12 December 2023

Procedure No.: EMA/SA/0000104642
Human Medicines Division

Initial Qualification Procedure - List of Issues
eGFR slope

Summary

The objective of this request is to seek a qualification opinion on the use of glomerular filtration rate (GFR) slope, i.e., the mean rate of change in GFR, as a validated surrogate endpoint for chronic kidney disease (CKD) progression in clinical trials for standard marketing authorization and indication extension approvals.

Scientific discussion

The presented validation approach for GFR slope as surrogate endpoint by the Applicant could be considered comprehensive and complete. It may allow qualifying GFR slope as surrogate endpoint in clinical trials. Use of individual patient data for analysis is acknowledged and availability of these data can be considered a strength of the validation approach. It is however noted that the proposed context of use is broad regarding e.g., trial designs, disease settings and target populations (including renally healthy patients).

The target population for use of the method is broad with 4 different disease areas proposed (CKD, diabetes/diabetic kidney disease, glomerular disease, cardiovascular disease (CVD)). It is currently unclear which population is targeted by the cardiovascular disease definition. A thorough discussion of differences between disease areas is currently missing and recommendations may differ for e.g., diabetes, as trials in this target population are of large size and dedicated studies addressing a population with high cardiovascular risk with also increased risk for renal disease may be included in the study programs. It is not clear if conclusions on surrogacy also hold for patients at risk for CKD. It is unclear if this population at risk is defined by an explicit threshold and the size of a database for this part of the proposed target population is currently not explained. The broad applicability of the approach may currently be questioned.

Regarding the analyses used for individual trial data, the extensive and well described work by the Applicant is noted (e.g., CKD-EPI Consortium Technical Report in Appendix C and Vonesh E et al., Stat Med 2019). The analysis for the validation approach used the same unified mixed effects model for GFR slope for all trials, using random effects slope and intercept terms for variability in GFR between patients. A shared parameter model was used to consider informative censoring by kidney failure with replacement therapy (KRFT) and death if a sufficient number of events was available. This simplified model allows estimating an acute effect on GFR slope. It is assumed that an acute effect is lasting up to 3 months but avoids making an assumption on the shape of the GFR curve for the first 3 months. The unified mixed effects model allows estimation of 'acute slope', chronic slope and total slope over the defined periods of 2 and 3 years (and change from baseline at 2 and 3 years). The rationale for using this model for all trials is noted. However, for application in future trials it is recommended that the analysis model and analysis of acute effects is tailored to the population and intervention. It is not fully clear if the conclusions on surrogacy would also be applicable for a tailored approach. It is not obvious that conclusions on Type 1 error and bias or GFR threshold to infer a beneficial effect on a clinical endpoint would be the same.

The trial level analysis comprises a dataset that includes data used in a previous publication by Inker and co-authors (Inker LA et al., J Am Soc Nephrol 2019), and a set of new studies resulting in 66 randomised comparisons from 17 interventions in 4 disease areas (CKD, diabetes/diabetic kidney disease, glomerular disease, CVD). For total GFR slope, a unified analysis method was applied to data over 3 years. The observed posterior median correlation was $R^2=0.98$ with Bayesian credible intervals from 0.85 to 1.00. The slope of the meta regression differs from 0, in principle supporting use of GFR slope as surrogate endpoint. For chronic slope, the posterior median R^2 only shows moderate association with clinical endpoints ($R^2=0.56$) and is lower than previously reported in the meta-

analysis mentioned above. The trial level surrogacy analysis for chronic slope exhibits sensitivity to acute effects and baseline levels of GFR. This is more obvious with the inclusion of additional datasets in the updated analysis. The weaker association as regards chronic slope was partly explained by the addition of two studies (FIDELIO-DKD and CREDENCE).

Overall results for the trial level surrogacy may indicate that the total slope is more robust than the chronic slope. For example, there is a relevant dependence of results for chronic slope on covariate defining subgroups (p. 63 briefing document). For total slope, the association between treatment effects and the clinical endpoint was well comparable across all subgroups by baseline GFR, causal disease, rate of progression on control, or baseline proteinuria. For chronic slope, there is some heterogeneity for results in the proposed disease areas. The association between chronic slope and clinical endpoints was best for glomerular disease (R2 0.99) and weaker for diabetes (R2 0.78), other CKD (R2 0.83) and CVD (R2 0.69). Information is missing on the criteria applied for categorising the studies into the four proposed subgroups by disease area and on how many patients without CKD were included in the studies.

Regarding model-based analysis of GFR slope, the work on analysis methods by the Applicant is acknowledged (CKD-EPI Consortium Technical Report in Appendix C and Vonesh E et al., Stat Med 2019). An aspect missing for the proposed approach is a discussion of the assumption that a linear model can adequately describe the GFR trajectories in all target populations and across all disease stages.. The impact of a decline with faster progression in advanced disease stage, close to renal failure or death is unclear. A discussion would have been expected on the observed trajectories considering potential factors like baseline GFR, disease, progression rate and potentially assessment schedule. For the trial level surrogacy analysis, the Applicant chose a 'one-size-fits-all' approach to the analysis method to allow using the same analysis method for all trials included. This approach may be questioned, as acute effects may have a relevant impact on acceptability of the proposed slope parameters (total and chronic slope). A tailored approach to the analysis based on the observed acute effect may be desirable. It is currently unclear if the conclusions on surrogacy would also hold if a tailored analysis for incorporating acute effects would have been used. It is acknowledged that the shared parameter models take informative censoring into account. However, choice and parametrisation of the model may depend on expectation of renal failure or death events. It is noted that a shared parameter was used only in a subset of trials (Table A-17, p. 131 briefing document). The Applicant should discuss the model choice in settings with low to intermediate numbers of expected for renal failure or death events and comment on interpretability of results from these models, specifically in settings with high risk of censoring events, i.e., in target populations with fast progression. For a Phase 3 trial, the proposal to use of a 2-slope linear mixed effects model with non-linear analysis may be problematic for a confirmatory analysis. In Phase 3 trials, the analysis should be pre-specified and convergence problems could pose a problem.

Concerning use of analysis models in future trials, it is currently unclear how the model-based analysis in a future trial would be impacted by intercurrent events as treatment discontinuations and missing data due to study drop-outs if applied in a Phase 3 trial. A description of an estimand is currently missing. Approaches to handle intercurrent events and missing data due to study drop-out should consider acute effects. Moreover, the strategy for handling intercurrent events would likely be dependent on the direction of acute effects. Using a treatment policy strategy may not be appropriate with acute effects of an intervention. The Applicant should also discuss robustness of the simulations with the assumed loss to follow-up of 2% per year, and if this assumption is realistic.

For future application of GFR slope parameters in clinical trials, the properties of GFR slopes in clinical trial settings are of importance. To explore this, the Applicant performed simulations to assess operating characteristics. From a regulatory perspective the risk for Type 1 error and bias are considered very important characteristics. For Applicants who consider using GFR slope, the advantages over use of alternative endpoints as time to GFR decline (e.g., 30% or 40% GFR decline or kidney failure) with a potentially simpler and more robust analysis is of importance. It is debatable if a minimum duration of follow-up should be defined. Regarding acute effects, it is obvious that longer duration of follow-up helps addressing impact of acute effects on Type 1 error and bias.
It is acknowledged that the Applicant provides some relevant discussion on results from the trial level surrogacy analysis, on the minimal clinically relevant GFR threshold, and on impact of acute effects

and heterogeneity across proposed disease areas. However, several questions regarding these aspects remain open and need further discussion. Regarding trial design, the CoU proposal for application of GFR slope currently allows large flexibility regarding the choice of the slope endpoint (total or chronic) and important trial design features such as e.g., follow-up duration and measurement frequency. Additionally, it is currently unclear how an analysis method and estimand should be defined when used in a future clinical trial.

It is currently not fully clear if use of GFR slope would have advantages in diseases with fast progression with high likelihood of censoring events or terminal events. It should be discussed whether the CoU may be restricted to disease entities for which studies with clinical endpoint are less feasible.

January 30, 2023

Applicant Reply

We thank the EMA for their response and questions. Below, we have included a detailed response to the issues raised. We look forward to the discussion and ability to further respond to the EMA's questions.

Our view is that regulatory qualification of GFR slope as a surrogate endpoint would indicate a general acceptance of the principle that GFR slope can be used as a primary endpoint in clinical trials for CKD progression for standard marketing authorization and indication extension approvals.

Based on the totality of evidence, we conclude that GFR slope can be considered to be a valid surrogate endpoint. This conclusion is based on the strong biological rationale, previously conducted epidemiological analyses, and the strong scientific evidence presented here demonstrating that treatment effects on GFR slope accurately predict treatment effects on the clinical endpoint across a broad range of study populations and treatment interventions. In addition, treatment effects on GFR slope are clinically meaningful. The results presented here can used to translate treatment effects on GFR slope to treatment effects on the clinical outcome across heterogeneous populations.

The analyses of GFR slope within a specific trial and setting does require careful consideration of factors specific to that context. Sponsors proposing to use GFR slope in a particular trial would be responsible for demonstrating to the regulatory agency the soundness of use of GFR slope in the study design and analysis plan for the specific setting in which the trial is conducted.

List of issues to be addressed in writing by 02 January 2023 and during the discussion meeting
Based on the coordinators' reports the Scientific Advice Working Party (SAWP) determined that the Applicant should discuss the following points, before advice can be provided:

Issues to be addressed <u>in writing and during the discussion meeting</u>

Questions regarding the analysis:

Population
1. **Please provide clarification on the definition of the four populations included in the target population (CKD, diabetes, glomerular disease, CVD).**

The appendix outlined our inclusion criteria. All studies had to indicate progression of CKD with the number of clinical kidney failure events varying by expected study size. These are indicated in the table below. In **Table R1**, we provide a more detailed description of disease categories.

Table R1: Detailed description of disease categories (expansion of Table A-5)			
Higher level disease category	**Basic disease category (N studies, N participants)**	**Study inclusion (all GFR > 15 and Follow up more than 12 months after first follow up measurement of urine protein or GFR**	**Definition**
Diabetes	Diabetes, not specified as DKD (11, 75464)	follow-up 1000 or more person-years and 30 or more clinical kidney failure events	Patients with diabetes where an inclusion criteria for the study did not require ACR > 30 or GFR > 90
	Diabetes with kidney disease (DKD) (10, 26552)	follow-up 500 or more person-years and 30 or more clinical kidney failure events	Patients with diabetes where an inclusion criteria for the study did require ACR > 30 or GFR > 90
CKD	CKD-Hypertension (3, 2621)	follow-up 500 or more person-years and 30 or more clinical kidney failure events	Patients with CKD (GFR > 90 or ACR > 30 with a diagnosis of hypertension
	Polycystic kidney disease (3, 1546)		Patients with PKD
	Other CKD (could not specify) (22, 15982)		Patients with CKD (GFR > 90 or ACR > 30 and the diagnosis was other or not specified
Glomerular	IgA nephropathy (7, 1037)	Clinical endpoint > 10 events	Patients with IgA nephropathy
	Lupus nephritis (1, 79)		Patients with Lupus nephritis
	Membranous nephropathy (1, 273)		Patients with Membranous nephropathy
	Focal segmental glomerulosclerosis (1, 138)		Patients with FSGS
Cardiovascular	High cardiovascular risk (3, 12788)	follow-up 1000 or more person-years and 30 or more clinical kidney failure events	Patients at high risk for CKD (diabetes, hypertension, cardiovascular disease)- not selected for having kidney disease
	Heart failure (4, 50843)		Patients with chronic heart failure enrolled in studies to evaluate treatments on chronic HR, not selected for having kidney disease

a. **Please clarify whether the studies summarized in the subgroup 'diabetes' only comprised patients with diabetic kidney disease or whether patients with GFR >90ml/min/1.73 m^2 were also included.**
b. **It should be clarified which entities were summarized under the group label 'CVD', and which proportion of patients in each of the 7 studies had GFR >90ml/min/1.73 m^2 at baseline.**

Reply to a and b

Table R2 lists the studies and their baseline characteristics of studies with CVD and diabetes, not specified as DKD

Table R2: Baseline characteristics of studies with CVD and diabetes without CKD										
Study name	Dis ease	Intervention	N	Baseline GFR	Baseline ACR	%GFR > 90 and ACR < 30	% GFR > 90	Hazard Ratio	Event % in control arm	Chr slope in control arm
ABCD(CCB)	DM	RASB v CCB	392	72.1	127	0.77	17.35	1.1(0.5,2.5)	5.6	-1.57
ABCD(BP)	DM	Low v Usual BP	392	72.1	127	0.77	17.35	1.4(0.6,3.3)	4.5	-1.67
ALTITUDE	DM	RASB vs Con	8150	58.4	284	0.04	11.72	1.1(0.9,1.3)	6.3	-3.71
ADVANCE(ACE)	DM	RASB vs Con	10876	78.3	15	20	28.84	1.3(0.9,1.8)	1.1	-1.42
ADVANCE(GLUC)	DM	Int Glu	10876	78.3	15	20	28.84	1(0.8,1.5)	1.3	-1.23
CANVAS	DM	SGLT2-I	10031	78.7	12	Data NA	Data NA	0.6(0.4,0.9)	0.9	-1.38
EMPA-REG	DM	SGLT2-I	6936	76.2	18	19.46	28.72	0.5(0.4,0.7)	3.4	-2.15
CAROLINA	DM	DPP-4 I	5985	78.7	10	23.16	30.03	1.4(0.9,2.1)	1.2	-1.33
EXAMINE	DM	DPP-4 I	5377	75.2	72	ACR NA	28.75	1(0.7,1.3)	3.1	0.13
Harmony	DM	GLP-1 A	8913	78.8	24	ACR NA	35.55	1.1(0.8,1.5)	2.0	-2.22
LEADER	DM	GLP-1 A	7533	65.1	20	0.00	39.56	0.8(0.7,1.1)	3.4	-2.81
TOPCAT	HF	MRA	3435	70.2	11	ACR NA	11.15	1.6(1.2,2.1)	4.4	-0.74
PARADIGM-HF	HF	RASB vs Con	8440	73.3	NA	ACR NA	17.37	0.8(0.6,1)	3.9	-2.18
CHARM-Added	HF	RASB vs Con	913	72.5	10	ACR NA	26.94	1.3(0.7,2.1)	5.5	-1.81
SPRINT	CV	Low v Usual BP	8885	75.0	13	17.83	20.45	1.7(1.2,2.4)	1.1	-0.76
ACCOMPLISH	CV	RASB+CCB	11482	74.6	NA	ACR NA	30.49	0.5(0.4,0.7)	3.7	-1.23
PEGASUS	CV	Antiplatelet	17782	82.6	NA	ACR NA	22.49	0.9(0.7,1.3)	0.8	-0.84
PLATO	CV	Antiplatelet	12679	78.8	24	ACR NA	43.32	1.4(0.9,2.1)	0.6	-1.03
Abbreviations: ACR NA, albumin:creatinine ratio unavailable or insufficient; CV, cardiovascular; Data NA, access to data lost due to nature of the agreement; DM, diabetes mellitus; N, sample size										

c. **The analysis as given in table 12 (p. 63 briefing document) should be newly performed excluding subjects with normal renal function at baseline. Results should be discussed considering the proposed context of use, which also apparently intends to include primary prevention of kidney disease.**

We agree with EMA that is important to ensure that the trial level associations are robust by disease severity. CKD is defined by both GFR and ACR, and many patients with high GFR (e.g. > 90) will have substantial progression as in the case of individuals with diabetes or PKD. As such, we took a broader look at this question. In the qualification opinion draft dossier, we showed several data that support that our main results are consistent across range of GFR and albuminuria.
*   Results were consistent when restricting the analysis to participants with urine albumin to creatinine ratio (ACR) > 30 mg/g, although precision was reduced given the smaller number of studies and participants (Qualification Opinion Draft Dossier [QOPD] Table 8)
*   Results were consistent after removing CVD studies, which are the studies with the highest proportion of participants with GFR > 90 ml/min per 1.73 m^2 (QOPD Table 9 and **Table R2**)

- Results for total slope were consistent for subgroups based on mean level of GFR, and results for total slope and chronic slope were consistent for subgroups based on rate of progression on the chronic slope in the control arm (QOPD Table 7)
- In response to question 2, we describe an expansion of our trial-level model that incorporates interaction terms with the study-level mean baseline GFR and mean baseline log transformed UACR. We demonstrated that the results for Total Slope remain robust to these disease severity factors.

We were not able to address the specific question of EMA within a short timeframe, that is to exclude participants with GFR > 90 ml/min/1.73m^2, as we do not have current access to all individual datasets. Some of the study level analyses were run on central servers to which we do not have access. In addition, we are concerned with taking a subset of the study population at this cut-off. GFR is estimated with error, with larger error variability at higher levels of GFR. Thus, application of an upper cut-off of 90 ml/min/1.73m^2 may mask subsequent progression due to regression to the mean, complicating interpretation of mean slopes. We could remove studies with a large proportion of participants who have GFR > 90 but this would essentially be the analysis we previously presented removing studies of CVD.

Given these considerations, we continue to support the current use of the broad set of studies and participants to provide the strongest evidence to support validity of the GFR slope as a primary surrogate endpoint. As we described in the QOPD (page 42), demonstration that treatment effects on a surrogate endpoint accurately predict treatment effects on the clinical endpoint across a broadly heterogeneous collection of studies strengthens the conclusions provided by a trial level analysis. Targeted subgroup analyses informs of the applicability of the results within key populations of interest.

Regarding EMA's question about context of use, we would like to clarify that we are not are not proposing GFR slope to be used to study primary prevention of CKD. The proposed context of use in for populations which have sufficient degree of progressive CKD to power for a slope analysis. In the summary section at the end of our responses, we suggest a revised context of use.

2. **Please discuss subgroup analyses results, e.g., regarding baseline GFR and target population including subjects at risk, considering potential differential acute effects in subgroups.**
As described above, in trial-level analyses evaluating the validity of total slope by subgroups, strong associations between treatment effects on total slope and treatment effects on the clinical endpoint were observed across the entire set of studies with results similar across all subgroups by baseline GFR, disease, rate of progression on control arm (QOPD Table 7), or baseline ACR (QOPD Table 8). In contrast, in trial-level analyses evaluating the chronic slope by subgroups, there was substantial variation by baseline GFR (Table 12).

In addition to these analyses, we expanded our trial-level model for evaluating the total and chronic slope as surrogate endpoints to incorporate continuous interaction terms with the study-level mean baseline GFR, mean baseline log transformed UACR, as well as the mean chronic slope in the control group, which we used as a marker of the average rate of CKD progression. (Of note, these three factors are interrelated, as trials with steeper average CKD progression tended to also have higher baseline UACR and lower baseline GFR). The goal of these analyses was to determine if our findings of the

associations between treatment effects on slope and treatment effects on the clinical endpoints differ by severity of CKD.

Our results are consistent with the subgroups analyses we had included in the qualification document. We also found that none of the three factors modified the meta-regression slope and intercept which relates treatment effects on the 3-year total slope to treatment effects on the clinical endpoint (**Figure R1**). On the other hand, the relationship of treatment effects on the chronic slope vs. treatment effects on the clinical endpoint did depend on each of the three factors; a given favorable treatment effect on the chronic slope (expressed as a difference in mean slopes, in ml/min/1.73m^2/year) translated to a greater reduction in the hazard for the clinical event for trials with higher baseline GFR, lower baseline UACR, or a less steep mean chronic slope. These findings further support the validity of the total slope as a surrogate endpoint, and also clarify the complications of using the chronic slope as a surrogate endpoint.

As we demonstrated in our published paper, there is substantial heterogeneity in acute effects across studies and by intervention and disease (QOPD Figure 5) (Neuen *et al.*, J Am Soc Nephrol 2022). Despite this, treatment effects on the Total Slope strongly associated with treatment effects on the clinical endpoint across the large collection of studies. The comparatively weaker trial-level association of the chronic slope with the clinical endpoint might be explained by its exclusion of the acute effects, and hence inability to capture the variation in acute effects across studies.

Chronic slope and simulations:

3. **Please elucidate the moderate association of chronic slope and clinical events in the FIDELIO DKD and CREDENCE studies (total slope data for these studies should also be provided). Study specific issues which may have contributed to the poor association between chronic slope and clinical outcome endpoints might be included in the discussion.**
4. **Please discuss the weaker association as regards chronic slope in the trial level surrogacy analysis when adding the FIDELIO-DKD and CREDENCE studies. A sensitivity analysis conducted excluding these studies (p. 54 briefing document) showed an increase in R^2 from 0.56 to 0.73. This finding may be considered at odds with the simulation results (p. 61 briefing document), which found that the relative efficiency of chronic slope is larger than for the time-to-event endpoints when the acute effect is negative. In both studies negative acute effects were observed.**

<u>**Reply to Q3 and Q4 together**</u>

We think it will be helpful to distinguish between the three types of analyses that are being discussed:
  1. The treatment effect on the chronic slope within each individual study
  2. The trial level analysis which examines the association of treatment effects on GFR slope compared to treatment effects on the clinical endpoint
  3. Relative efficacy of one endpoint vs another (in this case chronic slope vs time to events) in terms of increased power or reduced sample size
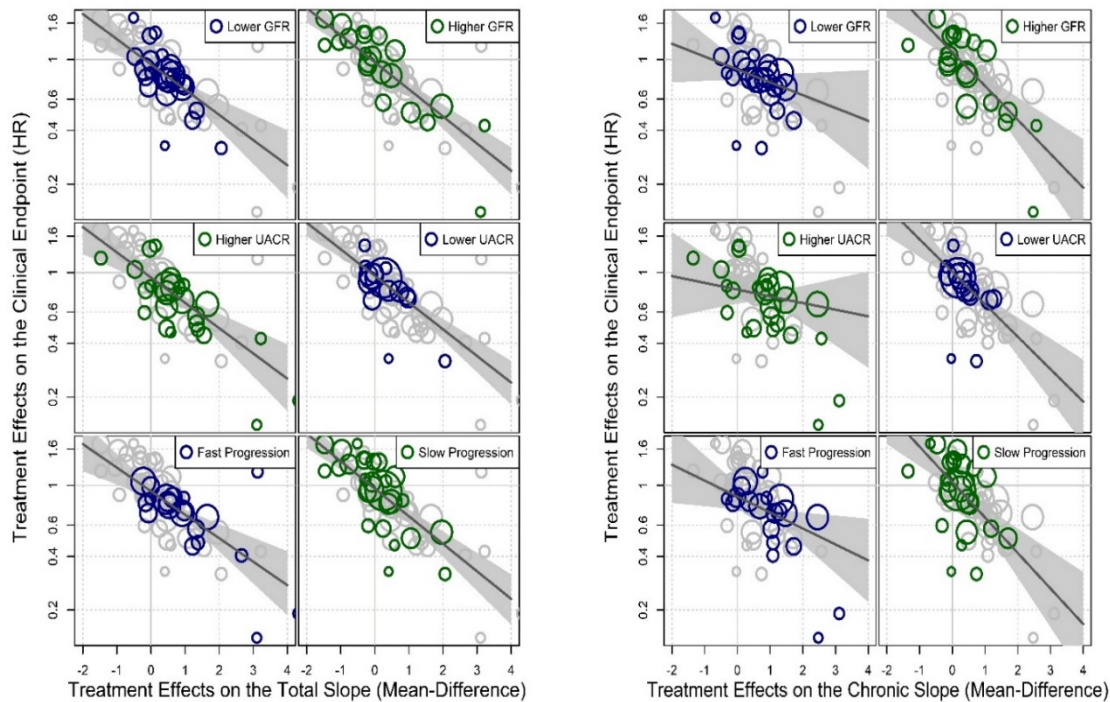
Figure R1: Displayed are trial-level associations between treatment effects on the clinical endpoint and treatment effects on the total slope. In the left-panels, displayed are the meta-regression line and credible bands when mean baseline GFR (top row), mean baseline ACR (middle row), and the mean control arm slope (bottom row) each are fixed at the 25th, 75th, and 25th percentiles for values of those variables observed in the data, respectively (baseline GFR: 40 mL/min per 1.73 m², baseline UACR: 1000 mg/g; control arm slope: -4 mL/min per 1.73 m²/year). In the right panels, displayed are the meta-regression line and credible bands when baseline GFR, baseline ACR, and the control arm slope are fixed at the 75th, 25th, and 75th percentile for values observed in the data, respectively (baseline GFR: 75 mL/min per 1.73 m²; baseline UACR: 71 mg/g; control arm slope: -1.5 mL/min per 1.73 m²/year). The coloured circles indicate trials that fall within a unit of the value used to define disease severity used to plot the relevant regression line. Green indicates less and blue more severe disease. The figure shows consistent performance of the total slope by each of the three factors considered, but steeper relationships between the treatment effects on the chronic slope and treatment effects on the clinical endpoint for higher baseline GFR, higher UACR and lower progression.

FIDELIO and CREDENCE both show benefit on the chronic slope (Analysis #1). The treatment effect of the chronic slope relative to the treatment effects on the clinical endpoint is larger compared to other studies. We indicate this by the ratio between the treatment effects on the chronic slope to the log hazard ratio (**Table R3**). Both are large studies, so that these comparatively large effects stood out in comparison to the random sampling errors in the estimated effects. Thus, in comparison to the other studies in the meta-regression, they appear as outliers (Analysis #2). Note that the strong effects of the treatments on the chronic slope in these two studies does not diminish the evidence for benefit of the treatments on the clinical endpoint.

Table R3: Ratio of Treatment effect on chronic slope compared to treatment effect on clinical endpoint for Fidelio and CREDENCE compared to similar studies										
Inter-vention	Disease	Study	N	eGFR	ACR	Tx effect on CS	Tx effect on TS3	Tx effect on CE (Log HR)	Ratio CS: log HR	Ratio TS: log HR
SGLT-2I	DM	CANVAS	10031	78.7 (18.8)	12 (7, 42)	1.18 (0.08)	0.25 (0.08)	-0.56 (0.23)	-2.11	-0.45
SGLT-2I	DM	CREDENCE	4399	55.9 (16.8)	927 (463, 1833)	2.45 (0.2)	1.65 (0.18)	-0.41 (0.1)	**-5.98**	-4.02
SGLT-2I	CKD-CNS	DAPA-CKD	4041	43.3 (12.4)	900 (500, 1900)	1.47 (0.16)	0.92 (0.15)	-0.36 (0.1)	-4.08	-2.56
SGLT-2I	DM	EMPA-REG	6936	76.2 (19.9)	18 (7, 72)	1.7 (0.12)	1.06 (0.1)	-0.68 (0.16)	-2.50	-1.56
MRA	DM	FIDELIO-DKD	5671	44.3 (12.6)	852 (446, 1634)	1.32 (0.13)	0.45 (0.12)	-0.18 (0.07)	**-7.33**	-2.50
MRA	HF	TOPCAT	3435	65.1 (18.6)	20 (7, 88)	0.15 (0.15)	-0.97 (0.16)	0.45 (0.15)	0.33	0.33
ERA	DM	SONAR	3659	42.5 (14.2)	483 (239, 979)	0.68 (0.19)	0.54 (0.17)	-0.27 (0.11)	-2.52	-2.00
Abbreviations: CE, clinical endpoint; CKD-CNS, other chronic kidney disease (could not specify); DM, diabetes mellitus; HF, heart failure; HR, hazard ratio; MRA, mineralocorticoid receptor antagonist; N, sample size; SGLT-2I, sodium-glucose cotransporter 2 inhibitor; Tx, treatment										

With respect to the question of the simulations vs. these data, we are not aware of a direct connection between the simulation results and the fact that these two trials were flagged as outliers in the trial level analysis of chronic slope. We note that the two concepts, the relative efficiency of the chronic slope versus the validity of the chronic slope as a surrogate endpoint, are distinct. In our simulations, the relative efficiency for the chronic slope pertains to a comparison of the required sample size of a clinical trial which uses the chronic slope as its primary endpoint vs. the required sample size based on the clinical endpoint, irrespective of the validity of the chronic slope as a surrogate. When the active treatment in a trial has a large negative acute effect, the relative efficiency of the chronic slope can appear favourable relative to the clinical endpoint because the negative acute effect, which goes in the direction unfavourable to the treatment, is ignored by the chronic slope. On the other hand, when there is a large negative acute effect, the greater accuracy the total slope compared to the chronic slope for predicting the treatment effect on the clinical endpoint in our trial level analysis suggests that the treatment effect on the chronic slope will tend to overstate clinical benefit since the chronic slope ignores the implications of the negative acute effect for the clinical endpoint.

5. **The simulations used KFRT and death as endpoint in the analysis.. Please discuss the inclusion of death as an endpoint in correlating eGFR slope and clinical endpoints in the epidemiological studies (where KFRT was used as endpoint) and in the trial-level analysis (KF with or without RT and doubling of serum creatinine used as endpoints). Please also discuss the feasibility of a trial level analysis without events based on doubling of serum creatinine, which is usually understood as corresponding to 57% GFR decline. Please provide information on the number of events based on this criterion and perform an analysis without these events when feasible.**

a) *The simulations used KFRT and death as endpoint in the analysis. Please discuss the inclusion of death as an endpoint in correlating eGFR slope and clinical endpoints in the epidemiological studies (where KFRT was used as endpoint) and in the trial-level analysis (KF with or without RT and doubling of serum creatinine used as endpoints*

We want to clarify that while the simulations accounted for the occurrence of death as a competing risk in analyses of other endpoints, the endpoints which we evaluated in the simulations (GFR slope and the clinical endpoint) did not include death.

We did not include death as a component of our composite clinical endpoint because death is not considered as part of the definition of CKD progression. It would not be reasonable to expect that drugs that decrease progression would also necessarily prevent mortality regardless of cause. Some trials may include renal death specifically as a component of a composite kidney endpoint, but renal deaths usually constitute a small fraction of all deaths. Most of the trials included in our database did not include information on renal death specifically.

In the epidemiological association papers, outcomes of KRFT and mortality were considered separately. In that paper, eGFR slope was more strongly associated with KFRT than mortality (Grams *et al.*, J Am Soc Nephrol 2019). We have obtained similar findings in analyses of joint shared parameter models of longitudinal eGFR measurements which treated death and KFRT as competing events, in which KFRT but not death was found to be strongly associated with the underlying eGFR slopes (Vonesh and Greene, Biometrics 2022).

b) ***Please also discuss the feasibility of a trial level analysis without events based on doubling of serum creatinine, which is usually understood as corresponding to 57% GFR decline. Please provide information on the number of events based on this criterion and perform an analysis without these events when feasible.***

We are able to conduct trial level analyses based on a secondary clinical endpoint of the composite of KFRT and GFR < 15, excluding creatinine doubling of serum creatinine (57% decline in GFR). **Table R4** shows the results. Key findings are:

1. When the secondary endpoint of KFRT or GFR < 15 is used as the outcome in place of the primary clinical endpoint, of KFRT/GFR < 15, doubling of SCr, the median trial-level R^2 for the chronic slope increases from 0.55 to 0.73 and decreases for total slope from 0.97 to 0.91. The posterior median root mean square error (RMSE) is reduced for the chronic slope, indicating better model fit, and remains similar for the total slope
   For total slope, we interpret that the modest reduction in the median R^2, with no change in the RMSE, reflects the fact that the estimated variation in the treatment effects across the 66 trials is less for the secondary clinical endpoint of KFRT or GFR < 15 than for the primary clinical endpoints that includes creatinine doubling. Lower variation across the study population tends to reduce the R^2.
   In contrast, the higher median R^2 and lower RMSE for the chronic slope for the secondary clinical endpoint of KFRT or GFR < 15 might reflect the fact that very large acute effects could contribute to a doubling of serum creatinine, a component of the our primary clinical endpoint. The treatment effects on the chronic slope (which ignore the acute effect) will be expected to better predict treatment effects on the secondary clinical endpoint as acute effects are likely to have less impact on this endpoint compared to the primary endpoint.

2. The estimated median intercept is now negative, with the credible intervals around the intercept now do not overlap 0. This suggests that even in the absence of a treatment effect on GFR slope, there is a predicted benefit on the clinical endpoint of

the secondary clinical endpoint of KFRT and GFR < 15. This result will require further investigation.

Table R4: Trial level results using endpoint of dialysis or GFR < 15 ml/min per 1.73 m^2					
	Event	**Meta-Regression Slope (95% BCI)**	**Intercept (95% BCI)**	**R^2 (95% BCI)**	**RMSE (95% BCI)**
Chronic slope	**Primary clinical endpoint**: KFRT/GFR < 15, doubling of SCr	-0.33 (-0.46, -0.20)	0.00 (-0.10, 0.10)	0.55 (0.24, 0.78)	0.19 (0.12, 0.27)
	Secondary clinical endpoint: KFRT/GFR < 15	-0.15 (-0.24, -0.05)	-0.10 (-0.19, -0.03)	0.73 (0.13, 0.98)	0.05 (0.02, 0.13)
Total slope at 3 years	**Primary clinical endpoint**: KFRT/GFR < 15, doubling of SCr	-0.35 (-0.42, -0.29)	-0.04 (-0.09, 0.01)	0.97 (0.83, 1.00)	0.05 (0.02, 0.12)
	Secondary clinical endpoint: KFRT/GFR < 15	-0.22 (-0.31, -0.12)	-0.10 (-0.18, -0.04)	0.91 (0.54, 0.99)	0.05 (0.02, 0.11)
Abbreviations: BCI, Bayesian credible intervals,					

We still consider the current analyses of our primary clinical endpoint of the composite of KFRT, GFR < 15, doubling of SCr as the main analyses for the following reasons

1. The precision, as indicated by the width the confidence intervals, of the trial level analyses of the secondary clinical endpoint is reduced relative to the primary clinical endpoint due to the reduction in the number of events. This is particularly problematic for subgroup analyses.
2. Many of the trials included in our study had fewer KFRT events to substantively contribute to the trial level analysis, so that the trial level analysis for KFRT alone is weighted more to the subset of studies with lower baseline GFR
3. Analyses of KFRT /GFR < 15 may be impacted to a greater extent by the competing risk of death than analyses of the composite clinical endpoint which includes doubling of serum creatinine.
4. The accepted and most commonly used clinical endpoint includes doubling of serum creatinine as a component of a composite kidney outcome in clinical trials. Thus including doubling of serum creatinine in our analyses increases generalizability as very few trials will only include KFRT as endpoint.

Nevertheless, we find this very supporting of our work and adds strength to consideration of chronic slope in the right circumstances.

6. **Please discuss the validity and impact of assumptions for acute effects in the simulation study regarding occurrence in first weeks and resolution until study end on results for the chronic slope analysis.**

The simulations manuscript examined relative efficiency of slope vs. analyses of time-to-event outcomes and a kind of Type 1 error, defined as the probability that a slope-based analysis would indicate a statistically significant benefit when there is no true effect on KFRT, under a range of models for GFR trajectories defined by the input parameters in Table 1 of that paper.

Three of these input parameters concerned the acute effect: 1) the mean acute effect, where we considered values of -2.5, -1.25, 0, +1.25, or +2.5 ml/min per 1.73 m^2 when

calibrated to a baseline GFR=42.5 ml/min per 1.73 m^2; 2) the variability of the acute effect, where we considered standard deviations of 0 or 1 ml/min per 1.73 m^2, and the attenuation of the acute effect as GFR declined. Our simulations considered two attenuation scenarios, one in which the acute effect attenuates fully to 0 when GFR reaches 15 ml/min per 1.73 m^2, and another with no attenuation of the acute effect. For simplicity, the published manuscript focused on the scenarios with full attenuation. The supplement included the alternative scenario where the acute effect does not attenuate (e.g. (Greene *et al.*, J Am Soc Nephrol 2019) Figure 5 and excel spreadsheet 'ASN.2019010009SupplementaryData2.xlsx')

Our simulations indicated little impact of the assumed variability of the acute effect, so the validity of our assumptions concerning the acute effect depends primarily on whether there is reason to question the values of the original input parameters for the mean acute effect and its attenuation. Our recent analyses suggest that the "true acute effects" across the 66 trials have a mean of roughly -0.33 ml/min/1.73m^2, with a standard deviation of approximately 1.56 ml/min/1.73m^2. This suggests most trials continue to have acute effects in the -2.5 and +2.5 ml/min per 1.73 m^2 range considered in our original simulations, indicating no reason to update the range that we originally considered. Our more recent analyses of acute effects across trials (Neuen et al JASN 2022) (Neuen *et al.*, J Am Soc Nephrol 2022) suggest that acute effects may attenuate at lower vs higher GFR levels as we originally hypothesized, but that they may not attenuate all the way to 0 by the time GFR declines to 15 ml/min per 1.73 m^2. We expect that the impact of a partial but not complete attenuation of the acute effect on relative efficiency and the type 1 error considered in the simulations would be subtle.

7. **Please discuss the use of unified mixed effects model in the trial level surrogacy analysis for GFR slope over the defined periods of 2 and 3 years. It is not fully clear if the conclusions on surrogacy would also be applicable for a tailored approach to model selection for future trials. It is not obvious that conclusions on Type 1 error and bias or GFR threshold to infer a beneficial effect on a clinical endpoint would be the same.**

We view the problem of establishing the validity of a surrogate endpoint as providing an answer to the broad-based question: Do treatment effects on the surrogate endpoint reliably predict treatment effects on the clinical endpoint across the class of clinical trials in which the surrogate endpoint might potentially be used? To answer this question across the extensive collection of trials in our analysis, we used a consistent analysis of GFR slope which stipulated that the transition between the acute and chronic phases occurred by 3 months. We used a reproducible algorithm which applied a shared parameter model to account for informative censoring by KFRT and death when the number of ESRD and death events are sufficiently large and simplified mixed models when the numbers of these events are insufficient.

For the more specific second question, i.e., when designing a specific trial, investigators will have the opportunity to tailor their design and analysis plan to reflect the circumstances of that trial. For example, the investigators may have information that indicate the period of the acute phase may differ from 3-months. Investigators may also tailor analytic decisions concerning how to address informative censoring, the correlation structure of longitudinal eGFR measurements, and intercurrent events. If so, improved performance might be obtained using a modified period for the acute effect. Because there is no reason to expect that tailoring the analysis for slope based on the

circumstances of the trial would worsen performance relative to the clinical endpoint, we expect that our results should provide a conservative assessment of the validity of slope-based endpoints that could be achieved under a more tailored (albeit less reproducible) approach.

Sensitivity analyses we have conducted suggest that the results under the simplified model we employed are relatively robust to issues such as designation of the timing of the change point (for details, please see answer to Question 9 and **Figure R2**) and model for handling informative censoring (Vonesh *et al.*, Stat Med 2019).

Questions regarding the CoU

8. **A treatment effect of 0.75ml/min/1.73m^2/year has been proposed as an important treatment effect. Please justify the clinical importance of this difference for different subpopulations of CKD patients, i.e., different baseline GFR and/or rate of progression.**
Our prior work published in JASN 2019 (Grams *et al.*, J Am Soc Nephrol 2019, Inker *et al.*, J Am Soc Nephrol 2019) and AJKD 2020 (Levey *et al.*, Am J Kidney Dis 2020) summarize the strength of evidence to support this change in GFR as clinically meaningful. We have summarized this in **Table R5**.

In the current analyses, we have provided the prediction interval for treatment effects on the clinical endpoint corresponding to a wide range of treatment effects on GFR slope, while noting 0.75 ml/min/1.73m^2/year as a particular example. Under the trial level model for the 66 randomized treatment comparisons, a treatment effect on the 3-year total slope of 0.75 ml/min/1.73m^2/year was associated with an average 18% lower hazard for the clinical end point (95% BCI, 15% to 22%). We believe that a treatment effect of this magnitude for the clinical endpoint would generally be regarded as clinically relevant, and comparable to the hypothesized effect sizes for a number of major randomized trials.

Based on the information provided in response to Question 2 above, this conversion applies across a wide range of levels of baseline GFR or baseline UACR. Importantly, while we cite an effect of 0.75 ml/min/1.73m^2/year as one example to illustrate our results, we do not advocate designating a one size fits all hypothesized effects size for slope analyses. An effect size of 0.75 ml/min/1.73m^2/year corresponds to a relative effect of 25% for a trial with a mean 3 ml/min/1.73m2/year mean decline in the control group, while an even smaller effect size of 0.5 ml/min/1.73 m^2 corresponds to a relative effect of 19% with a 99% positive predictive value based on the 3-year total slope.

Finally, at a recent Scientific Workshop sponsored by the US National Kidney Foundation and the Food and drug Administration, the multi-stakeholder group, including patients, agreed that surrogate endpoints including reduction in slope is clinically meaningful (Inker *et al.*, Am J Kidney Dis 2022).

Table R5: Justification to support GFR slope mean difference of 75 ml/min per 1.72 m3 as clinically meaningful		
Analysis	eGFR slope reduction (ml/min/1.73 m^2 per year mean difference)	Comment
Cohorts (Individual level analyses)– Reported in (Grams *et al.*, J Am Soc Nephrol 2019) and summarized in (Levey *et al.*, Am J Kidney Dis 2020)		
2-year median eGFR slope in median cohort	-0.68 for eGFR <60 stratum; -2.07 for eGFR $\geq$60 stratum	
Relative risk for ESKD	0.75 eGFR slope reduction estimated over 2 years: • HR 0.71 for GFR <60 • HR 0.70 for GFR ≥60)	• Stronger if estimated more than 2 years • Consistent across cohorts and subgroups (baseline eGFR, baseline ACR, diabetes status)
Absolute risk for ESKD	0.75 eGFR slope reduction: 1.6% AR reduction at 5 years for eGFR slope -5 in control group and baseline GFR 75	High risk for lower baseline eGFR

9. **Please discuss if the currently proposed CoU could provide guidance for the design of phase 3 studies using GFR slope as a primary endpoint, including:**
   The decisions for design of any single trial are large and must factor into decisions regarding if and how to use slope. We therefore would not propose including specific requirements for the design of phase 3 studies that use GFR slope in the CoU. We reinforce the concept that decisions regarding the use of total versus chronic slope, as well as length of follow-up, are important and should be done carefully and with consideration of the study population, treatment, and study design in the context of specific drug development program. GFR slope is likely only to be considered as the primary outcome in a particular trial if it provides clear advantages over the clinical endpoint in reducing the required sample size or follow-up duration, or if the regulatory goals for the trial are compatible with the use of a surrogate endpoint. We summarize below some of our current thinking on these issues. We anticipate that this might help EMA to work with sponsors in design of specific studies.

   a. **Minimal data, e.g., from phase 2 studies, required to optimally design a phase 3 study.**
      A successful trial design for a slope endpoint requires the following inputs:
      i) An estimate of the mean progression rate (expressed as the average GFR slope) and a rough estimate of the standard deviation of the chronic slope which may be expected in the control group of the trial. These quantities are required to perform power calculations for slope-based outcomes, and should be independent of the treatment being investigated. This information will generally not be estimable with adequate precision from typical phase 2 trials and instead will be obtained from prior studies with slope-based outcomes. We have included some relevant information in our two previous simulation papers (Greene *et al.*, Am J Kidney Dis 2014, Greene *et al.*, J Am Soc Nephrol 2019).
      ii) A projection for the direction and magnitude of the acute effect. Investigators should have some understanding of the direction and magnitude of the acute effect from previous studies (either phase 2 studies or similar interventions in later phase studies) and from knowledge of the physiology of the intervention.

iii) If an acute effect is expected, an understanding of the approximate timing of the acute effect can be helpful. In many cases, the approximate timing of the acute effect may be known from prior studies investigating the same class of medications or from understanding of the physiologically of the drug. When feasible, further assessment of the timing of the acute effect be obtained in a phase 2 trial before using slope in a major phase 3 trial could be useful.
An example of a design for a phase 2 study that could be used to identify timing of the change point is as follows: eGFR measurements to be done approximately every month for several months after the treatment is administered. We anticipate that with a feasible phase 2 sample size one would still be able to obtain useful information to help identify an approximate change point where the acute phase appears to end, and chronic phase begins, as well as on the size and direction of the acute effect using a simple repeated measures ANCOVA approach.

If the approximate timing is unknown, but the investigators have an understanding of the latest likely time that the acute effect is likely to occur by, the investigators have the option of using a similar model to the model we employed in this work which makes no assumptions of the pattern of GFR changes prior to this time point but assumes a linear mean progression afterwards.

In exploratory analyses, we have found that particularly for the total slope, estimated treatment effects are similar across the trials in our data base irrespective of whether the change point is placed at 3 months, 4 months or 6 months. **Figure R2** compares estimated treatment effects in 48 of our 66 studies based on a 2-slope linear mixed model without the shared parameter component. We have demonstrated similar robustness under the full shared parameter model for selected trials in a technical report we had included in our submission (Vonesh, CKD-EPI Website, 2022).
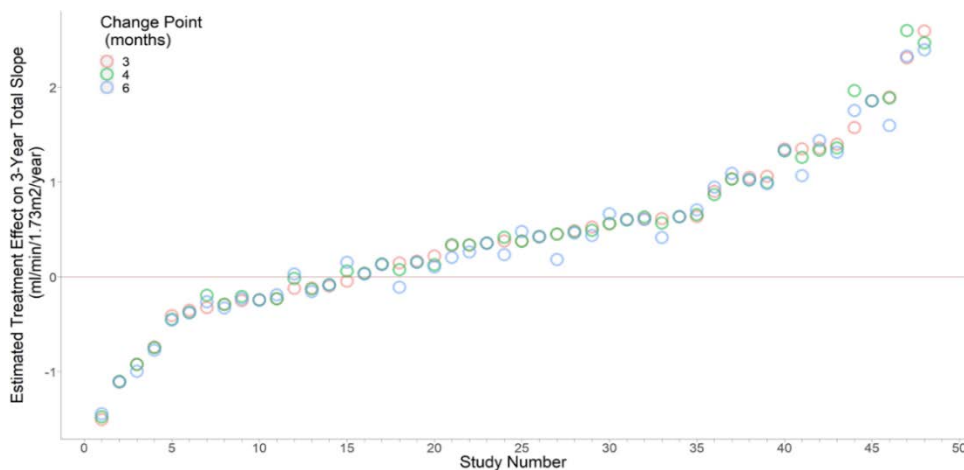
Figure R2: Comparison of estimated treatment effects using different change points to determine the acute effect. Effects are shown across 48 of the 66 studies. For 17 studies were not immediately available for analysis. The remaining study was FONT and evaluated steroids and cyclosporine in a small study of patients with FSGS. The treatment induced very large negative acute effects which attenuated at 6 months

b. Risk assessment for false positive / false negative conclusions, e.g., due to acute effect, based on preliminary phase 2 data.
We are not sure of the exact question of the EMA. We think they are asking our advice on how a sponsor can obtain an estimate of the acute effect from phase 2 trial, and the risks associated with either underestimating of overestimating the size of the acute effect from phase 2 data in designing a Phase 3 trial.

In our view, as long as total slope under a 2-slope model is used, under- or overestimating the acute effect in the phase 2 data should not substantially effect the risk of a false positive conclusion, as the analysis of the phase 3 trial will properly account for the size of the actual acute effect in the trial. The consequences of a mis-estimated acute effect in the design phase will play out in terms of power: If a negative acute effect is over-estimated or a positive acute effect is under-estimated the trial may be overpowered, while if a negative acute effect is underestimated or a positive acute effect is overestimated, the actual power may be lower than intended. Thus the burden of risk would be incurred primarily by the sponsor; from a regulatory perspective, there would be little risk that a mis-estimated acute effect in the design phase would increase the risk of a false positive conclusion.

c. Approach to study design, including (but not limited to) (primary) endpoint definition, study population; length of study; GFR assessment schedule;. assessment of GFR after study completion etc., to minimize the risk of false conclusions.
Study design features will need to be considered on a case-by-case basis. Several considerations are as follows:
1. Selection of endpoints: Slope should generally not be used in settings in which the use of slope would not substantially reduce the required follow-up time and/or the sample size relative to the clinical endpoint. The relative sample sizes and follow-up required by the two approaches should be assessed in the design phase prior to the initiation of the study.
Our previous simulation paper (Greene *et al.*, J Am Soc Nephrol 2019) can provide guidance on determining settings in which use of slope can increase statistical power, and on comparisons of power between the chronic slope and total slope. To avoid risk of a false conclusion of clinical benefit we would generally recommend that the total slope be the primary slope outcome. If a positive acute effect is expected, and there is some concern that the initial GFR increase may not persist, that it may be advisable to consider requiring benefit on both the chronic and total slopes.
2. Measurement schedule: for phase 3 trials we generally recommend two eGFR assessments at baseline, and that eGFR be assessed at least 6-month intervals for at least 2 years of follow-up. One or more additional eGFR measurements within the first 6 months will generally be useful for estimation of the acute effect and clarifying its timing. However, we would not encourage one-size fits all rules; the GFR measurement schedule can be evaluated and modified as per power calculations for each trial, as well as concerns related to participant burden. Additionally, a post-washout GFR measurement could be considered.

d. Relative efficiency and risk of Type 1 error of a GFR slope endpoint compared to clinical event-based endpoints to determine optimal choice of primary endpoints,
See responses to a, b and c above.

---

e. **Guidance on handling post randomisation observations and intercurrent events, as e.g., treatment withdrawal and adverse events**
We address this item in our response to Question 13 below.

f. **Choice of study population to entities in which longer-term outcome trials and use of clinical endpoints are difficult / not feasible to perform.**
As shown in our simulations paper, the primary settings in which slope-based analyses provide substantial savings relative to the clinical endpoint are in trials where the expected proportion of patients with clinical events within that time period is expected to be low, such as in relatively high baseline GFR (e.g., early stage CKD) in which the follow-up time to reach clinical events is prohibitively long. We would also expect slope analyses to provide savings relative to the clinical endpoint in some trials with intermediate or low levels of GFR which do not have a high event rates expected over the trial's follow-up period.

However, there may be appropriate indications for use of GFR slope in other settings. As an illustration, if a kidney outcome trial in one specific area has been done (for example diabetic kidney disease) a slope trial can be done in non-diabetic CKD to shorten timelines and make the drug faster available for non-diabetic CKD.

10. **Please discuss, whether a threshold analysis is needed to assure that in a chronic slope approach a detriment in the experimental arm can be excluded. Please also discuss whether a change from baseline analysis may fit this purpose and the implications on the minimum duration of a clinical trial.**
   a) *Please discuss, whether a threshold analysis is needed to assure that in a chronic slope approach a detriment in the experimental arm can be excluded*

   We may benefit from further clarification of this question with the EMA at our upcoming meeting. For the present, we note that our most recent analyses. Our most recent analyses suggest that the most appropriate threshold for inferring benefit based on the chronic slope should be informed by the size and direction of the acute effect.

   b) Please also discuss whether a change from baseline analysis may fit this purpose and the implications on the minimum duration of a clinical trial.

   We first wanted to acknowledge that the total slope is an analysis of a change from baseline using a predicted GFR at a specified time duration. We suspect the EMA is asking if a simple analysis that compares the mean change in eGFR from baseline to a measurement at the end of the study might be substituted for a slope-based analysis. This is an interesting question. This approach does have the advantage of simplicity compared to a slope-based analysis, but it has several limitations:

   1) The change from baseline approach requires a common end of study measurement time (relative to baseline) for all patients, irrespective of the calendar time different patients are enrolled. This precludes the opportunity to incorporate information from eGFR measurements later in follow-up for patients randomized early in the enrolment period. For most trials, this will constitute a major loss of statistical power. By contrast, a major advantage of slope-based analyses is that they are able to incorporate information from eGFR measurements until a common administrative end date for all patients, thus incorporating longer follow-up for the patients enrolled early in the trial.

2) A naïve change from baseline analysis is subject to bias from patients who drop out of the trial before the end of study measurement. By contrast, slope analyses can incorporate eGFR measurements for such patients up to the time of drop out.
3) A change from baseline analysis sacrifices information from eGFR measurements obtained between the initial baseline and end of follow-up times in the trial, sacrificing statistical power.
4) A naïve change from baseline analysis has no good way of dealing with the "truncation by death problem" (see our response to item 13 for more on that issue). By contrast, the truncation by death problem can be substantially reduced using a slope-based analysis in which a linear slope is assumed in the chronic phase.

This said, we do expect there may be a limited role for change from baseline analyses in certain phase 2 trials with a relatively short follow-up period (e.g., 2 years) and the analysis is conducted under a longitudinal model which incorporates eGFR measurements spaced over the follow-up period. In this strategy the intermediate eGFR measurements are incorporated to aid estimation of the treatment effect on the total eGFR change from baseline to the end of the trial for patients who drop out early, mitigating the second problem above. Such an analysis will have less power than a slope-based analysis, but there may be circumstances where the sacrifice in power is a reasonable trade-off for the gain in simplicity.

Questions regarding the statistical model and issues in future trials:

11. **Please discuss the assumption that a 2 linear slope mixed effects model can adequately describe the GFR trajectories in all target populations and across all disease stages. The impact of e.g., a decline with faster progression in advanced disease stage, close to renal failure or death is unclear.**

The GFR slope model can be applied across a broad range of GFR levels:
1. The shared parameter model used in our analyses includes a term which accounts for heterogeneity in GFR variability at high vs low levels of GFR.
2. The trial level subgroup analyses we have provided, along with the additional analyses provided in Collier *et al* (Collier *et al.*, Clin J Am Soc Nephrol In press) with baseline GFR and baseline UACR (**Figure R1**), indicates that treatment effects on the total 3-year GFR slope accurately predict treatment effects on the clinical endpoint across a wide range of stages and severity of kidney disease.
3. In Vonesh *et al* (Vonesh *et al.*, Stat Med 2019), we demonstrate GFR slopes according to CKD Stage at baseline. As shown in Figure 2, treatment effect appeared similar across different baseline GFR levels across stage.

As we said in question 7 above, we are not suggesting that this model needs to be used for all trials, and rather suggest that the model should be specified for the population being studied. See our response to question 7 below for additional details.

12. **Please discuss the proposed use of a shared parameter mixed effects model and elaborate on clinical interpretability of results from the model. Please discuss model choice in settings with different numbers expected for renal failure or death events with focus on settings with high risk of censoring events, i.e. in target populations with fast progression. The Applicant is invited to present exemplary study results and to provide guidance for interpretation by clinicians. Discussion should also cover alternative model-based approaches.**

This question encompasses several components. We divide our response into three parts, which address 1) the proposed use of our proposed shared parameter mixed effects model, considering settings with differing numbers of KFRT and death events, 2) the clinical interpretation of the model results, and 3) comparisons of alternative models for slope, focussing on a setting with a high rate of KFRT and death events. We use the IDNT trial as an exemplary study for the second and third parts of our response.

Scope of use of the shared parameter model:
The shared parameter model we have proposed has several features designed to support broad use across a wide range of settings with different levels of CKD severity:

a) Applicability to a wide range of trial designs with differing GFR measurement schedules and follow-up durations, including trials whose patients have varying follow-up durations due to staggered entry of patients into the trial and/or censoring events,
b) Accommodates acute effects,
c) Accounts for informative censoring by KFRT and death, with straightforward extensions that can be applied to handle artificial censoring by designated intercurrent events,
d) Accounts for variation in GFR trajectories through variation in intercepts, acute and chronic slopes across patients,
e) Accounts for heterogeneity in variability of individual GFR measurements across different levels of GFR and for heterogeneity in variability of GFR slopes between treatment groups.

It is straightforward to adapt the model to accommodate arbitrary patterns of GFR changes in the early months of follow-up to account for uncertainty in the timing of the acute effect. Extensions can also be used to investigate the severity of informative censoring as an issue in sensitivity analyses (Vonesh *et al.*, Stat Med 2019).
While the model has broad applicability, when a low rate of KFRT and deaths are expected during the trial, as might be the case for a trial focussing on early-stage CKD, it will often be reasonable to apply a simpler mixed effects model that does not explicitly account for these events. This is because the risk of bias from informative censoring is reduced when there are few KFRT or death events and the full shared parameter model may also incur convergence difficulties in this situation. It is straightforward to prespecify and implement an algorithm in which the analysis of the full shared parameter model reverts to a simpler mixed effects model without the shared parameter component when convergence cannot be achieved using the full model, as we have done for each of the 66 studies included in our analysis.

On the other hand, in trials where a very high proportion of patients are expected to reach KFRT during follow-up (e.g., ≥ 40%), the shared parameter remains useful, but in some cases may be better suited as a secondary analysis than as the primary analysis. This is because conceptual difficulties from truncation by death (see Question 13 response) and the sensitivity of the results to possible misspecification of the components of the model related to KFRT can increase in this situation. Such high event rates can occur in trials focussing on a patient population with highly advanced CKD (e.g., baseline GFR < 20 ml/min/1.73m^2) with a substantial rate of GFR decline. Fortunately, the high event rate in such trials will generally support use of the clinical endpoint as the primary outcome with analyses of slope providing a supporting analysis to facilitate interpretation.

IDNT Trial:

We use the IDNT trial to provide a detailed illustration of the clinical interpretation of the shared parameter model and for the comparison of alternative models (Vonesh *et al.*, Stat Med 2019). As requested, this study provides an example with a high rate of KFRT and death events, as the patients progressed quite rapidly (with a mean control group chronic GFR slope of approximately -6 ml/min/1.73m^2/year). Roughly 40% of the patients in the control group reached the clinical endpoint and roughly 30% reached KFRT or death.

Clinical interpretability. Each of the analyses presented in **Figure R3** are based on a 2-slope linear spline model, illustrated in the figure. The left panel provides a schematic of the 2-slope model and the right panel illustrates the fit of 2-slope model to least square mean changes in GFR to designated time points in the Irbesartan and control groups of the IDNT trial. The schematic illustrates following:

a) *The acute slope:* Rate of change in GFR during the first several months after randomization,

b) *The chronic slope:* Rate of change in GFR starting after the first several months after randomization,

c) *The total slope:* The change in GFR from baseline to a designated time towards the end of the follow-up period normalized for time.

The treatment effects on the different slope endpoints are defined as the differences in mean slopes between the randomized treatment groups.

The best fitting variation of the mixed model in Vonesh et al (labelled as Model 4 in that paper) provided estimated treatment effects (and 95% confidence intervals) on the chronic and 3-year total slopes of 0.90 (0.2, 1.57) and 0.55 (-0.11, 1.21) ml/min/1.73m^2/yr, respectively. Thus, on average, the treatment slowed the average rate of progression by 0.90 ml/min/1.73m^2/yr when only the chronic phase of the trial is considered, and by 0.55 ml/min/1.73m^2/yr over 3 years when both the acute and chronic phases are considered. The estimated treatment effect is smaller for the total slope than the chronic slope because the estimated acute effect is negative, in the direction unfavorable to the treatment. Based on the posterior median meta-regression coefficients of our trial-level meta-regression model, the estimated treatment effect on the 3-year total slope translates to an estimated hazard ratio for the clinical endpoint of $\exp(-0.04 - 0.35 \times 0.55) = 0.79$. This is similar to the estimated HR reported in the study's primary results paper for the IDNT's primary composite endpoint of doubling of serum creatinine, KFRT, or death of 0.80. In our own analyses, we obtained a HR of for the composite clinical endpoint (without death) of 0.77.
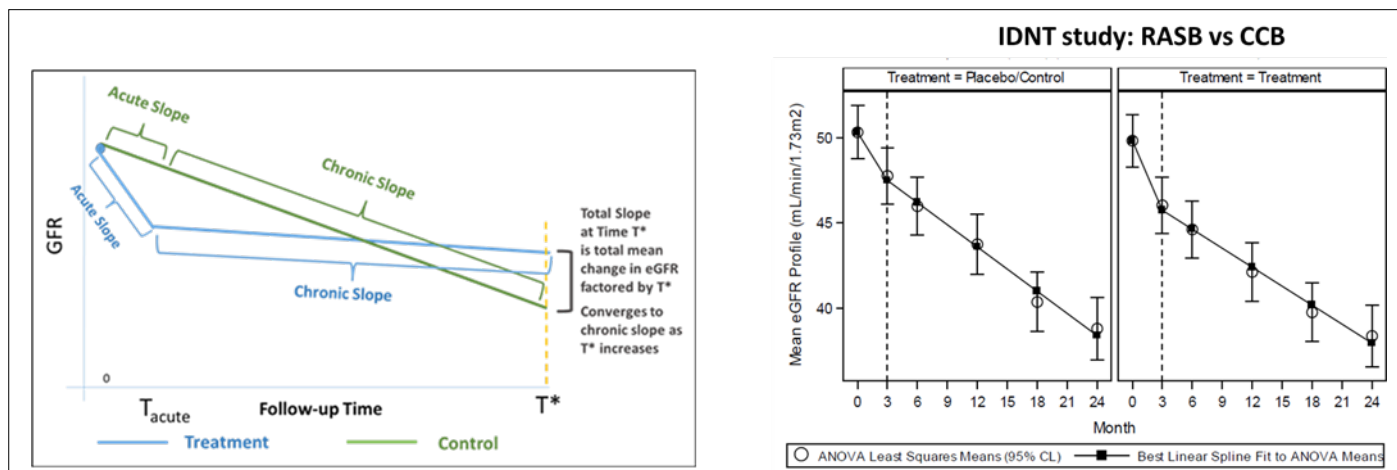
Figure R3: GFR slope by treatment arm. Left panel hypothetical example and right panel demonstration in an example study. eGFR, Estimated glomerular filtration rate; GFR, Glomerular filtration rate; RCT, Randomized controlled trial; T, Time.

Model Comparison (Vonesh *et al.*, Stat Med 2019) consider 4 variations on the 2-slope mixed model:

Model 1: A simplified linear mixed model without the shared parameter component (labelled "ignorable" in the manuscript, since this model assumes ignorable dropout),

Model 2: A shared parameter model in which the hazard function for KFRT or death are assumed to be related to the random intercept and slope coefficient of the slope component of the model,

Model 3: A shared parameter model in which the hazard function for KFRT or death are assumed to be related to the current mean GFR value,

Model 4: A shared parameter model in which the hazard function for KFRT or death are allowed to depend on both the random slopes and intercepts and the current mean GFR value.

Table 5 of Vonesh et al summarizes the maximum likelihood estimates for all the parameters of each model and provides model fit measures to compare the models. As shown in this table, the model includes a considerable number of parameters which characterize the joint distribution of the intercepts, acute and chronic slopes, the variation of individual GFR measurements as a function of the GFR level, the difference in GFR slope variance between treatment groups, and the relationship between KFRT and death with the GFR trajectories. All of the parameters shown are of potential interest for understanding the patterns of the GFR trajectories and their relationships with KFRT or death, but the focus for evaluating treatment benefit from a clinical perspective can be limited to the comparisons of the mean chronic and total slopes, presented in the lower part of the top panel of Table 6 in the paper and discussed above under Clinical Interpretability. The estimated treatment effects on the chronic and total slopes are generally similar across the 4 models, with the optimal fit provided by model 4 as noted above. We would expect even smaller variation between alternative shared

parameter models in CKD trials in which the KFRT event rate is lower than in the rapidly progressing IDNT patient population. Subsequent material in the article addresses additional diagnostics addressing model adequacy, including comparisons of robust and model-based standard errors.

13. **Please discuss the definition of an appropriate estimand in future confirmatory trials. Approaches to handle intercurrent events and missing data due to study drop-out should consider acute effects and their direction. Please elaborate on using a treatment policy strategy in settings with acute effects of an intervention.**

We agree that appropriate handling of intercurrent events and missing data can be critical for GFR slope analyses. Of note, missing data and early termination of follow-up due to competing clinical events is also a concern for time to event analyses. The standard accommodation is to censor for these events and assume that the censoring events are non-informative, but just as in analyses of slope this assumption can be problematic. Nevertheless, it is important to consider this carefully for GFR slope analyses.

Termination of GFR follow-up by KFRT and death are already incorporated in our proposed analyses under the shared parameter model. The study protocol should also include a list of additional prespecified intercurrent events which are expected to substantially compromise the interpretation of subsequent GFR values. These might include prohibited concomitant medications with very large acute effects and discontinuation of study medications with large acute effects. Acute effects to be considered would include both those that effect GFR hemodynamic and the filtration marker used to estimate GFR.

Two types of estimands can be considered to address such medication-related intercurrent events; 1) so-called intent-to-treat estimands in which all follow up GFR measurements are incorporated up to the end of the study, KFRT, or death, including those that occur after medication-related intercurrent events, and 2) on-treatment estimands, in which GFR measurements following medication-related intercurrent events are excluded.

Our own analyses have followed the intent-to-treat approach; thus the strong agreement we have observed between treatment effects on the total slope and treatment effects on the clinical endpoint was achieved in spite of any intercurrent events occurring in the trials contributing to our analysis. Nonetheless, on-treatment estimands may also be appropriate for the primary analysis in certain cases. The same approach we used in our recent publication in Biometrics can address informative censoring due to medication-related intercurrent events distinct from informative censoring due to KFRT and death (Vonesh and Greene, Biometrics 2022). In general, whichever approach (ITT or on-treatment) is used for the primary analysis, it will be useful to consider the other in a sensitivity analysis.

Another potential intercurrent event in CKD studies is acute kidney injury or disease. In some trials, GFR measurements during designated acute kidney injury episodes are excluded when defining the time to clinical endpoints. In principle, the same practice could be used for slope-based analyses in those trials.
We provide a more detailed statistical description of the intent-to-treat and on-treatment estimands below:

To facilitate discussion, we focus on the mean 3-year total slope for a trial with a 3-month acute phase, defined mathematically as:

$$(3/36)\times \theta_{acute} + (33/36)\times\theta_{chronic} \qquad\qquad \text{(EQ 1)}$$

where θ_{acute} represents the mean acute slope and $\theta_{chronic}$ represents the mean chronic slope, each expressed in ml/min/1.73m^2/year. This expression can be used to define two different conceptually coherent estimands, depending on how the chronic slope is interpreted for patients with medication-related intercurrent events during follow-up:

1) An intent-to-treat estimand, where $\theta_{chronic}=\left(\frac{1}{N}\right)\Sigma_i\theta_i$ and each θ_i is defined as the chronic slope for the ith patient starting at month 3 and continuing until the end of follow-up, KFRT, or death, whichever comes first, and N is the number of patients in the treatment group in which the mean chronic slope is being defined.

2) An on-treatment estimand, where again $\theta_{chronic}=\left(\frac{1}{N}\right)\Sigma_i\theta_i$ but now each θ_i is defined as the mean chronic slope starting at month 3 and continuing until the end of follow-up, KFRT, death or medication-related intercurrent event, whichever comes first.

These definitions are designed to work around philosophical challenges posed by the so-called truncation by death problem. The issue is that the interpretation of the linear combination in EQ 1 literally as a mean "3-year total slope" is clouded if a substantial portion of patients reach KFRT or death before year 3. This is because the mean 3-year total slope averages over "virtual GFRs" for patients who die or reach KFRT prior to 3-years. However, the algebraic interpretation of the estimand in EQ 1 as the linear combination of the acute and chronic slopes remains valid as long as patients remain past month 3, so that the chronic slope and hence the algebraic linear combination defined in EQ 1 are well defined. With this interpretation, our trial level analyses show that the algebraic linear combination of the mean acute and chronic slopes given by EQ 1 accurately predicts the treatment effect on the clinical endpoint.

A small portion of the broader truncation by death problem remains even with this algebraic interpretation, as the definition of the mean chronic slope includes θ_i in the average slope even for patients who experience KFRT or death prior to month 3. These slopes are defined mathematically under our proposed mixed effects models but are virtual constructs for these patients. However, the fraction of patients who die or reach KFRT prior to month 3 is usually less than 1% in CKD trials, which we believe is small enough that the philosophical challenges posed by truncation by death can be safely deemphasized for practical purposes.

14. **Please discuss using non-linear mixed effects model software for analysis for a confirmatory Phase 3 trial. The discussion should cover the need for pre-specification and potential convergence problems. The Applicant should also comment on the use of kappa values in the analysis models to account for heterogeneity in the treatment arms and parameters to account for heterogeneity of (baseline) GFR values in the trial level analysis in context of a pre-specified analysis.**
a) Please discuss using non-linear mixed effects model software for analysis for a confirmatory Phase 3 trial. The discussion should cover the need for pre-specification and potential convergence problems.

We have provided template SAS code using the PROC NLMIXED package which can be used to fit the shared parameter model in clinical trials with our Statistics in Medicine paper (Vonesh et al 2019) (Vonesh *et al.*, Stat Med 2019), and are committed to

supporting research teams in the application of this or similar code as needed for its use in clinical trials. During the design phase of a trial, it will be important to assess the rate of KFRT and death events and make a decision as to whether the primary analysis should be based on time to the clinical endpoint (or similar composite endpoint defined by clinical events), analysis of GFR slope under a shared parameter model that accounts for these informative censoring events of death or KFRT, or analysis of GFR slope using a simpler mixed effects model without explicitly accounting for KFRT and death as outcomes in a joint model. It is relatively straightforward to prespecify and implement an algorithm in which the analysis reverts to a simpler model when convergence cannot be achieved using the more complex model, as we have done for each of the 66 studies included in our analysis.

b) The Applicant should also comment on the use of kappa values in the analysis models to account for heterogeneity in the treatment arms and parameters to account for heterogeneity of (baseline) GFR values in the trial level analysis in context of a pre-specified analysis.

For current applications, we include kappa, which defines the ratio of slope standard variances between the treatment groups, as a parameter in the shared parameter model in order to improve statistical inference by accounting for a possible difference in slope variance in the intervention group compared to the control group. Similar to the parameters that define the relationship between slope and KFRT and death events, we do not view kappa as the primary target of inference, although it may be of secondary interests for explanatory purposes.

Summary and proposed revised context of use

In summary, we provide strong evidence that GFR slope can be viewed as a validated surrogate endpoint for CKD progression in clinical trials for standard marketing authorization and indication extension approvals. Our results support use of total slope across a broad range of settings. The duration of time over which the total GFR slope can be computed is dependent upon the presence and magnitude of an acute effect. The chronic slope may also be satisfactory in some situations. Decisions regarding the use of total versus chronic slope, as well as length of follow-up, are important and should be done carefully and with consideration of the study population, treatment, and study design in the context of specific drug development program.

Based on the set of issues and questions from EMA, we propose a revised context of use as follow:

- **General setting**: The proposed novel method, GFR slope, is intended to be used as a validated surrogate endpoint for CKD progression in clinical trials for standard marketing authorization and indication extension approvals.
- **Target population**: Broad population of patients with CKD or at risk for progression of CKD, including early and late disease and across cause of CKD.

References:

Collier, W. H., Inker, L. A., Haaland, B., *et al.* (In press). "Evaluation of Variation in the Performance of GFR Slope as a Surrogate End Point for Kidney Failure in Clinical Trials That Differ by Severity of Chronic Kidney Disease." Clin J Am Soc Nephrol.

Grams, M. E., Sang, Y., Ballew, S. H., *et al.* (2019). "Evaluating Glomerular Filtration Rate Slope as a Surrogate End Point for ESKD in Clinical Trials: An Individual Participant Meta-Analysis of Observational Data." J Am Soc Nephrol **30**(9): 1746-1755.

Greene, T., Teng, C. C., Inker, L. A., *et al.* (2014). "Utility and validity of estimated GFR-based surrogate time-to-event end points in CKD: a simulation study." Am J Kidney Dis **64**(6): 867-879.

Greene, T., Ying, J., Vonesh, E. F., *et al.* (2019). "Performance of GFR Slope as a Surrogate End Point for Kidney Disease Progression in Clinical Trials: A Statistical Simulation." J Am Soc Nephrol **30**(9): 1756-1769.

Inker, L. A., Grams, M. E., Guethmundsdottir, H., *et al.* (2022). "Clinical Trial Considerations in Developing Treatments for Early Stages of Common, Chronic Kidney Diseases: A Scientific Workshop Cosponsored by the National Kidney Foundation and the US Food and Drug Administration." Am J Kidney Dis **80**(4): 513-526.

Inker, L. A., Heerspink, H. J. L., Tighiouart, H., *et al.* (2019). "GFR Slope as a Surrogate End Point for Kidney Disease Progression in Clinical Trials: A Meta-Analysis of Treatment Effects of Randomized Controlled Trials." J Am Soc Nephrol **30**(9): 1735-1745.

Levey, A. S., Gansevoort, R. T., Coresh, J., *et al.* (2020). "Change in Albuminuria and GFR as End Points for Clinical Trials in Early Stages of CKD: A Scientific Workshop Sponsored by the National Kidney Foundation in Collaboration With the US Food and Drug Administration and European Medicines Agency." Am J Kidney Dis **75**(1): 84-104.

Neuen, B. L., Tighiouart, H., Heerspink, H. J. L., *et al.* (2022). "Acute Treatment Effects on GFR in Randomized Clinical Trials of Kidney Disease Progression." J Am Soc Nephrol **33**(2): 291-303.

Vonesh, E., Tighiouart, H., Ying, J., *et al.* (2019). "Mixed-effects models for slope-based endpoints in clinical trials of chronic kidney disease." Stat Med **38**(22): 4218-4239.

Vonesh, E. F. (2022). "CKD Consortium Technical Report 2022: Timing of acute treatment effects in clinical trials of chronic kidney disease." CKD-EPI Website, , from https://www.tuftsmedicalcenter.org/Research-Clinical-Trials/Institutes-Centers-Labs/Chronic-Kidney-Disease-Epidemiology-Collaboration/Surrogate-endpoints.

Vonesh, E. F. and Greene, T. (2022). "Biased estimation with shared parameter models in the presence of competing dropout mechanisms." Biometrics **78**(1): 399-406.