long-term) slope, 2) within-subject heteroscedasticity over time and 3) between-subject heteroscedasticity defined by a treatment-dependent effect on the random slope effects. A model within this class may serve as the primary analysis model in a CKD trial provided the analysis is carried out using maximum likelihood procedures together with the assumption that missing data are MAR. In Section 3, we extend this class of models to a class of shared parameter (SP) models by including an event-time model for patient dropout due to death or ESKD (dialysis or transplantation). Sensitivity analyses under alternative non-ignorable dropout mechanisms may be carried out by jointly modeling eGFR and patient dropout under different shared random effect specifications. Section 4 briefly describes estimation and inference based on maximum likelihood methodology. The models and methods are illustrated in Section 5 using data from the Irbesartan Diabetic Nephropathy Trial (IDNT) as provided to the workshop analytical team and described previously^{2-5,12-13} as well as to the Modification of Diet in Renal Disease study B (MDRD-B) which had been analyzed previously.⁷⁻⁸ Section 6 describes and illustrates an approach to sample size/power calculations for CKD clinical trials which one can implement using a SAS macro program made available through the repository Figshare.

2 | MIXED-EFFECTS MODELS FOR GFR SLOPES

Serial trends in eGFR can be modeled using either a two-slope linear spline mixed-effects model in which an acute and chronic slope are defined by a fixed knot or change point; or a simple linear mixed-effects model with a single common slope per treatment intervention. As the latter is a special case of the former, we need only consider a class of two-slope linear spline mixed-effects models. While models within this class are characterized by two slopes, there are actually four different slope-based parameters that describe the overall rate of decline in eGFR over time: an acute slope, a chronic slope, a change or delta slope and a time-dependent total slope. These parameters each describe different aspects of the overall eGFR trajectory that individuals experience over time. In contrast, under a simple linear mixed-effects model, there is a single slope per treatment intervention that describe the rate of decline in eGFR.

2.1 | A two-slope linear spline mixed-effects models for eGFR

Often in CKD trials, hemodynamic and other effects associated with various treatment interventions may result in an initial acute effect on eGFR shortly after the treatment is initiated (e.g., the first 3 to 12 months) that differs from the long-term treatment effect. When this occurs, the rate of decline in eGFR over time may be characterized by two slopes: an acute slope and chronic slope. To evaluate whether treatment interventions have such an acute effect, one can chose to fit a piecewise linear spline mixed-effects model to the eGFR measurements with a fixed knot or change point starting somewhere between 3 to 12 months. Ideally, the possible presence of an acute effect and where that effect ends and a more stable long-term effect starts would be determined in advance based on analyses from a phase 2 trial. In cases where this is not possible, one would need to perform preliminary analyses of the phase 3 data that would include some simple plots that are suggestive of where a fixed knot might be placed. Alternatively, assuming homogeneous within- and between-subject variability, one can fit a series of simple linear spline mixed-effects models with fixed knots varying between say 3 to 12 months and choose the knot corresponding to the model with the best fit as determined by Akaike's information criterion, AIC.

Once a fixed knot or change point has been determined, one can formulate a two-slope linear spline mixed-effects model as follows. Let t^* denote a fixed knot or change point and let $\max\{t_{ij} - t^*, 0\}$ be the truncated line function where t_{ij} are the time

points at which eGFR is measured for the i^{th} subject across p_i occasions ($j = 1, \dots, p_i$). Let Y_{ij} denote the response variable, eGFR, for the i^{th} subject on the j^{th} occasion and let X_i be a treatment group indicator variable with $X_i = 0$ if the i^{th} subject is randomized to a placebo control group and $X_i = 1$ if randomized to an active treatment group. Conceptually, the model may be written using a two-stage random coefficient regression formulation as

$$\begin{aligned} \text{Stage 1:} \quad Y_{ij} &= \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i} \max\{t_{ij} - t^*, 0\} + \epsilon_{ij}; \quad (i = 1, \dots, n; j = 1, \dots, p_i) \\ &= \begin{cases} \beta_{0i} + \beta_{1i}t_{ij} + \epsilon_{ij} & \text{when } t_{ij} < t^* \\ (\beta_{0i} - \beta_{2i}t^*) + (\beta_{1i} + \beta_{2i})t_{ij} + \epsilon_{ij} & \text{when } t_{ij} \geq t^* \end{cases} \\ \text{Stage 2:} \quad \begin{pmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \end{pmatrix} &= \begin{pmatrix} \beta_{0c} \\ \beta_{1c} \\ \beta_{2c} \end{pmatrix} (1 - X_i) + \begin{pmatrix} \beta_{0t} \\ \beta_{1t} \\ \beta_{2t} \end{pmatrix} X_i + \begin{pmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{pmatrix}; \quad (i = 1, \dots, n) \end{aligned} \quad (1)$$

where Stage 1 consists of n separate within-subject linear spline regression models with subject-specific (SS) regression parameters $\beta'_i = (\beta_{0i} \ \beta_{1i} \ \beta_{2i})$ for each subject and within-subject residuals ϵ_{ij} , while Stage 2 consists of a marginal multivariate linear regression model with population-average (PA) regression parameters, $\beta_c = (\beta_{0c} \ \beta_{1c} \ \beta_{2c})'$ and $\beta_t = (\beta_{0t} \ \beta_{1t} \ \beta_{2t})'$, for the control and treatment groups respectively, and between-subject random effects $b_i = (b_{0i} \ b_{1i} \ b_{2i})'$. By combining Stage 1 and 2, we end up with the linear spline mixed-effects model

$$Y_{ij} = (\beta_{0c} + \beta_{1c}t_{ij} + \beta_{2c} \max\{t_{ij} - t^*, 0\}) (1 - X_i) + (\beta_{0t} + \beta_{1t}t_{ij} + \beta_{2t} \max\{t_{ij} - t^*, 0\}) X_i + b_{0i} + b_{1i}t_{ij} + b_{2i} \max\{t_{ij} - t^*, 0\} + \epsilon_{ij}. \quad (2)$$

To complete the model specifications, the following assumptions are made.

(A.1) The $(\epsilon_{ij} | b_i)$ are conditionally independent normally distributed random variables with 0 means and within-subject variances σ_{ij}^2 that reflect either a common variance across time or a variance that varies with the mean over time.

(A.2) The SS random effects b_i are independent multivariate normal random vectors with 0 mean components and a 3×3 variance-covariance matrix Ψ_k for the k^{th} treatment group ($k = c$ for the control group and $k = t$ for the treatment group) that reflects either a uniform treatment effect (i.e., a common unstructured covariance matrix, Ψ , between treatment groups) or a treatment-dependent effect (i.e., a covariance matrix that differs between treatment groups most notably with respect to the variances and covariances of the random slope effects).

Although the model as formulated here assumes just two treatment groups, one can easily extend the model to include three or more treatment groups as well as other baseline covariates in Stage 2 of the two-stage formulation (1). Under Stage 1 of formulation (1), β_{0i} represents the i^{th} subject's intercept, β_{1i} is the subject's initial or acute slope up through but excluding time t^* , and β_{2i} is the subject's directional change in slope starting at time t^* . A second regression line is then formed for the i^{th} subject starting at time t^* with the term $(\beta_{0i} - \beta_{2i}t^*)$ representing an extrapolated intercept back to time 0 while the term $\beta_{3i} = (\beta_{1i} + \beta_{2i})$ represents a chronic or projected long-term slope. The change slope (or delta slope) β_{2i} is then seen simply as the difference between the chronic slope β_{3i} and the acute slope β_{1i} . In addition to the acute, chronic and change slopes, one can also define a time-dependent total slope at a specified time point $t > t^*$ (e.g., $t = \text{the total study time}$) as

$$\begin{aligned} \beta_{4i}(t) &= \{[\beta_{0i} + \beta_{1i}t + \beta_{2i}(\max(t - t^*, 0)) - \beta_{0i}]/t \\ &= \beta_{3i} - \beta_{2i}(t^*/t). \end{aligned} \quad (3)$$

This is simply the i^{th} subject's total rate of change in eGFR from time 0 to time t (i.e., the subject-specific mean change from baseline divided by t). As so defined, a subject's total slope converges to the chronic slope as $t \rightarrow \infty$ or as the change or delta slope $\beta_{2i} \rightarrow 0$. The regression parameters in Stage 2 of formulation (1) will have the same interpretation as the SS regression parameters of Stage 1 except that they represent the marginal or population-average (PA) intercept, acute slope, change slope, chronic slope and total slope.

The linear spline mixed-effects model defined by either (1) or (2) together with assumptions (A.1)-(A.2) forms a rather rich class of mixed-effects models from which to choose when evaluating slope-based endpoints in clinical trials of CKD. For example, one can assume homogeneity both within- and between-subjects by assuming the within-subject errors satisfy $\text{Var}(\epsilon_{ij}) = \sigma_{ij}^2 = \sigma^2$ for all (i, j) and the between-subject covariance structure satisfies $\text{Cov}(b_i) = \Psi$ for all i . In this case, one can fit the linear spline mixed-effects model using standard linear mixed models software (e.g., MIXED procedures in SAS or SPSS,

the lme package in R). Alternatively, one may wish to model within-subject heteroscedasticity over time assuming a subject-specific power-of-mean (SS-POM) structure with $\text{Var}(\epsilon_{ij}) = \sigma_{ij}^2 = \sigma^2(\mu_{ij}^2(\beta_i))^\theta$ where $\mu_{ij}(\beta_i) = E(Y_{ij}|\mathbf{b}_i)$ is the conditional mean as defined in Stage 1 of the two-stage model (1). Now, however, the model no longer fits within the class of linear mixed models one commonly encounters with commercial software as the SS regression parameters appear in both the mean and covariance structure. Consequently, one must then resort to the use of nonlinear mixed-effects software (e.g., NLMIXED in SAS or the nlme package in R).

Shown in Table 1 are candidate structures for σ_{ij}^2 and Ψ_k that allow for within- and between-subject homogeneity and/or heteroscedasticity, respectively. The most general form of the two-slope linear spline mixed-effects model would be one in which the Ψ_k are possibly different but unstructured covariance matrices resulting in 12 variance-covariance parameters (6 per treatment group) one would need to estimate. To reduce the number of covariance parameters, we elected to model between-subject heteroscedasticity assuming a uniform treatment effect on the random intercept effect b_{0i} (as would be expected for randomized clinical trials) and a common proportional treatment effect on the random slope effects b_{1i} and b_{2i} , respectively. This is accomplished by modifying the random-effects structure of the Stage 2 formulation in (1) by writing \mathbf{b}_i as

$$\mathbf{b}_i = \mathbf{Z}_i(\kappa)\mathbf{u}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (1 + X_i\kappa) & 0 \\ 0 & 0 & (1 + X_i\kappa) \end{pmatrix} \begin{pmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \end{pmatrix} = \begin{pmatrix} u_{0i} \\ (1 + X_i\kappa)u_{1i} \\ (1 + X_i\kappa)u_{2i} \end{pmatrix}. \quad (4)$$

where the \mathbf{u}_i are independent identically distributed $N(\mathbf{0}, \Psi)$ random effects. The parameter κ is a treatment-related proportionality constant in that it defines the heterogeneous between-subject covariance structure Ψ_k in Table 1. Specifically, $\phi_k = 1$ when $k = c$ (corresponding to the control group) and $\phi_k = (1 + \kappa)$ when $k = t$ (corresponding to the treatment group). By testing the hypothesis $H_0: \kappa = 0$ versus $H_A: \kappa \neq 0$, one can directly test whether there is a proportional treatment effect on the random slope effects or not. Under this specification, the active treatment group ($k = t$) will have lower variability with respect to the random slope effects (b_{1i}, b_{2i}) compared with the control group ($k = c$) whenever $-2 < \kappa < 0$ as these values result in values of ϕ_k^2 in the range of $0 \leq \phi_k^2 < 1$ with 0 occurring only when $\kappa = -1$ (in which case there is only a random intercept effect b_{0i} with common variance ψ_{11} for both treatment groups). It should be noted that by simply setting both θ and κ equal to 0, this model reduces to a piecewise linear spline mixed-effects model with homogenous within- and between-subject variability.

3 | SHARED PARAMETER MODELS

For clinical trials in diseases like CKD and idiopathic pulmonary fibrosis (IPF) where slope-based endpoints may serve as primary or key secondary endpoints, shared parameter (SP) models provide a valuable analytical tool for performing sensitivity analyses in the presence of informative censoring.^{6-8,11,14} Briefly, a shared parameter model is a random-effects dependent selection model in that it entails factoring the joint distribution of repeated measurements on a response variable of interest, say Y , and event times for dropout, say T , assuming these two outcome variables are conditionally independent given a set of shared random effects (i.e., the so-called shared parameters). Specifically, the shared parameter model corresponds to the factorization

$$\pi(\mathbf{Y}_i, T_i) = \int_b \pi(\mathbf{Y}_i, T_i | \mathbf{b}_i) \pi(\mathbf{b}_i) d\mathbf{b}_i = \int_b \pi(\mathbf{Y}_i | \mathbf{b}_i) \pi(T_i | \mathbf{b}_i) \pi(\mathbf{b}_i) d\mathbf{b}_i \quad (5)$$

where $\pi(\cdot)$ denotes a density function, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip_i})'$ is the intended but possibly unbalanced vector of repeated measurements (e.g., eGFR) on the i^{th} subject taken at times t_1, \dots, t_{p_i} , T_i is the event time of interest (e.g., time to dropout due to death or ESKD), and \mathbf{b}_i is a vector of latent random effects. Possible correlation (association) between the repeated measurements and event times is induced through the shared random effects. Under any given specification of this joint model, dropout may be classified as ignorable provided there is evidence that $\pi(T_i | \mathbf{b}_i) = \pi(T_i)$ since \mathbf{Y}_i and T_i are then independently distributed. Of course this only applies under the unverifiable assumption that the components of the joint model have all been correctly specified and that $\pi(T_i | \mathbf{b}_i)$ corresponds to the actual dropout mechanism.

In the CKD setting, SP models involve jointly modeling serial trends (i.e., slopes) in eGFR together with event times for censoring events like death, dialysis or transplantation. This entails jointly fitting a longitudinal component model, $\pi(\mathbf{Y}_i | \mathbf{b}_i)$, for eGFR and a survival component model, $\pi(T_i | \mathbf{b}_i)$, for the censoring event times using a set of shared random effects across the two models. Estimated parameters linked to the shared random effects determine to what degree, if any, changes in eGFR over time are associated with the potentially non-ignorable censoring events. When the association is significant (i.e., informative), a SP model can reduce bias in the estimated eGFR slopes due to informative censoring - at least to the extent that the specified

event-time survival model approximates the true underlying dropout mechanism. As such, several SP models may be used in a sensitivity analysis as a means for assessing what impact potential informative censoring may have on the estimation and comparison of eGFR slopes.

Here the longitudinal component model $\pi(Y_i|b_i)$ for eGFR is the linear spline mixed-effects model (2). Time to dropout due to death or ESKD (dialysis or transplantation) is modeled using a proportional hazards (PH) survival model defined by an underlying baseline hazard function and a set of SP covariates used to assess the presence or absence of informative censoring. Two common choices for the survival component model $\pi(T_i|b_i)$ are a fully parametric Weibull survival model and a semiparametric piecewise exponential (PE) survival model both of which may be classified as PH models.⁷⁻⁸ An example in which both survival models have been used in combination with a simple linear mixed-effects regression model for GFR (as measured by renal clearance of [¹²⁵I]iothalamate) has been presented previously based on the MDRD-B study.⁷

3.1 | Piecewise Exponential Survival Model

For the applications presented here, we chose to use the semiparametric piecewise exponential (PE) model as the survival component model for each of the SP models considered.^{7-8,15-16} Let T_i be the event time to death or ESKD, whichever occurs first. Events other than death or ESKD such as administrative censoring due to staggered entry or study completion as well as subject withdrawal for reasons unrelated to death or ESKD are assumed to be random censoring events independent of death or ESKD. Let C_i be the time to such independent censoring events and let $T_i^o = \min(T_i, C_i)$. We observe (T_i^o, δ_i) where $\delta_i = I(T_i \leq C_i)$ is the event indicator variable defined to be 1 if $T_i^o = T_i$ and 0 if $T_i^o = C_i$. Under a PE survival model, the time scale is partitioned into a sequence of m disjoint intervals, $(0, t_1], (t_1, t_2], \dots, (t_{m-1}, t_m]$ over which the baseline hazard function, $\lambda_0(t)$, is allowed to vary from interval to interval but which is assumed constant within an interval according to

$$\lambda_0(t) = \sum_{h=1}^m \lambda_{0h} I(t \in (t_{h-1}, t_h]) \quad (6)$$

where $\lambda_{0h} = \exp(\eta_{0h})$. The baseline hazard function (6) closely approximates that of the counting process style of the extended Cox PH model with the exception that the latter allows for an arbitrary baseline hazard rate.¹⁶⁻¹⁷ An advantage of the PE survival model is that it is semiparametric and, unlike the rank-ordered partial likelihood approach of a Cox model, is easily programmable under a SP model scenario.

By including the treatment covariate, X_i , along with say l baseline covariates, $U_{1i}, U_{2i}, \dots, U_{li}$, the overall PE hazard function in combination with the proportional hazards assumption may be written as

$$\lambda(t) = \lambda_0(t) \exp\{\eta_0 X_i + \eta_1 U_{1i} + \dots + \eta_l U_{li}\}. \quad (7)$$

One can easily extend the PE hazard function to include one or more time-dependent covariates by simply defining updated values for each time-dependent covariate at the start of a new interval and holding that value fixed for the duration of the interval. Therein lies one of the major advantages associated with using the PE survival model; namely the ability to model time-dependent covariates. Other advantages include the fact that 1) the PE survival model will very often result in HR estimates and standard errors similar to those obtained using the extended Cox PH model especially when the same intervals are used for both models; and 2) the PE survival model can be adapted to model baseline covariates that do not satisfy the proportional hazards assumption. For example, if the HR for two treatment groups does not satisfy the proportional hazards assumption, one can model a non-proportional hazards function by simply replacing the treatment group as a single covariate having a common HR across intervals with a treatment group by interval interaction term with time varying HR's across intervals.

3.2 | Shared Parameter Covariates

For sensitivity analyses, one can formulate a set of SP “covariates” for the survival component model $\pi(T_i|b_i)$ that, as functions of the random effects, can be used to assess what impact, if any, censoring events may have on the eGFR slope estimates. For example, in addition to the treatment indicator variable X_i , one might include the random intercept effect b_{0i} , the random acute slope effect b_{1i} , and the random change slope effect b_{2i} as time-independent latent “covariates” or predictors of dropout due to death or ESKD. Since $b_{3i} = b_{1i} + b_{2i}$ is a linear combination of b_{1i} and b_{2i} , one could just as easily replace b_{2i} with b_{3i} without changing the value of the overall likelihood function nor of the values of the linear spline mixed-effects model parameters. We elected to use b_{3i} as it provides greater insight into how subject-specific deviations from the treatment group chronic slope may be predictive of dropout.

Listed in Table 2 are the PE model components of four possible SP models. Excluding the treatment group indicator, each of the other covariates represents some latent patient-specific aspect of the longitudinal trend in eGFR under model (2). Strictly speaking, SP Model 1 with its lone covariate X_i is not a SP model as the survival component model $\pi(T_i|X_i)$ does not depend on any of the latent random effect parameters. It does, however, allow one to compare eGFR slopes under the assumption of ignorable dropout (i.e., under SP model 1, $\pi(Y_i, T_i) = \pi(Y_i)\pi(T_i)$ implying dropout is completely independent of eGFR) to the eGFR slopes under several non-ignorable dropout mechanisms. For example, the time-dependent covariate $\mu_{ih}(t_{h-1})$ under SP model 3 is the i^{th} subject's mean predicted eGFR at the start of the h^{th} interval of follow-up. It represents a subject's true but unobserved time-dependent eGFR measured without error. If the HR associated with this eGFR is found to be significantly less than 1.0 then this would be evidence of non-ignorable dropout in that subjects with higher but unobserved (missing) values of eGFR will be less likely to dropout compared with those with lower values. Similarly, evidence for non-ignorable dropout would occur if the HR associated with the chronic random slope effect b_{3i} is significantly less than 1.0 as subjects with higher values of b_{3i} (e.g., values > 0) will have a lower long-term rate of decline in eGFR compared with those with lower values of b_{3i} (e.g., values < 0).

4 | ESTIMATION AND INFERENCE

Under the linear spline mixed-effects model (2) and its assumptions (A.1)-(A.2) together with the SP model factorization (5) and PE survival model with baseline hazard function (6), maximum likelihood (ML) estimation is achieved by maximizing the integrated log-likelihood function

$$L(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\eta}; \mathbf{y}, \mathbf{T}) = \sum_{i=1}^n \log \int \pi(\mathbf{y}_i | \mathbf{b}_i) \pi(T_i^o | \mathbf{b}_i)^{\delta_i} \exp[-\Lambda(T_i^o | \mathbf{b}_i)]^{1-\delta_i} \pi(\mathbf{b}_i) d\mathbf{b}_i \quad (8)$$

where \mathbf{y} is the column vector of observed values $(\mathbf{y}_1, \dots, \mathbf{y}_n)$ of (Y_1, \dots, Y_n) , \mathbf{T} is the vector of observed event times (T_1^o, \dots, T_n^o) , $\boldsymbol{\beta}$ is the column vector of regression parameters $(\boldsymbol{\beta}_c, \boldsymbol{\beta}_t)$ and $\boldsymbol{\tau} = (\sigma^2, \theta, \text{Vech}(\boldsymbol{\Psi}), \kappa)$ the vector of variance-covariance parameters under (2), and $\Lambda(T_i^o | \mathbf{b}_i)$ is the conditional cumulative hazard function under the PE survival model. Maximizing (8) may be carried out using a second-order Newton-Raphson algorithm with ridging combined with an integral approximation like adaptive Gaussian quadrature (AGQ) to approximate the integral. For our particular class of models, we elected to use the Laplace approximation to perform numerical integration. We did so for the following reasons:

- 1) it significantly reduces the computation time that one would otherwise encounter with more computer-intensive numerical integration techniques (e.g., Gaussian quadrature with 10 quadrature points);
- 2) since the mean structure for the class of mixed-effects models (2) used to model eGFR is strictly linear in the random effects, it is anticipated that no more than a single quadrature point (i.e., the Laplace approximation) would be needed to provide an accurate approximation to the integrated log-likelihood function - a result which we confirmed by applying both the Laplace approximation and AGQ to the same SP model with exactly the same result;
- 3) under suitable regularity conditions, the use of Laplace's approximation yields estimates that are consistent upon order $O_p\{\max[n^{-1/2}, \min(p_i)^{-1}]\}$; ¹⁸
- 4) limited simulations suggest the Laplace approximation compares quite favorably with AGQ even for discrete logistic mixed-effects models where one faces greater challenges given the highly discrete nature of the data. ⁸

All estimation was carried out using the NLMIXED procedure in SAS 9.4. While there are a number of popular and flexible software packages for fitting SP models in both R and SAS ¹⁹⁻²¹, these do not, to the best of our knowledge, readily allow for the within-subject and between-subject heteroscedasticity assumptions as specified for the linear spline mixed-effects model portion of our SP models. Inference under the SP model was carried out using large sample theory consistent with maximum likelihood methodology. To assess whether inference might be affected by possible model misspecification with respect to the assumed variance-covariance structure for eGFR, inference was carried out using both model-based and robust standard errors the latter of which is computed using the empirical covariance matrix of parameter estimates. A default degrees of freedom of $n - v$ was used for computing p values and confidence limits where n is the number of subjects and v is the number of random effects.

4.1 | Starting values and other computational issues

Since nonlinear mixed-effects regression is required to fit the SP models, it is crucial that one have good starting values of the model parameters. For the linear spline mixed-effects model (2), one can start by assuming within- and between-subject homogeneity and use parameter estimates from say PROC MIXED as starting values for β_c , β_t , σ^2 and Ψ . One might then simply amend to these starting estimates a simple grid of possible starting values around 0 for θ and κ . Reasonably good starting values of β_c , β_t , σ^2 , Ψ and θ can also be obtained by outputting the predicted SS means $\hat{\mu}_{ij}$ from PROC MIXED, forming the weights $w_{ij} = \ln(\hat{\mu}_{ij}^2)$ and running PROC MIXED a second time using the LOCAL Power-of-X option of the REPEATED statement. Doing so will yield method-of-moment estimates of σ^2 and θ based on the variance function $\sigma^2 \exp(\theta w_{ij}) = \sigma^2 (\hat{\mu}_{ij}^2)^\theta$.

When determining starting values for the parameters of the PE survival model, one must first determine what the appropriate number of intervals, m , should be and what the width of each interval should be. Such a determination is data-dependent in that choosing the number of intervals and width of each interval will depend on the number of events and maximum length of follow-up for a given trial. We adopted the following guidelines to determine the number of partitioned PE intervals to be used for any given clinical trial. Following the work of Vittinghoff and McCulloch²², a general but conservative rule of thumb is that 5-10 events per predictor variable (EPV) is required to minimize bias and provide adequate confidence interval coverage when fitting a Cox PH model. When we apply this rule to the estimation of the baseline piecewise constant hazard rates $\lambda_{0h} = \exp(\eta_{0h})$ across $h = 1, \dots, m$ intervals, it would suggest that we use no more than m_0 intervals where m_0 is the smallest integer greater than $n_e/10$ with $n_e = \sum_{i=1}^n \delta_i$ being the total number of events. By using this as a starting point along with restricting the maximum width of any interval and further restricting the number of intervals to be no more than $m_1 = \min(m_0, 9)$, the following 4 steps were used to determine the number intervals and the width of each interval for a given CKD trial.

1. Restrict the maximum length of any interval to $W_{\max} = 6$ months,
2. Compute the maximum event time T_{\max} (censored or uncensored) and let m_2 is the smallest integer greater than T_{\max}/W_{\max} ,
3. Compute the number of intervals $m = \min(m_1, m_2)$
4. Compute the interval widths such that each interval will have approximately the same number of events.

The last step can be accomplished in SAS, for example, by fitting a Cox PH model to the event time data with treatment group as the sole covariate and using the BAYES statement of the PHREG procedure. Once the intervals have been determined, one can then fit the PE survival model using interval Poisson regression to get starting values for the baseline constant hazard rates λ_{0h} and a log hazard ratio for the treatment effect.¹⁵ A grid search of plausible starting values centered around 0 may then be used to get starting values for any of the SP covariates listed in Table 2.

In addition to having good starting values, one must also be concerned with an appropriate scaling of the model parameters. When the scaling of parameters varies by several orders of magnitude, one risks the problem of introducing numerical instability into the numerical optimization of the integrated log-likelihood function (8). This in turn can result in floating-point errors, overflows and eventual termination of the procedure. We have found that the initial starting value for σ^2 obtained either by a grid search or by using the method-of-moment estimates of σ^2 and θ as described above will result in values of σ^2 that are 3-4 orders of magnitude lower than the variance ψ_{11} of the random intercept effect. As this led to computational issues in many cases, we elected to fit the model in NLMIXED by rescaling $\sigma^2 = \sigma^2 \times 10^2$ and defining the intra-subject POM variance structure within NLMIXED to be $\text{Var}(\epsilon_{ij}) = \frac{\sigma^2}{100} (\mu_{ij}^2(\beta_i))^\theta$. The resulting estimate of σ^2 obtained from NLMIXED and summarized in this paper will therefore be 2 orders of magnitude higher than what it would be if scaled according to the eGFR unit of measurement.

Finally, for studies with very few events, say 15-30, the Weibull survival model may be preferred to the PE survival model as we have found in our experience with having fit these SP models to 56 studies, the latter will often fail to converge. For studies with fewer than 15 events, one should avoid using SP models due to possible numerical instability resulting in problems with convergence and/or problems with bias and lack of precision in the parameter estimates.

5 | EXAMPLES

To illustrate the use of the linear spline mixed-effects model (2) together with sensitivity analyses based on SP models in CKD trials, we fit the four SP models listed in Table 2 to data from the Irbesartan Diabetic Nephropathy Trial (IDNT) as provided to the workshop analytical team.^{2-5,12-13} We also fit SP models 1 and 2 to the MDRD-B study which had been analyzed previously

using SP models but with eGFR measured based on the renal clearance of [125 I]iothalamate and modeled assuming a simple linear mixed-effects model for eGFR.⁷⁻⁸

5.1 | The IDNT Study

The IDNT study was an international prospective randomized double-blind placebo-controlled trial in hypertensive patients with diabetic nephropathy due to type II diabetes mellitus. Patients were randomized into one of three treatment arms: a) those randomized to receive the angiotensin II receptor blocker (ARB) irbesartan; b) those randomized to receive the calcium channel blocker (CCB) amlodipine and c) those randomized to receive placebo control, with each group receiving additional antihypertensive agents as needed excluding the use of ACE inhibitors, ARB's or CCB's. Details of the study protocol including study design and baseline patient characteristics have been published previously.¹² For the purposes of illustration, we restricted our analyses to estimating and comparing eGFR slopes for the placebo control group and the irbesartan treatment group only.

The primary outcome in the IDNT study was time to a composite endpoint consisting of: i) doubling of baseline serum creatinine (equivalent to a 57% reduction in eGFR); ii) ESKD (as indicated by initiation of dialysis, transplantation or a serum creatinine of 6.0 mg/dL) and iii) death (all-cause mortality). A number of secondary endpoints were also evaluated including serum creatinine levels from which measures of eGFR were obtained at baseline (month 0), month 3, month 6, and approximately every 6 months thereafter. Shown in Table 3 are baseline eGFR values along with a summary of the number and percentage of patients who dropped out due to death or ESKD as a composite event. It should be noted that since eGFR is a direct function of serum creatinine, we confined our definition of ESKD to dialysis or transplantation as the inclusion of an observed serum creatinine of 6.0 mg/dL would not constitute an informative censoring event.

The first step in our analysis was to determine an appropriate knot to use when fitting the linear spline mixed-effects model (2). We did so by initially fitting a simple linear-spline mixed-effects model using PROC MIXED in SAS with a knot fixed at values ranging from 3 to 12 months. We did so assuming a constant within-subject variance σ^2 and a constant between-subject covariance matrix Ψ . We then output the SS predicted means $\hat{\mu}_{ij}$, computed weights $w_{ij} = \ln(\hat{\mu}_{ij}^2)$, and re-ran the mixed-effects model using the LOCAL Power-of-X option of the REPEATED statement corresponding to the variance function $\sigma^2 \exp(\theta w_{ij}) = \sigma^2 (\hat{\mu}_{ij}^2)^\theta$ so as to evaluate the fit of each knot using a one-time iterative estimate of the POM variance structure. The results based on rank ordered values of Akaike's information criterion, AIC, indicate a knot at 4 months provides the best fit (Table 4). Fixing the knot at 4 months, we then proceeded to fit the four SP models listed in Table 2 using starting values based on the methods described in Section 4.1. Shown in Table 5 are the model parameter estimates and their standard errors for the linear spline mixed-effects model (2) as well as the log hazard ratios for the PE survival model covariates (Table 2) under each of the four SP models. Included are -2 log likelihood values and AIC values for each of the SP models as well as formal likelihood ratio (LR) test p-values for comparing the full SP model 4 against each of the reduced SP models 1-3. Both SP models 1 and 3 show a significant lack of fit compared with SP model 4 while SP model 2 shows a comparable fit. Moreover, SP model 1 shows a significant lack of fit when compared to each one of the SP models 2-4 (p-values < 0.0001) providing reasonably strong evidence that dropout is non-ignorable. In terms of impact on the model parameters, we see from Table 5 that the acute and chronic slopes are all steeper under SP models 2-4 when compared against SP model 1. For example, the chronic slopes under SP model 2 are 9.5% and 17.3% lower for the control and treated groups, respectively, compared with those under SP model 1. We also observe that there is little evidence of a treatment-dependent effect on the random slope effects as estimates of κ are not significantly different for 0.

The adjusted downward slopes under SP models 2 and 4 result in an attenuation of the treatment effects (treated-control) on the change slopes, chronic slopes and total slopes at 3 and 4 years as shown in Table 6. While the treatment differences in the chronic slope remain significant under all four SP models, the total slope differences at 3 and 4 years lose their significance under SP models 2 and 4 reflecting the impact the acute effect has in combination with dropout due to death or ESKD. While the total slope differences are trending in the right direction (i.e., towards significance with increasing follow-up time), their attenuation under SP models 2 and 4 closely mirrors the attenuation seen in the chronic slope differences. In turn, the attenuation of a treatment effect on the chronic slopes may be explained by the significant increase in the rate of dropout due to death or ESKD associated with higher rates of decline in eGFR as measured by subjects with greater deviations (i.e., b_{3i}) below the population-averaged chronic slopes. In particular, we find under SP model 2 that for every 1.0 mL/min/1.73m²/year increase in b_{3i} there is an associated 32.5% reduction ($= 100\% \times (1 - \exp(-4.725/12))$) in the rate of dropout due to death or ESKD or, conversely, for every 1.0 mL/min/1.73m²/year decrease in b_{3i} there is an associated 48.3% increase in the rate of dropout due to death or ESKD. This association is clearly in evidence when one plots and compares the empirical Bayes estimates of b_{3i} versus time to dropout

T_i under SP models 1 and 2 as shown in Figure 1. Under SP model 1 there is no discernable correlation between \hat{b}_{3i} and T_i among patients who dropped out due to death or ESKD ($R^2 = 0.007$). This is consistent with its underlying assumption that dropout is independent of any trends in eGFR. In contrast, under SP model 2 wherein this assumption is relaxed, 46% of the variation in \hat{b}_{3i} among patients who dropped out due to death or ESKD may be explained by its linear relationship to T_i ($R^2 = 0.462$).

Not surprisingly, then, when the increased risk of dropout associated with higher subject-specific rates of decline in eGFR is ignored, ML estimates of the population-average rates of decline in eGFR will be biased upward leading to what appears to be a more favorable response profile as shown in Figure 2. Here we find the predicted mean eGFR profile under SP model 2 is attenuated slightly downward for both treatment arms as a result of providing greater weight to those individuals who dropout due to death or ESKD. In summary, SP model 2 provides greater weight to those who dropout due to death or ESKD than would otherwise be the case under the ignorable dropout assumptions imposed under SP model 1.

In terms of other modeling assumptions, we examined how robust our results are to model misspecification with respect to the covariance structure using a goodness-of-fit variance-covariance concordance correlation coefficient and a maximum likelihood (ML) discrepancy function similar to those implemented in the SAS procedure GLIMMIX.^{8,23-25} Letting $\hat{\Omega}$ and $\hat{\Omega}_R$ denote, respectively, the model-based and robust (empirical) variance-covariance matrices of the $s \times 1$ vector of estimated regression parameters $\hat{\beta}$ ($s = 6$) under model (2), the variance-covariance concordance correlation is given by

$$r(\hat{\omega}) = 1 - \frac{\|(\hat{\omega} - \mathbf{i})\|^2}{\|\hat{\omega}\|^2 + \|\mathbf{i}\|^2} \quad (9)$$

where $\hat{\omega} = \text{Vech}(\hat{\Omega}^{-1/2} \hat{\Omega}_R \hat{\Omega}^{-1/2})$ and $\mathbf{i} = \text{Vech}(\mathbf{I}_p)$. It has the property that $r(\hat{\omega}) = 1$ if and only if $\hat{\Omega} = \hat{\Omega}_R$ while $r(\hat{\omega}) = 0$ if and only if the components of $\hat{\omega}$ are orthogonal to those of \mathbf{i} . The ML discrepancy function,

$$\hat{d} = \{\ln |\hat{\Omega}| - \ln |\hat{\Omega}_R| + \text{trace}(\hat{\Omega}_R \hat{\Omega}^{-1}) - s\}, \quad (10)$$

is a measure of closeness between $\hat{\Omega}$ and $\hat{\Omega}_R$ with the property that \hat{d} will converge in probability to 0 given the variance-covariance structure under the linear spline mixed-effects model (2) is correctly specified. Values of these two statistics were $r(\hat{\omega}) = 0.994$ and $\hat{d} = 0.037$ under SP model 1 and $r(\hat{\omega}) = 0.992$ and $\hat{d} = 0.049$ under SP model 2 which, when combined with both model-based and robust standard error estimates and p-values for the treatment slopes and their differences shown in Table 7, suggest the assumed covariance structure under (2) is not unreasonable. Finally the assumptions of normality with respect to the distribution of eGFR and conditional independence of $Y_{ij}|\mathbf{b}_i$ and $T_i|\mathbf{b}_i$ also appear to be reasonable based on plots of the scaled conditional residuals shown in Figures 3-4.

In terms of the mean structure under model (2), marginal and conditional R^2 values as well a marginal and conditional concordance correlation coefficients, r_c , have been proposed as a means for assessing the goodness-of-fit of mixed-effects models.²⁴⁻²⁶ When the fixed-effects are known and specified in advance such as occurs in randomized controlled trials (RCT's), the conditional R^2 and r_c values are preferred measures as they take into account the adequacy of fit of both the random and fixed-effects.²⁴⁻²⁶ As with the case for marginal linear models, conditional R^2 and r_c values close to one provide users with an intuitive measure of how well the overall mixed-effects model fits the observed data. Conversely, the marginal R^2 and r_c values, while preferred in settings where the goal is to perform fixed-effects variable selection, should be avoided in a RCT setting as they tend to be undervalued due to the exclusion of explained variation associated with SS random effects.²⁵⁻²⁶ The conditional R^2 and r_c values for the IDNT study were $R^2 = 0.9198$ and $r_c = 0.9566$ for SP model 2 suggesting the linear spline mixed-effects model provides a reasonably good fit to the data at both the population level and subject-specific level as shown in Figures 5-6 for the control and treated groups, respectively.

5.2 | The MDRD-B Study

The MDRD-B study was a multi-centre study of diet and blood pressure control on progression of CKD among late stage CKD patients. A description of the study design and original results have been presented previously²⁷⁻²⁸ as have subsequent analyses utilizing both SP models and pattern mixture models.^{7-8,29} Both the original and subsequent analyses were carried out under the assumption of a simple linear mixed-effects model for eGFR with a single slope parameter per treatment group and homogeneous within- and between-subject variability.^{7-8,28-29} Here we re-analyze the data assuming the more general linear spline mixed-effects model (2) but restricted to comparing the two blood pressure treatment groups; the normal (control) blood pressure group and the low (treated) blood pressure group. A preliminary analysis similar to that performed for the IDNT study suggests that

an acute effect occurs over the first 6 months after which the rate of decline in eGFR stabilizes. Based on this, we fixed a knot at 6 months and fit SP models 1 and 2 to the data using the same algorithms and code used to analyze the IDNT study.

The results, as summarized in Table 8, show that there is a significant lack of fit when comparing SP model 1 to SP model 2 (likelihood ratio test chi-square=183.99, DF=3, p-value< 0.0001). This may be explained by the significantly lower risk of dropout due to death or ESKD associated with patients having latent random intercept (b_{0i}) and slope (b_{1i} , b_{3i}) effects that fall above 0 (i.e., patients having higher starting eGFR values and lower rates of decline in eGFR over time). These associations, in turn, account for the notably different estimates we obtain for the chronic slopes under the non-ignorable dropout mechanism of SP model 2 compared with the ignorable dropout mechanism of SP model 1. Specifically, the chronic slopes are 17% steeper under SP model 2 versus SP model 1 for both the control group (-0.650 mL/min/1.73m²/year lower under SP model 2) and treated group (-0.653 mL/min/1.73m²/year lower under SP model 2). Unlike with the IDNT study where the percent change in the estimated chronic slopes between SP models 1 and 2 differ according to treatment group leading to an attenuation of the treatment effects on the slopes, here the percent change is uniform across the two treatment groups such that the treatment effects on the chronic slope remain unchanged. Also unlike the IDNT study, here we find a significant treatment-dependent effect on the slope random effects. Specifically, the estimated proportionality parameter $\hat{\kappa}$ is significantly less than 0 resulting in lower standard errors for the treated group slopes compared with that of the control group slopes.

6 | SAMPLE SIZE AND POWER

Following a suggestion from a reviewer, a SAS macro program, %GFR_Slope_Power, was developed specifically for the purpose of determining sample size/power estimates for comparing slopes between two treatment groups based on the linear spline mixed-effects model (2). The program uses calculations based on an approximation to the power of an F -test for linear mixed-effects models described by Littell et al. (Chapter 12) and references therein.³⁰ The approximation, which can also be used for marginal linear models, entails the use of an “observed” F -statistic together with its numerator degrees of freedom to compute the noncentrality parameter needed for sample size/power calculations. To adapt their approach to the CKD clinical trial setting, we start by assuming a balanced study design with N subjects per treatment group ($n = 2N$ subjects in total) with eGFR measurements obtained at p specified measurement occasions or times t_1, t_2, \dots, t_p . Under these assumptions, it will be convenient to write model (2) as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$ where \mathbf{Y} is the $np \times 1$ vector of eGFR measurements, \mathbf{X} is the $np \times 6$ fixed-effects design matrix, $\boldsymbol{\beta}$ is the 6×1 vector of regression parameters, \mathbf{Z} is the $np \times nv$ random-effects design matrix ($v = 3$), \mathbf{b} is the $nv \times 1$ random-effects vector and $\boldsymbol{\epsilon}$ is the $np \times 1$ vector of within-subject residuals.

Using this framework and following Littell et al.³⁰, one can approximate power for testing a null hypothesis $H_0: \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ versus the alternative $H_1: \mathbf{L}'\boldsymbol{\beta} \neq \mathbf{0}$, by first creating a single “dummy” dataset with the hypothesized mean response profile, $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, taking the place of actual values of \mathbf{Y} . One would then “fit” the model to the “dummy” data using PROC MIXED and test the null hypothesis with CONTRAST or ESTIMATE statements. Based on the output, one would then estimate power as

$$\text{Power} = \Pr(F_{\{\text{NDF, DDF}, \delta\}} > F_{\{\text{NDF, DDF}, 0, \alpha\}}) \quad (11)$$

where $F_{\{\text{NDF, DDF}, \delta\}}$ is the distribution of the F -statistic, NDF=rank(\mathbf{L}) is the numerator degrees of freedom, DDF is the denominator degrees of freedom, $\delta = (\mathbf{L}'\boldsymbol{\beta})' \{ \mathbf{L}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{L} \} (\mathbf{L}'\boldsymbol{\beta})$ is the noncentrality parameter computed from the “observed” F -test statistic, $\boldsymbol{\Sigma}$ is the marginal covariance matrix and $F_{\{\text{NDF, DDF}, 0, \alpha\}}$ is the critical F value for testing the null hypothesis $H_0: \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ at a type I error of α . The marginal covariance matrix is given by $\text{Cov}(\mathbf{Y}) = \boldsymbol{\Sigma} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$ where \mathbf{G} is an $nv \times nv$ block diagonal matrix with $v \times v$ matrices $\boldsymbol{\Psi}_k$ along the diagonal, and \mathbf{R} is an $np \times np$ block diagonal matrix with $p \times p$ marginal within-subject matrices, \mathbf{R}_i , along the diagonal. The values of $\boldsymbol{\Psi}_k$ and \mathbf{R}_i will depend on what assumptions one is willing to make (see Table 1). Under the assumptions of homogeneity, $\boldsymbol{\Psi}_k = \boldsymbol{\Psi}$ and $\mathbf{R}_i = \sigma^2 \mathbf{I}_p$, while under the assumptions of heterogeneity, $\boldsymbol{\Psi}_k$ will depend on the treatment group and κ , and \mathbf{R}_i will be a $p \times p$ diagonal matrix with expected values of the SS-POM variances $\sigma^2(\mu_{ij}^2(\boldsymbol{\beta}_i))^\theta$ along the diagonal, i.e., $\sigma^2 E_b \{ (\mu_{ij}^2(\boldsymbol{\beta}_i))^\theta \}$. The program is setup to always run under the assumptions of heterogeneity for both the within-subject and between-subject variances and covariances since the cases of homogeneity can be achieved by simply setting one or both of the parameters, θ and κ , to 0.

To estimate power using the SAS macro, the user provides input values for N and the p measurement occasions along with hypothesized values of the parameters ($\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, κ , σ^2 , θ) from model (2) as well as other input variables including study accrual and follow-up periods, an annual dropout rate (assumed to be non-informative) and an option to include informative censoring when a subject's predicted mean eGFR falls below 15 mL/min.1.73m². The program provides the user with two options for approximating the marginal within-subject covariance matrix \mathbf{R} . Both options are predicated on the need to use a WEIGHT

statement with PROC MIXED in order to specify an approximation to \mathbf{R} . In particular, one can write the i^{th} subject's $p \times p$ within-subject diagonal covariance matrix as $\mathbf{R}_i(\mathbf{b}_i) = \sigma^2 \mathbf{W}_i(\mathbf{b}_i)^{-1}$ where $\mathbf{W}_i(\mathbf{b}_i)$ is a diagonal matrix of SS “weights” $w_{ij}(\mathbf{b}_i) = 1/(\mu_{ij}^2(\beta_i))^\theta$. By replacing $\mathbf{W}_i(\mathbf{b}_i)$ with a diagonal matrix \mathbf{W}_i of weights w_{ij} that do not depend on \mathbf{b}_i and which provide a reasonable approximation to $E_b\{(\mu_{ij}^2(\beta_i))^\theta\}$ one can approximate power based on $\mathbf{R}_i = \sigma^2 \mathbf{W}_i^{-1}$. One option would be to use a first-order Taylor series approximation to $\mathbf{R}_i(\mathbf{b}_i)$ by expanding the SS-POM function $(\mu_{ij}^2(\beta_i))^\theta$ about $\mathbf{b}_i = \mathbf{0}$ and replacing the SS “weights” with the PA “weights” $w_{ij} = 1/(\mu_{ij}^2(\beta))^\theta$ where $\mu_{ij}(\beta)$ is the marginal or population-averaged mean response under model (2). In this case, power based on $\mathbf{R}_i = \sigma^2 \mathbf{W}_i^{-1} = \sigma^2 \text{Diag}\{(\mu_{ij}^2(\beta))^\theta\}$ may be interpreted as the power required to detect differences in the β 's for an average or typical subject (i.e., a subject with $\mathbf{b}_i = E_b(\mathbf{b}_i) = \mathbf{0}$). As this approximation ignores subject-to-subject variation in the POM variances, the approximate power will tend to overestimate the true power with the degree of overestimation depending on the magnitude of Ψ_k (with bias $\rightarrow 0$ as $\Psi_k \rightarrow \mathbf{0}$).

A second more conservative option would be to estimate the marginal covariance matrix \mathbf{R}_i empirically based on a randomly generated sample of M random effects, \mathbf{b}_{mk} , within each treatment group ($m = 1, \dots, M; k = c, t$ for the control and treated groups) where the \mathbf{b}_{mk} are generated from a multivariate normal distribution with mean $\mathbf{0}$ and covariance Ψ_k . For a sufficiently large M and writing $\beta_i = [\beta_c + \mathbf{b}_i](1 - X_i) + [\beta_t + \mathbf{b}_i]X_i$, one can estimate the marginal weights, $w_{ij} = 1/E_b\{(\mu_{ij}^2(\beta_i))^\theta\}$, empirically as

$$w_{ij} = 1/\left\{ \frac{1}{M} \sum_{m=1}^M \left(\mu_{ij}^2([\beta_c + \mathbf{b}_{mc}](1 - X_i) + [\beta_t + \mathbf{b}_{mt}]X_i) \right)^\theta \right\}. \quad (12)$$

In either case, power calculations would be based on $\mathbf{R}_i = \sigma^2 \mathbf{W}_i^{-1}$ depending on which approximation to w_{ij} one chooses. We will refer to the PA “weights” $w_{ij} = 1/(\mu_{ij}^2(\beta))^\theta$ as population-averaged power-of-mean (PA-POM) weights and the empirically estimated weights given by (12) as the estimated marginal power-of-mean (EM-POM) weights.

A full description of the SAS macro and examples of its use are provided in Web Appendix A of the supplementary material. To illustrate, we used estimates of $\beta_{1c}, \beta_{1t}, \beta_{2c}, \Psi, \kappa, \theta$, and the rescaled variance $\sigma^2 = \sigma^2/100$, from the IDNT study (Table 5) as input to calculate power associated with incremental differences, $\beta_{3t} - \beta_{3c}$, between the chronic slope parameters of the treatment group versus control group holding the treated and control group acute slope parameters and their difference fixed as well as the delta (change) slope parameter and chronic slope parameter for the control group. The chronic slope differences, $\beta_{3t} - \beta_{3c}$, were varied from 0.75 to 1.50 mL/min/1.73m²/year in increments of 0.25 mL/min/1.73m²/year. Given these specifications, incremental differences in the delta (change) and total slopes between the treated and control groups are automatically determined. Moreover, in terms of clinical relevance, these slopes and their incremental differences can also be used to predict the average extended time to ESKD for patients receiving treatment. For example, assuming a mean eGFR of 15 mL/min/1.73m² is an indication of ESKD and that it occurs after the acute phase (something one can easily determine based on the input parameters), then the mean time to ESKD based on the population-average parameters from (2) would be

$$E(T_c) = \frac{(\beta_{0c} - \beta_{2c}t^*) - 15}{-\beta_{3c}}, \text{ for the control group.} \quad (13)$$

$$E(T_t) = \frac{(\beta_{0t} - \beta_{2t}t^*) - 15}{-\beta_{3t}}, \text{ for the treated group}$$

The extended time to ESKD due to treatment for the average or typical patient would then simply be $E(T_t) - E(T_c)$.

Power estimates based on incremental differences in the chronic slopes were computed for sample sizes of 300, 400, 500 and 600 subjects per group assuming a uniform accrual rate over a 36 month period with a minimum 24 month follow-up. Using the estimated parameters from SP model 2 of Table 5 as input along with EM-POM weights w_{ij} for the overall marginal covariance matrix \mathbf{R} and allowing for a 5% annual dropout rate together with informative censoring of subjects once their predicted mean eGFR falls below 15 mL/min/1.73m², it is estimated that a sample of 400 subjects per group would provide 76% power to detect a 1.00 mL/min/1.73m²/year difference in the population chronic slopes in favor of the treatment group (see Table A.7 of Web Appendix A). The power increases to 91% to detect a 1.25 mL/min/1.73m²/year difference between treated vs. control chronic slopes (Table A.7). Based on these differences the extended time to ESKD due to treatment for the average or typical patient would be 1.075 and 1.466 years, respectively, provided one is willing to extrapolate the assumed linear trajectories during the period of study into the future.

Alternatively, if one were seeking to power the study on the basis of the total slope given a chronic slope treatment effect of 1.25 mL/min/1.73m²/year, a sample of 500 patients per group would be required to detect a total slope treatment difference at 4 years of 0.957 mL/min/1.73m²/year with 83% power (see Table A.7). To demonstrate what impact the acute treatment effect has on power calculations for the total slope, we repeated the power calculations but with the treatment group's acute slope

reduced from -14.724 mL/min/1.73m²/year to -13.596 mL/min/1.73m²/year which corresponds to a 50% reduction in the acute treatment effect at 4 months (i.e., the mean difference in eGFR at 4 months between the treated and control patients goes from -0.756 mL/min/1.73m² to -0.38 mL/min/1.73m²). Under this alternative lower acute treatment effect, one would only require a sample of 400 patients per group to detect a total slope treatment effect through 4 years of 1.051 mL/min/1.73m²/year with 81% power or a total slope treatment effect through 3 years of 0.984 mL/min/1.73m²/year with 72% power (see Table A.8 of Web Appendix A). The lower sample size required to detect differences in total slopes based on a lower acute treatment effect reflects the need for conducting a phase 2 trial designed to estimate the magnitude and timing of a possible acute treatment effect prior to planning and conducting a phase 3 trial.

To illustrate what impact possible misspecification of the knot might have on study power for a given set of hypothesized parameters, we computed the power to detect a 1.00 to 1.25 mL/min/1.73m²/year difference in the chronic slopes based on a fixed sample of 400 subjects per treatment group with values of the knot ranging from 3 to 6 months. We used as input the same estimates of β_{1c} , β_{1t} , β_{2c} , Ψ , κ , θ , and $\sigma^2/100$ from the IDNT study used above to generate the results in Table A.7 of Web Appendix A but with a different starting seed for generating random effects. Also included for comparison are power estimates based on both the PA-POM and EM-POM approximations to the marginal weights w_{ij} used to specify the covariance matrix \mathbf{R} . As shown in Table 9, the power to detect differences in the chronic slope decreases slightly as one increases the assumed value of the knot from 3 to 6 months. This is not unexpected given the likely decreased number of observations one would have available for estimating and comparing chronic slopes during a shortened chronic phase. Also not unexpected are the more conservative estimates of power achieved by using the estimated marginal power-of-mean (EM-POM) weights to approximate the marginal variance matrix, \mathbf{R} , as these weights directly account for variation in the SS-POM variances. Finally, based on this limited exercise, it would appear that misspecifying the knot by as much as 3 months will have little impact on the estimated power with which to compare hypothesized values of the chronic slopes.

7 | DISCUSSION

In this paper we have described a class of mixed-effects models that may be used to evaluate treatment efficacy in clinical trials of CKD on the basis of slope-based surrogate endpoints as opposed to more traditional clinical event time endpoints. Treating the rate of decline in eGFR as an intermediary surrogate endpoint, a two-slope linear spline mixed-effects model for eGFR is proposed as a means for dealing with short-term acute treatment effects on the rate of decline as have been reported in a number of CKD trials.²⁻⁵ Although the model is parameterized in terms of two slope parameters, there are actually four slope parameters in total which describe different aspects of the rate of decline in eGFR over time: an acute slope, a change slope, a chronic slope and a time-dependent total slope. The model includes a subject-specific power-of-mean (SS-POM) variance structure to model within-subject heteroscedasticity and a proportionality parameter for the random slope effects to model between-subject heteroscedasticity. Finally, to address concerns over informative censoring resulting from dropout due to death or ESKD, we extend the model to a class of SP models by introducing a piecewise exponential survival model that, through the sharing of latent random effects, allows one to model eGFR adjusted for various non-ignorable dropout mechanisms.

While the class of models described here offer a relatively robust modeling approach to assessing treatment effects on the rate of decline in eGFR, there are a number of challenging factors that can impact one's ability to successfully fit these models. The first challenge stems from the fact the eGFR model alone is nonlinear in both the fixed and random effects when one assumes an intra-subject POM variance structure. Hence ML estimation will require the use of a nonlinear optimization routine combined with an integral approximation in order to maximize the integrated log-likelihood function. For some trials, especially trials with smaller patient sizes or trials with fewer and more widely dispersed measurements over time per subject, we found that convergence could not always be achieved with this model. This was due almost exclusively to modeling intra-subject variability via the SS-POM variance structure. To overcome this, we have found that one can replace the SS-POM structure with a power-of-time structure wherein within-subject heteroscedasticity is modeled assuming $\text{Var}(\epsilon_{ij}) = \sigma^2 \exp(\theta t)$. Under this power-of-time structure, the first- and second-order moments of eGFR are completely orthogonal to one another resulting in a strictly linear mixed-effects model having a within-subject variance structure that very often mimics the POM structure.³⁰

In cases where convergence is still an issue, one might consider two variants of the two-slope linear spline mixed-effects model (2). One variant would involve simplifying model (2) by setting $b_{12} = 0$ resulting in a single random slope effect. The variance of this single random slope effect would account for subject-to-subject variation in both the acute and chronic slopes - an assumption which may or may not be reasonable. In the second variant, one would assume an acute effect occurs within say

the first 3 months of follow-up. In this case, one can fit a modified one-slope linear mixed-effects model by simply excluding eGFR measurements occurring before 3 months. To improve power, one would include centered baseline eGFR values and their interaction with time as additional covariates in the model. Under this variant, one can estimate and compare chronic slopes between the two treatment groups based on a simple one-slope regression model. The model can also be used to estimate an acute treatment effect at 3 months by simply comparing the 3-month predicted mean eGFR values between the two treatment groups.

Another challenge with slope-based endpoints in CKD trials is determining whether the primary endpoint should be the chronic slope or the total slope based on a planned length of follow-up. Such a determination can be problematic especially in the presence of a relatively large acute effect as this could lead to different conclusions depending on the length of follow-up. Ideally, this is an issue that would be clarified in a phase 2 trial so that one can be confident that a phase 3 trial will be adequately designed (e.g., timing of eGFR measurements) and powered for either endpoint. One way to accomplish this would be to perform a 1 or 2 year phase 2 trial with monthly eGFR measurements on a smaller but reasonable number of patients. Such a trial would provide fairly precise within-subject trends for determining the magnitude and timing of any acute effects as well as providing estimates of the model parameters which can then be used to estimate sample size and power for a phase 3 trial.

The possibility of informative censoring raises a somewhat paradoxical challenge related to the use of SP models - namely that the effectiveness of a SP model is somewhat dependent on the number of events that occur within a clinical trial. If a study results in a relatively low number of events, fitting a SP model becomes problematic both at the inferential level and at the computational level. Too few events means there is very low power and precision with which to estimate hazard ratios associated with SP covariates. In that case fitting a SP model is unlikely to have any real impact on the estimation and comparison of eGFR related treatment effects such as differences in estimated total slopes. Moreover, too few events can lead to computational problems such as non-convergence or convergence to unstable estimates resulting from a non-positive definite Hessian matrix or a Hessian matrix with one or more negative eigenvalues (thus violating second-order optimality conditions). The paradox, of course, is that trials having a relatively high number of events leading to dropout are more susceptible to informative censoring which, in turn, requires greater scrutiny in the form of sensitivity analyses (such as the use of SP models) when assessing treatment effects.

Finally, sensitivity analyses other than those based on SP models might also be considered when assessing treatment effects in the presence of informative censoring. These include the use of pattern-mixture models and conditional linear mixed models.^{6,8,29,31-33} One approach would be to adopt a tipping point sensitivity analysis based on some form of a pattern-mixture model such as occurs, for example, when fitting a repeated measures ANCOVA model to select outcomes.⁸ However, some care is needed when undertaking these alternative analyses. For example, one may attempt a sensitivity analysis by imputing missing values following dropout under some form of non-ignorable missing data mechanism. If, however, dropout is the result of some terminal event like death, one must question whether this makes sense since outcomes following a terminal event are no longer possible.⁸

SOFTWARE CODE AND DATA ACCESSIBILITY

Access to the primary SAS code used to fit the models for the MDRD-B study including instructions for accessing the MDRD-B dataset are available through the repository Figshare. This is the same code used to fit the models for the IDNT study using the same basic dataset structure. Also included in Figshare is the SAS macro code for %GFR_Slope_Power and examples illustrating its use.

ACKNOWLEDGEMENTS

CKD-EPI investigators/collaborators (study acronyms/abbreviations are listed in Web Appendix B in the Supplement with other abbreviations):

AASK: Tom Greene, PhD; **ABCD:** Robert W. Schrier, MD; Raymond O. Estacio, MD; **ADVANCE:** Vlado Perkovic, MBBS, PhD; **AIPRI:** Giuseppe Maschio, MD; Francesco Locatelli, MD; **ALTITUDE:** Hans-Henrik Parving, MD, DMSc; **Bari:** Francesco Paolo Schena, MD; Manno Carlo, MD; **Bologna:** Pietro Zucchelli, MD; Boston: Barry M. Brenner; **canPREVENT:** Brendan Barret, MB, MSc, FRCPC; **Copenhagen:** Anne-Lise Kamper, MD, DMSc; Svend Strandgaard; **CSG:** Roger A. Rodby, MD; Richard D. Rohde, BS; Julia B. Lewis, MD; Edmund Lewis, MD; **EMPA-REG OUTCOME:** Christoph Wanner, MD; Maximilian von Eynatten, MD; **Fukuoka:** Ritsuko Katafuchi, MD; **Groningen:** Paul E. de Jong, MD, PhD; GG van

Essen, MD; **Guangzhou**: Fan Fan Hou, MD, PhD; Di Xie; **HALT-PKD**: Ronald D. Perrone, MD; Kaleab Z. Abebe, PhD; Godela Brosnahan, MD; **HKVIN**: Philip Li; CB Leung; CC Szeto; KM Chow, **IDNT**: Edmund Lewis, MD; Lawrence G. Hunsicker, MD; Julia B. Lewis, MD; **Lecco**: Francesco Locatelli, MD; Lucia Del Vecchio, MD; Simeone Andrulli; Claudio Pozzi MD; Leuven: Bart Maes, MD, PhD; **LNCS**: Julia B. Lewis, MD; Jamie Dwyer; Edmund Lewis, MD; John M. Lachin, ScD; **MADRID**: Marian Goicoechea, MD, PhD; Eduardo Verde; Ursula Verdalles; Jose Luño; **Madrid**: Manuel Praga, MD; Fernando Caravaca; Eduardo Gutierrez; Angel Sevillano; **MASTERPLAN**: Jack F.M. Wetzels, MD; Peter J Blankestijn; Arjan D. van Zuilen; Jan van den Brand; **MDRD Study**: Gerald Beck, PhD; Tom Greene; John Kusek; Garabed Eknayan; **Milan**: Claudio Ponticelli Montagnino, MD; Patrizia Passerini, MD; Gabriella Moroni, MD; Giuseppe Montogrino, MD; **New York**: Gerald B. Appel, MD; Gershon Frisch, MD; **ORIENT**: Fumiaki Kobayashi; Hirofumi Makino; Sadayoshi Ito; Enyu Imai, MD, PHD; **Hong Kong Lupus Nephritis**: Tak Mao Chan MD; **REIN 1**: Giuseppe Remuzzi, MD; Piero Ruggenenti, MD; **REIN 2**: Giuseppe Remuzzi, MD; Piero Ruggenenti, MD; **RENAAL**: Dick De Zeeuw, MD, PhD; Hiddo J Lambers Heerspink, PharmD, PhD; Barry M. Brenner, MD; William Keane, MD; **ROAD**: Fan Fan Hou, MD; **Rochester**: James Donadio, MD; Fernando C. Fervenza, MD, PhD; **SHARP**: Martin Landray, Will Herrington, Natalie Staplin, Colin Baigent; **STOP-IgAN**: Jürgen Floege, MD; Thomas Rauen, MD; Christina Fitzner; Ralf-Dieter Hilgers, PhD; **Strasbourg**: Imitiaz Jehan; Nish Chaturvedi; Neil Poulter; Thierry P. Hannedouche, MD; **SUN-MACRO**: Julia B. Lewis, MD; Jamie Dwyer; Edmund Lewis, MD; **Texas**: Robert D. Toto, MD; Victoria: Gavin .J. Becker; Benno U. Ihle, MBBS; Priscilla S. Kincaid-Smith MD, DSc

The planning and operations committee of the NKF-FDA-EMA workshop on Change in Albuminuria and GFR as End Points for Clinical Trials in Early Stages of Chronic Kidney Disease contributed to the design and critical review of these analyses. Planning and Operations Committees: Andrew Levey (Chair), Ron Gansevoort, Josef Coresh, Dick de Zeeuw, Kai-Uwe Eckardt, Hrefna Gudmundsdottir, Adeera Levin, Romaldas Maciulaitis, Tom Manley, Vlado Perkovic, Kimberly Smith, Norman Stockbridge, Aliza Thompson, Thorsten Vetter, Kerry Willis, and Luxia Zhang.

The study was funded by the National Kidney Foundation. A variety of sources have supported the RCTs included in the CKD-Epidemiology Collaboration. These funding sources include government agencies such as national institutes of health and medical research councils as well as foundations and industry sponsors. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The corresponding author had full access to all the data in the study and final responsibility for the decision to submit for publication.

ORCID

Edward Vonesh <https://orcid.org/0000-0002-7531-6160>

References

1. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983;39:499-503.
2. Levey AS, Inker LA, Matsushita K, et al. GFR decline as an end point for clinical trials in CKD: A scientific workshop sponsored by the National Kidney Foundation and the U.S. Food and Drug Administration. *Am J Kidney Dis*. 2014;64(6):821-835.
3. Inker LA, Lambers Heerspink HJ, Mondal H, et al. GFR decline as an alternative end point to kidney failure in clinical trials: a meta-analysis of treatment effects from 37 randomized trials. *Am J Kidney Dis*. 2014;64(6):848-859.
4. Lambers Heerspink HJ, Tighiouart, H, Sang Y, et al. GFR decline and subsequent risk of established kidney outcomes: a meta-analysis of 37 randomized controlled trials. *Am J Kidney Dis*. 2014;64(6):860-866, 2014.
5. Greene T, Tang C-C, Inker LA, et al. Utility and validity of estimated GFR-based surrogate time-to-event end points in CKD: A simulation study. *Am J Kidney Dis*. 2014;64(6):867-879.
6. Hogan JW, Roy J and Korkontzelou C. Tutorial in biostatistics: Handling drop-out in longitudinal studies. *Stat Med*. 2004;23:1455-1497.

7. Vonesh EF, Greene T and Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. *Stat Med*. 2006;25:143-163.
8. Vonesh EF. *Generalized Linear and Nonlinear Models for Correlated Data: Theory and Applications Using SAS*. Cary, NC: SAS Institute Inc.; 2012.
9. Lysaght MJ, Vonesh EF, Gotch F et al. The influence of dialysis treatment modality on the decline of remaining renal function. *ASAIO Trans*. 1991;37:598-604.
10. Moist LM, Port FK, Orzol SM et al. Predictors of loss of residual renal function among new dialysis patients. *J Am Soc Nephrol*. 2000;11:556-564.
11. Follmann, D. and Wu, M. An approximate generalized linear model with random effects for informative missing data. *Biometrics* 1995;51:151-168.
12. Rodby RA, Rohde RD, Clarke WR, et al. The Irbesartan Type II Diabetic Nephropathy Trial: study design and baseline patient characteristics. *Nephrol Dial Transplant*. 2000;15:487-497.
13. Lewis EJ, Hunsicker LG, Clarke WR, et al. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. *N Engl J Med*. 2001;345(12):851-860.
14. Thabut G, Crestani B, Porcher R, Richeldi L. Missing data in IPF trials: do not let methodological issues undermine a major therapeutic breakthrough. *Eur Respir J*. 2015;46:607-614.
15. Holford TR. The analysis of rates and of survivorship using log-linear models. *Biometrics* 1980;36:299-305.
16. Allison PD. *Survival Analysis Using the SAS System: A Practical Guide*. 2nd ed.. Cary NC: SAS Institute Inc., 2010.
17. Therneau, TM. *A Package for Survival Analysis in S*. Technical Report 53: Section of Biostatistics, Mayo Clinic, Rochester, MN; 1994.
18. Vonesh, EF. A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika* 1996;83:447-452.
19. Rizopoulos, D. JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. *J Stat Softw*. 2010;35(9):1-33.
20. Rizopoulos, D. *Joint Models for Longitudinal and Time-to-Event Data: with Applications in R*. Boca Raton: Chapman and Hall/CRC, 2012.
21. Garcia-Hernandez, A. and Rizopoulos, D. %JM A SAS macro to fit jointly generalized mixed models for longitudinal data and time-to-event responses. *J Stat Softw*. 2018;84(12):1-29, doi 10.18637/jss.v084.i12.
22. Vittinghoff E and McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol*. 2007;165:710-718.
23. Satorra A and Bentler PM. Some robustness properties of goodness of fit statistics in covariance structure analysis, 549-554. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*. Washington D.C.: American Statistical Association, 1986.
24. Vonesh EF, Chinchilli VM and Pu K. Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics* 1996;52:572-587.
25. Vonesh EF and Chinchilli VM. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York, NY: Marcel Dekker, Inc., 1997.
26. Orelan JG and Edwards LJ. Fixed-effects variable selection in linear mixed models using R^2 statistics. *Comput Stat Data Anal*. 2008;52:1896-1907.
27. Beck GJ, Berg RL, Coggins CH et al. Design and statistical issues of the Modification of Diet in Renal Disease Trial. *Control Clin Trials* 1991;12:566-586.

28. Klahr S, Levey AS, Beck GJ, et al. for the Modification of Diet in Renal Disease Study Group. The effects of dietary protein restriction and blood pressure control on the progression of chronic renal disease. *N Engl J Med*. 1994;330:877-884.
29. Li J and Schluchter MD. Conditional mixed models adjusting for non-ignorable dropout with administrative censoring in longitudinal studies. *Stat Med*. 2004;23:3489-3503.
30. Littell RC, Milliken GA, Stroup WW and Wolfinger RD. *SAS for Mixed Models*, 2nd ed. Cary, NC: SAS Institute Inc., 2006.
31. Little RJA. Modeling the dropout mechanism in repeated-measures studies. *J Am Stat Assoc*. 1995;90:1112-1121.
32. Schluchter MD, Greene T and Beck GL. Analysis of change in the presence of informative censoring: application to a longitudinal clinical trial of progressive renal failure. *Stat Med*. 2001;20:989-1007.
33. Wu MC and Bailey KR. Analyzing changes in the presence of informative censoring caused by death and withdrawal. *Stat Med*. 1988;7:337-346.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

Source of Variability	Description	Structure	Variances and Covariances
Within-Subject σ_{ij}^2	Homogeneity	Constant	$\sigma_{ij}^2 = \sigma^2$
	Heterogeneity	Power-of-mean (POM)	$\sigma_{ij}^2 = \sigma^2(\mu_{ij}^2(\beta_i))^\theta \dagger$
Between-Subject Ψ_k	Homogeneity	Uniform Treatment Effect	$\Psi_k = \Psi = \begin{pmatrix} \psi_{11} & \psi_{12} & \psi_{13} \\ \psi_{21} & \psi_{22} & \psi_{23} \\ \psi_{31} & \psi_{32} & \psi_{33} \end{pmatrix}$
	Heterogeneity	Proportional Treatment Effect on Random Slope Effects	$\Psi_k = \begin{pmatrix} \psi_{11} & \phi_k \psi_{12} & \phi_k \psi_{13} \\ \phi_k \psi_{21} & \phi_k^2 \psi_{22} & \phi_k^2 \psi_{23} \\ \phi_k \psi_{31} & \phi_k^2 \psi_{32} & \phi_k^2 \psi_{33} \end{pmatrix}^{\dagger\dagger}$

$\dagger \mu_{ij}(\beta_i) = E(Y_{ij}|\mathbf{b}_i) = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i} \max\{t_{ij} - t^*, 0\}$
 $\dagger\dagger \phi_k = \begin{cases} 1 & \text{for } k = c \text{ (control group)} \\ (1 + \kappa) & \text{for } k = t \text{ (treatment group)} \end{cases}$ is a proportional treatment effect on the random slope effects.

TABLE 1 Possible variance-covariance structures.

Dropout Mechanism	SP Model	SP Covariates	PE Hazard Function
Ignorable dropout	1	X_i	$\lambda_0(t) \exp\{\eta_0 X_i\}$
Non-ignorable dropout	2	$X_i, b_{0i}, b_{1i}, b_{3i}$	$\lambda_0(t) \exp\{\eta_0 X_i + \eta_1 b_{0i} + \eta_2 b_{1i} + \eta_3 b_{3i}\}$
	3	$X_i, \mu_{ih}(t_{h-1})^\dagger$	$\lambda_0(t) \exp\{\eta_0 X_i + \eta_4 \mu_{ih}(t_{h-1})\}$
	4	$X_i, b_{0i}, b_{1i}, b_{3i}, \mu_{ih}(t_{h-1})^\dagger$	$\lambda_0(t) \exp\{\eta_0 X_i + \eta_1 b_{0i} + \eta_2 b_{1i} + \eta_3 b_{3i} + \eta_4 \mu_{ih}(t_{h-1})\}$

$\dagger \mu_{ih}(t_{h-1}) = \beta_{0i} + \beta_{1i}t_{h-1} + \beta_{2i} \max\{t_{h-1} - t^*, 0\}$

TABLE 2 Piecewise exponential model components of four SP models.

Outcome	Control Group (n=563)	Treatment Group (n=572)
	Mean±SD ($n_e, \%$) [†]	Mean±SD ($n_e, \%$)
Baseline eGFR mL/min/1.73m ²	50.55±20.01	49.83±18.98
Censoring Events of Death or ESKD	134 (24%)	116 (20%)
Time to Event = $T_i^0 = \min(T_i, C_i)$	30.80±12.66	31.78±13.26

[†] n_e =number of events, %=percentage of subjects with an event

TABLE 3 Baseline eGFR and percentage of subjects with censoring events in the IDNT study. SD=standard deviation.



Fixed Knot (Month)	AIC	Rank
3	47903.6	7
4	47878.7	1
5	47879.8	2
6	47890.9	6
7	47882.5	4
8	47880.6	3
9	47889.3	5
10	47904.6	8
11	47923.1	9
12	47940.7	10

TABLE 4 Rank order selection of fixed knots based on preliminary analysis of the IDNT study.

		Dropout Mechanism and SP Model			
		Ignorable	Non-ignorable		
Model	Parameter	1	2	3	4
Component					
$\pi(\mathbf{y}_i \mathbf{b}_i)$	σ^2	3.501 (0.718)	3.355 (0.676)	3.198 (0.668)	3.214 (0.653)
	θ	0.917 (0.027)	0.922 (0.027)	0.928 (0.027)	0.927 (0.027)
	ψ_{11}	325.825 (15.849)	327.977 (15.932)	325.050 (15.878)	326.968 (15.891)
	ψ_{21}	2.260 (1.454)	1.831 (1.373)	2.056 (1.533)	1.936 (1.400)
	ψ_{22}	1.088 (0.239)	1.034 (0.215)	1.194 (0.268)	1.084 (0.223)
	ψ_{31}	-3.032 (1.514)	-2.265 (1.425)	-2.758 (1.595)	-2.391 (1.456)
	ψ_{32}	-0.935 (0.242)	-0.863 (0.214)	-1.039 (0.274)	-0.912 (0.223)
	ψ_{33}	0.876 (0.253)	0.794 (0.221)	0.984 (0.286)	0.845 (0.232)
	κ	-0.061 (0.070)	0.028 (0.074)	-0.070 (0.069)	-0.019 (0.077)
	β_{0c} (mL/min/1.73m ²)	50.609 (0.812)	50.642 (0.814)	50.609 (0.811)	50.629 (0.813)
	β_{0t} (mL/min/1.73m ²)	49.741 (0.802)	49.734 (0.805)	49.701 (0.801)	49.709 (0.803)
	β_{1c} (mL/min/1.73m ² /month)	-1.012 (0.102)	-1.038 (0.100)	-1.015 (0.103)	-1.025 (0.101)
	β_{1t} (mL/min/1.73m ² /month)	-1.213 (0.096)	-1.227 (0.098)	-1.214 (0.097)	-1.215 (0.097)
	β_{2c} (mL/min/1.73m ² /month)	0.593 (0.107)	0.579 (0.105)	0.583 (0.108)	0.565 (0.106)
	β_{2t} (mL/min/1.73m ² /month)	0.878 (0.100)	0.834 (0.101)	0.868 (0.101)	0.830 (0.101)
	β_{3c} (mL/min/1.73m ² /month)	-0.419 (0.020)	-0.459 (0.020)	-0.433 (0.020)	-0.460 (0.021)
	β_{3t} (mL/min/1.73m ² /month)	-0.335 (0.018)	-0.393 (0.020)	-0.346 (0.018)	-0.384 (0.021)
$\pi(T_i^0 \mathbf{b}_i)$	η_0 (Trt effect)	-0.202 (0.127)	-0.186 (0.198)	-0.207 (0.128)	-0.200 (0.156)
	η_1 (b_{0i} effect)		-0.087 (0.008)		-0.023 (0.037)
	η_2 (b_{1i} effect)		-0.308 (0.181)		-0.149 (0.188)
	η_3 (b_{3i} effect)		-4.725 (0.698)		-3.264 (0.962)
	η_4 ($\mu_{ih}(t_{h-1})$ effect)			-0.079 (0.006)	-0.065 (0.038)
Model Fit	-2 Log Likelihood	50801.53	50466.50	50542.31	50464.06
	LR test p-value [†]	< 0.0001	0.11783	< 0.0001	-
	AIC ^{††}	50851.53	50522.50	50594.31	50522.06

[†] LR tests comparing SP models 1-3 to the full SP model 4, ^{††} Akaike's information criterion, AIC

TABLE 5 Parameter estimates (standard errors) from the IDNT study for the SP models under ignorable and non-ignorable dropout mechanisms. LR=Likelihood Ratio, SP=Shared Parameter.

		Dropout Mechanism and SP Model			
		Ignorable		Non-ignorable	
Slope	Δ mL/min/1.73m ² /year	1	2	3	4
Acute	$\beta_{1t} - \beta_{1c}$	-2.42 (-5.54, 0.70)	-2.27 (-5.37, 0.82)	-2.39 (-5.55, 0.77)	-2.28 (-5.38, 0.82)
Change	$\beta_{2t} - \beta_{2c}$	3.42 (0.17, 6.68)	3.06 (-0.15, 6.27)	3.43 (0.14, 6.72)	3.18 (-0.05, 6.41)
Chronic	$\beta_{3t} - \beta_{3c}$	1.01 (0.38, 1.63)	0.79 (0.14, 1.44)	1.04 (0.40, 1.68)	0.90 (0.23, 1.57)
Total (3 years)	$\beta_{4t}(3) - \beta_{4c}(3)$	0.63 (0.01, 1.25)	0.45 (-0.20, 1.10)	0.66 (0.03, 1.29)	0.55 (-0.11, 1.21)
Total (4 years)	$\beta_{4t}(4) - \beta_{4c}(4)$	0.72 (0.12, 1.32)	0.53 (-0.10, 1.16)	0.76 (0.14, 1.37)	0.64 (-0.01, 1.28)
SP Covariates		Comparison			
Treatment Group	HR(Treated:Control)	0.82 (0.64, 1.05)	0.83 (0.56, 1.23)	0.81 (0.63, 1.04)	0.82 (0.60, 1.11)
b_{0i}^{\dagger}	HR (b_{0i}) per unit \uparrow		0.92 (0.90, 0.93)		0.98 (0.91, 1.05)
b_{1i}^{\ddagger}	HR (b_{1i}) per unit \uparrow		0.97 (0.95, 1.00)		0.99 (0.96, 1.02)
b_{3i}^{\ddagger}	HR (b_{3i}) per unit \uparrow		0.67 (0.60, 0.76)		0.76 (0.65, 0.89)
$\mu_{ih}(t_h - 1)^{\dagger}$	HR($\mu_{ih}(t_{h-1})$) per unit \uparrow			0.92 (0.91, 0.94)	0.94 (0.87, 1.01)

\dagger unit of measure = mL/min/1.73m²; \ddagger unit of measure = mL/min/1.73m²/year

TABLE 6 Treatment differences in slopes (Δ =Treated-Control, 95% CI's) and hazard ratios (HR, 95% CI's) from the IDNT study for SP covariates under ignorable and non-ignorable dropout mechanisms. SP=Shared Parameters, CI=Confidence Interval, HR=Hazard Ratio.

		Dropout Mechanism and SP Model					
Unit of measure		Ignorable (SP Model 1)			Non-ignorable (SP Model 2)		
Slope	mL/min/1.73m ² /year	Estimate	Model-based SE (p-value)	Robust SE (p-value)	Estimate	Model-based SE (p-value)	Robust SE (p-value)
Acute	β_{1c}	-12.14	1.224	1.258	-12.45	1.203	1.240
	β_{1t}	-14.56	1.154	1.124	-14.72	1.171	1.148
Change	$\beta_{1t} - \beta_{1c}$	-2.42	1.591 (0.1290)	1.597 (0.1304)	-2.27	1.580 (0.1501)	1.575 (0.1490)
	β_{2c}	7.11	1.281	1.336	6.94	1.255	1.306
	β_{2t}	10.54	1.199	1.175	10.01	1.213	1.196
Chronic	$\beta_{2t} - \beta_{2c}$	3.42	1.658 (0.0392)	1.684 (0.0422)	3.06	1.637 (0.0616)	1.642 (0.0624)
	β_{3c}	-5.03	0.238	0.262	-5.51	0.242	0.269
	β_{3t}	-4.02	0.217	0.227	-4.72	0.242	0.259
	$\beta_{3t} - \beta_{3c}$	1.01	0.319 (0.0016)	0.341 (0.0032)	0.79	0.331 (0.0175)	0.358 (0.0278)

TABLE 7 Robust inference on treatment slopes (SE's) and their differences (SE's and p-values) from the IDNT study under ignorable and non-ignorable dropout mechanisms. SE=Standard Error, SP=Shared Parameter.

		Dropout Mechanism and SP Model			
		Ignorable (SP Model 1)		Non-ignorable (SP Model 2)	
Slopes	mL/min/1.73m ² /year	Estimate	SE (p-value)	Estimate	SE (p-value)
Acute	β_{1c}	-5.524	0.753	-5.453	0.739
	β_{1t}	-5.568	0.622	-5.377	0.625
	$\beta_{1t} - \beta_{1c}$	-0.044	0.971 (0.9635)	0.076	0.960 (0.9371)
Change	β_{2c}	2.401	0.838	1.680	0.809
	β_{2t}	2.999	0.704	2.146	0.707
	$\beta_{2t} - \beta_{2c}$	0.589	1.087 (0.5885)	0.466	1.056 (0.6595)
Chronic	β_{3c}	-3.123	0.357	-3.773	0.368
	β_{3t}	-2.578	0.266	-3.231	0.295
	$\beta_{3t} - \beta_{3c}$	0.545	0.443 (0.2197)	0.542	0.461 (0.2404)
Source of Variability	Parameter	Estimate	SE (p-value)	Estimate	SE (p-value)
Within-Subject	σ^2	3.495	1.112	3.153	0.998
	θ	0.847	0.056 (< 0.0001)	0.862	0.056 (< 0.0001)
Between-Subject	κ	-0.264	0.077 (0.0007)	-0.234	0.078 (0.0028)
	ψ_{11}	16.951	1.853	17.040	1.857
	ψ_{21}	0.464	0.209	0.423	0.203
	ψ_{22}	0.245	0.056	0.238	0.052
	ψ_{31}	-0.505	0.233	-0.356	0.224
	ψ_{32}	-0.211	0.057	-0.187	0.052
	ψ_{33}	0.246	0.066	0.221	0.059
SP Covariates	Comparison	Estimate	95% CI	Estimate	95% CI
Treatment Group	HR(Treated:Control)	0.95	(0.65, 1.41)	0.94	(0.43, 2.04)
b_{0i}^{\dagger}	HR (b_{0i}) per unit \uparrow			0.77	(0.71, 0.84)
b_{1i}^{\ddagger}	HR (b_{1i}) per unit \uparrow			0.87	(0.81, 0.93)
b_{3i}^{\ddagger}	HR (b_{3i}) per unit \uparrow			0.55	(0.47, 0.65)
Information Criteria	-2 Log Likelihood	10935.21		10751.22	
	AIC	10981.21		10803.22	

† unit of measure = mL/min/1.73m²; ‡ unit of measure = mL/min/1.73m²/year

TABLE 8 Summary statistics from the MDRD-B study on treatment slopes and their differences, variance-covariance parameters, hazard ratios with respect to SP covariates and likelihood-based information criteria under ignorable and non-ignorable dropout mechanisms. SP=Shared Parameters, SE=Standard Error, HR=Hazard Ratio, CI=Confidence Interval, AIC=Akaike's information criterion.

Knot	PA-POM Weights		EM-POM Weights	
	$\Delta = 1.00$	$\Delta = 1.25$	$\Delta = 1.00$	$\Delta = 1.25$
3	0.815	0.947	0.784	0.930
4	0.794	0.936	0.760	0.916
5	0.783	0.929	0.745	0.906
6	0.776	0.925	0.736	0.900

TABLE 9 Estimated power to detect a 1.00 to 1.25 mL/min/1.73m²/year difference, $\Delta = (\beta_{3t} - \beta_{3c})$, in the chronic slopes for a sample of 400 subjects per treatment group assuming different knots and different approximations to the marginal POM weights, w_{ij} . PA-POM=Population-Average Power-of-Mean, EM-POM=Estimated Marginal Power-of-Mean.

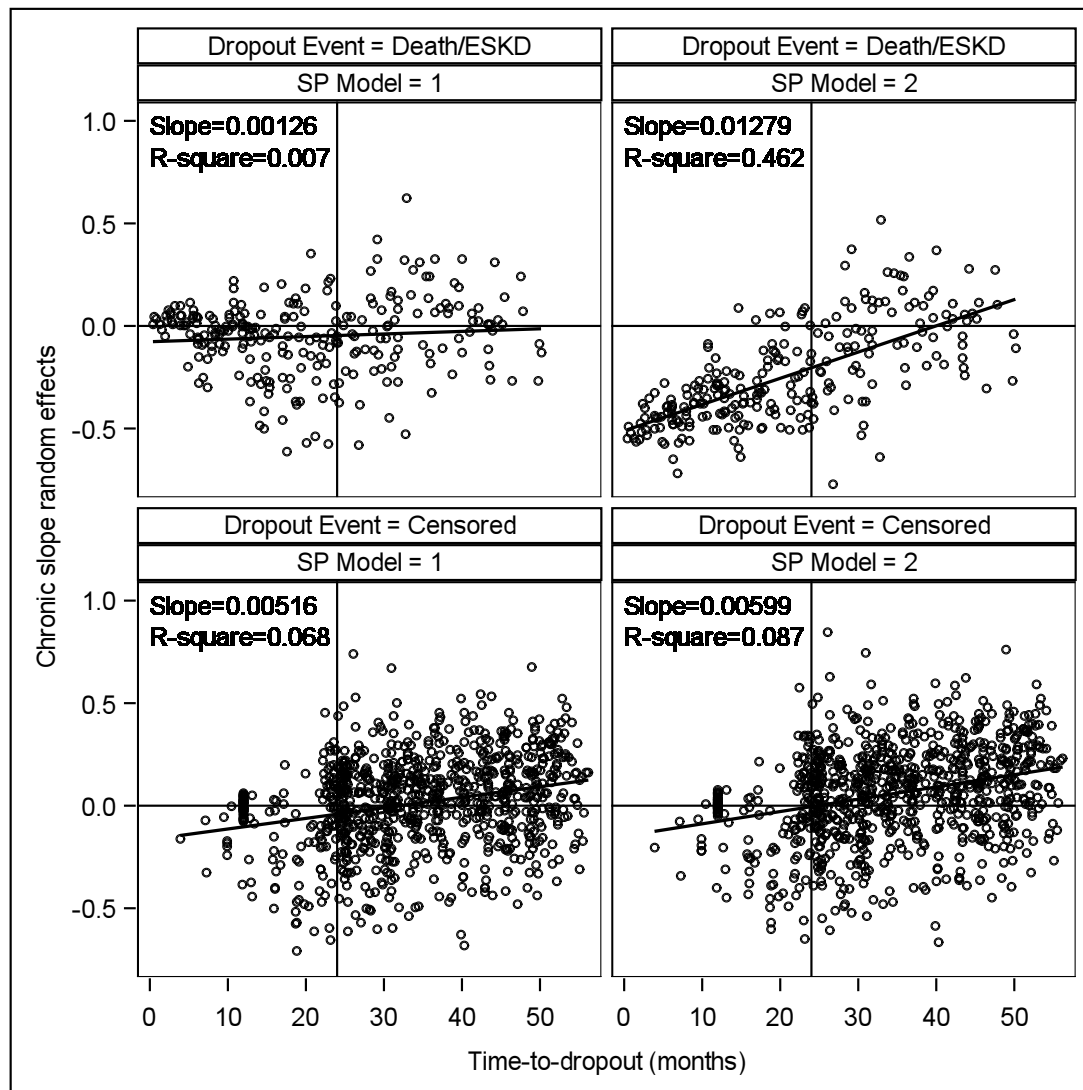


FIGURE 1 Chronic slope random effects \hat{b}_{3i} versus time to dropout T_i by event status and SP model for the IDNT study. The slope and R-square values are crude measures of association corresponding to a simple linear regression of \hat{b}_{3i} versus T_i . The vertical reference line at 24 months corresponds to the minimum planned follow-up time for all patients following the end of recruitment with administrative censoring occurring on December 31, 2000.

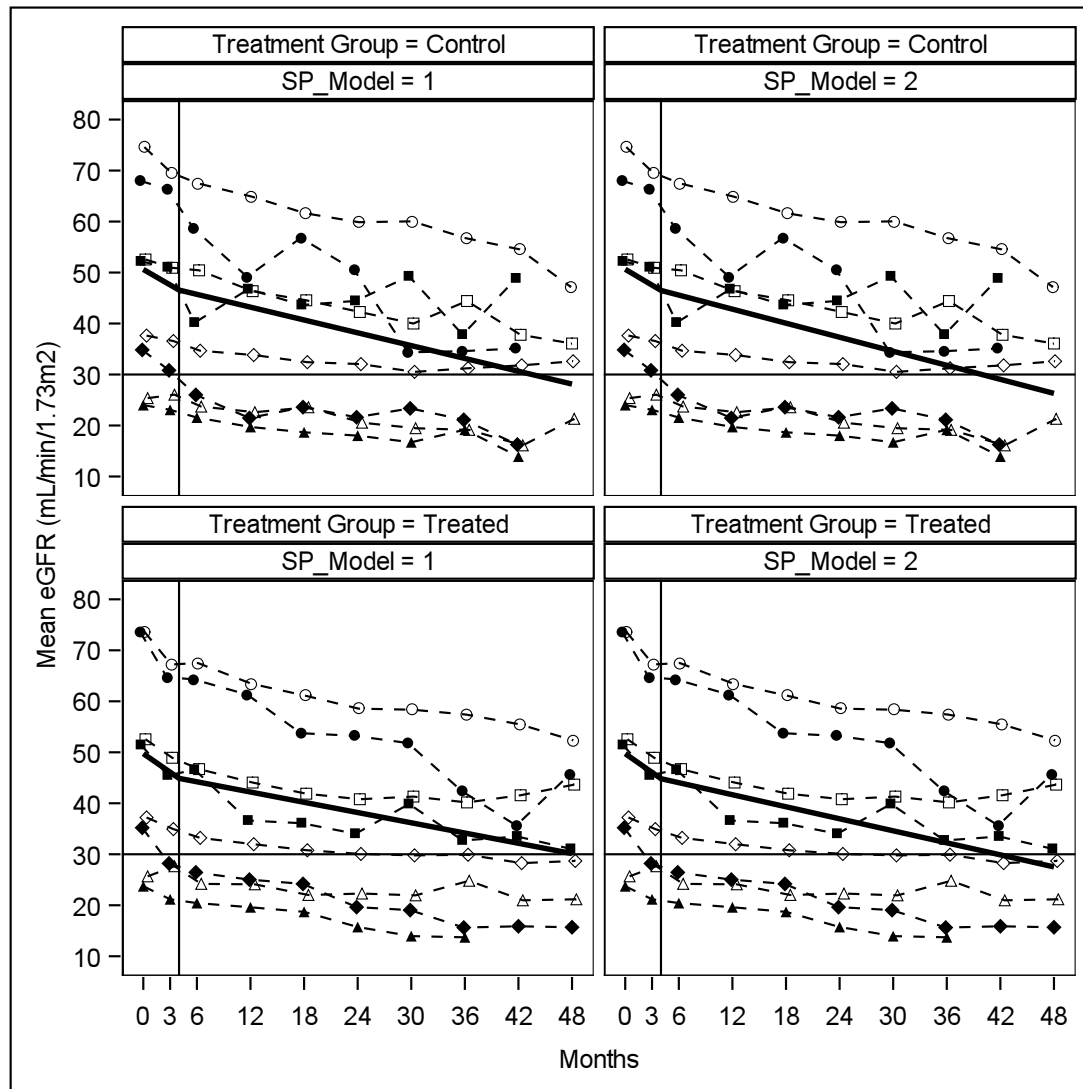


FIGURE 2 Observed mean eGFR values from the IDNT study according to what stage of CKD patients are in at baseline and what their dropout status is compared to the overall predicted mean eGFR profile by treatment group. The circle, square, diamond, and triangle are observed means for CKD Stages 1-2, 3a, 3b and 4, respectively. The filled (open) symbols are the observed means for subjects with (without) the composite dropout event of death or ESKD. The solid lines are the predicted mean eGFR profiles from the linear spline mixed-effects portion of the respective SP models.

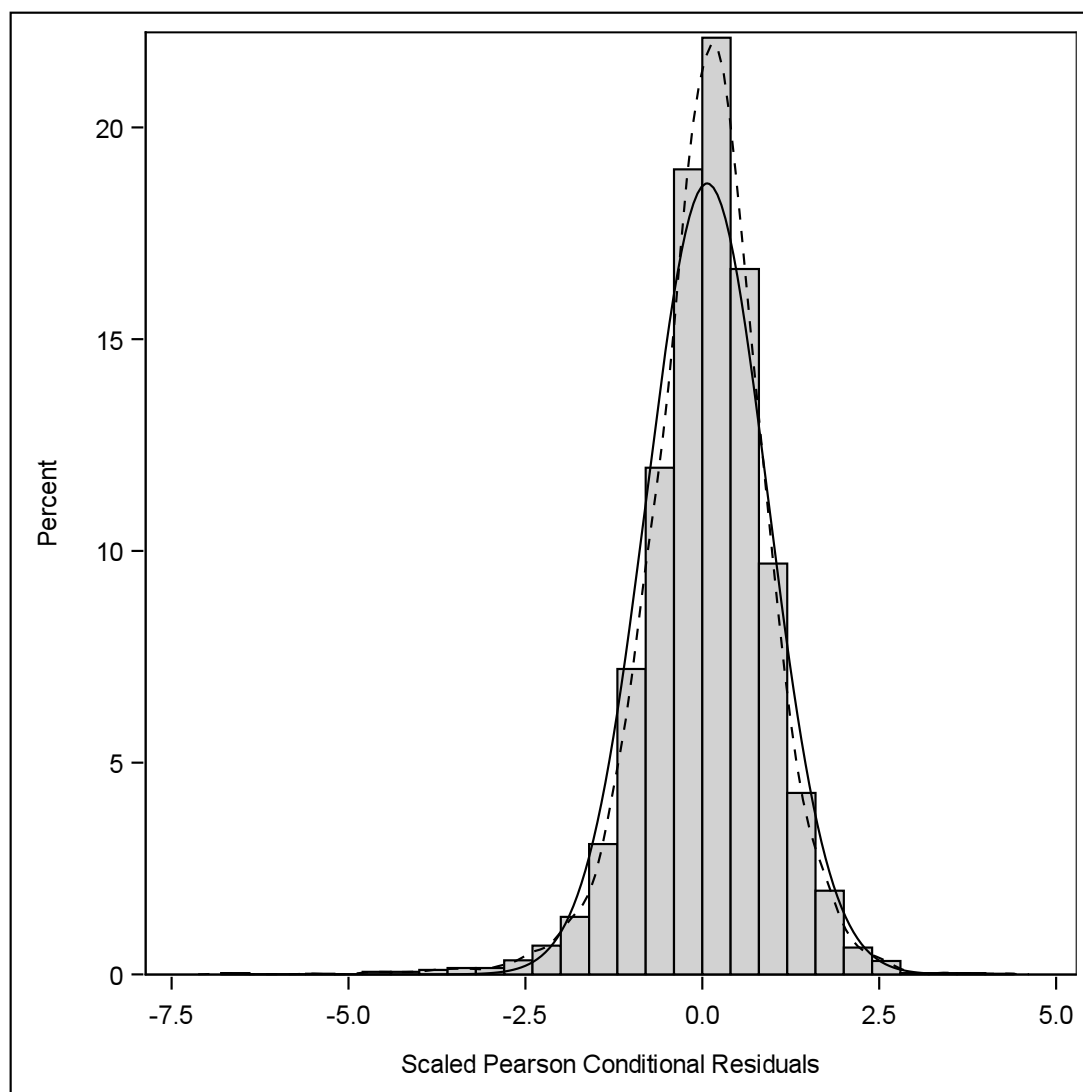


FIGURE 3 Histogram of the scaled conditional Pearson residuals, $\{y_{ij} - \mu(\hat{\beta}_i)\}/\hat{\sigma}_{ij}$, under SP model 2 of the IDNT study. Superimposed are the estimated Gaussian (solid curve) and Kernel (dashed curve) probability density curves suggesting the assumption of normality is not unreasonable. There are 3 outlying values below 5 standard deviations of the assumed conditional mean of 0 but given this plot is based on 6,628 observations, such extreme values are not unexpected.

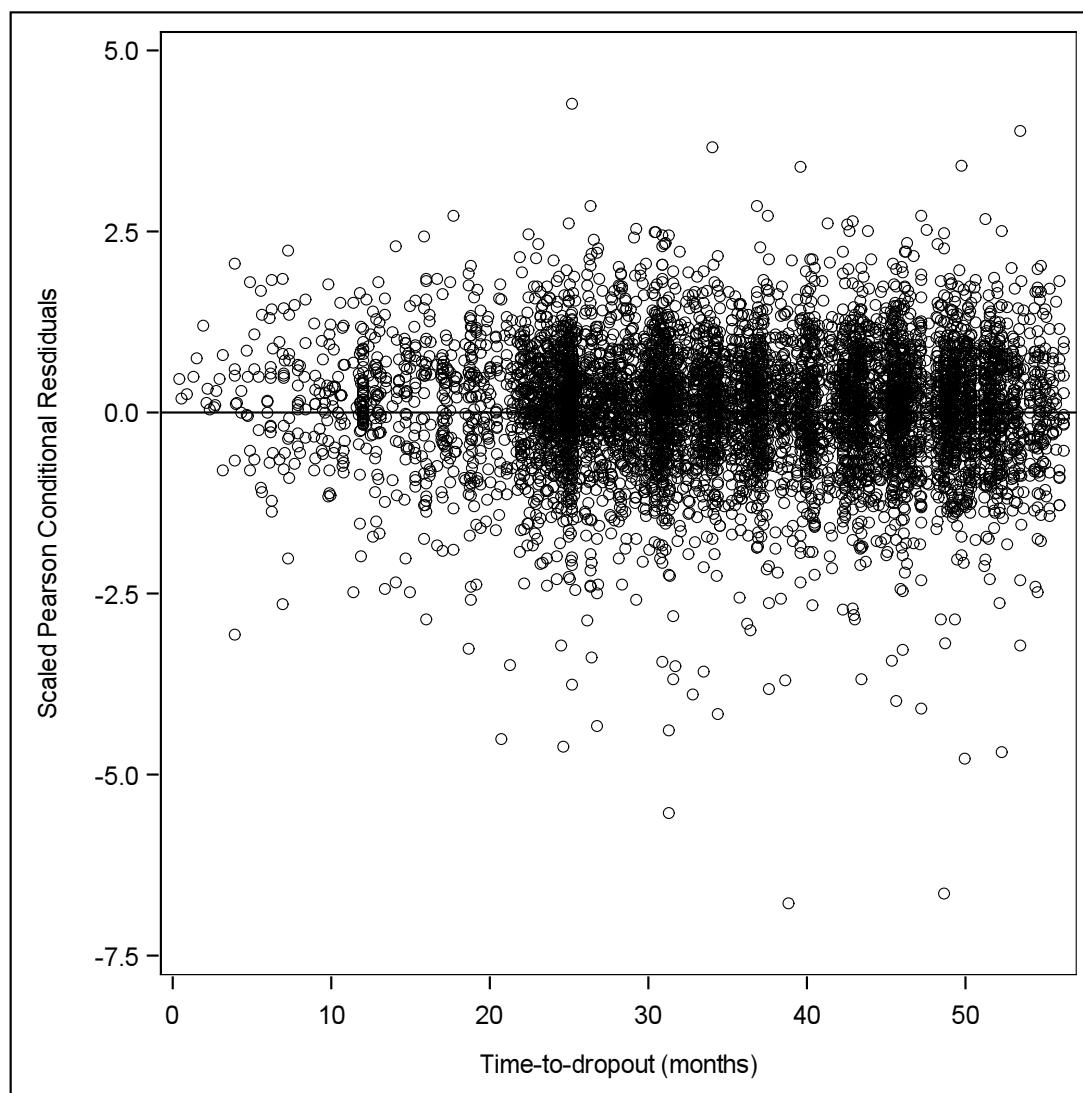


FIGURE 4 Scaled conditional Pearson residuals, $\{y_{ij} - \mu(\hat{\beta}_i)\} / \hat{\sigma}_{ij}$, versus time to dropout under SP model 2 of the IDNT study. There is no graphical evidence to suggest the assumption of conditional independence between $Y_{ij} | \mathbf{b}_i$ and $T_i | \mathbf{b}_i$ is violated.

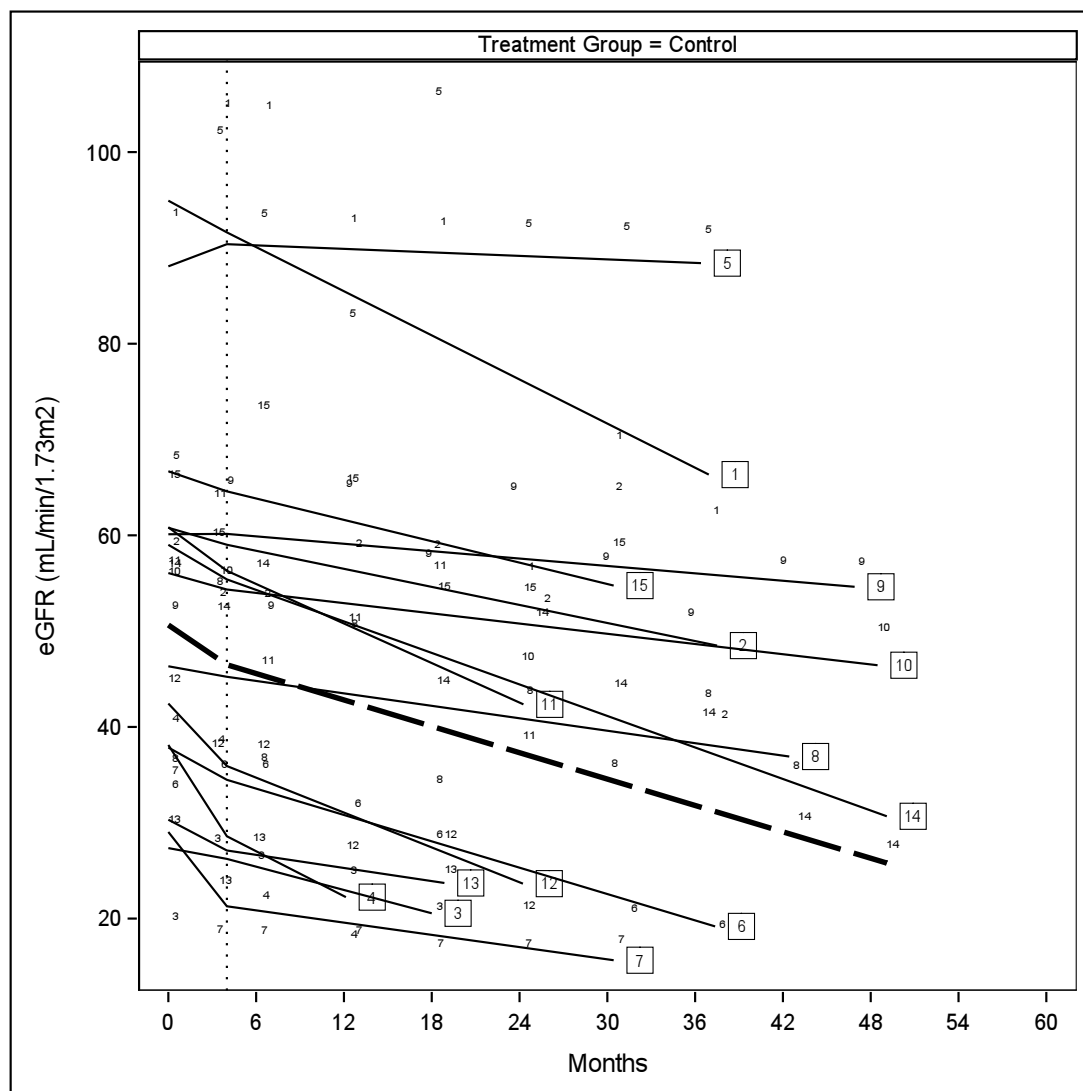


FIGURE 5 Subject-specific mean eGFR profiles and observed eGFR values under the IDNT study for 15 randomly selected subjects from the control group under SP model 2. The observed eGFR values are displayed according to the randomly assigned subject ID values which are labeled 1-15. The boxed numbers align with the subject-specific predicted mean eGFR profiles. The dotted vertical line marks the linear spline knot at 4 months while the bold dashed line marks the population-average predicted mean eGFR profile.

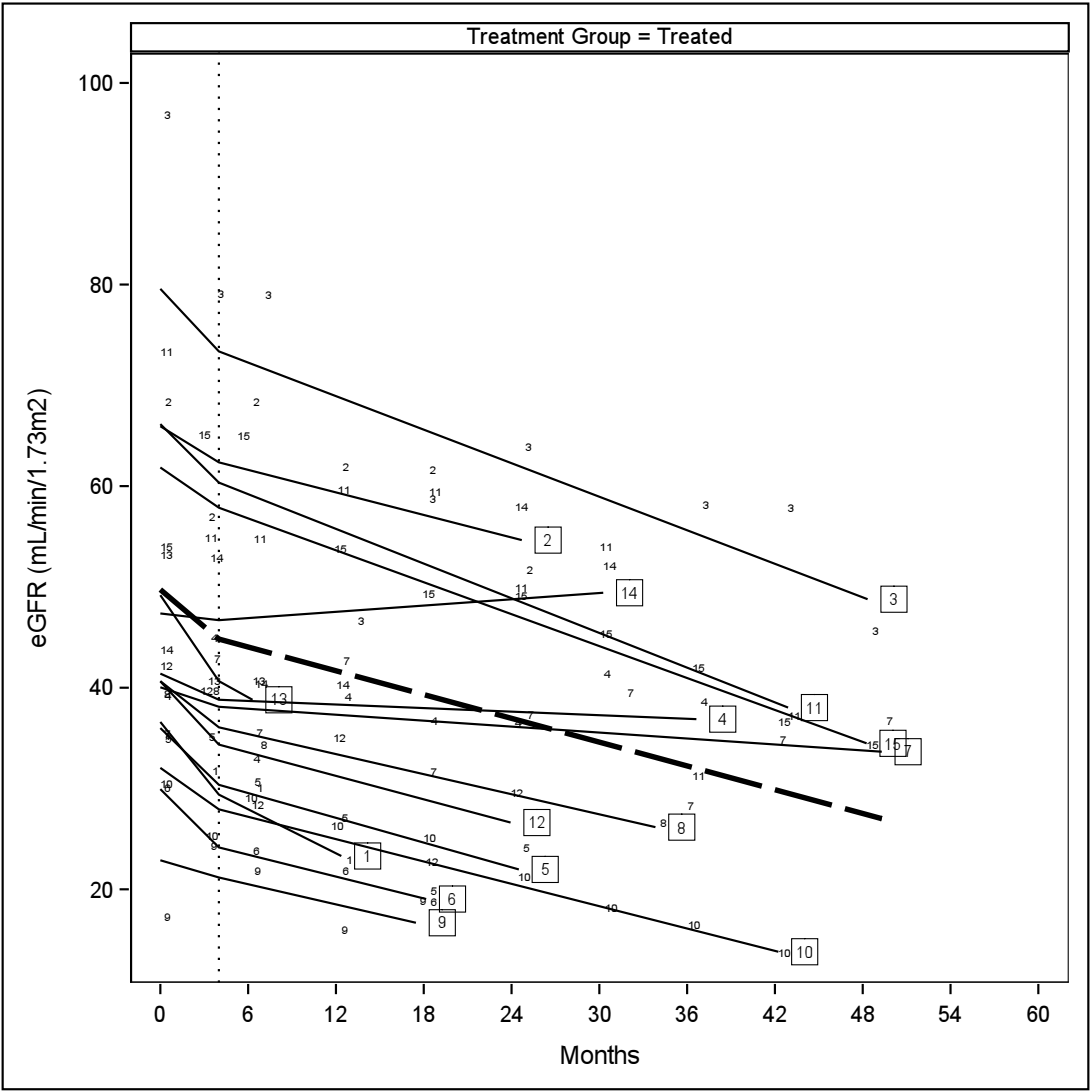


FIGURE 6 Subject-specific mean eGFR profiles and observed eGFR values under the IDNT study for 15 randomly selected subjects from the treated group under SP model 2. The observed eGFR values are displayed according to the randomly assigned subject ID values which are labeled 1-15. The boxed numbers align with the subject-specific predicted mean eGFR profiles. The dotted vertical line marks the linear spline knot at 4 months while the bold dashed line marks the population-average predicted mean eGFR profile.