A multicentre international randomized parallel group double-blind placebo-controlled clinical trial of EMPAgliflozin once daily to assess cardio-renal outcomes in patients with chronic KIDNEY disease

EMPA-KIDNEY

Data Analysis Plan (EDMS #6290)

(Standard Operating Procedure 11)

Version History

Version	Description	
1.0	First released version	Created by: Natalie Staplin, Will Herrington, Richard Haynes, Jonathan Emberson, Colin Baigent (30MAY2019) Reviewed by Boehringer Ingelheim (sponsor) (JUN2019) Review and agreement by Steering Committee (23OCT2019) Released for use by Will Herrington (01NOV2019)
1.1	Inclusion of Appendix II, subgroup analyses by kidney disease cause, KDIGO risk & analyses using different % declines in eGFR. Clarifications on handling of missing values in subgroup analyses and definitions of tertiles.	Proposed to Steering Committee (30MAR2021) Updated by Will Herrington, Richard Haynes (21JUN2021) Reviewed by Boehringer Ingelheim (sponsor) (JUL2021) and reminder of key changes provided to Steering Committee (26OCT2021) Released for use by Will Herrington (12NOV2021)
1.2	Clarify planned presentations in the primary results publication, including presenting all components of primary outcome as tertiary assessments, emphasis on the overall population for safety assessments, & typographical corrections.	Updated by Will Herrington, Richard Haynes (20MAY2022) Provision to Boehringer Ingelheim (sponsor) & Steering Committee (24MAY2022) Released for use by Will Herrington (14JUN2022)

This is a controlled document. Distribution and approval is to be managed using the Electronic Document Management System.

Contents

1 Introduction

The purpose of this Data Analysis Plan is to define, before unblinding of the treatment allocation, all randomized analyses to be presented in the primary publication of the EMPA-KIDNEY trial results. Exploratory assessments are also listed, but will generally not be presented in the primary publication. Additional pre-specified analyses and summaries required for regulatory submission will be detailed in separate analysis plans. The nature of further post-hoc analyses including those related to subsequent publications cannot be specified in detail but, where appropriate, a general analytical approach is set out. The current latest version of SAS® will be used for analyses. Primary and secondary outcomes have been defined in SOP9a (Approach to recording of outcomes, adjudication and primary/secondary outcome definitions; EDMS #5452) and will be adjudicated using SOP9b (Adjudication procedures; EDMS #6062).

2 Baseline characteristics

In order to assess balance of baseline characteristics between randomized arms, the following variables recorded at randomization (or at Screening) will be presented for each of the empagliflozin and placebo groups:

- Age
- Sex
- Region (Europe, North America, China and Malaysia, Japan)
- Race (all regions: White, Black, Asian, Mixed, Other)
- History of prior disease (presence vs. absence): diabetes mellitus* (overall and by type), cardiovascular disease, heart failure, and peripheral arterial disease
- Blood pressure (systolic and diastolic separately)
- Body mass index
- Laboratory values at Randomization:
    - CKD-EPI estimated glomerular filtration rate (eGFR) (mL/min/1.73m^2): as a continuous variable and in the following categories <30, ≥30<45, ≥45 estimated from central enzymatic creatinine (or local creatinine where central value unavailable)
    - Urinary albumin:creatinine ratio (mg/g): as a continuous variable and in the following categories <30, ≥30≤300, >300, or missing (based on central measurement)
    - Glycosylated haemoglobin (HbA1c) (mmol/mol): as a continuous variable and in the following categories <39 (normoglycaemia), ≥39<48 (pre-diabetes), ≥48<75 (well-controlled diabetes), ≥75 (poor glycaemic control), or missing
    - N-terminus prohormone of brain natriuretic peptide (NT-proBNP) as a continuous variable
    - Haematocrit as a continuous variable
- Medication (yes vs. no): renin-angiotensin system (RAS) inhibitors, beta-blockers, diuretics (overall, and loop, thiazide, mineralocorticoid receptor antagonists, and other considered separately), lipid lowering medication, antiplatelet therapy, anticoagulants, insulin,

sulphonylureas, biguanides, GLP-1 receptor agonists, DPP-4 inhibitors, and other antidiabetic drugs

- KDIGO risk category: low, moderate or high (3 categories combined) versus very high
- Cause of kidney disease: diabetic, glomerular, hypertensive/renovascular, other and unknown

* Diabetes at randomization is defined as participant-reported history of diabetes of any type, a diabetes-related adverse event, use of glucose-lowering medication, or baseline HbA1c ≥48 mmol/mol at Randomization visit.

For continuous variables, mean (standard deviation) will be presented unless the variable has a skewed distribution, in which case median (interquartile range) will be used. For all categorical variables, the number and percentage of participants in the category will be presented. All possible categories will be displayed, zero-filled where necessary, the category 'missing' will only be displayed (e.g. in footnotes) if there are actually missing values. Extra detail on baseline characteristics was provided in baseline paper[1] and may be included in other publications (see Section 5.4 for details of how missing values are handled in subgroup analyses).

3 Definitions for efficacy and safety endpoints

Unless otherwise specified, all analyses will involve an intention-to-treat comparison among all randomized participants of the effects of allocation to empagliflozin versus placebo during the scheduled treatment period (i.e. all participants will be included irrespective of whether they take none, some or all of their allocated treatment). The event-free survival time will be calculated as the time from the date of randomization to either the date of the first occurrence of the event of interest or the censoring date for those who do not have such an event (see section 5.5 for censoring rules).

In accordance with the protocol, all deaths and events initially reported as hospitalization for heart failure, myocardial infarction, stroke, liver injury, ketoacidosis, lower limb amputation, and acute kidney injury will be subject to adjudication as set out in SOP 9b: Adjudication procedures (EDMS #6062). For those events that are subject to adjudication, analyses will include all confirmed and unrefuted events.

3.1 Estimands

For the efficacy and safety endpoints (with the exception of all cause hospitalizations and annual rate of change in eGFR), the estimand of interest will be the hazard ratio of the first occurrence of the endpoint in the target population for participants allocated empagliflozin relative to those allocated placebo, conditional on the baseline covariates specified for the minimized randomization algorithm, ignoring any non-fatal intercurrent events and in the hypothetical absence of death from any cause not included in the endpoint (see section 5.1.1 for analysis methods).

For all-cause hospitalizations the estimand of interest will be the hazard ratio of all occurrences of the endpoint in the target population for participants allocated empagliflozin relative to those allocated placebo, conditional on the baseline covariates specified for the minimized randomization algorithm, ignoring any non-hospitalized non-fatal intercurrent events and in the hypothetical absence of death from any cause.

For annual rate of change in eGFR the estimand of interest will be the difference in the mean annual rate of change in eGFR in the target population allocated empagliflozin compared to the target population allocated placebo, conditional on the baseline covariates specified for the minimized randomization algorithm, ignoring any non-fatal intercurrent events (except end-stage kidney disease [ESKD], defined as the initiation of maintenance dialysis or receipt of a kidney transplant) and in the hypothetical absence of ESKD or death from any cause.

3.2 Hypotheses

For all statistical tests (other than tests for heterogeneity or trend), the null hypothesis will be that the effect of allocation to empagliflozin on the endpoint of interest in the target population is the same as the effect of allocation to placebo (and hence the alternative hypothesis will be that the effect of allocation to empagliflozin is not the same as the effect of allocation to placebo).

3.3 Primary assessments

The primary endpoint will be the time to the first occurrence of the composite outcome of:

(i) Kidney disease progression (defined as ESKD, a sustained decline in eGFR to <10 mL/min/1.73m^2, renal death, or a sustained decline of ≥40% in eGFR from randomization); or

(ii) Cardiovascular death.

The term 'sustained' will be taken to mean that it is either:

(a) Measured at two consecutive scheduled study Follow-up Visits (at least 30 days apart); or

(b) Measured at the last scheduled study Follow-up Visit or the last scheduled visit before death (or withdrawal of consent or loss to follow-up).

If the conditions for a sustained decline in eGFR are met, the date of the event will be the date of the earlier of the two eGFR measurements.

The analysis method for the primary assessment is described in section 5.1.1.

3.4 Secondary efficacy assessments

The key secondary efficacy outcomes are:

(i) Time to first occurrence of hospitalization for heart failure or cardiovascular death;

(ii) Time to occurrences of all-cause hospitalisations (first and recurrent combined); and

(iii) Time to death from any cause.

The other secondary efficacy outcomes are:

(iv) Time to first occurrence of kidney disease progression;

(v) Time to cardiovascular death; and

(vi) Time to first occurrence of cardiovascular death or ESKD.

The analysis method for the secondary efficacy assessments is described in section 5.1.1, with the exception of the key secondary outcome of recurrent all cause hospitalizations which is given in section 5.1.2.

3.5 Tertiary efficacy assessments

The tertiary efficacy assessments are:

(i) Time to components of kidney disease progression defined as:

(a) Time to first occurrence of a composite of ESKD, a sustained decline in eGFR to <10 mL/min/1.73m^2, or renal death;

(b) Time to first occurrence of: (i) ESKD, (ii) a sustained decline in eGFR to <10 mL/min/1.73m^2, and (iii) a sustained decline of ≥40% in eGFR from randomization (each considered separately);

(ii) Annual rate of change in eGFR, calculated separately:

(a) For the period from baseline to the final follow-up visit (i.e. “total slope”);

(b) For the period from 2 months to the final follow-up visit (i.e. “chronic slope”);

(iii) Time to first occurrence of ESKD or death from any cause combined;

(iv) Time to first occurrence of kidney disease progression or death from any cause combined;

(v) Time to death from particular categories of causes: including cardiovascular (coronary death, other cardiac [including heart failure and sudden cardiac death not known to be coronary], stroke, other cardiovascular and presumed cardiovascular [presented in combination and separately]) and non-cardiovascular (renal, infection, cancer, other medical, and non-medical [presented in combination and separately]);

(vi) Time to first occurrence of a major cardiovascular event (defined as the composite of cardiovascular death, myocardial infarction, stroke or hospitalization for heart failure);

(vii) Time to new-onset diabetes mellitus (defined as clinical diagnosis, commencement of glucose-lowering treatment, or HbA1c ≥48 mmol/mol measured by central laboratory on at least one occasion) among participants without diabetes at randomization*, overall and separately among those with normoglycaemia or “pre-diabetes” (defined as HbA1c <39 mmol/mol [normoglycaemia] and ≥39 to <48 mmol/mol [pre-diabetes], respectively);

(viii) Time to first self-reported episode of gout;

(ix) Subgroup analyses of the primary composite outcome (see section 3.5.1 below for details).

* Diabetes at randomization is defined as participant-reported history of diabetes of any type, use of glucose-lowering medication or baseline HbA1c ≥48 mmol/mol at Randomization visit.

The analysis method for the tertiary efficacy assessments is described in section 5.1.1, with the exception of annual rate of change in eGFR which is described in section 5.1.3.

3.5.1 Subgroup analyses

Subgroup analyses are planned for the primary composite outcome. Exploratory subgroup analyses are also pre-specified in section 3.8:

Subgroup subcategories are based on randomization values of:

a. History of prior disease:
   i. Diabetes mellitus (presence vs absence);
   ii. Cardiovascular disease (presence vs absence);
   iii. Heart failure (presence vs absence);
   iv. Peripheral arterial disease (presence vs absence);
   v. Primary renal diagnosis (4 groups: diabetic; glomerular; hypertensive/renovascular; other and unknown combined);

b. Patient characteristics;
   i. Age (split by tertiles of baseline distribution);
   ii. Sex (male vs female);
   iii. Region (Europe, North America, China and Malaysia, Japan);
   iv. Blood pressure (split by tertiles of baseline distribution [systolic & diastolic]);
   v. Body mass index (split by tertiles of baseline distribution);

c. Laboratory values at Randomization (as defined in section 2):
   i. HbA1c (split by tertiles of baseline distribution);
   ii. eGFR (<30, ≥30<45, ≥45 mL/min/1.73m^2);
   iii. Urinary albumin:creatinine ratio (<30, ≥30≤300, >300 mg/g);
   iv. NT-proBNP (split by tertiles of baseline distribution);
   v. Haematocrit (split by tertiles of baseline distribution);
   vi. KDIGO risk category (low, moderate and high [3 categories combined] versus very high)

d. Medication use at randomization:
   i. RAS-inhibition (yes vs no);
   ii. Beta-blocker (yes vs no);
   iii. Diuretics (yes vs no).

The analysis method for these subgroup analyses is detailed in section 5.3. The subgroup analyses of the primary composite outcome which are of key interest are those involving subdivision by: (a) baseline diabetes status, (b) baseline eGFR, and (c) urinary albumin:creatinine ratio. These results will be prioritised for presentation in the primary publication, with other subgroups considered supplementary. The exact cutpoints for age, HbA1c, blood pressure, body mass index, NT-proBNP and haematocrit at randomization are based on approximate tertiles; these were specified prior to unblinding and were

presented in the baseline paper[1]. Other exploratory subgroups may also be considered in later publications.

3.6 Safety assessments

All adverse events are coded using MedDRA version 20.1. The safety outcomes are:

(i) Time to first occurrence of an SAE due to:
- (a) Urinary tract infection, overall† and separately by sex;
- (b) Genital infection, overall† and separately by sex;
- (c) Hyperkalaemia;
- (d) Acute kidney injury;
- (e) Dehydration;

(ii) Time to first occurrence of an AE of Special Interest (AESI):
- (a) Liver injury, both overall† and separately by cause (defined as ALT or AST ≥5x Upper Limit of Normal [ULN] or the combination of ALT or AST ≥3x ULN with bilirubin ≥2x ULN; measured in the same blood sample);
- (b) Ketoacidosis, both overall† and separately by baseline diabetes status;
- (c) Lower limb amputations (overall† and by level [i.e. toe, forefoot, foot, below knee or above knee]);

(iii) Time to first occurrence of another AE relevant to the study question:
- (a) Bone fracture, both overall† and separately by site (long bones versus non-long bones) and aetiology (i.e. distinguishing those resulting from high and low impact trauma, and other causes);
- (b) Severe hypoglycaemia (defined as low blood sugar causing severe cognitive impairment which requires assistance from another person for recovery);
- (c) Symptomatic dehydration (defined as whether or not a participant has experienced symptoms they attribute to dehydration, such as feeling faint or fainting);

(iv) Time to first occurrence of hospitalization by specific causes (events are to be categorised according to their MedDRA Primary System Organ Class (SOC));

(v) Time to first occurrence of SAEs both overall† and, separately, by Primary SOC category;

(vi) Discontinuation of study treatment overall† and by various causes (including SAEs, non-serious adverse events, and other reasons [using Primary SOC categories for SAEs and non-serious adverse events]);

(vii) Changes in weight and systolic and diastolic blood pressure from baseline.

† Presentation and interpretation of results will emphasize safety assessments overall.

All adverse events and hospitalization endpoints will be analysed as described in section 5.1.1, discontinuation of study treatment will use the methods in section 5.1.5 and changes in weight and blood pressure will be analysed using methods outlined in section 5.1.4.

3.7 Laboratory assessments

Biochemical efficacy assessments using central laboratory measurements will be:

- Difference in mean urinary albumin:creatinine ratio calculated as a weighted average over all post-randomization time points (urine collected at the Randomization visit, 2 months, 18 months and the Final-follow-up visit)
- Difference in mean HbA1c calculated as a weighted average over all post-randomization time points, overall and by those with and without diabetes separately (blood collected at the Randomization visit, 2 months, 18 months and the Final-follow-up visit).

For all participants, biochemical safety assessments using the local laboratory results will be:

- Difference in mean potassium calculated as a weighted average over all post-randomization time points (measured at each scheduled Follow-up Visit)
- Elevations in ALT/AST in various categories (ALT or AST ≥5x ULN, ALT or AST ≥3x ULN with bilirubin ≥2x ULN in the same blood sample measured at each scheduled Follow-up Visit).

In the subset of UK participants, biochemical safety assessments using local laboratory values collected at 18 months of follow-up will also be conducted on:

- Sodium
- Corrected calcium
- Phosphate
- Haematocrit
- Haemoglobin.

See sections 5.1.4 and 5.1.5 for details of how these biochemical markers will be analysed.

3.8 Exploratory assessments

Exploratory assessments planned for other publications include (but are not limited to) assessments of the effect of:

a) Allocation to empagliflozin versus placebo on subgroups (using the categories set out in section 3.5.1) for the following outcomes:
   i. Time to kidney disease progression;
   ii. Time to cardiovascular death or hospitalization for heart failure; and
   iii. Annual rate of change of eGFR.

b) Allocation to empagliflozin versus placebo on:
   i. Mean eGFR at each scheduled visit;
   ii. Subcategories of SAEs by MedDRA High Level Group Terms (HLGTs); and
   iii. The primary outcome by year since Randomization;

c) Effect of stopping study treatment on:

i. Mean eGFR (using the surviving UK participants with a 4-week post-Final Follow-up blood draw); and
ii. Urinary albumin:creatinine ratio (using the surviving UK participants with a measurement 4 weeks after Final Follow-up).

d) Allocation of empagliflozin versus placebo on time to the first occurrence of Kidney disease progression or Cardiovascular death (overall and for kidney disease progression alone), where kidney disease progression utilises the following alternative thresholds:
   i. sustained ≥50% decline in eGFR
   ii. sustained ≥57% decline in eGFR (i.e. approximately consistent with a doubling of creatinine); and
e) Allocation of empagliflozin versus placebo on time to occurrences of hospitalizations for heart failure (first and recurrent combined)

The analysis methods to be used for these exploratory assessments are described in section 5.1.6. Additional exploratory assessments may be conducted after unblinding if it is deemed appropriate.

4 Analysis sets for efficacy and safety analyses

For all outcomes, the primary analysis will compare the outcome from randomization to the end of the scheduled treatment period (see section 5.5) among all those participants who are allocated at randomization to receive empagliflozin versus all those allocated to receive matching placebo (i.e. "intention-to-treat" analyses).[2-4]

For the AESIs (see section 3.6), additional analyses will subsequently be performed to compare the effects of empagliflozin among only those participants who are recorded as being adherent to study treatment at the time of the event (or the preceding follow-up visit) compared to those who are not. However, the emphasis in interpretation will be on the intention to treat analyses. Unless otherwise indicated, all analyses will be based on the first occurrence of the specified outcome.

No further on-treatment analyses for either efficacy or safety will be presented in the primary publication, but may be conducted as part of the regulatory submission process (these will be pre-specified in a separate analysis plan).

5 Statistical methodology for efficacy and safety analyses

5.1 Methods of analysis

5.1.1 Analyses of time to first event

Cox proportional hazards regression adjusted for the variables specified in the protocol for the minimization algorithm (age, sex, prior diabetes, eGFR, urinary albumin:creatinine ratio, and region

(footnote[1]) will be used to estimate the hazard ratio associated with allocation to empagliflozin versus placebo (with the Wald chi-square statistic used to both test significance and generate an asymptotic 95% confidence interval).[5] Any ties will be handled using Breslow's method. Kaplan-Meier estimates for the time to each of the primary and secondary outcomes will also be calculated. If any regression models fail to converge, the hazard ratio and its confidence interval will instead be estimated from a Cox model adjusted only for treatment allocation. For any outcome where there are insufficient numbers of events to reliably estimate a hazard ratio (e.g. fewer than 5 participants with the event), Fisher's exact test will be used to compare the number of participants affected in each arm. A participant may contribute to more than one analysis if they have events of more than one type (e.g. hospitalization for heart failure followed by ESKD).

5.1.2 Recurrent event analyses

For the key secondary outcome of all-cause hospitalizations (and relevant exploratory analyses), a semi-parametric joint frailty model will be used[6]. The approach will jointly model:

a) The hazard function for recurrent all-cause hospitalizations conditional on the patient specific random frailty; and
b) The hazard function for time to death conditional on the patient specific random frailty.

It will be assumed that the patient specific random frailty follows a gamma distribution with mean 1 and variance θ, where θ is the correlation between the recurrent events. Piecewise constant hazards will be assumed for both hazard functions to allow estimation of the likelihood by Gaussian quadrature, with follow-up time split into five equally sized intervals.[7] Hazards ratios for the effect of treatment on the rate of recurrent all-cause hospitalizations and the rate of death will be calculated by the model, but only the former will be formally tested. The joint frailty model will be adjusted for the prognostic variables used in the minimization algorithm (as described previously in section 5.1.1, with the exception that continuous versions of age, eGFR and urinary albumin:creatinine ratio will be used to improve convergence). If the adjusted model fails to converge, then the hazard ratio for recurrent all-cause hospitalizations will instead be estimated from a joint frailty model adjusted only for treatment allocation (and defaults to a parametric model using a Poisson distribution for recurrent event component and an exponential distribution for the death component if this fails to resolve convergence issues).

The methods outlined here would also be used if any exploratory analyses for other recurrent event outcomes (e.g. hospitalization for heart failure) are conducted.

5.1.3 Annual rate of change in eGFR

The annual rate of change in eGFR across the whole study (i.e. the "total" eGFR slope) will be compared between all those allocated to empagliflozin and all those allocated to placebo using shared parameter models.[8] The approach will jointly model:

[1] Analyses will fit the factors as they were included into the algorithm, in their categorical format, except region (which was not effectively included in the algorithm for most participants).

(a) The annual rate of change in eGFR using a linear mixed model with random effects for each patient's slope and intercept; and
(b) The time to event for ESKD or death using a Weibull survival model in which the scale parameter is assumed to be linearly related to the random effects from the linear mixed model. This allows for the dependence between annual rate of change in eGFR and time to ESKD or death (i.e. those with faster rates of change in eGFR will generally have a shorter time to ESKD or death).

The shared parameter model will be adjusted for the prognostic variables used in the minimization algorithm (as described previously in section 5.1.1). If the adjusted shared parameter model does not converge, then the difference in the annual rate of change in eGFR will instead be estimated from a shared parameter model adjusted only for treatment allocation. The eGFR slope from the 2 month scheduled Follow-up Visit until the last scheduled visits (known as the "chronic" slope) will be assessed using similar methods.

5.1.4 Continuous outcomes

For outcomes with only post-randomization measurements planned (i.e. sodium, corrected calcium, phosphate), mean values of measurements will be compared by t-tests. Where measurements were planned at randomization and a single follow-up visit (e.g. haematocrit and haemoglobin), comparisons of baseline-adjusted mean follow-up values between the allocated treatment arms will be performed using ANCOVA adjusted for each patient's value at randomization.[9]

Where randomization and multiple post-randomization measurements were planned (i.e. all other continuous outcomes), baseline-adjusted mean follow-up values averaged over time (with weights proportional to the amount of time between visits) will be compared between the allocated treatment arms using a mixed model repeated measures (MMRM) approach. The model will include fixed, categorical effects of treatment allocation, time treatment-by-time interaction, and the prognostic variables used in the minimization algorithm (in the same categories used in the minimization process) along with continuous effects of baseline (randomization) measurement and baseline-by-time interaction. The within-person error correlations will be assumed to be unstructured. Baseline-adjusted mean values at each of the follow-up times will also be presented (but not formally tested for differences between the allocated treatment arms).

As urinary albumin:creatinine ratio is not normally distributed a log transformation will be applied before analysis.

5.1.5 Categorical outcomes

For categorical outcomes, the effect of allocated treatment on the number of randomized participants with at least 1 event will be compared using chi squared tests, unless any of the expected cell counts

in the 2x2 contingency table are less than 5, in which case Fisher's exact test will be used. No adjustment for other covariates will be made.

5.1.6 Exploratory analyses

The effect of allocation to empagliflozin versus placebo on mean eGFR at each scheduled visit will be estimated using the MMRM approach previously outlined in Section 5.1.4, as will the effect of stopping study treatment on mean eGFR and urinary albumin:creatinine ratio (but restricted to the surviving UK participants with a measurement about 4 weeks after Final Follow-up).

Analyses of the effect of allocation to empagliflozin versus placebo on subcategories of SAEs by MedDRA HLGTs will use the Cox proportional hazards model as described in Section 5.1.1.

To assess whether the effect of allocation to empagliflozin versus placebo on the primary outcome differs over time, follow-up time will be split into categories by year since Randomization and the significance of the time x treatment allocation interaction will be assessed.

Technical documentation to accompany this Data Analysis Plan may also be added as an appendix, before any unblinding, if additional methodological details for the approaches described in section 5 are found to be required.

5.2 Multiplicity adjustments

As a formal interim efficacy analysis performed by the DMC is planned during the trial (see DMC charter [EDMS #5708] for details), the required alpha-level for the primary outcome in the final analysis will be adjusted as per the Hwang-Shih-DeCani alpha-spending function (γ=-8). For example, with 60% of the primary outcomes accrued at the time the interim analysis, the final two-sided p-value would need to be <0.0497 to be deemed statistically significant.

If the primary assessment shows a significant benefit of empagliflozin, the key secondary outcomes will be assessed with their p-values corrected for multiple testing using the Hochberg "step-up" procedure that controls the familywise error rate. To account for the formal interim efficacy analysis, the Hwang-Shih-DeCani alpha-spending function (γ=0) will be used to specify the familywise error rate. Other secondary outcomes will be assessed without adjustment for multiplicity at a nominal level of $\alpha = 0.05$ (two-sided). Further details are provided in Section 2.5.2.2 of the protocol.

If based on the results of the interim analysis (which will be based in large part on local creatinine results), stopping criteria are met and the DMC recommend stopping the trial for efficacy, substantial further data will be collected as participants will attend their final follow-up visits and provide latest information on their renal status and blood for central creatinine analysis (to apply the definition of sustained). It may take a several weeks/months to see all participants at a final follow-up visit at each site, as the trial aims to establish large sites. All analyses will be based on final database using the

results from the final follow-up visits. Note the information fraction used in the alpha-spending function for the primary and key secondary outcomes will be based on the number of primary outcome events observed at the time of the DMC analysis.

If a formal interim efficacy analysis is not conducted, for example for operational reasons, the primary outcome will be assessed without adjustment for multiplicity and the key secondary outcomes controlled at the familywise error rate of 5.0%.

For the tertiary efficacy, subgroup, safety, laboratory and exploratory analyses, allowance in their interpretation will be made for multiple hypothesis testing,[2,3] taking into account the nature of events (including timing, duration and severity) and evidence from other studies. In addition to the pre-specified comparisons, other post-hoc analyses may be performed with due allowance for their exploratory and, perhaps, data-dependent nature.

5.3 Subgroup analyses and tests for heterogeneity

Tests for heterogeneity of the proportional effect observed in subgroups, through the inclusion of relevant interaction terms (with main effects if not already included in model) in Cox models for time to event outcomes and the shared parameter model for annual rate of change in eGFR, will be used to determine whether the proportional effects in specific subcategories are clearly different from the overall effect. If, however, three or more patient categories can be arranged in some meaningful order then assessment of any trend will be made. Tests of trend will be conducted for age, blood pressure, body mass index, HbA1c, eGFR, urinary albumin:creatinine ratio, NT-proBNP and haematocrit by assessing the significance of the interaction term between treatment allocation and the relevant linear continuous factor.

5.4 Handling of missing and incomplete data

Missing eGFR values will be handled as detailed below. Participants with missing values relevant to subgroup analyses (e.g. body mass index) will be included in the subgroup containing the median value.

5.4.1 Handling of missing central eGFR values

eGFR measured at scheduled visits will be calculated using the CKD-EPI formula,[10] irrespective of whether the creatinine is measured centrally or locally. Appendix I provides definitions of when scheduled Follow-up Visits are expected and how follow-up periods are defined for the purposes of analyses. If multiple central eGFR measurements are available in any one scheduled follow-up period, then the eGFR closest to the ideal follow-up day will be used to define the eGFR for that particularly scheduled Follow-Up Visit period. The eGFR measured at the Final Follow-up Visit will be usually be the last of the eGFRs to be included in analyses, and will be included irrespective of whether or not it is the eGFR closest to the ideal follow-up day. This may result in two eGFRs in the final scheduled follow-up period.

eGFR will be estimated from creatinine measured in the central laboratory wherever possible, but where a central laboratory eGFR measurement is expected (e.g. because a scheduled Follow-up Visit was completed whilst the participant was alive) but missing, the local blood creatinine measurement closest to the ideal follow-up day within the scheduled Follow-up visit period (if one exists) will be used to estimate the local eGFR in its place. In instances where a local Follow-up Visit eGFR is used, percentage change in eGFR will be calculated relative to the local Randomization Visit eGFR measurement and the local eGFR closest to the ideal follow-up day of the next scheduled Follow-up Visit period with a local measurement used to assess the definition of sustained (see Appendix II for details of how this is to be implemented).

Sensitivity analyses where the ≥40% decline and <10 mL/min/1.73m^2 eGFR components of the Kidney Disease Progression outcome are based solely on central laboratory eGFR measurements, and separately solely with local measurements, will be conducted.

Those without any Randomization Visit eGFR measurement (central or local) will still be included in the ≥40% decline in eGFR component of the Kidney Disease Progression outcome by using the latest pre-randomization locally measured eGFR value. Sensitivity analyses excluding these participants from the ≥40% decline in eGFR component of the Kidney Disease Progression outcome (but still counting them in the other components of the primary outcome) will be conducted.

eGFR slope analyses will use all eGFR measurements (excluding those measured after developing ESKD). The main analysis for eGFR slopes is restricted to only central eGFR measurements. Sensitivity analyses using only local eGFR measurements (both with and without the local eGFR measurements planned for 4 weeks after the end of scheduled study treatment in surviving UK participants) will be performed.

5.5 Censoring schema for time-to-event endpoints

5.5.1 Date of censoring for intention-to-treat analyses

It is the aim at each scheduled Follow-up Visit to ascertain all components of the primary outcomes, including the Final Follow-up Visit, with subsequent adjudication of any recorded deaths. Follow-up Visits are usually conducted by direct participant interview and central blood sampling. Occasionally follow-up will need to be by interview with their relative, carer or doctor, or by medical record review (e.g. paper, electronic or registry). During such indirect follow-ups, the date last known to be alive is recorded. Every effort to collect a central blood sample is made at each scheduled visit (direct or not), but for those known to be alive who do not provide a blood sample, the most recent local blood creatinine measurement can be recorded.

Censoring dates for those who withdraw consent or who are lost to follow-up will be derived from information collected at their most recent Follow-up Visit before consent withdrawal or loss to follow-up. Otherwise, the censoring date will be the date of death or the date of the Final Follow-up Visit (except

when the Final Follow-up Visit is indirect, in which case the date last known to be alive will be used). Outcomes recorded as starting after the Final Follow-up Visit will not be included in any within-trial analyses.

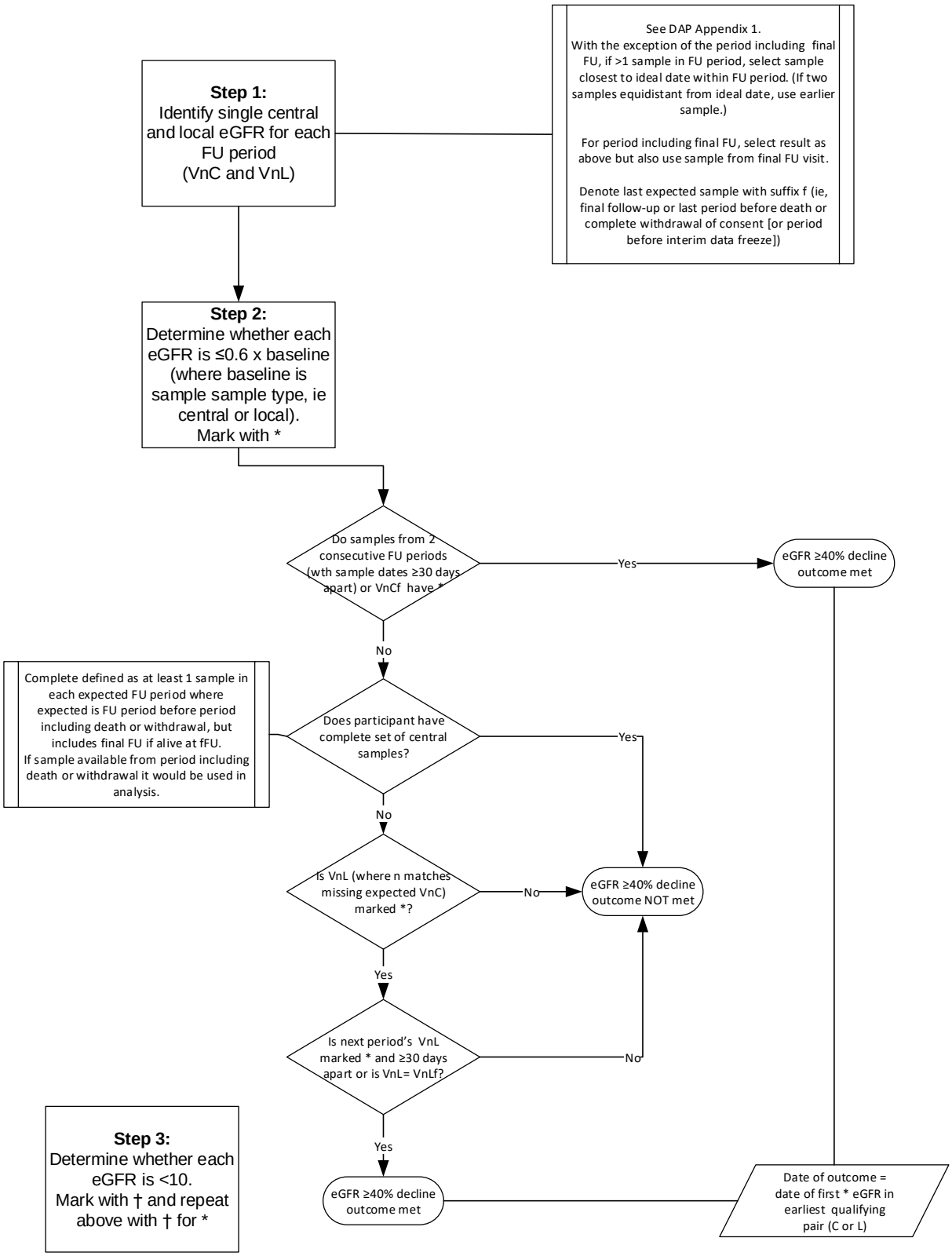
5.5.2 Date of censoring for on-treatment analyses

A censoring date for on-treatment analyses may need to be derived for safety assessments (see Section 3.6). For those participants compliant with any dose of study treatment at their Final Follow-up Visit (or their last Follow-up Visit before death, withdrawal of consent or loss to follow-up), their on-treatment analysis censoring date will be their intention-to-treat analysis censoring date. Where participants are recorded as stopping study treatment and are never recorded as restarting, their on-treatment analysis censoring date will be the date they were recorded as stopping treatment, plus an additional 7 days to allow for any residual effect of treatment.

6 Appendix I: Definitions of Study Scheduled Follow-up Visit period windows for analysis purposes

Scheduled Follow-up Visits relative to the Randomization Visit date			
Visit number	**Follow-up month**	**Follow-up period**	**Ideal follow-up day**
1	2	≥1, <121 days	60 days
2	6	≥121, < 271 days	180 days
3	12	≥271, <451 days	360 days
4	18	≥451, <631 days	540 days
5	24	≥631, <811 days	720 days
6	30	≥811, <991 days	900 days
7	36	≥991, <1171 days	1080 days
8	42	≥1171, <1351 days	1260 days
9	48	≥1351, <1531 days	1440 days
10…	54...	≥1531, <1711 days…	1620 days

7 Appendix II: Algorithm used to substitute local creatinine values when central samples are missing



8 References

1. EMPA-KIDNEY Collaborative Group. Design, recruitment, and baseline characteristics of the EMPA-KIDNEY trial. *Nephrol Dial Transplant* 2022 Mar 3; gfac040. doi: 10.1093/ndt/gfac040. Online ahead of print.
2. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer* 1976; **34**(6): 585-612.
3. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *British Journal of Cancer* 1977; **35**(1): 1-39.
4. Peto R, Peto J. Asymptotically efficient invariant test procedures. *J Roy Stat Soc* 1972; **135**(2): 185-207.
5. Cox DR. Regression Models and Life-Tables. *J R Stat Soc B* 1972; **34**(2): 187-+.
6. Rogers JK, Yaroshinsky A, Pocock SJ, Stokar D, Pogoda J. Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. *Statistics in Medicine* 2016; **35**(13): 2195-205.
7. Liu L, Huang X. The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine* 2008; **27**(14): 2665-83.
8. Vonesh EF, Greene T, Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine* 2006; **25**(1): 143-63.
9. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *BMJ* 2001; **323**(7321): 1123-4.
10. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009; **150**(9): 604-12.