Effects of empagliflozin versus placebo on kidney outcomes: Protocol and outline analysis plan for an individual participant-level data renal meta-analysis

1 Summary

Four large placebo-controlled trials have tested the effects of empagliflozin on kidney outcomes in populations with prior cardiovascular disease, heart failure and chronic kidney disease (CKD). The mechanisms of kidney benefits are incompletely understood. It is possible that the presence of intraglomerular hypertension - of which albuminuria is a marker - may be a pre-requisite for benefits of empagliflozin on kidney disease progression, and it remains possible that other mechanisms of kidney benefit exist. The aim of this document is to define key questions and outline the statistical methodology for a meta-analysis that can explore the effects of empagliflozin on kidney disease.

2 Introduction

Sodium glucose co-transporter-2 (SGLT-2) inhibitors were originally developed to treat hyperglycaemia in patients with diabetes.[1] Thirteen large placebo-controlled trials have now reported and a meta-analysis has shown that SGLT-2 inhibitors reduce risk of kidney disease progression by about two-fifths and risk of acute kidney injury by about a quarter. These kidney benefits appear to be present irrespective of diabetes status.[2] Results from the EMPA-KIDNEY trial showed kidney benefits were not modified by the level of eGFR (down to 20 mL/min/1.73m^2 and perhaps lower), but showed relative benefits are larger in patients with higher levels of albuminuria.[3] It is likely that lowering intraglomerular pressure is a key mechanism by which empagliflozin affords renoprotection. It remains possible, however, that other mechanisms of kidney benefit exist. The presence of albuminuria is considered a marker of abnormally raised intraglomerular pressure (as well as a marker of intrinsic glomerular disease which can develop in the absence of intraglomerular hypertension). Reductions in albuminuria when commencing empagliflozin are considered to be due to favourable changes to intraglomerular haemodynamics (and are accompanied by a reversible acute dip in eGFR). It is also relevant to assess whether larger acute dips in eGFR on commencing an SGLT-2 inhibitor reflect a greater effect of study treatment on intraglomerular pressure (which may yield larger kidney benefits), and/or predispose to risk of acute kidney injury. This is important to assess, as large dips have been considered a reason to stop SGLT2 inhibitors, when in reality, the dips may predict kidney benefits. Another question of interest is whether benefits differ by heart failure status at baseline, as those with CKD-associated cardiorenal syndromes may particularly benefit from SGLT-2 inhibition, or indeed be at increased risk of acute kidney injury. Progression of CKD among patients with heart failure, who generally exhibit low levels of albuminuria, may be due to different mechanisms to progression of intrinsic kidney disease.[4]

All such analyses described above will optimally be performed using individual participant-level data. The aim of this document is to define key renal questions which need to be addressed, and provide statistical methodology to be used in this exploratory meta-analysis.

2.1 Outcomes

Kidney effects will be assessed using outcomes based on (i) progression of CKD, and (ii) acute kidney injury.

(i) Kidney Disease Progression assessed using:
   a. A categorical outcome based on the EMPA-KIDNEY definition.*
   b. eGFR slope, which is a surrogate of Kidney Disease Progression with particular advantages when trial follow-up is short and when assessing for differences between subgroups.[5] Methods based on EMPA-KIDNEY explorations will be used.

(ii) Acute kidney injury assessed using:
   a. Adverse event data and definitions from a recent renal meta-analysis.†
   b. "Abrupt decline in kidney function" – an exploratory outcome defined as a ≥50% increase in serum creatinine (measured by either a local or central laboratory) compared with the most recent laboratory value (taken less than one year previously)

2.2 Key subgroups

All these kidney outcomes will be assessed overall and subdivided by characteristics recorded at baseline. These include assessments by:

(i) Diabetes status (defined by self-report or investigator-report, HbA1c≥6.5% or use of anti-diabetic treatment at baseline, wherever possible [or EMPA-REG OUTCOME recruit]): yes vs no
(ii) Level of albuminuria: uACR <30 vs ≥30,<300 vs ≥300<1000 vs ≥1000 mg/g
(iii) uACR categories by diabetes status: no diabetes, uACR<200 vs no diabetes, uACR≥200 vs diabetes, uACR <200 vs diabetes, uACR ≥200 mg/g
(iv) Level of kidney function: eGFR (CKD-EPI) <30 vs ≥30<45 vs ≥45<60 vs ≥60 mL/min/1.73m^2
(v) Primary kidney disease: diabetic vs hypertensive/renovascular vs glomerular vs other/unknown vs no reported disease (where possible to distinguish reliably)‡
(vi) Predicted (rather than directly measured) acute eGFR dip based on multivariable model (see section 7 for outline model details): <-8% vs ≥-8%, <-7% vs ≥-7%, <-5% vs ≥-5%§
(vii) Evidence for heart failure (defined as EMPEROR trial recruit, or in the case of a non-EMPEROR trial recruit: self-reported heart failure or the combination of NT-proBNP >300 pg/mL and NYHA classification ≥2 at baseline**): yes vs no
(viii) NT-proBNP: tertiles of distribution

2.3 Other exploratory assessments

Further outcomes may be added to this list in due course due to relevance (e.g. hyperkalaemia) and in response to new findings from randomized or non-randomized data. Similarly, other subgroups may be explored post-hoc (e.g. by measures of age/frailty, by race/ethnicity, by baseline blood

---

* End-stage kidney disease (ESKD); the initiation of maintenance dialysis or receipt of a kidney transplant), a sustained decrease in the eGFR to less than 10 or 15 ml per minute per 1.73 m^2 (as per trial use), a sustained decrease from baseline in the eGFR of at least 40%, or death from kidney failure. The EMPA-KIDNEY definition of sustained will be used, wherever possible.

† Based on the MedDRA Preferred Term of Acute Kidney Injury (adjudicated where possible, otherwise unadjudicated).

‡ Restricted to trials with such data.

§ Subgroups created based on approximately equal number of abrupt decline in kidney function events in each group, followed by the subdivision of the highest predicted acute dip group (< -7%) into two

** EMPA-KIDNEY measured NT-pro BNP and assessed for heart failure symptoms in all participants at baseline. Note that in UKHARP3, mean eGFR was 35 mL/min/1.73m^2 and mean NT-proBNP was ~250 pg/mL at baseline with >95% having an NT-proBNP <300 pg/mL (Haynes *et al.* Circulation 2018; 138:1505-1514).

pressure, by co-medication use). Subgroup definitions based on clinical definitions may be collapsed if numbers of events are low in some subcategories of patients.

3 Trial eligibility

Randomized trials of empagliflozin versus placebo in which at least 1000 participants with a scheduled follow-up duration of at least 1 year. These include: EMPA-REG OUTCOME,[6,7] EMPEROR-REDUCED,[8] EMPEROR-PRESERVED,[9] and EMPA-KIDNEY.[3]

4 Statistical methods

Statistical analyses performed with the knowledge of the main trial results should be interpreted in that context and as exploratory. Analyses will first be performed in EMPA-KIDNEY data to develop the methods and will then be "replicated" in the other trial dataset using individual participant-level data. General approaches are set out below.

Analyses will include all participants who were randomly assigned to the different study treatment arms, irrespective of whether or not they remained adherent with their allocated study treatment, e.g., follow-up data from patients with premature treatment discontinuation will be included in the analyses. Such 'intention-to-treat' comparisons are used to provide unbiased assessment of moderate effects of a treatment on relatively common outcomes.[10] Empagliflozin 25mg and 10mg arms from the EMPA-REG OUTCOME trial will be combined and compared to matching placebo.

These analyses will be done for each trial separately and combined using an Inverse Variance Weighted (IVW) approach. Each separate Cox model will be adjusted for age (continuous), sex, diabetes status, eGFR (categorical: <30, ≥30 to <45, ≥45 to <60, ≥60 mL/min/1.73m^2) and uACR (categorical: <30, ≥30 to <300, ≥300 to <1000, ≥1000 mg/g). 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.

The proportional effects of empagliflozin in various different subgroups will be estimated, and compared with the overall effect seen in all patients using standard χ^2 tests for heterogeneity or, where appropriate, χ^2 test for trend (with consideration for additional terms if necessary to address possible non-linearity). Development of a multivariable model for a post-randomization acute dip in eGFR among participants assigned to active empagliflozin enables subgroup analyses by predicted size of an acute eGFR dip, including those allocated to placebo. This approach has the advantage that it allows the comparison of subgroups with similar 'predicted change' in both empagliflozin and placebo arms, and as such, the subgroup analyses remain randomized. The acute dip-based subgroup analyses will only be conducted if a reliable predictive model can be developed. Outline details of models of size of an acute eGFR dip are provided in a statistical methodology appendix (section 7).

The annual rate of change in eGFR (i.e. eGFR slopes) will be compared between all those allocated to empagliflozin and all those allocated to placebo using shared parameter models.[5] The trial-specific differences in annual rate of change in eGFR will then be meta-analysed using an IVW approach. The shared parameter approach will jointly model:

- The annual rate of change in eGFR using a linear mixed model with random effects for each participant's slope and intercept; and
- 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 same variables as Cox models using the principles applied in EMPA-KIDNEY. 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 an alternative approach (e.g. a shared parameter model adjusted only for treatment allocation).

Analyses will include chronic (i.e. long-term) and total slope analyses, with emphasis on results from chronic slope analyses due to the relative short follow-up of the included trials. Importantly, absolute differences in eGFR slope derived from shared parameters models are limited in their ability to assess effect modification between subgroups, as any differences are determined both by any difference in baseline progression rate plus the effect of the intervention between subgroups. To address this key limitation, exploratory analyses will also assess effect modification on a relative scale.

5 Responsibilities, data handling and funding.

The Renal Studies Group at the University of Oxford initiated this renal meta-analysis and drafted this analysis plan. It is primarily responsible for analyses. The University of Oxford and Boehringer Ingelheim will establish a data sharing agreement which will allow both parties to independently perform analyses and enable independent replication, where desirable. Initial funding is provided by the UK Medical Research Council and may be sought from other non-industry funders.

6 Meta-analysis collaborative group (proposed)

The meta-analysis will be overseen by a collaborative group of key personnel involved in the running of the large empagliflozin trials which have recruited people without diabetes, in addition to others with clinical, meta-analysis and statistical expertise:

- EMPA-KIDNEY: William Herrington, Richard Haynes, Natalie Staplin, Zhaojing Che, Parminder Judge, Alistair Roddick, Martin Landray, Jonathan Emberson & Colin Baigent (CTSU, University of Oxford), and Jennifer Green (DCRI)
- EMPA-REG OUTCOME & EMPA-KIDNEY: Christoph Wanner (University of Würzburg)
- EMPEROR Steering Committee representative: Stefan Anker, Javed Butler, Milton Packer, Faiez Zannad (EMPEROR Steering Group)
- Boehringer Ingelheim clinicians: Sibylle Hauske, Dominik Steubl, Martina Brückmann
- Boehringer Ingelheim statistical oversight: Svenja Seide and Dan Massey

7 Appendix I: Statistical strategy for development of multivariable risk models

Development of this multivariable model for a post-randomization acute dip in eGFR will be developed using data from participants assigned to active empagliflozin.

Dependent variables: Percent change in eGFR between 0 and about 2 months/4 weeks.

Examples of relevant covariates to be considered (measured at baseline) will include: trial, age, sex, eGFR, uACR, diuretic treatment, systolic blood pressure, renin-angiotensin system inhibitor use, HbA1c, history of prior atherosclerotic disease, history of diabetes, history of heart failure, diastolic blood pressure, left ventricular ejection fraction, hematocrit, NT-proBNP, anthropomorphic measures (e.g. weight, body mass index, height, waist and hip circumference), lifestyle factors (e.g. smoking, alcohol), ethnicity, and region.

Exploratory analyses:

- The distribution of dependent variables will be assessed and any necessary transformations applied
- The linearity of the associations between continuous predictors and dependent variables will also be assessed. Where there is evidence of non-linearity predictors will be categorized (e.g. using standard subgroupings or equally sized subgroups)

Model Building:

- Univariable analyses will be used to reduce the number of predictors considered in multivariable models, with p values of generally <0.2 used to select significant predictors (although age, sex will be taken forward regardless of p value in univariable analyses)
- Backwards selection will be used to remove variables from the model until all predictors meet the inclusion criterion (generally using $p<0.05$ as a selection criterion)

Model performance:

- Measures of discrimination and calibration will be calculated

Internal validation:

- Bootstrapping will be used to internally validate the risk model and calculate the model performance measure with adjustment for optimism (e.g. to account for potential overfitting)

Model updating:

- Regression coefficients in the risk model may be recalibrated based on results of internal validation stage

8 Version control

Version #	Date	Authors	Comments
1.0	30th July 2020	Will Herrington, Richard Haynes, Natalie Staplin, Jonathan Emberson	First issued version (EDMS #6835)
1.1	21st October 2020	Will Herrington, Colin Baigent	Incorporating review comments from Colin Baigent
2.0	11th April 2023	Will Herrington, Richard Haynes, Colin Baigent, Jonathan Emberson, Natalie Staplin	Kidney only meta-analysis focus. For Collaborative review.
2.1	21th July 2023	Will Herrington	Revision following comments from Boehringer Ingelheim
2.2	3rd October 2024	Zhaojing Che, Natalie Staplin & Will Herrington	Removal of section 2.3 as no analyses of effects of albuminuria were conducted as subsequent EMPA-KIDNEY analyses completed after release of this protocol published in LDE in January 2024 show effects of empagliflozin on chronic slope are not attenuated at low levels of albuminuria (i.e. the rationale of the exploration no longer remained). Update of the key subgroups to add cut-offs for acute dips; remove KFRE (as trials represented different extremes of risk); and addition of new subgroup based on combined diabetes status and uACR category to match KDIGO 2024 SGLT2i recommendation categories. Minor updates to appendix I to document model approach.

9 References

1. Bakris GL, Fonseca VA, Sharma K, Wright EM. Renal sodium-glucose transport: role in diabetes mellitus and potential clinical implications. Kidney Int 2009;75(12):1272-1277. DOI: 10.1038/ki.2009.87.
2. Nuffield Department of Population Health Renal Studies Group & the SMART-Consortium. Impact of diabetes on the effects of sodium glucose co-transporter-2 inhibitors on kidney outcomes: collaborative meta-analysis of large placebo-controlled trials. Lancet 2022. DOI: 10.1016/S0140-6736(22)02074-8.
3. EMPA-KIDNEY Collaborative Group. Empagliflozin in Patients with Chronic Kidney Disease. N Engl J Med 2023;388(2):117-127. DOI: 10.1056/NEJMoa2204233.
4. Packer M, Siddiqi TJ, Butler J. Drugs that slow the progression of diabetic kidney disease: are renoprotective effects attenuated in heart failure? Eur Heart J 2023. DOI: 10.1093/eurheartj/ehad158.
5. 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):143-163. DOI: 10.1002/sim.2249.
6. Wanner C, Inzucchi SE, Lachin JM, et al. Empagliflozin and Progression of Kidney Disease in Type 2 Diabetes. N Engl J Med 2016;375(4):323-34. DOI: 10.1056/NEJMoa1515920.
7. Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. N Engl J Med 2015;373(22):2117-28. DOI: 10.1056/NEJMoa1504720.
8. Packer M, Anker SD, Butler J, et al. Empagliflozin in Patients With Heart Failure, Reduced Ejection Fraction, and Volume Overload: EMPEROR-Reduced Trial. J Am Coll Cardiol 2021;77(11):1381-1392. DOI: 10.1016/j.jacc.2021.01.033.
9. Anker SD, Butler J, Filippatos G, et al. Empagliflozin in Heart Failure with a Preserved Ejection Fraction. N Engl J Med 2021;385(16):1451-1461. DOI: 10.1056/NEJMoa2107038.
10. Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. The Lancet 2001;357(9253):373-380. DOI: 10.1016/s0140-6736(00)03651-5.