EMPA-KIDNEY Urine Biomarker Substudy
Data Analysis Plan (EDMS 8205)

Version History

Version	Description	Authors
1.0	First released version	Created by Greco Mark B. Malijan, Natalie Staplin, Parminder Judge, Rebecca Sardell & William G. Herrington (NOV-DEC2023) Reviewed by Michael Shlipak and Joe Ix (09JAN2024) Sent to Steering Committee for Comment (16JAN2024) Released for use by William G. Herrington (30JAN2024)

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

TABLE OF CONTENTS

Contents

1 RELEVANT PROCEDURAL DOCUMENTS

Document title	EDMS #
EMPA-KIDNEY Protocol	5434
EMPA-KIDNEY Data Analysis Plan v1.2 (SOP 11)	6290
EMPA-KIDNEY Post-Trial Follow-up Data Analysis Plan v1.0	7987

2 ABBREVIATIONS

Abbreviation	Definition
α-1M	Alpha-1 microglobulin
AKI	Acute kidney injury
CKD	Chronic kidney disease
CKD-EPI	Chronic Kidney Disease – Epidemiology Collaboration
DAP	Data analysis plan
DKK-3	Dickkopf-3
EGF	Epidermal growth factor
eGFR	Estimated glomerular filtration rate
IFCC	International Federation of Clinical Chemistry
IL-18	Interleukin-18
KDIGO	Kidney Disease: Improving Global Outcomes
KFRE	Kidney Failure Risk Equation
KIM-1	Kidney injury molecule-1
LOD	Limit of detection
MCP-1	Monocyte chemoattractant protein-1
MMRM	Mixed model repeated measures
NGAL	Neutrophil gelatinase-associated lipocalin
RAS	Renin-angiotensin system
RDR	Regression dilution ratio
SGLT-2	Sodium-glucose co-transporter-2
uACR	Urinary albumin-creatinine ratio
uCr	Urine creatinine
UMOD	Uromodulin
YKL-40	Human cartilage glycoprotein-39

3 INTRODUCTION

This document provides the Data Analysis Plan (DAP) for the EMPA-KIDNEY Urine Biomarker substudy which primarily aims to evaluate the effects of sodium-glucose co-transporter-2 (SGLT-2) inhibitor empagliflozin on urinary markers of kidney tubular stress/injury, inflammation or fibrosis (Table 1) compared to placebo. The main trial design, baseline characteristics of included participants, and main trial results have been previously reported. (1,2) The purpose of this DAP is to specify the details of the prioritized randomized analyses to be presented in the initial publication(s) of the substudy. The nature of all analyses cannot be fully specified in detail, but, where appropriate, the general analytical approaches follow those set out in the EMPA-KIDNEY main DAP version 1.2 (SOP11; EDMS #6290) (3).

Table 1. Urine biomarkers included in the substudy

Biomarker Group	Biomarker
Tubule Injury and Inflammation	Interleukin-18 (IL-18)
	Kidney injury molecule-1 (KIM-1)
	Neutrophil gelatinase-associated lipocalin (NGAL)
Stress, Ischemia, and Repair	Dickkopf-3 (DKK-3)
	Human cartilage glycoprotein-39 (YKL-40)
	Monocyte chemoattractant protein-1 (MCP-1)
Tubular Function	Alpha-1 microglobulin (α-1M)
Tubular Reserve and Atrophy	Epidermal growth factor (EGF)
	Uromodulin (UMOD)

4 STUDY DESIGN & BASELINE CHARACTERISTICS

In this substudy, urine biomarker measurements are performed at the Randomization visit among all participants in trial sites outside China (n ≈ 5,500). Measurements are repeated at 2 and 18 months post-randomization among a nested subcohort of participants (n ≈ 2,600) selected for diabetes mellitus status (present *vs* absent) and severity of albuminuria (uACR <200 mg/g *vs* ≥200 mg/g). Hence, longitudinal measurements of urine biomarker are available for a subset of participants within the substudy.

In order to assess the balance of baseline characteristics between the randomized arms of the urine biomarker substudy in the nested subcohort population, the following variables, documented at the Randomization visit (or at Screening), will be presented for the empagliflozin and placebo groups. All participants with at least one valid urine biomarker measurement will be included, and missing urine biomarker values will be imputed using the methods set out in Section 6.1. The following variables are a subset of the pre-specified characteristics in the main DAP (SOP11; EDMS #6290) with addition of variables relevant to urine biomarkers.

a. Participant characteristics:
    i. Age (continuous and categorised: <60; ≥60 to <70; ≥70 years);
    ii. Sex (male *vs* female);
    iii. Race (White, Black, Asian, Mixed or Other);
b. History of prior disease:
    i. Diabetes mellitus (presence *vs* absence);
    ii. Primary cause of kidney disease (diabetic kidney disease; hypertensive/renovascular disease; glomerular disease; other or unknown combined[*]);
    iii. Cardiovascular disease (presence *vs* absence);
c. Clinical measurements:
    i. Body mass index (continuous and categorised: <25; ≥25 to <30; ≥30 kg/m^2);
    ii. Systolic blood pressure (continuous and categorised: <130; ≥130 to <145; ≥145 mmHg);
    iii. Diastolic blood pressure (continuous and categorised: <75; ≥75 to <85; ≥85 mmHg);
d. Laboratory measurements:
    i. CKD-EPI—(4) estimated glomerular filtration rate (eGFR) (continuous and categorised: <30; ≥30 to <45; ≥45 mL/min/1.73 m^2);
    ii. Urinary albumin:creatinine ratio (uACR) (continuous and categorised: <30; ≥30 to ≤300; >300 to <1000; ≥1000 to <2000; ≥2000 mg/g);
    iii. Urine creatinine (uCr) in mg/L (continuous);
    iv. Glycosylated haemoglobin (IFCC HbA1c) (continuous and categorised: <39 [normoglycaemia]; ≥39 to <48 [pre-diabetes]; ≥48 to <75 [well-controlled diabetes]; ≥75 [poor glycaemic control] mmol/mol);
    v. N-terminus pro B-type natriuretic peptide (NT-pro BNP) (continuous and categorised: <110; ≥110 to <330; ≥330 ng/L);
    vi. Haematocrit (continuous and categorised: <37%; ≥37% <41%; ≥41%);
e. Urine biomarker measurements (continuous and categorised into approximate thirds of the distribution):
    i. Alpha-1 microglobulin (α-1M) in ng/mL;
    ii. Dickkopf-3 (DKK-3) in pg/mL;
    iii. Epidermal growth factor (EGF) in pg/mL;

[*] Other kidney diseases include tubulointerstitial disease, familial/hereditary nephropathies, other systemic disorders, and miscellaneous renal disorders. Glomerular disease is further sub-categorised into focal segmental glomerulosclerosis, IgA nephropathy, membranous nephropathy, minimal change disease, and other glomerular diseases.

iv. Interleukin-18 (IL-18) in pg/mL;
v. Kidney injury molecule-1 (KIM-1) in pg/mL;
vi. Monocyte chemoattractant protein-1 (MCP-1) in pg/mL;
vii. Neutrophil gelatinase-associated lipocalin (NGAL) in pg/mL;
viii. Uromodulin (UMOD) in ng/mL;
ix. Human cartilage glycoprotein-40 (YKL-40) in pg/mL;

f. Medication use at randomization:
   i. Renin-angiotensin system (RAS) inhibitor (yes *vs* no);
   ii. Any diuretic use (yes *vs* no; analyses by type [loop *vs* thiazide *vs* other potassium-sparing *vs* mineralocorticoid receptor antagonist]);
   iii. Any lipid-lowering medication (yes *vs* no);
   iv. Any medication for diabetes (yes *vs* no);

g. Baseline risk score
   i. Predicted 5-year end stage kidney disease (ESKD) risk based on the Kidney Failure Risk Equation (KFRE) (5,6) (continuous and categorised: <5%; ≥5 to <20%; ≥20%);

Baseline characteristics by randomized arm in the entire substudy cohort will be presented. A comparison of the baseline characteristics of all participants randomised in the EMPA-KIDNEY trial, the entire urine biomarker substudy cohort, and the nested subcohort population will also be provided. To support the observational analyses, the baseline characteristics of participants in the entire substudy population divided into approximate thirds of the distribution of each urine biomarker will be presented.

The variables that will be presented in publications will include all those listed above, with those provided in the primary versus subsidiary tables to be designated based on the relevance to the respective publication. For continuous variables, mean (standard deviation) will be presented unless the variable has a skewed distribution in which case median [interquartile range] will be presented. For categorical variables, the number and percentage of participants in the category will be presented. All possible categories will be provided and zero-filled where necessary. The category “missing” will only be presented if there are actual missing values.

5 RANDOMIZED ASSESSMENTS

The analyses of the effects of allocation to empagliflozin versus placebo on urine biomarkers will involve an intention-to-treat comparison among all randomized participants with at least

one valid biomarker measurement during follow-up. Handling of missing data including urine biomarker measurement is described in Section 6.1.

5.1. 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 parameter of interest (e.g. urine biomarker concentration) in the target population is the same as the effect of allocation to placebo (hence, the alternative hypothesis will be that the effect of allocation to empagliflozin is not the same as the effect of allocation to placebo).

5.2. Primary Randomized Assessment

The primary randomized assessment will be the effect of allocation to empagliflozin on the baseline-adjusted study-average concentration of urine biomarkers during follow-up compared to placebo in the nested subcohort population. Details pertaining to methods of analysis for the primary assessment are described in Section 6.2.1.

5.3. Secondary Randomized Assessment

The secondary randomized assessments will include the evaluation of the effect of allocation to empagliflozin on the baseline-adjusted concentration of urine biomarkers compared to placebo at the following time-points where measurements are made:

- 2 months
- 18 months

5.4. Tertiary Randomized Assessments

Tertiary randomized assessments will include the evaluation of whether any relative effects of allocation to empagliflozin on urine biomarker concentrations are modified by the pre-selected subgroups defined by baseline characteristics. These include the key subgroups from the main DAP, including:

- Diabetes mellitus status (present *vs* absent);
- CKD-EPI eGFR (<30; ≥30 to <45; ≥45 mL/min/1.73 m^2);
- Level of albuminuria (uACR <30; ≥30 to ≤300; >300 to <1000; ≥1000 to <2000; ≥2000 mg/g);
- Primary cause of kidney disease (diabetic kidney disease; hypertensive/renovascular disease; glomerular disease; other or unknown combined)

5.5. Additional Exploratory Analyses

Additional exploratory randomized analyses will include the following.

1. Using the entire substudy population, evaluation whether the absolute and relative effects of allocation to empagliflozin versus placebo on 1) acute eGFR dip (between randomization and 2 month follow-up), 2) eGFR chronic slope (between 2 months and final follow-up), and 3) kidney disease progression[†] are modified by baseline levels of urine biomarkers (with subgroups split into approximate thirds)
2. Using the nested subcohort population, estimation of the proportion of treatment effect of empagliflozin explained by urine biomarker concentration at 2 months on the annualised rate of change of eGFR chronic slope[†]
3. Using the nested subcohort population, evaluation of whether any relative effects of allocation to empagliflozin on urine biomarker concentrations are modified by other subgroups defined by baseline characteristics of interest including the following:
    - Age (<60; ≥60 to <70; ≥70 years);
    - Sex (male *vs* female);
    - Race (Asian, Black, Mixed/Other, White);
    - Baseline concentration of other urine biomarkers (with subgroups split into approximate thirds)

5.6. Additional Observational Analyses

To help interpret the implications of the randomized assessments, observational analyses on the associations between usual levels of each urine biomarker and the following outcomes will be estimated:

- Time to kidney disease progression
- Acute eGFR dip
- eGFR chronic slope
- Time to first occurrence of a serious adverse event for AKI

Other observational analyses will be explored but are beyond the scope of this DAP.

[†] As defined in the EMPA-KIDNEY DAP v 1.2 (EDMS #6290)

6 STATISTICAL METHODOLOGY

6.1. Handling of Missing or Undetectable Biomarker Concentration and Other Missing Data

Missing baseline urine biomarker concentrations will be imputed using the average observed value in both treatment arms combined. Where urine biomarker concentrations that fall below the lower limit of detection (LOD) are not available, single imputation using the expected mean value of samples below the lower LOD from the distribution of the log_2-transformed biomarker values will be used. Sensitivity analyses using only the values that fall within the LOD will also be performed.

Handling of missing data including eGFR values will follow the methods set in the main DAP.

6.2. Methods of Analysis

6.2.1. Primary randomized assessment

Urine biomarker concentrations will be indexed to urine creatinine and analysed as log_2-transformed continuous variable. To evaluate the effect of allocation to empagliflozin on study average urine biomarker concentrations, mixed model repeated measurement (MMRM) analyses will be performed. The urine biomarker concentration at 2 and 18 months will serve as the outcome. The model will include the fixed, categorical effect of treatment allocation, time, treatment-by-time interaction, and the prognostic variables specified in the minimization algorithm (age, sex, prior diabetes, eGFR, uACR, and region, in the same categories used in the minimization process). Because of the lower stability of albumin and potentially other proteins at higher freezer storage temperatures (7), additional adjustment for initial freezer storage temperature (categorical: -20°C *vs* lower than -20°C) will be performed. The model will also be adjusted for the baseline log_2-transformed urine biomarker:creatinine ratio (as a continuous variable) and baseline biomarker-by-time interaction. The treatment-by-time interaction will be used to estimate the mean log_2-transformed urine biomarker:creatinine ratio at each follow-up time for each treatment arm, conditional on the other factors in the model. The within-person correlations are assumed to be unstructured. The models assume that any missing urine biomarker:creatinine ratio values can be predicted by the non-missing urine biomarker:creatinine ratio data for other individuals together with the other covariates in the model (i.e., that they are ‘missing in random’).

A weighted average of these baseline-adjusted mean follow-up values will be used (with the weights proportional to the amount of time between visits[‡]) to calculate the study average log_2-transformed urine biomarker:creatinine ratio in each treatment arm. These values are then back transformed to give the geometric means of study average urine biomarker concentration in each treatment arm as well as the relative differences in the geometric means. For the primary randomized assessment, the relative difference in study average urine biomarker concentrations between treatment arms will be presented.

6.2.2. Secondary randomized assessments

Similar MMRM analyses used in the primary randomized assessment will be performed for the secondary randomized assessments. The treatment-by-time interaction term will be used to estimate the mean log_2-transformed urine biomarker:creatinine ratio at 2 months and 18 months for each treatment arm, conditional on other factors in the model. The relative difference in urine biomarker concentrations between treatment arms at 2 and 18 months will be presented.

6.2.3. Tertiary randomized assessments

Subgroup analyses on the effects of allocation to empagliflozin on urine biomarker concentrations on follow-up will be facilitated by fitting relevant interaction terms for the respective subgroup and treatment-by-time interaction terms in the MMRM analyses. This method will assess whether the proportional effects in specific subgroups are statistically different from the overall effect. In the subgroup analysis involving the level of albuminuria, the use of three categories (i.e., uACR <30, ≥30 to ≤300, and >300 mg/g) will be prioritised. Exploratory analyses concerning more severe albuminuria using five categories (i.e., uACR <30, ≥30 to ≤300, >300 to <1000, ≥1000 to <2000, and ≥2000 mg/g) will also be performed.

6.2.4. Additional exploratory analyses

Analyses of the effects of allocation to empagliflozin on the acute eGFR dip (between randomization and 2-month follow-up) and eGFR chronic slope by baseline urine biomarker concentration will follow the principles specified in Section 5.1.3. of the EMPA-KIDNEY DAP v1.2. Absolute differences in chronic slopes will be calculated, which will then be used to estimate the relative differences across subgroups (to enable direct tests of any differences of the effects of empagliflozin between subgroups) by dividing the absolute effect and its 95% confidence interval by the mean slope in the placebo arm.

Similarly, analyses of the effects of allocation to empagliflozin on time to kidney disease progression by baseline characteristics including urine biomarker concentration will follow the

[‡] Using the definitions of scheduled follow-up visit window periods given in the EMPA-KIDNEY DAP v1.2 (EDMS #6290)

principles specified in Section 5.1.1. of the EMPA-KIDNEY DAP v1.2. Additional outcomes from the EMPA-KIDNEY Post-Trial Follow-up study (EDMS #7987) (8) will be incorporated in these analyses.

To quantify the proportion of treatment effects by empagliflozin on eGFR chronic slope explained by urine biomarker concentrations at 2 months, the landmark method (9,10) will be used. Linear regression models with and without adjustment for the 2 month values of urine biomarkers will be compared. The proportion of treatment effect explained by on-study urine biomarker concentrations will be estimated by the following:

$$1 - \frac{regression\ coefficients\ for\ the\ overall\ treatment\ effect\ adjusted\ for\ urine\ biomarker}{regression\ coefficients\ for\ the\ overall\ treatment\ effect\ unadjusted\ for\ urine\ biomarker}$$

Bias-corrected and accelerated bootstrap intervals with 10,000 replications will be used to construct the 95% confidence intervals. To quantify the change in strengths of association between treatment allocation and outcomes after adjustment for biomarker concentrations, changes in Wald X^2 statistics will be presented.

The methods specified in Section 6.2.3. will be used in the subgroup analyses by other baseline characteristics of the effects of allocation to empagliflozin on urine biomarker concentrations on follow-up.

6.2.5. Observational analyses

The evaluation of the association between usual urine biomarker concentration and kidney outcomes will involve the entire substudy population. The exposure variable will be log_2-transformed urine biomarker:creatinine ratio. Cox proportional hazards regression models and linear regression models will be used for time-to-event outcomes and eGFR slopes, respectively. The models will be exploratory and developed based on example principles set out in Table 2.

Table 2. Approach to model development in the prioritised observational analyses

Model	Covariates	Rationale
M1	age, sex, race, prior diabetes, prior cardiovascular disease, primary kidney disease, BMI, SBP	Adjustment for presumed confounders
M2	M1 covariates and eGFR	Assessment of the relevance of each urine biomarker on kidney outcomes for a given level of eGFR and uACR considered separately and together
M3	M1 covariates and uACR	
M4	M1 covariates, eGFR, and uACR	
M5	M4 covariates and other urine biomarkers	Assessment of the relevance of each urine biomarker on kidney outcomes for given levels of other urine biomarkers (having adjusted for

		confounders and kidney function [i.e., eGFR and uACR]) with priority for biomarkers of tubular reserve/atrophy

Relevant analyses will be used to explore the shape of associations. The combination of measurement error and natural within-person variability of urine biomarker concentration means that analyses using baseline measurements of urine biomarkers will underestimate the relevance of long-term average ("usual") urine biomarker levels on the risk of kidney outcomes (11). Correction for regression dilution bias will be performed by dividing the log hazard ratios and corresponding standard error of the mean associated with the baseline urine biomarker by an estimate of the regression dilution ratio (RDR). This approach allows the quantification of the relevance of "usual" urine biomarker concentrations to risk but does not affect the assessment of the statistical significance of the associations. RDR plots will be constructed and used to establish the optimum method to estimate RDRs (e.g., with preference for Rosner parametric method (12)).

6.3. Interpreting Results and Addressing Multiplicity of Testing

Interpretation of results will take into consideration the number of subgroups studied and the biological rationale. To correct for familywise error rates in the primary randomized assessment, p values for the relative study-average difference will be compared against the critical thresholds from the Holm procedure (13) to determine whether they are considered to be statistically significant.

6.4. Censoring Schema for Time-to-Event Endpoints

Censoring schema will follow the methods set in the main DAP. Briefly, 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 before consent withdrawal or loss to follow-up. Otherwise, the censoring date will be the date of death or the date of final follow-up visit.

6.5. Software

R 4.3.2/RStudio 2023.09.1 and SAS 9.4 will be used to prepare the analytical datasets and to perform the analyses.

Further technical documentation to accompany this Data Analysis Plan may be added as an appendix if additional methodological details for the approaches detailed in Section 6 will be required.

7 REFERENCES

1. EMPA-KIDNEY Collaborative Group. Empagliflozin in Patients with Chronic Kidney Disease. N Engl J Med. 2022 Nov 4;1–11.
2. The EMPA-KIDNEY Collaborative Group. Design, recruitment, and baseline characteristics of the EMPA-KIDNEY trial. Nephrol Dial Transplant. 2022 Jul 1;37(7):1317–29.
3. EMPA-KIDNEY Collaborative Group. EMPA-KIDNEY Data Analysis Plan v 1.2 [Internet]. 2022 [cited 2023 Oct 30]. Available from: https://www.empakidney.org/downloads/empa-kidney_dataanalysisplan_v1p2.pdf
4. Levey AS, Stevens LA, Schmid CH, Zhang Y (Lucy), Castro AF, Feldman HI, et al. A New Equation to Estimate Glomerular Filtration Rate. Ann Intern Med. 2009 May 5;150(9):604–12.
5. Tangri N, Grams ME, Levey AS, Coresh J, Appel L, Astor BC, et al. Multinational assessment of accuracy of equations for predicting risk of kidney failure: a meta-analysis. JAMA. 2016 Jan 12;315(2):164–74.
6. Major RW, Shepherd D, Medcalf JF, Xu G, Gray LJ, Brunskill NJ. The Kidney Failure Risk Equation for prediction of end stage renal disease in UK primary care: An external validation and clinical impact projection cohort study. PLoS Med. 2019 Nov 6;16(11):e1002955.
7. Herrington W, Illingworth N, Staplin N, Kumar A, Storey B, Hrusecka R, et al. Effect of Processing Delay and Storage Conditions on Urine Albumin-to-Creatinine Ratio. Clin J Am Soc Nephrol CJASN. 2016 Oct 7;11(10):1794–801.
8. EMPA-KIDNEY Collaborative Group. EMPA-KIDNEY Post-Trial Follow-Up Data Analysis Plan v1.0 [Internet]. Available from: https://www.empakidney.org/downloads/empa-kidney-ptfu-dap-v1-0-2023-05-02.pdf
9. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med. 1992 Jan 30;11(2):167–78.
10. Simes RJ, Marschner IC, Hunt D, Colquhoun D, Sullivan D, Stewart RAH, et al. Relationship between lipid levels and clinical outcomes in the Long-term Intervention with Pravastatin in Ischemic Disease (LIPID) Trial: to what extent is the reduction in coronary events with pravastatin explained by on-study lipid levels? Circulation. 2002 Mar 12;105(10):1162–9.
11. Clarke R, Shipley M, Lewington S, Youngman L, Collins R, Marmot M, et al. Underestimation of Risk Associations Due to Regression Dilution in Long-term Follow-up of Prospective Studies. Am J Epidemiol. 1999 Aug 15;150(4):341–53.
12. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med. 1989;8(9):1051–69.
13. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scand J Stat. 1979;6(2):65–70.