Boehringer Ingelheim

Protocol for non-interventional studies based on existing data

Document number:	c03856813-11
BI Study number:	1245.97
BI Investigational Product(s):	Jardiance® (empagliflozin), Synjardy® (empagliflozin/metformin hydrochloride)
Title:	Post-authorisation safety study to assess the risk of urinary tract malignancies in relation to empagliflozin exposure in patients with type 2 diabetes: a multi-database European study
Protocol version identifier:	8.0
Date of last version of protocol:	07 July 2022
PASS:	Yes
EU PAS register number:	EUPAS16424
Active substance:	A10BK03 empagliflozin, A10BD20 metformin and empagliflozin
Medicinal product:	Jardiance, Synjardy
Product reference:	EMEA/H/C/002677, EMEA/H/C/003770
Procedure number:	EMEA/H/C/002677/MEA; EMEA/H/C/003770/MEA
Joint PASS:	No
Research question and objectives:	The aim of the study is to assess the risk of urinary tract malignancies in patients initiating empagliflozin compared to patients initiating a DPP-4 inhibitor.

Countries of study:	The United Kingdom, Sweden, and Finland
Authors:	[redacted]
Marketing authorisation holder(s):	
MAH contact person:	
EU-QPPV:	
Signature of EU-QPPV:	*The signature of the EU-QPPV is provided electronically*
Date:	17 May 2023

Page 1 of 111

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. LIST OF ABBREVIATIONS

AT	As-treated analysis
ATC	Anatomical Therapeutic Chemical Classification System
AvoHILMO	Register of primary health care visits
BI	Boehringer Ingelheim International GmbH
BMI	Body mass index
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CPRD	Clinical Practice Research Datalink
CV	Cardiovascular
DDD	Defined daily dose
DPP-4	Dipeptidyl peptidase-4
eGFR	Estimated glomerular filtration rate
EMA	European Medicines Agency
EMR	Electronic medical record
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
FDA	Food and Drug Administration
GLP-1	Glucagon-like peptide-1
GP	General practitioner
GPP	Good Pharmacoepidemiology Practices
GPV	Good Pharmacovigilance Practices
HbA1c	Glycated haemoglobin A1c
HES	Hospital Episode Statistics
HF	Heart failure
HILMO	Care register for health care
HR	Hazard ratio
HUS	Helsinki University Hospital region
ICD-10	International classification of diseases
IPTW	Inverse-probability of treatment weights

ITT	Intention-to-treat analysis
MI	Myocardial infarction
MSM	Marginal structural model
NCC	Nested case-control
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NSAID	Non-steroidal anti-inflammatory drugs
ONS	Office for National Statistics
PASS	Post-safety study
PH	Proportional hazards
PID	Personal identification number
PS	Propensity score
PSUR	Periodic Safety Update Report
PY	Person-years
RMP	Risk management plan
RR	Relative risk
SEAP	Statistical/Epidemiological Analysis Plan
SGLT-2	Sodium glucose co-transporter-2
SID	Study identification number
SIR	Standardized incidence ratio
T2D	Type 2 diabetes mellitus
THIN	The Health Improvement Network database
UK	United Kingdom
US	United States
UT	Urinary tract
UTI	Urinary tract infection

2. RESPONSIBLE PARTIES

Principal investigator

[redacted]

Participating institutions

[redacted]

Sponsor
Boehringer Ingelheim International GmbH

The study investigators at [redacted] ([redacted] Company) share responsibility with Boehringer Ingelheim International GmbH (BI) for the design of the study. The investigators are responsible for conducting the study in a manner that meets regulatory standards, conducting analyses, and preparing scientific reports. The study shall be conducted as described in the approved protocol. The authors will not develop or implement any deviation or change to the protocol without prior review by BI.

The financial sponsor of this study is BI. The sponsor is responsible to assure study progress. BI is also responsible for communicating with the European Medicines Agency (EMA) (called 'the agency') about the study protocol, the progress of the study, and study results.

3. ABSTRACT

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®, Synjardy			
Name of active ingredient: Empagliflozin (A10BK03) Metformin and empagliflozin (A10BD20)			
Protocol date: 16-Sep-2015	**Study number/ Document number:** 1245.97 / c03856813-10	**Version/Revision:** 8.0	**Version/Revision date:** 17-May-2023
Title of study:	Post-authorisation safety study to assess the risk of urinary tract malignancies in relation to empagliflozin exposure in patients with type 2 diabetes: a multi-database European study		
Rationale and background:	Empagliflozin is an oral blood glucose lowering drug belonging to the sodium glucose co-transporter-2 (SGLT-2) inhibitor class. SGLT-2 inhibitors promote the renal excretion of glucose and help lower elevated blood glucose levels in patients with type 2 diabetes. As a part of the risk management plans for empagliflozin and empagliflozin/metformin hydrochloride agreed upon by the European Medicines Agency (EMA), a post-authorisation safety study (PASS) will be performed to assess the risk of urinary tract malignancies in patients initiating empagliflozin, compared to patients initiating dipeptidyl peptidase-4 (DPP-4) inhibitors. Prior to the development of the protocol for this study, a feasibility assessment of three databases in Europe (Clinical Practice Research Datalink (CPRD) and data with potential for linkage (Hospital Episode Statistics (HES), and Office for National Statistics (ONS) mortality statistics) in the United Kingdom (UK) and the national registers in Sweden and Finland) was performed with the aim to detail the minimum data requirements for the study and to assess, for each potential data source, the data quality, completeness of recorded information for urinary tract malignancies, and possibility of linkage with cancer registries, mortality registries, and other data sources. Initially, this study included two countries, the UK (CPRD GOLD) and Sweden, with follow-up until 31st December, 2019. In order to increase the accumulated patient years and number of patients initiating the study drugs as well as accounting for the long latency period related to risk of malignancies, the study follow-up was first extended until 31st December, 2020 and Finnish national registers were included. In this version of the protocol (Version 7.0), CPRD Aurum was included and the study period extended to 31st December, 2021 for both CPRD databases, further increasing the sample size and accumulated person-years, and, consequently, increasing study power.		

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®, Synjardy			
Name of active ingredient: Empagliflozin (A10BK03) Metformin and empagliflozin (A10BD20)			
Protocol date: 16-Sep-2015	**Study number/ Document number:** 1245.97 / c03856813-10	**Version/Revision:** 8.0	**Version/Revision date:** 17-May-2023
Research question and objectives:	**Research Question:** The aim of the study is to assess the risk of urinary tract malignancies in patients initiating empagliflozin (free or fixed-dose combination) compared to patients initiating a DPP-4 inhibitor. **Primary objectives:** (i) To estimate the adjusted hazard ratios (aHRs) and incidence rates (IRs) of all urinary tract cancers (bladder, renal and other urinary tract cancers) in patients initiating empagliflozin compared to patients initiating a DPP-4 inhibitor (ii) To estimate the aHRs and IRs of bladder cancer in patients initiating empagliflozin compared to patients initiating a DPP-4 inhibitor (iii) To estimate the aHRs and IRs of renal cancer in patients initiating empagliflozin compared to patients initiating a DPP-4 inhibitor **Secondary objective:** To estimate the aHRs and IRs of each primary outcome (all urinary tract cancers, bladder cancer, renal cancer) with respect to: a. increasing cumulative dosage of empagliflozin exposure b. dosage of empagliflozin prescribed per use (e.g., 10mg vs 25mg of empagliflozin) c. time since first dose of empagliflozin exposure **Further objectives:** The following objectives will be explored if adequate sample size is available. (i) To estimate the aHRs and IRs of all urinary tract cancers with respect to empagliflozin initiation among sub-groups defined by age, sex, relevant comorbidities, concurrent use of metformin and other relevant treatments (ii) To estimate the aHRs and IRs of non-renal, non-bladder urinary tract cancers in patients initiating empagliflozin compared to patients initiating a DPP-4 inhibitor Non-renal, non-bladder urinary tract cancers (referred to as other urinary tract cancers) includes ureteral and urethral cancers. The impact of the COVID-19 pandemic on prescription patterns, incidence of diagnosed cancer, diabetes related healthcare utilization and overall cancer screening will be evaluated and reported the final study report.		

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®, Synjardy			
Name of active ingredient: Empagliflozin (A10BK03) Metformin and empagliflozin (A10BD20)			
Protocol date: 16-Sep-2015	**Study number/ Document number:** 1245.97 / c03856813-10	**Version/Revision:** 8.0	**Version/Revision date:** 17-May-2023
	Further, a summary of the literature review on the latest knowledge about the associations between metformin use and the risk of bladder, renal and other urinary tract cancers will be made.		
Study design:	This is a non-interventional, comparative, cohort safety study based on European healthcare databases**.** The databases for this study are constructed from linked prescription, hospital, general practitioner, cancer and death registration records. The study will use an "incident users" design and compare new users of empagliflozin to new users of DPP-4 inhibitors. The index date will be defined as the date on which each identified new user receives the index prescription for empagliflozin, or a DPP-4 inhibitor. Follow-up will start 6 months after the index date (12 months in sensitivity analyses) to account for an empirical induction/promotion period. To minimize channelling bias, individuals will be matched with similar treatment and clinical history at index date by deriving propensity scores (PS). PS-matched cohorts will be derived for the comparison from the study base by matching patients who initiate empagliflozin with patients who initiate DPP-4 inhibitors at index date. To further gain power, the highest possible matching ratio in each country-level dataset will be used. This will be implemented using a variable ratio approach in all countries, where exposed patients are matched to as many control patients as possible, up to a maximum of 3 matched controls. Thus, an observational cohort is constructed to study the risk of urinary tract cancer among the two treatment groups. Depending on sample size, combination exposures with metformin will be accounted for through separate PS, stratification, or adjustment.		

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®, Synjardy			
Name of active ingredient: Empagliflozin (A10BK03) Metformin and empagliflozin (A10BD20)			
Protocol date: 16-Sep-2015	**Study number/ Document number:** 1245.97 / c03856813-10	**Version/Revision:** 8.0	**Version/Revision date:** 17-May-2023
Population:	The broad study population will include all patients from CPRD GOLD and CPRD Aurum in the UK and patients in the nationwide registers in Sweden and Finland, who have purchased at least one prescription of empagliflozin, or DPP-4 inhibitor drugs during the study period 2014-2020 (2021 for CPRD). Patients initiating empagliflozin, or DPP-4 inhibitor fixed-dose combinations with metformin will be included. ***Index date:*** Date of first purchase/prescription of empagliflozin, or a DPP-4 inhibitor during follow-up (2014-2020; 2014-2021 for CPRD). Patients initiating DPP-4 inhibitor fixed-dose combinations with metformin will also be included. The study population will be selected from the broad study population with the following inclusion and exclusion criteria: Inclusion criteria: • Diagnosis of type 2 diabetes (T2D) • Age over 18 years at index date • At least 1 year of membership in the database prior to index date; in CPRD, at least 1 year of continuous up-to-standard registration in the CPRD prior to the index date Exclusion criteria: • Patients with any cancer (excluding non-melanoma skin cancer) prior to index date • Diagnosis of type 1 diabetes or other specific non-T2D • Use of any SGLT-2 inhibitor or any DPP-4 inhibitor (including free or fixed-dose combinations) prior to index date • Use of fixed-dose combinations of SGLT-2 inhibitors with DPP-4 inhibitors Diagnosis of end stage renal disease or receipt of renal dialysis prior to index date		
Variables:	**Outcome variables:** • All urinary tract cancers • Bladder cancer • Renal cancer • Non-renal, non-urinary bladder urinary tract cancers		

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®, Synjardy			
Name of active ingredient: Empagliflozin (A10BK03) Metformin and empagliflozin (A10BD20)			
Protocol date: 16-Sep-2015	**Study number/ Document number:** 1245.97 / c03856813-10	**Version/Revision:** 8.0	**Version/Revision date:** 17-May-2023
	Definition of exposure: The primary definition of exposure for as-treated (AT) analysis is whether an individual is exposed to one of the following two index drug groups during the study period starting from 6 months after the index date and ending at treatment discontinuation of index drug: • Empagliflozin, or • DPP-4 inhibitor (sitagliptin, saxagliptin, linagliptin, vildagliptin, or alogliptin). Patients initiating empagliflozin, or DPP-4 inhibitor in fixed-dose combinations with metformin will be included. Note that use of any of these drugs prior to index date is an exclusion criterion. Secondary definitions of exposure include • cumulative dosage of empagliflozin use, • daily dosage of empagliflozin prescribed per use (10mg vs 25 mg), and • time since first dose of empagliflozin. The following covariates, at a minimum, will be accounted for as variables for exact or PS matching and stratification, and in sensitivity analyses, in the risk set definition in nested-case-control (NCC) design and as covariates to derive inverse-probability of treatment weight (IPTW) in marginal structural models (MSMs) and/or model adjustment. These covariates will be identified based on codes recorded prior to the index date (exclusive of index date) unless otherwise specified below. • Socio-demographic variables (age, sex, year of index date, length of available look-back time) • Concomitant medications (insulin use, other oral glucose lowering medication use, diabetes treatment complexity, other non-diabetes medications) • Comorbidities such as o Diabetic complications (including latest HbA1c value) o Urinary tract-related diseases o Cardiovascular diseases • BMI, smoking, alcohol use (closest to index date) Additional covariates will be included according to data availability are, such as socioeconomic position and laboratory measurements.		

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®, Synjardy			
Name of active ingredient: Empagliflozin (A10BK03) Metformin and empagliflozin (A10BD20)			
Protocol date: 16-Sep-2015	**Study number/ Document number:** 1245.97 / c03856813-10	**Version/Revision:** 8.0	**Version/Revision date:** 17-May-2023
Data sources:	The following databases are country-specific data sources from which a linked master database for each country will be constructed. **The UK:** CPRD GOLD and CPRD Aurum optionally complemented with HES (Outpatient data, and Admitted Patient Care data), and ONS mortality sources. **Sweden:** The following nationwide population-based registers will be accessed: The National Patient Register, the Swedish Prescribed Drug Register, the Swedish Cancer Register, the Causes of Death Register, the National Diabetes Register, and the Longitudinal integration database for health insurance and labor market studies. **Finland:** The following nationwide population-based registers will be accessed: the Finnish prescription register/the registry for reimbursed medications, the Care register for health care (HILMO), the Register of primary health care visits (AvoHILMO), and the Finnish cancer register. Additionally, four regional electronic medical record (EMR) databases will be accessed to obtain information on laboratory measures. Initially, the country-level datasets will be analysed separately. Thereafter, country-level estimates will be pooled using meta-analysis techniques to gain power and produce as narrow confidence intervals (CI) as possible. If patient numbers accumulating in these data sources are below the expected counts, alternatives will be evaluated as described in Section 8.4.4		

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®, Synjardy			
Name of active ingredient: Empagliflozin (A10BK03) Metformin and empagliflozin (A10BD20)			
Protocol date:	**Study number/ Document number:**	**Version/Revision:**	**Version/Revision date:**
16-Sep-2015	1245.97 / c03856813-10	8.0	17-May-2023
Study size:	The study size will be driven by the uptake of empagliflozin and DPP-4 inhibitors for the treatment of T2D to improve glycaemic control in adults in the UK, Sweden, and Finland. Study-size and power computations were carried out for the meta-analysis of the primary outcomes. Originally, it was expected that by the end of 2019, approximately 18,000 empagliflozin-treated patients will accumulate in the CPRD and Swedish databases. However, based on the conservative estimates from the third monitoring report (2019), the expected sample size was higher than anticipated with approximately 25,000 empagliflozin initiators eligible for inclusion in the study. Further, additional 38,000 patients were expected by the end of 2020 with the inclusion of Finnish databases, for a total of approximately 78,000 by the end of the study period. Updated estimates in the fifth interim report (2021) indicated that a total of 57,000 empagliflozin initiators were expected to be eligible for inclusion in the study in CPRD-GOLD, Sweden, and Finland by the end 2020. New estimations including the inclusion of CPRD-Aurum and the extension of the study period to the end of 2021 for both CPRD databases suggest that a total of 67,490 empagliflozin initiators could be eligible for inclusion across all three countries. Furthermore, the observed numbers of patients initiating DPP-4 inhibitors in each country suggest that achievable matching ratios would be 1:3 in UK-CPRD, 1:1 in Sweden, and 1:2 in Finland. Power to detect relative risks of 1.5 and 2 was determined at the meta-analysis level, for each of the study outcomes, based on the expected number of individuals exposed to empagliflozin eligible for inclusion in the study for each country (approximately 16,887 in UK-CPRD (GOLD and Aurum), 32,334 in Sweden, and 18,269 in Finland) and the observed IRs of urinary tract cancers, bladder cancer and renal cancer in country-level study populations as reported in the fifth interim report (UK-CPRD: 111.5, 66.7, 26.5 per 100 000 PY: Sweden: 164.5, 104.1, 58.0 per 100 000 PY; Finland: 161.5, 86.8, 71.6 per 100 000 PY; ,respectively for urinary tract cancer, bladder cancer and renal cancer). Power calculations performed for the primary comparison between empagliflozin users and DPP-4 inhibitors indicated that the power at the meta-analysis level for detecting a hazard ratio of 1.5 would be 93.6% for urinary tract cancer and 84,0% for bladder cancer. Power for detecting a hazard ratio of 2.0 for renal cancer would be 97.0%.		

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®, Synjardy			
Name of active ingredient: Empagliflozin (A10BK03) Metformin and empagliflozin (A10BD20)			
Protocol date: 16-Sep-2015	**Study number/ Document number:** 1245.97 / c03856813-10	**Version/Revision:** 8.0	**Version/Revision date:** 17-May-2023
	In addition, anticipated study sample size would ensure sufficient power for the meta-analysis to detect the targeted effect size of < 1.6 for urinary tract cancers and bladder cancer and the targeted effect size of < 2.0 for renal cancer, with 80% power under the expected matching ratios in each study country. Results from the second and third interim analysis showed that not enough patients will have accumulated in the UK CPRD or Swedish registers by the planned end of data collection to reach the desired sample size. For this reason, the protocol was first amended to expand the study population by including data from the Finnish national registries and extend the study period until 31st December 2020. Following results from the fifth interim analysis, the protocol was further amended to also include the CPRD Aurum database and to extend the study period to the end of 2021 for both CPRD databases.		

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®, Synjardy			
Name of active ingredient: Empagliflozin (A10BK03) Metformin and empagliflozin (A10BD20)			
Protocol date: 16-Sep-2015	**Study number/ Document number:** 1245.97 / c03856813-10	**Version/Revision:** 8.0	**Version/Revision date:** 17-May-2023
Data analysis:	For the primary, secondary, and further objectives, the main data analysis will be conducted in two stages: (i) construction of the PS-matched cohort by modelling the exposure to empagliflozin vs. DPP-4 inhibitors and (ii) estimating the effect of exposure to empagliflozin on the urinary tract (bladder, renal and other) cancers using aHR compared to those exposed to DPP-4 inhibitor and their respective IRs. Additional sensitivity analyses will be performed to validate the robustness of alternative definitions of outcome, exposure, and covariates, to assess reduction of bias due to matching and presence of detection/diagnostic bias, and to evaluate the current use of empagliflozin exposure using an AT analysis. Patients starting a combination of empagliflozin, and metformin (whether a fixed-dose or free combination) will be compared with patients starting a combination of a DPP-4 inhibitor and metformin. Depending on sample size, combination exposures with metformin will be accounted for through separate PS, stratification, or adjustment. IRs (crude and adjusted) will be presented for each exposure group and stratified by relevant variables using the Poisson regression approach. Relative risks will be presented as hazard ratios (HRs) adjusted for relevant variables using the Cox's model with the time- varying covariate approach. The aHRs and IRs will be presented along with 95% CIs for the risk estimates. Country-level estimates will be entered into meta-analysis models to compute pooled estimates. Sensitivity analyses will be performed using the NCC design and marginal structural models (MSM). Whereas the two-stage modelling (or PS approach) is designed to capture all exposed individuals along with a matched sample of comparator groups, the NCC will capture all cases (bladder, renal or other urinary tract cancers) along with a matched sample of controls (individuals with no cancer event) among empagliflozin initiators and the comparator group. For each case, the controls will be sampled randomly from a corresponding risk set defined relevant characteristics of the case. MSMs will be used to adjust for time-dependent confounding throughout follow-up by estimating inverse-probability of treatment weights (IPTW). HRs using the Cox's models will be presented from both sensitivity approaches along with 95% CIs. Odds ratios from conditional logistic regression model will also be presented for the NCC design. Meta-analysis of the country-level estimates will also be carried out for the meta-analyses where relevant		

Name of company: Boehringer Ingelheim			
Name of the finished medicinal product: Jardiance®, Synjardy			
Name of active ingredient: Empagliflozin (A10BK03) Metformin and empagliflozin (A10BD20)			
Protocol date: 16-Sep-2015	**Study number/ Document number:** 1245.97 / c03856813-10	**Version/Revision:** 8.0	**Version/Revision date:** 17-May-2023
Milestones:	Jardiance (empagliflozin) was granted EU marketing authorisation in May 2014. Launch in the UK and Finland was in August 2014; launch in Sweden took place November 2014. Synjardy (empagliflozin/metformin hydrochloride) was granted EU marketing authorisation in May 2015. Protocol version 1 was submitted to the EMA for review in September 2015, version 2 in February 2016, version 3 in June 2016, version 4 in March 2019, and protocol version 5.0 in Sep 2019. Registration in the EU PAS register occurred in December 2016 (EUPAS16424) and was updated following protocol endorsement by the EMA. The number of users in each treatment cohort have been monitored. Based on the available patient numbers and the event rates observed, a decision will be made to proceed with adjusted, treatment-stratified analyses at the interim report stage. Any decision based on the results of the interim reports concerning the length of the study period or the use of other data sources will be approved via a protocol amendment. Taking into account the database lags, the analysis of Swedish and Finnish data through December 2020 will be performed between Q3 2022 and Q3 2023. Analysis of CPRD data through December 2021 will be performed in Q3/Q4 2023 and the final report will be finalised by Q2 2024. Progress reports including monitoring of users in each data source were sent to EMA annually up to 2021.		

4. AMENDMENTS AND UPDATES

Amendment number	Date	Section of study protocol	Amendment or update	Reason
8.0	17 May 2023	5	Update of planned date for final study report	Delayed CPRD approval of study collaborators agreement
7.0	28 June 2022	Multiple	To include data from the Clinical Practice Research Datalink (CPRD) Aurum database in the study.	European Medicines Agency (EMA) requested to reconsider the inclusion of CPRD Aurum as an additional study database in the UK to capture as many CPRD UK patients as possible, increase the sample size, and improve the study power in this country.
		Multiple	Study period extended until 31st December 2021, for the CPRD (GOLD and Aurum).	To increase the sample size, and improve study power in the UK.
		Multiple	To not include data from the Cancer registry from the United Kingdom (UK) in the study.	Data linkage for a sensitivity analysis (Section 8.7.2.3) is not available in time for the final report submission.
		Multiple	To update the power calculations.	EMA requested to update the power calculations for the UK, taking into account the inclusion of CPRD Aurum as additional country-specific data source, and for Finland, taking into account that 1:2 matching ratio may be feasible in this country.

Amendment number	Date	Section of study protocol	Amendment or update	Reason
		8.7.1.4	To briefly describe the appropriate statistical methods.	EMA requested to briefly describe the appropriate methods to combine results obtained from the main analysis performed in the different data sources, including the meta-analytical techniques such as the meta-analysis model and analysis and handling of heterogeneity or describe the process to identify the most appropriate method.
		8.7.2.3	Add a sensitivity analysis for the meta-analysis	To assess the effect of the meta-analysis model choice on the final results for the main analysis of the primary outcomes.
		Multiple	Other minor updates.	To align updates holistically.
6.0	26 January 2022	Multiple	To not include data from the Clinical Practice Research Datalink (CPRD) Aurum database in the study.	The study will have adequate power to demonstrate a significant difference assuming the target hazard ratios with patients from the data sources mentioned in the study (i.e., national registers in Sweden and Finland and CPRD GOLD database). Therefore, additional data from the Aurum database is not needed.

Amendment number	Date	Section of study protocol	Amendment or update	Reason
				Information on Aurum database was removed from this protocol.
		Multiple	To use the highest possible matching ratio between initiators of empagliflozin and (dipeptidyl peptidase-4) DPP-4 inhibitors.	It was observed in the last interim report that there were not enough patients for 1:3 matching in Finland and Sweden. Therefore, 1:1 or highest possible matching ratio will be used in each country.
		Multiple	To remove sodium glucose co-transporter-2 (SGLT-2) inhibitors as a comparison group from the study analyses.	It was observed in the last interim report that the yearly number of patients initiating SGLT-2 inhibitors other than empagliflozin decreased during the study period (2014-2020). The total number of SGLT-2 inhibitors was considerably lower than that of empagliflozin initiators (a 1:1 matching cannot be performed), therefore only patients initiating DPP-4 inhibitors will be used as a primary comparison group.
		8.7.2	Impact of COVID-19 on study results.	The impact of the COVID-19 pandemic on the study will be assessed.
		8.7.2	Impact of metformin use on urinary tract malignancies.	The impact of metformin use on urinary tract malignancies will be

Amendment number	Date	Section of study protocol	Amendment or update	Reason
				assessed.
		8.7	Meta-Analysis	Data from Finland and Sweden cannot be transferred out of the country of origin. Therefore, pooling of all data and pooled analyses cannot be conducted. Information on pooled analyses was removed and instead meta-analysis was added to the protocol where relevant.
		8.4.3	Findata	In Finland, all the data will be delivered to and analyzed in the remote access server environment provided by the national permit authority Findata. The data holders will extract the relevant data and deliver it to Findata for pseudonymization.
		Multiple	Other minor updates.	To align updates holistically.
5.0	23 August 2019	Multiple	Inclusion of new study country, Finland.	To increase of the size of the patient population.
		Multiple	Using DPP-4 inhibitors as the primary comparison group and other SGLT-2 inhibitors as the secondary comparison group.	Comparison with other SGLT-2 inhibitors is expected to be underpowered due to the low uptake of SGLT-2 inhibitors other than empagliflozin.
		Multiple	Extending the study	Increasing the study

Amendment number	Date	Section of study protocol	Amendment or update	Reason
			period until 31st December 2020.	sample size and accounting for the long latency period related to risk of malignancies.
		9.5	Updated power calculations.	The power calculations have been updated as a result of extending the study period.
		9.4.4	Additional information on potential data sources included.	Additional data sources can be included if the study size will be too small with the current data sources.
		Multiple	Details provided on how fixed-dose combination drugs are processed in the study.	Several fixed-dose combination drugs have become available after the previous protocol version was approved.

5. MILESTONES

Milestone	Planned Date	Actual date
Protocol endorsed by the European Medicines Agency (EMA)	Q3 2016	19 Sept 2016
Start of data collection	Expected during Q3 2016, depending on final protocol approval by the EMA	21 Oct 2016
Interim report 24 months	Q1 2017 Will be based on Swedish and the UK data available 24 months after use of empagliflozin is first captured (November 2016)	13 Mar 2017
Interim report 36 months	Q1 2018 Will be based on Swedish and the UK data available 36 months after use of empagliflozin is first captured (November 2017)	02 Apr 2018
Interim report 48 months	Q1 2019 Will be based on Swedish and the UK data available 48 months after use of empagliflozin is first captured (November 2018).	19 Mar 2019
Interim report 60 months	Q1 2020 Will be based on Swedish, the UK and Finnish data available 60 months after use of empagliflozin is first captured (November 2019)	13 May 2020
Interim report 72 months	Q1 2021 Will be based on Swedish, the UK, and Finnish data available 72 months after use of empagliflozin is first captured (November 2020)	23 June 2021

Milestone	Planned Date	Actual date
End of data collection	Q4 2021 Taking into account the database lags, analysis of Swedish and Finnish data through December 2020 will be performed between Q3 2022 and Q3 2023. Analysis of CPRD data through December 2021 will be performed in Q3/Q4 2023.	-
Registration in the EU PAS register[1]	Q4 2016	30 Nov 2016
Final report of study results	Q2 2024	-
Registration of study results in the EU PAS register	Three months following approval of final study report	-

Note: Approvals by data protection, data custodian, ethics, and scientific review bodies are pending. Timelines may be impacted by approvals of these bodies, duration of contract reviews, and availability of data and staff at research institutions once contracts and approvals are finalised.

Sponsor to submit to EMA the results of the interim and final study reports within the earliest corresponding PSURs.

[1] Website for the EU PAS Register: encepp.eu/encepp_studies/indexRegister.shtml.

6. RATIONALE AND BACKGROUND

Jardiance (empagliflozin), a highly potent and selective inhibitor of the sodium glucose co-transporter 2 (SGLT-2), was approved in Europe in May 2014 for the treatment of type 2 diabetes (T2D) to improve glycaemic control in adults. Synjardy (empagliflozin/metformin hydrochloride) was approved in Europe in May 2015. SGLT-2 is highly expressed in the kidney; as the predominant glucose transporter, it is responsible for the reabsorption of glucose from the glomerular filtrate back into the circulation. Empagliflozin improves glycaemic control in patients with T2D by reducing renal glucose reabsorption [R14-4617].

The recommended starting dose is 10 mg empagliflozin once daily. In patients tolerating empagliflozin 10 mg once daily who have an estimated glomerular filtration rate (eGFR) ≥ 60 mL/min/1.73 m2 and need tighter glycaemic control, the dose can be increased to 25 mg once daily. The maximum daily dose is 25 mg [R14-4617].
In Europe, empagliflozin/metformin hydrochloride is available in the following fixed-dose combinations:

- 5 mg empagliflozin plus 850 mg or 1000 mg metformin hydrochloride
- 12.5 mg empagliflozin plus 850 mg or 1000 mg metformin hydrochloride

The recommended dose for the fixed-dose combinations is one tablet twice a day [R17-0153].
Currently, little information is available on whether there is a potential increased risk of cancers of the urinary tract associated with empagliflozin use. Clinical data with empagliflozin has not identified or established a potential mechanism for the development of malignancies. However, “urinary tract malignancies” are listed in the Risk Management Plan (RMP) as a potential risk. The inclusion of renal cancer as a potential risk was based on preclinical toxicology findings and clinical cases of bladder cancer observed with other SGLT-2 inhibitors.

As a part of the RMPs for empagliflozin and empagliflozin/metformin hydrochloride agreed upon by the European Medicines Agency (EMA), a post-authorisation safety study (PASS) will be performed to assess the risk of urinary tract malignancies in incident users of empagliflozin, compared to incident users of dipeptidyl peptidase-4 (DPP-4) inhibitors, which is an alternative treatment option to empagliflozin and is prescribed at a similar diabetes progression stage. Prior to the development of the first version of the protocol for this study, a feasibility assessment of three databases in Europe (the CPRD [GOLD and Aurum] and data with potential for linkage (HES and ONS mortality statistics) in the United Kingdom (UK) and the national registers in Sweden and Finland) was performed with the aim to detail the minimum data requirements for the study and to assess, for each potential data source, the data quality, completeness of recorded information for urinary tract malignancies, and possibility of linkage with cancer registries, mortality registries and other data sources. The first version of the protocol included two study countries: Sweden and the UK. with follow-up until 31st December 2019. In version 6.0 of the protocol, the study follow-up was extended until 31st December, 2020 and Finland national registers were included to increase both accumulated patient-years and the number of patients initiating the study drugs.

In this version of the protocol (Version 7.0), additional steps were taken to increase sample size and follow-up for the UK. CPRD Aurum was added as an additional database to be pooled with CPRD-GOLD and the study period for both CPRD databases was extended to 31st December, 2021.

6.1 EPIDEMIOLOGY OF URINARY TRACT MALIGNANCIES IN DIABETES PATIENTS

Based on data from the national cancer registries throughout the Nordic countries, the crude incidence rates (IRs) for urinary tract malignancies (bladder and renal cancers) in the Nordic countries in 2013 was calculated as 43.0 per 100,000 person-years (PY) for the total population and 82.8 per 100,000 PY for individuals over 40 years of age (Table 1) [R15-4873]. As the study population consists of T2D patients who initiate second- or third-line treatment, the majority of the patients are expected to be over 40 years old. The crude incidences of bladder cancer in the UK, as reported through cancer registries to GLOBOCAN, were lower, as compared to the Nordic countries, but rates for renal cancer were comparable (Table 1) [R15-4860]. In both regions, incidences are higher among men.

Table 1 Crude incidence rates per 100,000 person-years for urinary tract cancers, Nordic countries 2013 and the UK 2012

	Nordic Countries			**The UK**		
	Male	**Female**	**Total**	**Male**	**Female**	**Total**
Total Population						
Bladder	43.7	14.6	29.1	20.5	7.6	14
Renal	17.6	10.2	13.9	19.6	11.5	15.5
UT (bladder + renal)	61.3	24.9	43.0	40.1	19.1	29.5
Age ≥ 40 years						
Bladder	87.3	27.7	56.5	47.2	16.5	31.2
Renal	34.3	18.8	26.3	44.0	23.9	33.5
UT (bladder + renal)	121.6	46.5	82.8	91.3	40.4	64.7

Source: R15-4873; R15-4860
UK=The United Kingdom; UT=Urinary tract

T2D seems to be associated with an increased risk of several types of cancer when compared to the general population. Reviews of recent studies and meta-analyses indicate increased risks for liver, pancreas, colorectal, kidney, bladder, endometrial and breast cancers, as well as non-Hodgkin's lymphoma in patients with T2D [R10-6494, R12-2439, R12-0469]. A population-based cohort study on patients hospitalized with T2D diagnoses in Denmark showed a higher risk of kidney cancers (with the standardized incidence ratios (SIRs) of 1.4 (95% confidence interval (CI); 1.2–1.6) in males and 1.7 (95% CI; 1.4–1.9) in females) and liver cancers (with the SIRs of 4.0 (95% CI; 3.5–4.6) in males and 2.1 (95% CI; 1.6–2.7) in

females) associated with T2D, but not a higher risk for bladder cancer [R15-4869]. Another Swedish inpatient register-based cohort indicated a higher risk of kidney cancer associated with T2D in both women (SIR = 1.7, 95% CI; 1.4-2.0) and men (SIR = 1.3; 95% CI; 1.1-1.6) throughout the duration of follow-up (1-25 years) [R10-6630]. Crude IRs of bladder cancer in patients with T2D range from 38 per 100,000 PY in women from Denmark [R12-1334] to 140-150 per 100,000 PY in men in Sweden and Denmark [R12-0368, R12-1334]. Meta-analyses of observational studies suggest that incidence of kidney cancer in patients with T2D is approximately 40% higher than that in the general population [R13-1812]. The available crude incidence estimates of renal cancer in men with T2D range between 40 and 50 per 100,000 PY in most studies.

Co-existence of hyperinsulinemia and hyperglycaemia in T2D and associated obesity may have a higher associated risk for cancers [R10-6494]. However, there may be additional risks from antidiabetic treatments and insulins, albeit these may be non-significant risks [R10-6494]. Other risk factors of urinary tract cancer are smoking amount and duration]R15-4868], different occupational and environmental exposures [R15-4865], history of kidney/ureter stones [R15-4863], sex, chronic bladder irritation and infections, genetics and family history, chemotherapy and radiation therapy [R15-4866; R15-4856].

6.1.1 Incidence of urinary tract malignancies in the study population in the UK and Sweden observed in the fifth interim report of this study

The country-specific IRs of urinary tract malignancies were estimated in the study population identified in the UK, Sweden and Finland, from 1st August 2014 to 31st December 2020 (2019 for Sweden).

In the UK CPRD GOLD, during a total of 147,020 patient years, 164 urinary tract cancer events were recorded in the study population with an incidence rate of 111.5 (95% CI, 95.7-130.0) per 100,000 patient years. The incidence rate of bladder cancer was 66.7 (95% CI, 54.7-81.2) per 100,000 patient years with a total of 98 events. For renal cancer, the incidence rate was 26.5 (95% CI, 19.4-36.3) per 100,000 patient years with a total of 39 events.

In Sweden, during a total of 210,329 patient years, 346 urinary tract cancer events were recorded in the study population with an incidence rate of 164.5 (95% CI, 148.1-182.8) per 100,000 patient years. The incidence rate of bladder cancer was 104.1 (95% CI, 91.2-118.9) per 100,000 patient years with a total of 219 events. For renal cancer, the incidence rate was 58.0 (95% CI, 48.6-69.3) per 100,000 patient years with a total of 122 events.

In Finland, during a total of 256,947 patient years, 415 urinary tract cancer events were recorded in the study population with an incidence rate of 161.5 (95% CI, 146.7-177.8) per 100,000 patient years. The incidence rate of bladder cancer was 86.8 (95% CI, 76.1-99.0) per 100,000 patient years with a total of 223 events. For renal cancer, the incidence rate was 71.6 (95% CI, 62.0-82.7) per 100,000 patient years with a total of 184 events.

6.2 EPIDEMIOLOGY OF DIABETES AND DIABETES TREATMENT PATTERNS IN THE UK

The prevalence of diabetes has increased in the UK from 2.8% in 1996 to 4.3% in 2005, and the incidence has increased from 2.7 per 1,000 PY in 1996 to 4.4 per 1,000 PY in 2005. During the period 1996-2005, a change in oral glucose lowering medication use has occurred, predominantly from sulfonylureas to metformin [R11-5320]. Moreover, since 2005-2006, the use of thiazolidinediones has decreased due to concerns about cardiovascular safety, which led to suspension of the rosiglitazone marketing authorisation in the European Union in 2010 [R12-1620]. Together with the introduction in the market of DPP-4 inhibitors, this has changed the selection of second-line treatment regimens, as shown in two studies performed in the UK. One was a cohort study performed in the CPRD from 2000 to 2010, which found that the combination of metformin and DPP-4-inhibitors represented 0.7% of all second-line regimens in 2007, but DPP-4 inhibitors were prescribed in 20.2% of all second-line regimens in 2010 [R14-5249]. On the other hand, the combination of metformin and thiazolidinediones (pioglitazone or rosiglitazone) represented 34% of the all second-line regimens in 2007 but only 9.8% in 2010 [R14-5249]. The other study was performed in The Health Improvement Network (THIN) database, where the annual incidence of prescriptions of thiazolidinediones decreased from 1.2 per 1,000 PY in 2007 to 0.8 per 1,000 PY in 2009, at the same time that "other glucose lowering drugs," including DPP-4 inhibitors, increased from 0.2 per 1,000 PY to 1.1 per 1,000 PY [R14-5244].

As part of the feasibility assessment, patient counts for users of empagliflozin and other glucose lowering medications were extracted from CPRD GOLD database in the UK for the period January 2014 through May 2015 (Table 2). During this time, metformin was the most frequently used glucose lowering medication followed by sulphonylureas, and DPP-4 inhibitors; utilization of empagliflozin was still low.

Table 2 Patient counts by Glucose Lowering Medication category in the UK, Sweden, and Finland

Drug category	**The UK[1]**	**Sweden[2]**	**Finland[3]**
Biguanides	149,987	235,342	246,444
Sulphonylureas	64,006	35,890	9,381
Combinations	2,654	2,552	39,937
Alpha glucosidase inhibitors	424	703	None
Thiazolidinediones	9,849	1,535	4,565
DPP-4 inhibitors	30,918	13,430	94,722
Empagliflozin	108	373	24,968
Others, excl. insulins	9,088	15,151	1,867

1 Data from CPRD GOLD, January 2014- July 2015 (end date varies slightly by GP practice). 2 2014. 3 2017

6.3 EPIDEMIOLOGY OF DIABETES AND DIABETES TREATMENT PATTERNS IN SWEDEN AND FINLAND

The total age-standardized prevalence of both pharmacologically and non-pharmacologically treated diabetes was 4.69% in 2012 in Sweden. The age-standardized prevalence of pharmacologically treated diabetes increased from 4.19% and 2.99% in 2005/2006 to 5.08% and 3.46% in 2012/2013 in men and women, respectively [R15-4858]. In Finland, 4.3% of women and 7.4% of men aged 45-74 years are diagnosed with diabetes [P19-01624]. In 2016, 5.7% of the Finnish population were dispensed a diabetes drug with a special reimbursement for the costs.

As part of the feasibility assessment, patient counts for users of empagliflozin and other glucose lowering medications were extracted from the Prescribed Drug Register in Sweden for 2014 (Table 2). During this time, metformin was, by far, the most frequently used glucose lowering medication, followed by sulphonylureas and DPP-4 inhibitors; utilization of empagliflozin was still low.

More recent numbers in Finland indicate that metformin was the most frequently used glucose lowering medication in 2016 (Table 2). Empagliflozin use has had a major increase: the number of patients in Finland has increased from 1,007 in 2015 to 24,968 in 2017 [P19-01956].

6.4 RISK OF BLADDER CANCER IN PREVIOUS TRIALS OF SGLT-2 INHIBITORS

In a clinical trial comparing empagliflozin with placebo, three (0.1%) and nine (0.4%) cases of bladder cancer were observed in persons exposed to 10 mg and 25 mg empagliflozin, respectively, while five cases (0.2%) were observed among those with exposure to placebo [P19-01957]. Further, dapagliflozin and canagliflozin, in comparison to placebo, did not increase the risk of bladder cancer in clinical trials [R19-2814, P19-07186]. The hazard ratios (HRs) for dapagliflozin, and canagliflozin were 0.57 (95% CI 0.35-0.93) and 0.90 (95% CI 0.47-1.72), respectively. Therefore, contrary to the preclinical toxicology findings, SGLT-2 inhibitors did not show an increased risk of bladder cancer in clinical trials.

6.5 RATIONALE FOR THE PROTOCOL AMENDMENT

1. The CPRD Aurum database is added to the study data sources. EMA requested to reconsider the inclusion of CPRD Aurum as an additional study database in the UK to capture as many CPRD UK patients as possible, increase the sample size, and improve the study power in this country and in the study overall.
2. The study period is extended until 31st December, 2021, for CPRD databases (GOLD and Aurum). To increase the sample size and improve study power in the UK, the study period is extended to 31st December, 2021, for the CPRD databases.
3. Data from the Cancer registry from the United Kingdom (UK) is not included in the study. Data linkage for a sensitivity analysis (Section 8.7.2.3) is not available in time for the final report submission.

4. Power calculations are updated using assumptions made from the data available in the 5th Interim Report (2021). EMA requested to update the power calculations for the UK considering (1) the inclusion of CPRD Aurum as additional country-specific data source, and (2) using a 1:2 matching ratio for Finland as this matching ratio should be feasible in this country whilst 1:3 will not.
5. Additional details regarding the meta-analysis method are included. EMA requested to briefly describe the appropriate methods to combine results obtained from the main analysis performed in the different data sources. This includes describing the meta-analytical techniques such as the meta-analysis model and analysis, describing the handling of heterogeneity, and/or describe the process to identify the most appropriate method for running the meta-analysis.

7. RESEARCH QUESTION AND OBJECTIVES

Research Question
The aim of the study is to assess the risk of urinary tract malignancies in patients initiating empagliflozin compared to patients initiating a DPP-4 inhibitor.

Objectives:
Primary objectives:

I. To estimate the adjusted hazard ratios (aHRs) and incidence rates (IRs) of all urinary tract cancers (bladder, renal and other urinary tract cancers) in patients initiating empagliflozin compared to patients initiating a DPP-4 inhibitor
II. To estimate the aHRs and IRs of bladder cancer in patients initiating empagliflozin compared to patients initiating a DPP-4 inhibitor
III. To estimate the aHRs and IRs of renal cancer in patients initiating empagliflozin compared to patients initiating a DPP-4 inhibitor

Secondary objective:
To estimate the aHRs and IR of each primary outcome (all urinary tract cancers, bladder cancer, renal cancer) with respect to:

a) increasing cumulative dosage of empagliflozin
b) dosage of empagliflozin prescribed per use (10mg vs 25mg)
c) time since first dose of empagliflozin

Further objectives:
The following objectives will be explored if adequate sample size is available.

i. To estimate the association of aHRs and IRs of all urinary tract cancers with respect to empagliflozin initiation among sub-groups defined by age, sex, relevant comorbidities, concomitant use of metformin and other relevant treatments
ii. To estimate the aHRs and IRs of non-renal, non-bladder urinary tract cancers in patients initiating empagliflozin compared to patients initiating a DPP-4 inhibitor

Non-renal, non-bladder urinary tract cancers (referred to as other urinary tract cancers) includes ureteral and urethral cancers
The impact of the COVID-19 pandemic on prescription patterns, incidence of diagnosed cancer, diabetes related healthcare utilization and overall cancer screening will be evaluated and reported the final study report.

Further, a summary of the literature review on the latest knowledge about the associations between metformin use and the risk of bladder, renal and other urinary tract cancers will be made.

8. RESEARCH METHODS

8.1 STUDY DESIGN

This is a non-interventional, comparative, cohort safety study based on European healthcare databases and includes databases from the UK, Sweden and Finland.

The study will use an "incident users" design and compare new users of empagliflozin to new users of DPP-4 inhibitors. The index date will be defined as the date on which each identified new user receives the index prescription for empagliflozin, or a DPP-4 inhibitor. The incident-user design avoids comparing a population predominantly composed of first-time users of a newly marketed drug such as empagliflozin with a population of prevalent users of an older drug who may have stayed on the comparator treatment for a longer time and be less susceptible to the events of interest. To avoid the inclusion of prevalent users, patients will be required to have no exposure to empagliflozin or a DPP-4 inhibitor during the available look-back (pre-index) period [R13-1120, R14-4378].

DPP-4 inhibitors have been selected as the comparator group for several reasons. First, the National Institute for Health and Care Excellence (NICE) appraisal of dapagliflozin (an SGLT-2 inhibitor) recommended that dapagliflozin should be used as described for DPP-4 inhibitors. The NICE Evidence Review Group considered that, overall, "DPP-4 inhibitors are the key comparators for dapagliflozin in both the dual therapy and triple therapy settings" [R13-5134]. Second, DPP-4 inhibitors, SGLT-2 inhibitors, and thiazolidinediones have similar indications and target population, while dual therapy with GLP-1 (glucagon-like peptide-1) analogues has a restricted target population [P14-17374] and metformin and sulfonylureas are typically used earlier in the treatment paradigm. Finally, the use of thiazolidinediones has decreased in recent years, given increasing concerns about their safety (including bladder cancer), and at the same time, use of DPP-4 inhibitors increased, making second-line regimens with DPP-4 inhibitors the most common second-line regimens after metformin with sulfonylurea [R14-5244, R14-5249].

Empagliflozin is usually a second- or third-line treatment for T2D; thus, it is expected that few patients with T2D initiating empagliflozin will be treatment naïve. For the majority of patients, empagliflozin will be added to an existing treatment (e.g., added to metformin), or patients will be switched to empagliflozin (e.g., from metformin plus an oral glucose lowering medication other than the study drugs to metformin plus empagliflozin) due to disease progression, treatment failure, or side effects that may be related to study outcome. Analyses will account for treatment complexity (monotherapy, dual combination therapy, or triple combination therapy) to achieve a fair comparison between groups [R13-1120, R14-4378] and concomitant use of metformin (whether a fixed-dose or free combination). Despite the selection of comparators with similar utilization patterns to empagliflozin, the cohorts created with incident users of DPP-4 inhibitor drug group may have inherent channelling bias wherein the treatment assignment is influenced by a range of factors such as disease progression, comorbidity, treatment history, age, and sex. Additionally, some of these variables lead to increased or decreased risk of cancer. To minimize channelling bias, individuals are matched with similar treatment and clinical history at index date by deriving

propensity scores (PS) conditional on factors affecting the treatment and outcome [R13-1120, R14-4378]. Propensity-score-matched cohorts will be derived from the study base by matching patients who initiate empagliflozin with patients who initiate DPP-4 inhibitors at index date with the highest possible matching ratio (1:1, 1:2, or 1:3 using Greedy matching methods). Depending on sample size, combination exposures with metformin will be accounted for through separate PS, stratification, or adjustment. Additional analyses will explore the presence of detection/diagnostic bias as screening for urinary tract signs or symptoms may be more frequently among empagliflozin users. The potential for diagnostic bias will be evaluated and addressed through consideration of the frequency of urine dipstick testing (as captured through albuminuria tests) and stage of cancer at the time of diagnosis (as more early-stage cancers in one group would be suggestive of diagnostic bias).

The primary, secondary, and further analyses utilize a cohort design which will allow direct estimation of the IRs and aHRs of multiple outcomes of interest among new users of empagliflozin compared with new users of a DPP-4 inhibitor. The covariate information will be assessed during the time preceding treatment initiation (varying length look-back (pre-index) period) and during follow-up and will include all historical information available for each patient up until occurrence of outcome or censoring (see Section 8.3.3). Follow-up will start 6 months after the index date (12 months in sensitivity analyses) to account for an empirical induction/promotion period (see Section 8.3.5). In the context of data sources such as the CPRD and the Swedish and Finnish national registers, the use of a cohort design has more advantages than limitations compared with the use of a nested case-control (NCC) design—see the appendix discussion in Schneeweiss (2010) [R13-1120] and Patorno et al. (2014) [R14-4378]. Thus, an observational cohort is constructed to study the risk of urinary tract malignancies among the two treatment groups. Several sensitivity analyses will be conducted including analyses using a NCC design (given the rarity of the outcomes) and analyses using marginal structural models (MSM) to control for time-varying confounders towards switching or discontinuation of primary exposure.

The country-level datasets will be analysed separately. Country-level effect size estimates (IRs, aHRs, etc) will be pooled using suitable meta-analysis methods to compute overall estimates.

8.2 SETTING

8.2.1 Study populations

The source population will include all patients with T2D from the UK, Sweden, and Finland. In the UK, the population will include patients with T2D from the CPRD databases of longitudinal medical records collected from the UK primary care practices (CPRD GOLD and CPRD Aurum). Sensitivity analyses may include a subset of the CPRD GOLD and CPRD Aurum database eligible for linkage to Hospital Episode Statistics (HES) data and/or to Office of National Statistics (ONS) mortality data if linked data with sufficient sample size during the study period are available (see Section 8.4.1). In Sweden and in Finland, patients with T2D in the nationwide registers will be included. All patients with T2D who have purchased at least one prescription of empagliflozin, or a DPP-4 inhibitor during the study

period (2014-2021 for CPRD GOLD and Aurum, 2014-2021 for Sweden and Finland) will be included in the study. Identification of diabetes will be based on medical/prescription records and all non-T2D patients will be excluded.

8.2.2 Study period

The study will include individuals who have purchased at least one prescription of empagliflozin, or DPP-4 inhibitors, from 1st August 2014 through 31st December 2020 (31st December 2021 for CPRD). The starting of the study period corresponds to the time of drug licensing (second half of 2014 in all three countries) and subsequent introduction into clinical practice. The number of users in each treatment cohort will be monitored at 24 months after empagliflozin use starts being captured in the data sources (November 2016) and annually thereafter until 2021. The maximum follow-up time for an individual will be seven years for CPRD GOLD and Aurum and 6 years for Sweden and Finland.

8.2.3 Index date

The index date will be defined as the date of first purchase/prescription of empagliflozin, or a DPP-4 inhibitor, during the study period (2014-2020).

8.2.4 Baseline and look-back (pre-index) period

To characterise the empagliflozin, or DPP-4 inhibitor cohorts at the time of study drug initiation, all information available during the look-back (pre-index) time period will be collected. The look-back time period is defined as the time period ending on the day before the index date. Since all cohort members are required by inclusion criteria to have at least 1 year of data before the index date (baseline period), the look-back period will include at least 365 days during which covariates can be evaluated. For some study participants, data on covariates might be available beyond 1 year prior to the index date; in these cases, all available information will be considered for covariate classification related to diabetes, diabetes medications, and concomitant chronic conditions. For concomitant medications for diseases other than diabetes, the look-back time period will be limited to 180 days prior to the index date.
If the distribution of the duration of look-back time period is different among empagliflozin, or DPP-4 inhibitor categories of look-back time will be created using indicator variables. Those indicator variables will then be used as covariates in the multivariable regression models for outcome prediction, and for PS development, to control for possible differences in availability of information between the empagliflozin and comparator cohorts.

8.2.5 Inclusion and exclusion criteria

The study population will be selected from the broad study population with the following inclusion and exclusion criteria:

Inclusion criteria:

- Diagnosis of T2D

  - To identify the diagnosis in the UK, lists of codes that have been previously used to identify patients with T2D in the CPRD [P14-16383] will be utilized (Annex 3).
  - For Swedish registers, identification of diabetes will be based on prescription records and the National Diabetes Register; all non-T2D patients will be excluded. The final algorithm to identify patients with T2D in each data source, which might include medication codes and glucose/glycated haemoglobin test results, will be described in the statistical/epidemiological analysis plan (SEAP).
  - In Finnish registers, identification of patients with T2D will be based on prescription records and diagnoses available in various data sources; all non-T2D patients will be excluded. More specifically, all persons with dispensations of empagliflozin, or a DPP-4 inhibitor will be identified. Patients with non-T2D will be identified with the following criteria: special reimbursement of diabetes drugs with an ICD-10 diagnosis for a non-T2D, type 1 diabetes mellitus (T1D), pre-existing T1D before pregnancy, gestational diabetes, malnutrition-related to diabetes mellitus, neonatal diabetes mellitus, other specified diabetes mellitus, diabetes mellitus due to underlying condition, and drug- or chemical-induced diabetes mellitus. The specific diagnosis codes will be included in the SEAP.
- Age over 18 years at index date.
- At least 1 year of membership in the medication database prior to index date; in CPRD, at least 1 year of continuous up-to-standard registration in the CPRD prior to the index date.
  - This is the minimum look-back period and is informative of the history of treatments and other medical conditions in the recent past. Some of the important covariates to be accounted for in matching at index date are derived from the look-back period.

Exclusion criteria:

- Patients with any cancer (excluding non-melanoma skin cancer) recorded at any time prior to the index date (i.e., during the available look-back time).
- Diagnosis of T1D or other specific non-T2D.
- Use of any SGLT-2 inhibitor or any DPP-4 inhibitor (including free and fixed-dose combinations) recorded at any time prior to index date (i.e., during the available look-back time).
- Use of fixed-dose combinations of SGLT-2 inhibitors with DPP-4 inhibitors.
- Diagnosis of or procedures related to end stage renal disease, receipt of renal dialysis recorded, or eGFR ≤15 ml/min at any time prior to index date (i.e., during the available look-back time).

8.3 VARIABLES

The final list of variables and operational definitions will be presented in a separate SEAP to be developed prior to the start of data analysis.

8.3.1 Exposures

Empagliflozin (current ATC code A10BK03, previous ATC code A10BX12) is the exposure of interest, free or fixed-dose combination with metformin (ATC code: A10BD20).

The DPP-4 inhibitor comparator exposure group will consist of the following drugs, free or fixed-dose combination with metformin (See Section 8.1 and Annex 4).

- Alogliptin
- Linagliptin
- Saxagliptin
- Sitagliptin
- Vildagliptin

If additional DPP-4 inhibitors are introduced in any of the study countries during the study period, they will also be considered to be members of the corresponding comparator group. Note that use of any of these index drugs prior to index date is an exclusion criterion. Information on whether patients received prior oral glucose lowering therapy or if they were "added on" or "switched to" empagliflozin, or a DPP-4 inhibitor at the time of inclusion in the study will be collected. Patients will be classified according to their treatment complexity as receiving monotherapy, dual combination therapy, or triple combination therapy and according to use of metformin combination therapy (fixed-dose or free combination). (See Section 8.1)

The primary definition of exposure, i.e., as-treated analysis (AT), is defined as an ongoing exposure (free or fixed-dose combination with drugs other than DPP-4 inhibitors and SGLT-2 inhibitors) to one of the following two index drugs or drug groups: empagliflozin or DPP-4 inhibitors (Table 3, Annex 4). Exposure is defined as use of these drugs during the study period starting from 6 months after the index date (see Sections 8.2.3 and 8.3.5)
As a typical prescription length is one month in the UK and three months in Sweden and Finland, multiple prescriptions are required for a patient to have continuous ongoing treatment at 6 months after index date. Therefore, patients with only one prescription are unlikely to contribute to the AT analysis.

Secondary definitions of exposure include cumulative dosage of empagliflozin use, daily dosage of empagliflozin prescribed per use (10mg vs 25mg), and time since first dose of empagliflozin (Table 3, Annex 4).
Sensitivity analyses of exposure include intention-to-treat (ITT) analyses.

Table 3 Exposure definitions for primary and secondary analyses

Exposure definition	**Description**	**Time-dependent or Fixed at index date**
Current exposure to study treatment (as-treated analysis)	Indicator of current use of empagliflozin, or a DPP-4 inhibitor, either alone or as a combination.	Time-dependent
Exposure to Empagliflozin (Intention-to-treat)	Drug purchased/prescribed at index date, which is one of empagliflozin, or a DPP-4 inhibitor.	Fixed at index date
Cumulative dosage of empagliflozin use	Time-dependent cumulative sum of drug consumption based on the daily defined dosage of empagliflozin since entry into the study cohort.	Time-dependent
Daily dosage of empagliflozin per use	The dosage of empagliflozin prescribed per daily use (10mg vs 25mg).	Time-dependent
Time since first dose of empagliflozin	Time-dependent cumulative sum of duration since the first use of (exposure to) empagliflozin.	Time-dependent
Treatment complexity	Monotherapy, dual combination therapy, or triple combination therapy (based on predefined groups).	Fixed at index date
Switch	Switch from the drug purchased/prescribed on index date to any other drug in the primary exposure.	Time-dependent
Concomitant use of insulin	Was empagliflozin an add-on to insulin (yes/no).	Fixed at index date
Concomitant use of metformin	Was empagliflozin an add-on to metformin (yes/no).	Fixed at index date
Concomitant use of metformin	Concomitant use of metformin either as fixed-dose combination or free combination.	Time-dependent
Ever used other oral glucose lowering medication	Use of other oral glucose lowering drugs based on predefined drug groups in the past.	Time-dependent
Ever used metformin	Was metformin ever used – taking value ‘yes’ as soon as one prescription with metformin is purchased.	Time-dependent

Exposure definition	Description	Time-dependent or Fixed at index date
Ever used sulfonylureas	Was sulphonylureas ever used – taking value 'yes' as soon as one prescription with sulphonylureas is purchased.	Time-dependent
Ever used thiazolidinediones	Was thiazolidinediones ever used – taking value 'yes' as soon as one prescription with thiazolidinediones is purchased.	Time-dependent
Ever used GLP-1	Was GLP-1 ever used – taking value 'yes' as soon as one prescription with GLP-1 is purchased.	Time-dependent
Ever used insulin	Was insulin ever used – taking value 'yes' as soon as one prescription with insulin is purchased.	Time-dependent

GLP=glucagon-like peptide

8.3.2 Outcomes

8.3.2.1 Primary outcomes

The outcomes of interest for this study are urinary tract cancers, bladder cancer, and renal cancer. These outcomes will be identified through CPRD GOLD and CPRD Aurum data and the national cancer registries in Sweden and Finland improving confidence in the accuracy and validity of these outcome diagnoses (See Section 8.4).

Date of diagnosis of the first incidence of urinary tract cancers as specified in CPRD GOLD and CPRD Aurum, or in the Swedish or Finnish cancer registry after the entry into the study will be used as the primary outcome date. The urinary tract cancer definition will include malignant neoplasm and carcinoma in situ of the urinary tract. Only the first occurrence of a urinary tract cancer will be captured. Patients with a prior history of cancer, including a prior urinary tract cancer, will be excluded.

In the Swedish and Finnish datasets urinary tract cancer events will be identified from cancer registers using ICD-10 and ICD-O-3 codes (Annex 5). The reporting of all cancer cases is compulsory, and the completeness and quality of the register data are considered good for scientific research [R15-4864, R15-4867, R15-4857, R15-4871].

In CPRD, urinary tract cancer events will be identified in CPRD GOLD and CPRD Aurum using READ and SNOMED codes (Annex 5). A study evaluating the validity of cancer diagnoses in CPRD compared with cancer registry data concluded that recording of cancer diagnoses in the two sources was generally consistent. The predictive value of a CPRD GOLD diagnosis of urinary tract cancer was 92%. Taking all records of cancer in the general

practitioner (GP) data, 4% did not occur in the cancer registry data. Differences were generally due to different dates of diagnosis or tumour types [R15-4862].

Sensitivity analyses based on other outcome definitions as well as using a subset of CPRD GOLD and CPRD Aurum with linkage to HES data (Outpatient, HES OP; Admitted Patient Care, HES APC) data, will be performed to address some of the accuracy and validity issues in the primary outcomes (see Section 8.7.2).

8.3.2.2 Further Outcomes

The further outcome of interest for this study are non-renal, non-bladder urinary tract cancers (referred to as other urinary tract cancers). Non-renal, non-urinary bladder urinary tract cancers include ureteral and urethral cancers.

8.3.3 Covariates

Variables potentially associated with urinary tract cancers, such as socio-demographic variables including age, sex, socioeconomic status, body mass index, smoking and alcohol consumption, laboratory measurements, concomitant medications, comorbidities, and duration of look-back period (see Section 6.1), will be identified for study participants as recorded prior to the index date (see Annex 6 for list of covariates to be included in the study). With the exception of exposure to toxic chemicals such as arsenic, known risk factors for urinary tract cancers are captured in CPRD and the national registers from Sweden and Finland.

Study participants will also be classified by indicator variables on the calendar time of cohort entry (by quarter) whether the index treatment (empagliflozin, or DPP-4 inhibitor) was added to existing medication (adding on), or if the index treatment was initiated as a replacement for another antihyperglycemic medication (switching to empagliflozin, or DPP 4 inhibitor), and whether this treatment was received as monotherapy or as dual or triple combination therapy. A variable indicating whether or not patients were receiving insulin at the index date will also be created.

The confounding due to covariates in this section are to be accounted for at minimum by design or analysis when estimating the risk of urinary tract cancers with the defined exposures (Table 4). Accounting for by design means specifying of variables for exact or propensity matching, risk set definition in NCC approach and/or covariates for deriving inverse-probability of treatment weight (IPTW) in MSMs. Accounting for by analyses means stratifying on and adjusting for these covariates in the analytical models.
The examples of covariates in Table 4 describe the comorbidities, treatment history, and socio-demographic variables that will considered for inclusion. These variables along with times of occurrence/diagnosis of comorbidities, times of prescription/dispensing of drugs and times of recording of the socio-demographic variables in combination with the exposure and outcome variables define the minimal required dataset.

Table 4 Baseline covariates to be considered for inclusion alternatively or simultaneously in primary, secondary, and further analyses

Covariate	**Exact matching**	**PS matching**	**Stratification**	**Risk set (NCC) / covariate in MSM[3]**	**Model adjustment[4]**
Database variable (for CPRD only)[1]					
Patient database (Aurum vs GOLD)	Yes	Yes			
Socio-demographic variables					
Age	Yes	Yes	Yes	Yes	Yes
Sex	Yes	Yes	Yes	Yes	Yes
Socioeconomic status (if available)		Yes		Yes	Yes
Calendar year of index date		Yes		Yes	Yes
Duration of look-back period		Yes		Yes	Yes
Insulin concomitant (or cumulative duration of) use		Yes	Yes	Yes	Yes
Other oral antihyperglycemic medication use		Yes		Yes	Yes
Comorbidities					
Diabetic complications - diabetic nephropathy- diabetic neuropathy - peripheral vascular diseases - lower limb severity		Yes		Yes	Yes
Urinary tract-related comorbidities		Yes		Yes	Yes
Kidney or genitourinary stones		Yes		Yes	Yes

Covariate	**Exact matching**	**PS matching**	**Stratification**	**Risk set (NCC) / covariate in MSM[3]**	**Model adjustment[4]**
Prior history of UTI or pyelonephritis		Yes		Yes	Yes
Liver disease		Yes		Yes	Yes
eGFR[2]		Yes		Yes	Yes
Renal impairment		Yes		Yes	Yes
Prior ICU admission		Yes		Yes	Yes
Pancreatitis		Yes		Yes	Yes
BMI (when available)		Yes		Yes	Yes
Smoking (when available)		Yes		Yes	Yes
Alcohol use (when available)		Yes		Yes	Yes
HbA1c (when available)		Yes		Yes	Yes
Albuminuria testing		Yes		Yes	Yes
Cardiovascular Diseases (HF, Hypertension, stroke, MI)		Yes		Yes	Yes
Blood pressure measurements (when available)		Yes		Yes	Yes
Hypertension		Yes		Yes	Yes
Stroke		Yes		Yes	Yes
MI		Yes		Yes	Yes
Autoimmune disease		Yes		Yes	Yes
COPD		Yes		Yes	Yes
Time since first diabetes diagnosis		Yes		Yes	Yes
Other non-diabetes medications					
Antihypertensives/diureti		Yes			Yes

Covariate	Exact matching	PS matching	Stratification	Risk set (NCC) / covariate in MSM[3]	Model adjustment[4]
cs					
Non-steroidal anti-inflammatory drugs (NSAIDs)		Yes			Yes
Oral steroids		Yes			Yes
Statins, fibrates		Yes			Yes
Lipid modifying agents		Yes			Yes
Zoledronic acid		Yes			Yes
Antibiotics		Yes			Yes

1 This variable will be present only for CPRD data. It will indicate which database (GOLD or Aurum) the patient data was extracted from-

2 Time-dependent eGFR estimate will be computed with the CKD-Epi equation from serum creatinine measurements, where available.

3 Sensitivity analyses

4 Covariates considered for inclusion if not balanced after PS matching

BMI=body mass index; COPD=chronic obstructive pulmonary disease; eGFR=estimated glomerular filtration rate; HbA1c=glycated haemoglobin A1c; HF=heart failure; ICU=intensive care unit; MI=myocardial infarction; MSM=marginal structural model; NCC=nested case-control; PS=propensity score; UK CPRD=The United Kingdom Clinical Practice Research Datalink; UTI=urinary tract infection.

8.3.4 Censoring of follow-up time

The following events lead to censoring: all-cause mortality, emigration, patient transfer, meeting specific exclusion criteria (see exclusion criteria in Section 8.2.5) and end of study period. Follow-up times will be censored at any of these events, whichever occurs first. In the primary AT analysis discontinuation of index drug treatment is an additional censoring variable. Treatment discontinuation is defined either by

- A switch from the index drug to any other drugs (empagliflozin, DPP-4 inhibitor, or other SGLT-2 inhibitor), or to a fixed-dose combination of a SGLT-2 inhibitor with a DPP-4 inhibitor (identified from the ATC class A10BD) plus a 3-month period following the switch, or
- As the stop date plus a 3-month period following it. The stop date is the end date of the first continuous exposure period of the index drug, where continuous treatment is defined as having consecutive prescriptions separated by 30 days or less (Figure 1).

The 3-month period after the switch or the stop date is used to allow for a delay in the diagnosis of cancer [R16-2539]. In a sensitivity analysis this time period is extended to 6 months.

Switch to a fixed-dose combination including the same index drug and drugs other than DPP-4 inhibitors and SGLT-2 inhibitors is not considered a censoring variable (for example, switch from sitagliptin to a fixed-dose combination including sitagliptin and pioglitazone).

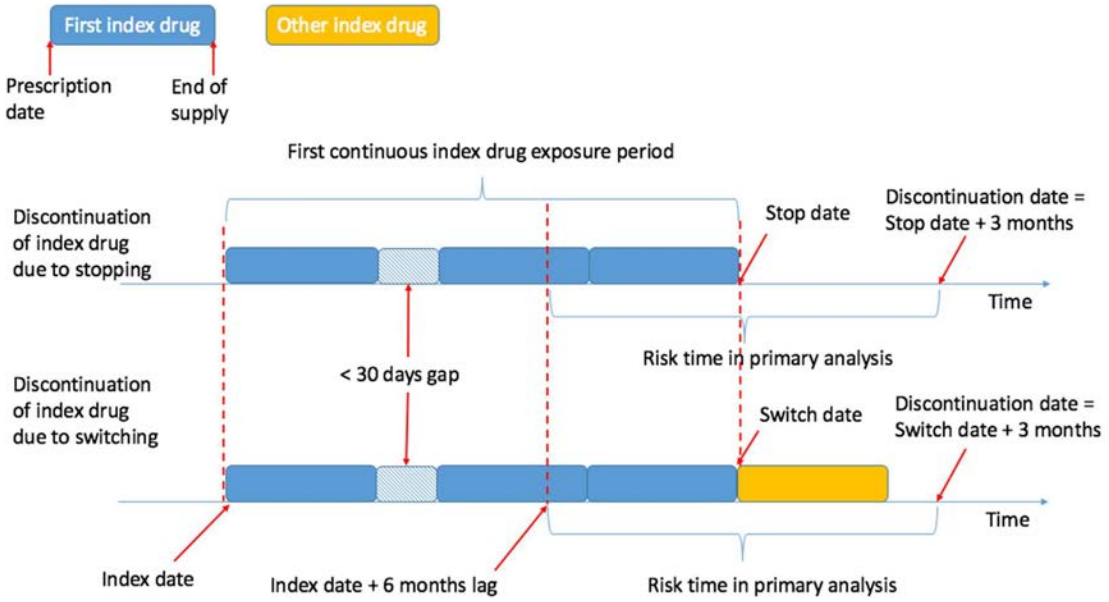

Figure 1 Discontinuation of index drug due to stopping of index drug or switching of index drug. In the primary analysis follow-up is censored at treatment discontinuation.

8.3.5 Follow-up time

For the primary AT analyses, the follow-up time (time at risk) will begin 6 months after the index date (lag period to account for empirical induction/promotion period of cancer) and end at urinary tract (bladder, renal or other) cancer or at any of the censoring events including discontinuation of index drug treatment.

For the sensitivity ITT analyses, the follow-up time will begin 6 months after the index date and end at urinary tract (bladder, renal or other) cancer or any of the censoring events excluding treatment discontinuation of index drug.
Patients with censored follow-up time will not be able to re-enter study cohort at a later time point. Follow-up will not be censored if oral or injectable glucose lowering drugs other than the index drugs are prescribed in addition to empagliflozin, a DPP-4 inhibitor, or a SGLT-2 inhibitor after the index date.

The lag period of 6 months is to account for an empirical induction/promotion period. Information in the first 6 months after the index date (i.e., first 6 months of follow-up) will be excluded, including patients with less than 6 months of follow-up and all occurrences of cancer within the first 6 months of follow-up, in order to consider a biologically meaningful induction/promotion time window. Sensitivity analyses will expand the lag period to 12 months.

8.4 DATA SOURCES

Prior to the development of the protocol for this study, a feasibility assessment of three databases in Europe (CPRD [GOLD and Aurum] and data with potential for linkage [HES (Outpatient and Admitted Patient Care), and ONS mortality statistics] in the UK and the national registers in Sweden and Finland) was performed with the aim to detail the minimum data requirements for the study and to assess, for each potential data source, the data quality, completeness of recorded information for urinary tract malignancies, and possibility of linkage with cancer registries, mortality registries and other data sources. The minimal data requirements for data sources which were assessed are listed in Table 5.

Table 5 Minimum data requirements and data availability for the UK and the national registries in Sweden and Finland.

	The UK			**Sweden**					**Finland**						
	CPRD (GOLD and Aurum)	**HES**	**Mortality**	**Patient**	**Prescription**	**Cancer**	**Cause of**	**Diabetes**	**HILMO**	**AvoHILMO**	**Special Reimb**	**Prescriptions**	**Cause of**	**Cancers**	**EMR databa**
Age	x	x	x	x	x	x	x	x	x	x	-	x	x	x	x
Sex	x	x	x	x	x	x	x	x	x	x	-	x	x	x	x
Medication	x	-		x[3]	x	-	-	x[5]	-	-	-	x	-	x	-
DDD	x	-	-	x	x	-	-	-	-	-	-	x	-	-	-
Purchase date	x[1]	-	-	x	x	-	-	-	-	-	-	x	-	-	-
Diagnoses	x	x	-	x[4]	-	x	-	x	x	x	x[7]	x[9]	-	x	-
Date of diagnosis	x	x	-	x	-	x	-	-	x	x	x[8]	-	-	x	-
Time of death	-	-	x	-	-	x	x	-	-	-	-	-	x	x	-
Cause of death	-	-	x	-	-	x	x	-	-	-	-	-	x	x	-
Smoking	x	-	-	-	-	-	-	x	x[6]	x[6]	-	-	-	-	x
BMI	x	-	-	-	-	-	-	x	x[6]	x[6]	-	-	-	-	x
HbA1c	x	-	-	-	-	-	-	x	-	-	-	-	-	-	x

x = available in the register, - = not available in the register

1 Prescription date; 2 Month and year; 3 Treatment given in hospitals (no full coverage); 4 No primary care data; 5 Only type of medication; 6 Partial availability based on applicable diagnosis codes; 7 Only chronic diseases included in the special reimbursement scheme; 8 Start and end dates of the special reimbursement; 9 In case a special reimbursement was included in the purchase

AvoHILMO=Register of primary health care visits, BMI=body mass index, CPRD=Clinical Practice Research Datalink, DDD=Defined daily dose, EMR=electronic medical record, HbA1c=Glycated haemoglobin A1c, HES=Hospital Episode Statistics (Outpatient data, and Admitted Patient Care data), HILMO=Care register for health care; UK =The United Kingdom

8.4.1 The United Kingdom Clinical Practice Research Datalink (CPRD)

The UK CPRD data contains the anonymised longitudinal medical records managed by GPs working in the National Health Service (NHS) primary care setting. The data contains demographic information, diagnoses, prescriptions, tests and referrals. This study will use data extracted from CPRD GOLD [R16-1231] and CPRD Aurum [R19-1534].

CPRD GOLD
CPRD GOLD data [R16-1231] cover approximately 8.8% of the UK population, including practices in England, Wales, Scotland and Northern Ireland. Data has been collected prospectively since 1987. Approximately 54% of CPRD GOLD practices take part in the English linkage program (CPRD GOLD-HES subset), which includes patient-level linkage to the inpatient data (HES), allowing for hospitalization diagnoses and procedures to be captured [R19-2928]. Records in both settings include the unique NHS number, which is used for linkage purposes. In a similar way, CPRD GOLD data is linked to the UK mortality data (including date and cause of death). Protocols using CPRD GOLD and linked data are subject to approval by an Independent Scientific Advisory Committee.

CPRD Aurum
CPRD Aurum was established in October 2017 [R19-1534], and therefore, its feasibility could not be assessed before starting the present study.

CPRD Aurum data cover approximately 13% of the population in England (as of September 2018) [R19-1534]; however, recruitment of additional practices is ongoing. Similar to CPRD GOLD, the data available in CPRD Aurum can be linked with other data sources using the NHS number. In total, 93% of the participating practices take part in the linkage program (R19-2928). The same variables are available in CPRD GOLD and in CPRD Aurum.

The main difference between CPRD Aurum and CPRD GOLD is that the data has been recorded from different practices using different software [R19-1534]. As practices migrate from one software to the other, they will also migrate from the GOLD database to the Aurum database. Duplicate historical data will then be available for patients from these practices in both GOLD and Aurum. This possibility will be taken into account as both CPRD GOLD and CPRD Aurum data are incorporated in this study. The data holder (CPRD) will provide a list of migrated (GOLD to Aurum) practices. As Aurum will provide the longest follow-up, the data of patients from migrating practices present in both databases will be removed from the GOLD data but retained in the Aurum data.

Linkage
Based on the feasibility assessment before this study, the UK data sources evaluated (CPRD GOLD and data with potential for linkage - HES, ONS mortality statistics) met the minimal data requirements (Table 5). Similar results were expected for the CPRD Aurum. The following UK data sources are considered for linkage with CPRD: HES (Outpatient data and Admitted Patient Care data), and Office for National Statistics (ONS) mortality sources. In the UK, the full study data including follow-up period until the end of 2021, will be available

in March 2022. The primary analyses will be based on CPRD GOLD and CPRD Aurum data; sensitivity analyses may include linked HES (Outpatient and Admitted Patient Care data) and ONS data, if linked data with sufficient sample size during the study period are available (Section 8.7.2.3.3).

To determine renal impairment, eGFR values will be extracted from CPRD. Additionally, information on important covariates such as proxies of socioeconomic position, number of albuminuria tests, blood pressure measurements will be extracted.

8.4.2 Swedish national registries

Sweden has a well-developed population-wide register system with longitudinal follow-up data. The persons are identified in the registers with a unique personal identification number (PID) and thus the records can be linked for research purposes on subject level between the registers. Permissions to use the data may be received upon providing a research plan to the authorities responsible for the registers. The protocol is also subjected to relevant Ethical Committee´s review and approval. The data permit processes and timelines for data permit granting by the authorities responsible for the registers vary between different registers and countries, as well as the requirements for the local research collaboration and the publication of results [R15-4861].

The nationwide prescription register contains information on outpatient medication purchases in pharmacies. All prescribed medicines purchased in community pharmacies are included irrespective of reimbursement status.
Based on the feasibility assessment, the Sweden national registers met the minimal data requirements (Table 5). The following data sources will be used to construct the Sweden cohort by linkage: The National Patient Register, the Swedish Prescribed Drug Register, the Swedish Cancer Register, the Causes of Death Register, the National Diabetes Register, and the Longitudinal integration database for health insurance and labour market studies.

To determine renal impairment, eGFR values will be extracted from the National Diabetes Register. Information on important covariates such as number of albuminuria tests and blood pressure measurements will also be extracted from the National Diabetes Register. Further, proxies of socioeconomic position will be obtained from the Longitudinal integration database for health insurance and labour market studies.
The study data will be extracted from the relevant registers using the following general process. First, patients in the study cohort with the unique PID number are identified from the prescription register. A unique dummy study identification number (SID) is created for each PID by the prescription register holder. The list with the PID-SID pairs will be provided from the prescription register holders to the other register holders. Each register holder extracts the relevant data according to the study protocol and links the data to the PID-SID list. Subsequently, the register holders decode the data by destroying the key between the PID and SID permanently and provides the de-identified study data to the applicant. Thus, only de-identified data will be provided for performing the analysis of the study data. In Sweden, the full study data including follow-up period until the end of 2020 will be available during in March 2022.

8.4.3 Finnish data sources

An additional, feasible data source in Europe includes the national registers in Finland, which were evaluated in the feasibility assessment before this study.

As in Sweden, Finland has a well-developed population-wide register system with longitudinal follow-up data. The persons are identified in the registers with a unique PID and thus the records can be linked for research purposes on subject level between the registers. Further, the data obtained from registers can be linked with data obtained from the EMR databases.

The following Finnish nationwide data sources will be included: Care register for health care (HILMO), Register of primary health care visits (AvoHILMO), Special reimbursement register, Prescription register, Cause of death register, Cancer register, Population register centre, and Statistics Finland (Table 5). HILMO and AvoHILMO include diagnoses associated with healthcare encounters in specialized and primary care, respectively. The Special reimbursement register will be used for identification of chronic comorbidities. This register includes patients who are entitled for special refunds of medicine expenses based on their chronic conditions. The Finnish nationwide Prescription register contains information on outpatient medication purchases in pharmacies. The register contains information on all purchased prescribed medicines within the reimbursement scheme. Cause of death register includes time of death and the cause of death classified according to the ICD categorization. Cancer register includes information on all diagnosed cancers, details of the tumour, and description of treatment.

The study data will be extracted from these registers using the general process as described above for Sweden (Section 8.4.2). Information about highest level of education will be obtained as a proxy for socioeconomic status from Statistics Finland, and information on patient's place of residence and migration will be obtained from Population register centre.

Further, to obtain data on important potential confounders, specifically BMI, blood pressure, smoking, and laboratory measures such as HbA1c and eGFR, EMR databases will also be included in the study. Four regional EMR databases (Helsinki, Espoo, Vantaa, and Helsinki University Hospital [HUS] area), covering both primary and secondary care in the capital area of Finland, will be contacted separately to obtain access to their data. The data obtained from different data sources will be linked as described previously (Section 8.4.2).

Permissions to use the data may be received upon providing a research plan to the authorities responsible for the registers. The protocol is also subjected to relevant Ethical Committee´s review and approval. The data permit processes and timelines for data permit granting by the authorities responsible for the registers vary between different registers and countries, as well as the requirements for the local research collaboration and the publication of results [R15-4861].
In Finland, all the data will be delivered to and analysed in the remote access server environment provided by the national permit authority Findata. The data holders will extract the relevant data and deliver it to Findata for pseudonymization. Before data delivery to the

Findata server environment, the data holders will collect and manage data according to their own standards. After the data is delivered to the Findata server environment by the data holders, [redacted] will process the data.

[redacted] will use the study identification numbers for data linkage on individual level. [redacted] will have access to pseudonymized raw data without PINs. Therefore, the researchers at [redacted] will have access to data where individuals cannot be directly identified.

In Finland, the full study data including follow-up period until the end of 2020 will be available in June 2022.

8.4.4 Additional potential data sources

Beside the data sources evaluated originally in the UK, Sweden, and Finland, none of the holders of other databases have been contacted for participation in the study so far, and further feasibility assessments are required to insure the validity of urinary cancer case identification. It is important for the choice of the appropriate data source to have the appropriate information on cancer and important risk factors included.

Potential additional data sources for this study have been assessed (Annex 1). In addition to those described above, Danish nationwide registers, PHARMO Database Network (Netherlands), Spanish regional registers, and Maccabi Healthcare Services database (Israel) are feasible for inclusion in the study.

8.5 STUDY SIZE

The study size will be driven by the uptake of empagliflozin and the primary comparator (DPP-4 inhibitors) in the UK, Sweden, and Finland. Originally, it was expected that by the end of 2019, approximately 18,000 empagliflozin-treated patients will accumulate in the CPRD and Swedish databases. However, based on the conservative estimates from the fifth interim analysis (Annex 1), the expected sample size is higher than anticipated. Taking into account the addition of Finland, the addition of CPRD Aurum, and the extension of the study period (end of 2020 for Sweden and Finland, end of 2021 for CPRD), a total of approximately 67,490 empagliflozin users are expected to be eligible for inclusion in the study (Table 6). However, the number of patients available for analysis may be lower. Given the observed patient numbers, it is expected that the maximum achievable matching ratios will be 1:3 for UK CPRD (GOLD and Aurum), 1:1 for Sweden, and 1:2 for Finland.

Table 6 Observed and expected number of patients eligible for inclusion in the study population, by study drug group (including fixed-dose combinations with metformin), by study country and overall, using an arithmetic growth assumption.

Year	Study drug group	Finland	Sweden	CPRD GOLD	CPRD Aurum4	Pooled
2014[1]	Empagliflozin	0	0	5		5
	DPP 4-inhibitors	5,824	3,214	2,555		11,593
	Other SGLT-2 inhibitors	127	662	683		1,472
2015[1]	Empagliflozin	91	629	157		877
	DPP 4-inhibitors	14,130	10,305	6,851		31,286
	Other SGLT-2 inhibitors	320	1,537	2,065		3,922
2016[1]	Empagliflozin	2,201	2,181	897		5,279
	DPP 4-inhibitors	11,391	12,126	5,648		29,165
	Other SGLT-2 inhibitors	1,463	1,183	1,740		4,386
2017[1]	Empagliflozin	3,725	4,811	1,391		9,927
	DPP 4-inhibitors	7,402	13,677	4,577		25,656
	Other SGLT-2 inhibitors	1,475	1,191	1,493		4,159
2018[1]	Empagliflozin	4,293	8,738	1,536		14,567
	DPP 4-inhibitors	6,959	12,798	4,158		23,915
	Other SGLT-2 inhibitors	1,610	1,030	1,470		4,110
2019[1]	Empagliflozin	5,182	10169	1,846		17,197
	DPP 4-inhibitors	6,158	11,673	3,622		21,453
	Other SGLT-2 inhibitors	1,853	1,061	1,665		4,579
2020[2]	Empagliflozin	2,777	5,806	1,541		10,124
	DPP 4-inhibitors	2,215	4,688	2,210		9,113
	Other SGLT-2 inhibitors	1,201	653	1,369		3,223
2021[3]	Empagliflozin			969		
	DPP 4-inhibitors			1,051		
	Other SGLT-2 inhibitors			726		
Total	Empagliflozin	18,269	32,334	8,342	8,545	67,490
	DPP 4-inhibitors	54,079	68,481	30,672	59,283	212,515

Year	Study drug group	Finland	Sweden	CPRD GOLD	CPRD Aurum4	Pooled
	Other SGLT-2 inhibitors	8,049	7,317	11,211	25,277	51,854

1 Observed number of patients.

2 Observed number of patients until 30th June for Sweden and Finland and until 31st December for UK CPRD GOLD, taking into account the extension of the study period to the end of 2021 for UK CPRD.

3 Expected number of patients (arithmetic growth assumption given observed numbers in previous years) between 1st January and 30th June, taking into account the 6 months minimum follow-up.

4 Only aggregate level data up (2014-2020) of was available for CPRD Aurum. Patient numbers were extrapolated based on assumptions derived from CPRD GOLD data, taking into account the extension of the study period to the end of 2021

DPP-4=dipeptidyl peptidase-4 inhibitors; SGLT-2=sodium glucose co-transporter-2 inhibitors; UK=the United Kingdom

The power to detect relative risks of 1.5 and 2 was determined for each of the study outcomes for each individual country (Table 7) and for the final random-effects meta-analysis (Table 8). For the single-country analyses, power computations were based on HRs derived from Cox models. For the meta-analysis, simulations were used to estimate country-level HRs which were then entered into random-effects meta-analysis models as described in Section 8.7.1.4. All power computations were based on the expected number of individuals exposed to empagliflozin eligible for inclusion in the study and the observed country-level IRs of urinary tract cancers, bladder cancer and renal cancer in the study population as reported in the 5th Interim Report. The expected averages of follow-up time were 2.4 years for UK CPRD, 2.2 for Sweden, and 2.4 for Finland, as reported in the 5th Interim Report. Matching ratios of empagliflozin initiators to initiators of DPP-4 inhibitors were fixed at 1:3 for UK CPRD, 1:1 for Sweden, and 1:2 for Finland. These are the highest achievable matching ratios given expected patient numbers. For computing the meta-analysis power, these assumptions were used to simulate for each country separately individual level time-to-event data from each drug cohort following an exponential distribution given the expected number of empagliflozin initiators, the expected matching ratio, the observed incidence rates as the baseline rate in the DPP-4 cohort. The hazard ratio for each simulation run was determined by drawing from a normal distribution with a mean of the natural logarithm of the target hazard ratio (1.5 or 2.0) and a standard deviation of 0.1. These individual time-to-event data were then entered into a Cox proportional hazard model for each country to compute a simulated hazard ratio. For each simulation, the country-level hazard ratios were entered into a random-effects meta-analysis level, and the resulting p-value of the pooled estimate was retained. Simulations were run 500 times and power was computed as the proportion of simulations for which the p-value of the pooled estimate was inferior to 0.5.
The power calculations show that anticipated study sample size would ensure sufficient power at the meta-analysis level to detect the targeted effect size of < 1.6 for urinary tract cancers and bladder cancer and the targeted effect size of < 2.0 for renal cancer, with at least 80% power and a type 1 error probability of 5% under the expected matching ratios. Power at the meta-analysis level to detect an effect size of 1.5 is expected to be of 93.6% for urinary tract cancer and 84.0% for bladder cancer. Power at the meta-analysis level to detect an effect size of 2.0 is expected to be of 97.0% for renal cancer.

Table 7 Power for different effect sizes for single-country analyses based on Cox models with a type 1 error probability of 5%, given country-level observed incidence rates, expected numbers of empagliflozin initiators, expected follow-ups, and expected matching ratios.

Effect size (Hazard ratio)	**Urinary tract cancer**	**Bladder cancer**	**Renal cancer**
UK CPRD[1]			
1.5	63.7%	43.1%	20.3%
2.0	99.1%	91.7%	55.9%
Sweden[2]			
1.5	85.1%	66.6%	42.9%
2.0	>99.9%	99.0%	89.4%
Finland[3]			
1.5	80.6%	54.3%	46.7%
2.0	99.9%	96.8%	93.4%

1 Includes both CPRD GOLD and Aurum; Observed IRs of 111.5, 66,7, and 26.5 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected follow-up of 2.4, matching ratio of 1:3, and expected number of empagliflozin initiators of 16,360.

2 Observed IRs of 164.5, 104.1, and 58.0 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected follow-up of 2.2, matching ratio of 1:1, and expected number of empagliflozin initiators of 32,334.

3 Observed IRs of 161.5, 86.8, and 71.6 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected follow-up of 2.4, matching ratio of 1:2, and expected number of empagliflozin initiators of 18,269.

UK CPRD=The United Kingdom Clinical Practice Research Datalink

Table 8 Power of meta-analyses for a type 1 error probability of 5% for different effect sizes, given country-level observed incidence rates, expected number of empagliflozin initiators, expected average follow-ups, and expected matching ratios.

Effect size (Relative risk / hazard ratio)	**Urinary tract cancer**	**Bladder cancer**	**Renal cancer**
1.5	93.6%	84.0%	68.2%
2.0	>99.9%	99.6%	97.0%

For UK CPRD (GOLD and Aurum): observed IRs of 111.5, 66,7, and 26.5 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected follow-up of 2.4, matching ratio of 1:3, and expected number of empagliflozin initiators of 16,360.

For Sweden: observed IRs of 164.5, 104.1, and 58.0 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected follow-up of 2.2, matching ratio of 1:1, and expected number of empagliflozin initiators of 32,334.

For Finland: observed IRs of 161.5, 86.8, and 71.6 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected follow-up of 2.4, matching ratio of 1:2, and expected number of empagliflozin initiators of 18,269.

The minimum detectable effect sizes for each outcome were computed given the observed and expected IRs, average follow-ups, matching ratios, and number of empagliflozin initiators for single-country analyses as well as for a random-effects meta-analysis (Table 9). At the meta-analysis level, the minimum detectable effect sizes were 1.37 for urinary tract cancers, 1.47 for bladder cancer, and 1.62 for renal cancer.

Table 9 Minimum detectable effect sizes (Hazard Ratios) given country-level observed incidence rate, expected sample sizes, expected average follow-ups, and matching ratios for single-country analyses and meta-analyses for a type 1 error probability of 5%.

	Urinary tract cancer	**Bladder cancer**	**Renal cancer**
UK CPRD[1]	1.62	1.82	2.40
Sweden[2]	1.46	1.60	1.85
Finland[3]	1.50	1.70	1.78
Meta-analysis	1.37	1.47	1.62

1 Includes both CPRD GOLD and Aurum; Observed IRs of 111.5, 66,7, and 26.5 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected follow-up of 2.4, expected number of empagliflozin initiators of 16,360, and matching ratio of 1:3.

2 Observed IRs of 164.5, 104.1, and 58.0 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected follow-up of 2.2, expected number of empagliflozin initiators of 32,334, and matching ratio of 1:1.

3 Observed IRs of 161.5, 86.8, and 71.6 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected follow-up of 2.4, expected number of empagliflozin initiators of 18,269, and matching ratio of 1:2.

UK CPRD=The United Kingdom Clinical Practice Research Datalink

The study sizes for each individual country required to detect a relative risk of 1.5 and 2.0, with 80% power under the expected country-level matching ratios and observed country-level IRs were computed for each study country individually (Table 10) and for a random-effects meta-analysis (Table 11). At the meta-analysis level, the required country-level sample sizes to detect an effect size of 1.5 for urinary tract cancer are 9,076, 17,377, and 9,818 for UK CPRD, Sweden, and Finland, respectively. For detecting an effect size of 2.0 for renal cancer at the meta-analysis level, the required sample sizes are 6,534, 12,510, and 7,069 for UK CPRD, Sweden, and Finland, respectively.

Table 10 Required sample sizes in each study country to estimate different effect sizes for the three primary endpoints in the single-country analyses with at least 80% power and a type 1 error probability of 5%, given country-level incidence rates, expected average follow-ups, and expected matching ratios.

Primary endpoints	**HR=1.5**			**HR=2**		
	UK CPRD[1] (1:3)	Sweden[2] (1:1)	Finland[3] (1:2)	UK CPRD[1] (1:3)	Sweden[2] (1:1)	Finland[3] (1:2)
All Urinary tract cancers	25,810	28,151	17,999	7,072	8,665	5,276
Bladder cancer	41,550	44,523	33,543	11,846	13,707	9,835
Renal cancer	104,474	79,991	40,665	29,788	24,630	11,924

1 Includes both CPRD GOLD and Aurum; Observed IRs of 111.5, 66,7, and 26.5 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively and expected follow-up of 2.4.

2 Observed IRs of 164.5, 104.1, and 58.0 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively expected follow-up of 2.2.

3 Observed IRs of 161.5, 86.8, and 71.6 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively expected follow-up of 2.4.

HR=Hazard ratio; UK CPRD=The United Kingdom Clinical Practice Research Datalink

Table 11 Required sample sizes in each study country to estimate different effect sizes for the three primary endpoints in the meta- analyses with at least 80% power and a type 1 error probability of 5%, given country-level incidence rates, expected average follow-ups, and expected matching ratios.

Primary endpoints	**HR=1.5**			**HR=2.0**		
	UK CPRD[1] (1:3)	Sweden[2] (1:1)	Finland[3] (1:2)	UK CPRD[1] (1:3)	Sweden[2] (1:1)	Finland[3] (1:2)
All Urinary tract cancers	9,076	9,076	9,076	2,561	4,903	2,770
Bladder cancer	15,505	29,687	16,773	4,058	7,769	4,390
Renal cancer	25,252	48,350	27,318	6,534	6,534	6,534

1 Includes both CPRD GOLD and Aurum; Observed IRs of 111.5, 66,7, and 26.5 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively and expected follow-up of 2.4.

2 Observed IRs of 164.5, 104.1, and 58.0 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively expected follow-up of 2.2.

3 Observed IRs of 161.5, 86.8, and 71.6 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively expected follow-up of 2.4.

HR=Hazard ratio; UK CPRD=The United Kingdom Clinical Practice Research Datalink

The average follow-up time required to achieve 80% power to detect HRs of 1.5 and 2.0 with the expected number of patients (Table 6) and observed IRs was evaluated for each of the study outcomes for the country-level analyses (Table 12) and for a random-effects meta-analysis (Table 13).

Table 12 Required follow-up time (years) to achieve 80% power in country-level analyses with a type 1 error probability of 5% for different effect sizes for the three primary endpoints given country-level observed incidence rates, expected sample sizes, and expected matching ratios

Matching	Hazard ratio	Urinary tract cancer	Bladder cancer	Renal cancer
UK CPRD[1]				
1:3	1.5	3.03	4.73	11.13
	2.0	1.22	1.71	3.54
Sweden[2]				
1:1	1.5	1.99	2.86	4.73
	2.0	0.96	1.23	1.81
Finland[3]				
1:2	1.5	2.38	4.00	4.75
	2.0	1.05	1.53	1.75

1 Includes both CPRD GOLD and Aurum; Observed IRs of 111.5, 66,7, and 26.5 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected number of empagliflozin initiators of 16,360, and matching ratio of 1:3.

2 Observed IRs of 164.5, 104.1, and 58.0 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected number of empagliflozin initiators of 32,334, and matching ratio of 1:1.

3 Observed IRs of 161.5, 86.8, and 71.6 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected number of empagliflozin initiators of 18,269, and matching ratio of 1:2.

UK CPRD=The United Kingdom Clinical Practice Research Datalink

Table 13 Required follow-up time (years) in each study country to achieve 80% power for the meta-analysis with a type 1 error probability of 5%for different effect sizes for the three primary endpoints given country-level observed incidence rates, expected sample sizes, and expected matching ratio

Matching	Hazard ratio	Urinary tract cancer	Bladder cancer	Renal cancer
UK CPRD[1]				
1:3	1.5	1.38	2.05	3.08
	2.0	0.74	0.91	1.20
Sweden[2]				
1:1	1.5	1.38	2.04	3.07
	2.0	0.74	0.91	1.20
Finland[3]				
1:2	1.5	1.50	2.22	3.35
	2.0	0.81	0.99	1.31

1 Includes both CPRD GOLD and Aurum; Observed IRs of 111.5, 66,7, and 26.5 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected number of empagliflozin initiators of 16,360, and matching ratio of 1:3.

2 Observed IRs of 164.5, 104.1, and 58.0 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected number of empagliflozin initiators of 32,334, and matching ratio of 1:1.

3 Observed IRs of 161.5, 86.8, and 71.6 per 100,000 person-years for UTC, bladder cancer, and renal cancer respectively, expected number of empagliflozin initiators of 18,269, and matching ratio of 1:2.

UK CPRD=The United Kingdom Clinical Practice Research Datalink

Based on the feasibility assessment before this study and the interim reports, utilization of empagliflozin in the UK and Sweden is still low. Accrual of empagliflozin users is monitored in the data sources as described in Section 8.7.1.7. The number of users in each treatment cohort has been monitored at 24 months after empagliflozin use started being captured in the UK and Swedish data sources (November 2016) and has and will be monitored annually thereafter until November 2020. The numbers were reported in annual interim reports in the 1st quarter of years 2017-2021 and the results of the interim reports were submitted to EMA by the Sponsor within the earliest corresponding PBRER. Initially, this study included two countries, the UK (CPRD GOLD) and Sweden, with follow-up until 31st December 2019. In order to increase the accumulated patient years and number of patients initiating the study drugs as well as accounting for the long latency period related to risk of malignancies, the study follow-up was first extended until 31st December, 2020 and Finnish national registers were included. Additionally, in this version of the protocol (Version 7.0) CPRD Aurum was included and the study period for both CPRD databases was extended to 31st December, 2021 in order to further increase the sample size and follow-up, and, consequently, study power.

8.6 DATA MANAGEMENT

Full audit trail starting from raw data obtained from register holders and ending with statistical tables and graphs in reports will be maintained. Data management, tabulations, graphics, and statistical modelling will be carried out with R data analyses language (http://www.r-project.org). R language is described more detailed in report "R: Regulatory Compliance and Validation Issues: A Guidance Document for the Use of R in Regulated Clinical Trial Environments" (http://www.r-project.org/doc/R-FDA.pdf). Source code of data management and data analyses is kept for inspection for five years after publication of results. The study may be inspected by the sponsor's study independent representative(s) or inspected by the competent authorities. A SEAP detailing all statistical analyses will be drawn and used as a basis for analysis.

Data from multiple country sources will be sent utilizing password protected media. All study related datasets will be stored in secured server environment. Access to data will be permitted only to study statisticians and data managers in line with the data permits. All data used for this study will be anonymous such that no study individual can be directly identified. Data will be regularly backed-up and stored in a separate secure location. All data and back-ups will be maintained within the EU.

8.7 DATA ANALYSIS

The final approach to data analysis will be presented in a separate SEAP to be developed prior to the start of data analysis. The SEAP will include a detailed algorithm for identifying T2D patients, detailed methods for covariate selection, checking for empirical equipoise, model checking for outcome models, methods for time-dependent exposure and time-varying confounding, the approaches for handling missing data, the assessment of the potential impact of unmeasured confounding, and the critical sensitivity analyses. The SEAP will be made available upon request to EMA and will be included or appended in the study report.

8.7.1 Main analysis

For the primary, secondary and further objectives, the main data analysis will be conducted in two stages: (i) construction of the PS-matched cohort by modelling the exposure to empagliflozin vs. DPP-4 inhibitors and (ii) estimating the effect of exposure to empagliflozin on the urinary tract (bladder, renal and other) cancers using aHRs and IRs compared to those exposed to DPP-4 inhibitors. Additional sensitivity analyses will be performed to validate the robustness of alternative definitions of outcome, exposure, covariates and reduction of bias due to matching.

8.7.1.1 Matching at index date

For all individuals, PS for the comparison of empagliflozin vs DPP-4 inhibitors will be computed at the index date as the probability to receive empagliflozin conditional on treatment and clinical history up to index date within each stratum. The PS will be estimated

using a binary logistic regression model. Depending on sample size, combination exposures with metformin will be accounted for through separate PS, stratification, or adjustment.

The selection of variables to be included in the PS models (see Section 8.3.3) will be based on scientific basis from previous studies, clinical/epidemiological significance and examination of exposure group differences in the distribution of each covariate. The covariates included in the PS model should ideally be associated with the outcomes of interest. Simulation studies show that variables that are unrelated to the exposure but are related to the outcome should always be included in the estimation of PS [R12-1913]. Inclusion of these variables increases the precision of the estimated effect of exposure without increasing bias. In contrast, inclusion of variables that are related to the exposure but not to the outcome can decrease precision of the estimated effect of exposure without decreasing bias. The variables listed in Annex 6 are potential candidates for inclusion in the PS model. Depending on sample size, combination exposures with metformin and treatment complexity (monotherapy, dual combination therapy, triple combination therapy) will be accounted for through separate PS, stratification, or adjustment.

Matching will be performed for empagliflozin initiators with patients initiating DPP-4 inhibitors within each country database. Matching will be carried out in several rounds to achieve the highest possible matching ratio given the available data in each country (up to 1:3 matching) [P12-07757]. In the first round of matching, for each empagliflozin initiating individual i with a particular index date TiE, all individuals initiating other DPP-4 inhibitors within three months before or after TiE will be chosen. Of these, one individual with the closest PS to that of i within a predefined PS interval will be selected as a match (Greedy matching method). The predefined PS interval will be calculated based on the standard deviation of the logit of PS (δ) multiplied by some coefficient k (say, kδ, when k=0.2). The value of the coefficient k will be calibrated to ensure at least 99% of those exposed to empagliflozin have matches from each comparator. The selected comparators are then removed from the pool of the possible comparators for further selection (sampling without replacement). Once the first round of matching is done, additional rounds of matching will be carried out to identify a second and possibly a third match for each already matched for each empagliflozin initiating individual (variable ratio matching with 1:3 as the maximum ratio). The number of matched DPP-4 inhibitors, and hence the overall matching ratio, will therefore be maximized within each country. Matching by time period, based on index dates, enables better control of time-varying confounders including changing prescription patterns for empagliflozin and potential confounding.

8.7.1.2 As-treated (AT) analysis

For the incident users of empagliflozin along with their respective matches, the primary analysis will be performed using the AT approach. This corresponds to censoring individuals who discontinue use of the index drug, i.e., either switch from the index drug to other drugs (empagliflozin, DPP-4 inhibitors, or other SGLT-2 inhibitors) during follow-up, or stop using the index drug (see Section 8.3.4).

8.7.1.3 Population summary and descriptive analyses

Univariate and bivariate distribution of exposures, outcomes and relevant covariates will be presented using absolute and relative frequencies for categorical variables and summary measures for continuous variables (mean, standard deviation, median, min, max). Bivariate distributions for outcomes and covariates will be with reference to primary exposure definitions.

Absolute risks (IRs) for the defined outcomes will be estimated within each exposure category and key covariates using Poisson regression. The IRs will be presented along with 95% CIs. Other descriptive summaries such as number of events and person-time for calculating the IR will be included.

Kaplan-Meier plots, with 95% CIs, for survival probabilities will be presented for all urinary tract cancers, bladder cancers and renal cancers with primary exposures using the AT definition of follow-up times.

8.7.1.4 Comparative analyses

Relative risks (HRs) for the defined outcomes will be estimated using Cox models with time-varying covariates if model pre-requisites are fulfilled. HRs will be presented along with 95% CIs.

The Cox models will include the exposure and the comparator group within the model. This will account for multiple testing and for retaining same variables in variable selection (see Section 8.7.1.6). The models are defined in terms of covariates adjusted for in addition to the primary exposure: (i) baseline model with covariates adjusted for with values at index date (ii) model with time-dependent covariates wherein change in covariates over the follow-up period after index date are also incorporated into the model.

Country-specific analysis:
The country-specific Cox models will be fit with single-country level datasets. The analyses will use datasets from the UK, Sweden, and Finland, separately.

Meta-analysis:
Country-level effect sizes will be pooled using meta-analysis models to compute pooled effect size estimates. Prior to conducting meta-analyses, heterogeneity across the countries in terms of study design and conduct will be assessed. Specifically, the variability in population characteristics (including data availability), propensity score matching variables as well as propensity score balance will be examined.

Meta-analyses will be carried out for analyses where the country-level effect size is available for all three study countries. Pooled effect sizes will be estimated using random-effects models. The inverse-variance method will be used where appropriate.

Statistical heterogeneity between country-level estimates will be assessed using:

- Cochran's Q test
- I^2 statistic and its 95% CI
- The τ^2 statistic and its 95% CI

For each meta-analysis model, results will be presented in a forest plot including at least the following data:

- Data source.
- Effect size and 95% CI for each study included in the analysis (on a log scale).
- The weights allocated to each study.
- The combined estimated effect size with 95% CI (on a log scale) for the random-effects model.
- Value of Cochran's Q statistic, with the number of degrees of freedom and p-value.
- The proportion of variability explained by heterogeneity (I^2).
- The between-study variance τ^2.

If high levels of heterogeneity exist, the population characteristics and statistical models in the different country-level analyses will be reviewed and the possible source of heterogeneity discussed. A sensitivity analysis for the main analysis of the primary outcomes will be carried out using fixed-effect models to ascertain the effect of between-study heterogeneity on the meta-analysis result.

Further details regarding the meta-analysis models and heterogeneity will be defined in the study SEAP.

8.7.1.5 Variable selection

All potential confounders, exact and PS matching variables will enter the outcome model only after being selected using predefined variable selection criteria to be detailed in the SEAP.

8.7.1.6 Monitoring of accrual of empagliflozin users

Accrual of empagliflozin users, free or fixed-dose combinations with metformin, has been monitored starting at 24 months after use of empagliflozin was first captured in the UK and Swedish data sources (November 2016), and the numbers will be monitored annually until 2020. The purpose of this monitoring is to estimate the event rates for the outcomes of interest to confirm the event rates described in the literature and used in the sample size analyses. Finland was added in protocol version 5.0 and the final report will include data from the UK CPRD GOLD, CPRD Aurum, and National registers in Sweden and Finland.

The available power to estimate the association between empagliflozin use and the outcomes will be examined. If there is insufficient power (<80%) because, given the current event rates, the number of new users of empagliflozin accrued up to that point is too low to yield acceptable precision, decisions will be made about extending the length of the study period, the need to use additional country data sources, and/or changing the primary study design to a nested case-control study.

8.7.1.7 Handling of missing data

A high frequency of missing values is not expected for most variables, with the possible exception of lifestyle and laboratory parameters. For the medical history conditions/ comorbidities to be collected for inclusion in the PS, the absence of a code for a condition will be interpreted as an absence of the event. If a study variable is totally missing from a database (CPRD GOLD or Aurum, Sweden or Finland), it is excluded from the analysis of the pooled data. If a variable is missing for only some of the patients a missing data category will be added and utilized in the analysis.

8.7.2 Further analysis

All of the following analyses will be first carried out at the single country-level. Pooled effect sizes will then be estimated by entering country-level effect sizes into meta-analysis models as described in Section 8.7.1.4.

8.7.2.1 Stratified analysis

Depending on sample size, stratified analyses may be performed by estimating the HR from the fully adjusted model (with all variables after variable selection, excluding the stratifying variable) within the sub-groups of relevant variables (both matched by design or otherwise).

The analyses will be stratified at least by the following sub-groups.

- Concomitant metformin use (either as fixed-dose or free combination)
- Treatment complexity (monotherapy, dual combination therapy, triple combination therapy)

Stratification by additional sub-groups, including age, sex, and concurrent use of metformin, may also be performed if adequate sample size is available.

8.7.2.2 Analyses with respect to secondary exposure definitions

The AT exposure definition will be used for the primary analysis. Other time-dependent empagliflozin exposure definitions from Section 8.3.1 (current use, cumulative dosage, dosage per use and time since first use) will be utilized in secondary analyses.

8.7.2.3 Sensitivity Analyses

Additional sensitivity analyses may be defined in the SEAP.

8.7.2.3.1 Sensitivity analyses: study design

As sensitivity analyses to the two-stage modelling, two additional data analyses approaches will be implemented: NCC design and MSMs.

Whereas the two-stage modelling (or PS approach) is designed to capture all exposed individuals along with a matched sample of comparator group, the NCC will capture all cases (bladder, renal or other urinary tract (UT) cancers) along with a matched sample of controls

(individuals with no cancer event) among empagliflozin initiators and the comparator group. For each case, five controls will be sampled randomly from a corresponding risk set defined by relevant characteristics of the case. HRs from Cox model and odds ratios from conditional logistic model along with 95% CIs will be presented from the NCC approach.
Whereas the two-stage modelling is used to control for channelling bias at index date, MSMs will be used to adjust for time-dependent confounding throughout follow-up by estimating IPTW. HRs using the Cox's models will be presented along with 95% CIs.

8.7.2.3.2 Sensitivity analyses: exposures

For the incident users of empagliflozin along with their respective matches, a sensitivity analysis will be performed using the ITT approach. This corresponds to assuming individuals do not discontinue treatment with the index drug. Thus, follow-up is not censored at switching drug or stopping use of index drug during follow-up, but instead assuming continuous use until end of follow-up.

8.7.2.3.3 Sensitivity analyses: outcomes

Sensitivity outcomes include

- Non-renal, non-bladder UT cancers
- Cancers of the UT from the linked records in the UK, including CPRD GOLD, CPRD Aurum, and HES (Outpatient and Admitted Patient Care)), if possible (if linked data with sufficient sample size during the study period is available)
- Neoplasms of uncertain or unknown behaviour (includes non-malignant and excludes carcinomas in situ) of the UT

All sensitivity outcomes will be presented using descriptive tables and adjusted IRs as stratified by primary exposure.

In addition, the potential for diagnostic bias will be evaluated and addressed through descriptive analysis of the frequency of urine dipstick testing (as captured through albuminuria tests) and stage of cancer at the time of diagnosis (as more early-stage cancers in one group would be suggestive of diagnostic bias).

8.7.2.3.4 Sensitivity analyses: follow-up times

Careful consideration of right-censoring and left-truncation of follow-up time allows the exploration of cancer latency and mechanism of empagliflozin as an initiator versus a promoter of cancer. Follow-up times will be left-truncated for 12 months to exclude all cancers occurring within 12 months from the index date. Information in the first 12 months after the index date (i.e., first 12 months of follow-up) will be excluded, including patients with less than 12 months of follow-up and all occurrences of cancer within the first 12 months of follow-up, in order to consider a biologically meaningful latency time window. The possible delay in the cancer diagnostic pathway is handled in the primary analysis by extending the discontinuation time up to 3 months after switching or stopping the index drug. [See Section 8.3.4] In a sensitivity analysis this 3-month delay time will be extended up to 6 months.

8.7.2.3.5 Sensitivity analyses: extension of follow-up for UK-CPRD data

To increase sample size and follow-up and thus, power, the study period was extended to the end of 2021 for UK-CPRD data. Results from the COVID-19 impact assessment in the 5th IR suggested that 2020 saw fewer healthcare encounters and UTC diagnoses than expected in the CPRD-GOLD data. In this sensitivity analysis, the impact of including study years 2020 and 2021 in the UK-CPRD analyses will be assessed by running the main analysis of the primary outcomes including only data for the 2014-2019 time-period.

8.7.2.3.6 Sensitivity analyses: meta-analysis

The meta-analysis will be run using a random-effects model under the assumption that there exists statistical heterogeneity between the different study countries in the potential effect of the study-drug on UTC cancer risk. In this sensitivity analysis, the effects of this assumption will be assessed by carrying out the meta-analysis using a fixed-effect model.

8.7.2.4 COVID-19 impact assessment

The impact of COVID-19 on prescription patterns, incidence of diagnosed cancer, diabetes related healthcare utilization, and overall cancer screening will be assessed. Prescription patterns will be investigated by computing the average number of prescriptions of study drugs across drug initiators for each quarter of each study year, separately for each study drug. IRs for UT cancers across the entire source population will be computed for each study year separately. Diabetes related healthcare utilization will be assessed by computing the average number of diabetes related healthcare visits for each study year. Additional details will be provided in the SEAP. Impact on overall cancer screening in the study countries will be assessed by a literature review.

8.7.2.5 Metformin impact assessment

The current knowledge about the associations between metformin exposure and the risk of bladder, renal and other UT cancers will be assessed by a review of the literature.

8.8 QUALITY CONTROL

The study will be conducted as specified in this protocol and SEAP. All revisions to the protocol shall be properly documented as protocol changes/amendments and when necessary, such protocol amendments will be delivered to register holder(s) whenever amendment(s) to the data permissions are required.

The study protocol has been written by following the Code of Conduct by the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) [R15-4870] that provides a set of rules and principles for post-authorisation studies with regard to the best practices and transparency, thereby promoting scientific independence of such studies. ENCePP is a project led by the EMA to further strengthen the post-authorisation monitoring of medicinal products in Europe by facilitating the conduct of multi-centre, independent, post-authorisation studies focusing on safety and on benefit/risk. The study will

be registered to the ENCePP's E-register. The results of this study will also be published on the same site.

The study protocol also follows the key elements of the Guideline for Good Pharmacoepidemiology Practices by International Society for Pharmacoepidemiology [R11-4318], and the recent draft Guidance for Industry and FDA Staff "Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets" [R15-4859].

All programs for data management and data analyses will be written by study statistician(s). Quality control check of these programs will be carried out by a statistician other than the one who writes the program. All processes from data management leading to dissemination of study results will undergo quality control checks for programs, result tables and written text. A detailed audit trail of all documents (programs, result tables, reports) along with quality control processes will be maintained.

8.9 LIMITATIONS OF THE RESEARCH METHODS

There are several methodological challenges to conducting epidemiological studies evaluating the comparative safety of glucose lowering medications [P12-13528, P14-17457, R14-4378, R13-1120, R10-6638].

The last (fifth) interim report (ANNEX 1) presented sample sizes of the study populations. It was noted that SGLT-2 inhibitor users were few and so that group was removed from analyses as stated in the amendments and updates, Section 4.

The actual use of prescribed or purchased drugs, especially oral glucose lowering medications cannot be verified with certainty. Drug exposure will be defined by algorithmic approaches based on the amount dispensed/prescribed in a prescription. This may be subject to misclassification. Sensitivity analyses will be performed to assess the impact of the exposure definitions on the main results.

The time-dependent nature of the exposure also creates a challenge due to time-dependent confounding. Although PS matching is used to control for imbalances between exposed and unexposed cohorts at the time of index date, other factors potentially associated with both exposure and outcome after index date are not controlled by PS matching. This may be tackled by use of causal models (e.g., marginal structural models, inverse-probability of weighting approach) as sensitivity analysis.

The AT approach used as the primary analysis may result in exclusion of cancer cases due to censoring at treatment discontinuation of the index drug. This will be addressed in sensitivity analyses as an ITT approach will be conducted. The ITT approach is conservative and is subject to exposure misclassification, but it ensures the evaluation of all cases occurring during the study period. For this reason, we will conduct sensitivity analyses considering a longer time period after treatment discontinuation. Finally, we will apply a 6-month lag

period after the index date to the follow-up time in both the AT and ITT analyses to exclude subjects with only a short exposure time.

Determining the appropriate induction/promotion time window is challenging and typically only empirically based. Sensitivity analyses will evaluate a longer latency time window (12 months).

Although many variables may be included in the study, there will be relatively limited information (including age, sex, previous hospitalizations, chronic diseases) about risk factors connected to malignancies of UT from the available databases. For example, information on exposure to toxic chemicals is not available in the study databases. We attempt to control for the confounding bias by using baseline and time-dependent variables of potential confounders in the statistical analyses.

A potential limitation of using PS matching is that any exposed patient without a matched control would be excluded from the analyses. This will be tackled by applying a wider matching caliber for the PS for all unmatched cases after the first round of matching.

8.9.1 Bias

Potential source of difference in cancer incidence may arise due to channelling bias, detection bias and confounding. The channelling bias is addressed through both two-stage modelling approach wherein individuals are first matched on PS conditional on variables that affect exposure and outcome. Confounding will be minimized through model adjustment after carefully selecting variables using the variable selection criteria. However, there may be confounding due to unobserved variables and/or incomplete adjustment. Detection bias is inherent in cancer studies, especially those where the outcome is based on cancer notification. Although the notification is complete, early stages may go undetected. In studies including exposures with publicized safety concerns, diagnostic bias is also a concern, as patients with the exposure may be more likely to receive screenings for the outcome. The potential for diagnostic bias will be evaluated and addressed through consideration of the frequency of urine dipstick testing (as captured through albuminuria tests) and stage of cancer at the time of diagnosis (as more early-stage cancers in one group would be suggestive of diagnostic bias).

Bias due to definitions of exposure, definition of follow-up times and other relevant covariates will be addressed through sensitivity analyses. As mentioned in Section 8.9, the AT approach used as the primary analysis is subject to omission of cancer cases due do censoring after treatment switch or treatment discontinuation. An ITT approach will therefore be used in sensitivity analyses.

8.9.2 Generalizability

In the current study three population-based data sources (CPRD GOLD, CPRD Aurum, and Swedish and Finnish national registries) will be used. CPRD GOLD provides data entered by primary care practitioners in a routine clinical care setting and covers approximately 8.8% of

the UK population and CPRD Aurum about 13% of England population The patients are broadly representative of the UK population in terms of age, sex and ethnicity. In Sweden, the national registries cover practically all patients with T2D. In Finland, the coverage of the EMR data is approximately 26.7%, and study population is broadly generalizable. Therefore, the study results can be generalized to similar patients with T2D in other geographic settings, including most industrialized countries.

9. PROTECTION OF HUMAN SUBJECTS

To ensure the full data protection of patients, all the research data in each country is anonymised. The implications of the General Data Protection Regulation (EU) 2016/679 on the national legislations, during the course of the study, will be considered. Approval from relevant Ethical/Research Review Boards will be required before conducting the study.

10. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

Based on current guidelines from the International Society for Pharmacoepidemiology [R11-4318] and the EMA [R13-1970], non-interventional studies such as the one described in this protocol, conducted using medical chart reviews or electronic claims and health care records, do not require expedited reporting of suspected adverse events/reactions. Specifically, as stated in section VI.C.1.2.1 of Guideline on Good Pharmacovigilance Practices (GVP), Module VI – Management and Reporting of Adverse Reactions to Medicinal Products, for non-interventional study designs, which are based on use of secondary data, reporting of adverse reactions is not required.

The data generated in the course of the study will be monitored by the BI responsible person.

When an observation is identified that may qualify as a special safety issue or that may have implications for the benefit-risk balance of empagliflozin, appropriate BI functions will be notified according to BI standard operating procedures.

11. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

Study reports will be prepared using a template following the Guideline on Good Pharmacovigilance Practices (GVP), Module VIII, Section B.6.3 [R13-5420]. The principal and co-investigators will write the annual interim reports and the final study report. The reports will be delivered to the Sponsor. The Sponsor will submit the interim and the final reports to EMA in line with the milestones (see Section 5) and report the results within the earliest corresponding PSURs; the final report will be submitted to the EMA as a type II variation. The study results will be published in the EU PASS registry.

The principal and co-investigators will co-author scientific manuscript(s) of the results to be published. The publication strategy is defined in the research agreement between the principal investigators and the Sponsor. A summary of the main results of the study, whether positive or negative and including results from prematurely terminated studies, will always be made available to the public. Per Section V of Guidelines for Good Pharmacoepidemiology Practices (GPP) [R11-4318] and the Guideline on Good Pharmacovigilance Practices, Module VIII, Section B.7 [R13-5420]. The outcome of a study will always be presented in an objective and truthful manner providing a comprehensive and accurate description of the findings. In no way shall the interpretation and presentation of the results be aimed towards any commercial, financial or personal interests.

The Sponsor is entitled to view the final results and interpretations thereof prior to submission for publication and to comment in advance of submission as agreed in the research contract and without unjustifiably delaying the publication.

12. REFERENCES

12.1 PUBLISHED REFERENCES

P12-07757 Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ, Schneeweiss S. One-to-many propensity score matching in cohort studies. Pharmacoepidemiol Drug Saf 2012 ; 21(Suppl 2) ; 69-80.

P12-13528 Lo Re V, 3rd, Haynes K, Ming EE, Wood Ives J, Horne LN, Fortier K, et al. Safety of saxagliptin: rationale for and design of a series of postmarketing observational studies. Pharmacoepidemiol Drug Saf 2012; 21 (11): 1202-15.

P14-16383 Macedo AF, Douglas I, Smeeth L, Forbes H, Ebrahim S. Statins and the risk of type 2 diabetes mellitus: cohort study using the UK clinical practice pesearch datalink. BMC Cardiovasc Disord. 2014; 14: 85.

P14-17374 NICE. Type 2 diabetes: the management of type 2 diabetes. NICE clinical guideline 87. National Institute for Health and Care Excellence; 2014. Available at: website: nice.org.uk/guidance/cg87/resources/guidance-type-2-diabetes-pdf. Accessed 09 October 2014.

P14-17457 Rawson NS. Review of the quality of observational studies of the association between rosiglitazone and acute myocardial infarction. J Popul Ther Clin Pharmacol 2014; 21 (2): e214-32.

P19-01624 Tuomi T, Antikainen A, Hokasalo M, Härmä-RodriguezS, Ilanne-Parikka P et al., Finnish Medical Society Duodecim. Diabetes. Current Care Guideline, 2018.

P19-01956 Statistical database Kelasto [Internet]. The Social Insurance Institution of Finland [cited 2019 August 21]. Available from: http://raportit.kela.fi/ibi_apps/WFServlet?IBIF_ex=NIT137AL&YKIELI=E

P19-01957 Boehringer Ingelheim. Endocrinologic and metabolic drugs advisory committee briefing document. EMPA-REG OUTCOME® Trial. 19 May 2016.

P19-07186 The CANVAS Program (CANagliflozin cardioVascular Assessment Study). Conference presentation at American Diabetes Association (ADA) 77th Scientific Sessions. Available from: https://www.georgeinstitute.org/sites/default/files/canvas-study-results-ada-2017.pdf

R10-6494 Vigneri P, Frasca F, Sciacca L, Pandini G, Vigneri R. Diabetes and cancer. Endocrine-Related Cancer, 2009; 16: 1103-23.

R10-6630 Lindbald P, Chow WH, Chan J, Bergström A, Wolk A, et al. The role of diabetes mellitus in the aetiology of renal cell cancer. Diabetologia 1999; 42: 107-12.

R10-6638 Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, Pollak M, Regensteiner JG, Yee D. Diabetes and cancer: a consensus report. Diabetes Care 2010; 33 (7): 1674-1685.

R11-4318 ISPE, International Society for Pharmacoepidemiology. Guidelines for Good Pharmacoepidemiological Practices. Revision 2: April 2007. www.pharmacoepi.org/resources/guidelines_08027.cfm.

R11-5320 Gonzalez EL, Johansson S, Wallander MA, Rodriguez LA. Trends in the prevalence and incidence of diabetes in the UK: 1996-2005. J Epidemiol Community Health 2009; 63 (4): 332-6.

R12-0368 Larsson SC, Andersson SO, Johansson JE, Wolk A. Diabetes mellitus, body size and bladder cancer risk in a prospective study of Swedish men. Eur J Cancer 44:2655-2660.

R12-0469 Larsson SC, Orsini N, Brismar K, Wolk A. Diabetes mellitus and risk of bladder cancer: a meta-analysis. Diabetologia 2006; 49: 2819-23.

R12-1334 Carstensen B, Witte DR, Friis S. Cancer occurrence in Danish diabetic patients: duration and insulin effects. Diabetologia 2012; 55:948-958.

R12-1620 EMA. European Medicines Agency recommends suspension of Avandia, Avandamet and Avaglim (press release). 23 September 2010. Available at: website: ema.europa.eu/ema/index.jsp?curl = pages/news_and_events/news/2010/09/news_detail_001119.jsp&mid = WC0b01ac058004d5c1. Accessed 09 October 2014.

R12-1913 Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. Am J Epidemiol. 2006 Jun 15; 163 (12): 1149-56.

R12-2439 Onitilo AA, Engel JM, Glurich I, Stankowski RV, Williams GM, Doi SA Diabetes and cancer I: risk, survival, and implications for screening. Cancer Causes Control 2012; 23:967-981.

R13-1120 Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol Drug Saf 2010; 19 (8): 858-68.

R13-1812 Bao C, Yang X, Xu W, Luo H, Xu Z, Su C, Qi X. Diabetes mellitus and incidence and mortality of kidney cancer: a meta-analysis. J Diabetes Complications 2013;27:357–364.

R13-1970 EMA. Guideline on good pharmacovigilance practices (GVP). Module VI – Management and reporting of adverse reactions to medicinal products. European Medicines Agency; 22 June 2012. Available at: website: emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129135.pdf. Accessed 7 February 2013.

R13-5134 NICE. Dapagliflozin in combination therapy for treating type 2 diabetes. NICE technology appraisal guidance 288. National Institute for Health and Care Excellence; 2013. Available at: website: nice.org.uk/guidance/ta288/resources/guidance-dapagliflozin-in-combination-therapy-for-treating-type2-diabetes-pdf. Accessed 09 October 2014.

R13-5420 EMA. Guideline on good pharmacovigilance practices (GVP). Module VIII – Post-authorisation safety studies. European Medicines Agency; 9 July 2012. Available at: website: emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129137.pdf. Accessed 7 February 2013.

R14-4378 Patorno E, Patrick AR, Garry EM, Schneeweiss S, Gillet VG, Bartels DB, et al. Observational studies of the association between glucose-lowering medications and cardiovascular outcomes: addressing methodological limitations. Diabetologia 2014; 57 (11): 2237-50.

R14-4617 Boehringer Ingelheim International GmbH. Jardiance (empagliflozin) summary of product characteristics. 2014. Available at: website: ema.europa.eu/ema/index.jsp?curl = pages/medicines/human/medicines/002677/human_med_001764.jsp&mid = WC0b01ac058001d124. Accessed 09 October 2014.

R14-5244 Leal I, Romio SA, Schuemie M, Oteri A, Sturkenboom M, Trifiro G. Prescribing pattern of glucose lowering drugs in the United Kingdom in the last decade: a focus on the effects of safety warnings about rosiglitazone. Br J Clin Pharmacol 2013; 75 (3): 861-8.

R14-5249 Morgan CL, Poole CD, Evans M, Barnett AH, Jenkins-Jones S, Currie CJ. What next after metformin? A retrospective evaluation of the outcome of second-line, glucose-lowering therapies in people with type 2 diabetes. J Clin Endocrinol Metab 2012; 97 (12): 4605-12.

R15-4856 Kreiger N, Marrett LD, Dodds L, Hilditch S, Darlington GA. Risk factors for renal cell carcinoma: results of a population-based case-control study. Cancer Causes Control 1993; 4(2):101-110.

R15-4857 Korhonen P, Malila N, Pukkala E, Teppo L, Albanes D, Virtamo J. The Finnish Cancer Registry as Follow-Up Source of a Large Trial Cohort. Acta Oncologica. 2002; 41: 381-388.

R15-4858 Jansson SPO, Fall K, Brus O, Magnuson A, Wändell P, Östgren CJ, Rolandsson O. Prevalence and incidence of diabetes mellitus: a nationwide population-based pharmaco-epidemiological study in Sweden. Diab Med 2015.

R15-4859 FDA, Food and Drug Administration. Guidance for Industry and FDA Staff. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets. February 2011. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM243537.pdf

R15-4860 Forman D, Bray F, Brewster DH et al., eds [Internet]. Cancer Incidence in Five Continents, Vol. X (electronic version) Lyon, IARC [cited 2015 June 23]. Available from: http://ci5.iarc.fr/Default.aspx

R15-4861 Furu K, Wettermark B, Andersen M et al. The Nordic countries as a cohort for pharmacoepidemiological research. Basic Clin. Pharmacol. Toxicol. 2010, 106:86-94.

R15-4862 Dregan A, Moller H, Murray-Thomas T and Gulliford MC. Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study. Cancer Epidemiol. Elsevier Ltd; 2012 Oct;36(5):425–9.

R15-4863 Chow WH, Lindblad P, Gridley G, Nyren O, McLaughlin JK, Linet MS, Pennello GA, Adami HO, Fraumeni JF. Risk of urinary tract cancers following kidney or ureter stones. J Natl Cancer Inst 1997; 89(19):1453-1457.

R15-4864 Boggon R, van Staa TP, Chapman M, Gallagher AM, Hammad TA, Richards MA. Cancer recording and mortality in the General Practice Research Database and linked cancer registries. Pharmacoepidemiol Drug Saf. 2013 Feb; 22(2): 168–75.

R15-4865 Boffetta P, Jourenkova N, Gustavsson P. Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons. Cancer Causes Control 1997; 8(3):444-472.

R15-4866 Burger M, Catto JWF, Dalbagni G, Grossman HB, Herr H, Karakiewicz P, Kassouf W, Kiemeney LA, Vecchia C la, Shariat S, Lotan Y. Epidemiology and risk factors of urothelial bladder cancer. Eur Urol 2013; 63(2):234-241.

R15-4867 Barlow L, Westergren K, Holmberg L and Talbäck M. The completeness of the Swedish Cancer Register – a sample survey for year 1998. Acta Oncologica. 2009; 48: 27-33.

R15-4868 Zeegers MPA, Tan FE, Dorant E, Brandt PA van den. The impact of characteristics of cigarette smoking on urinary tract cancer risk: a meta-analysis of epidemiologic studies. Cancer 2000; 89(3):630-639.

R15-4869 Wideroff L, Gridley G, Mellemkjaer L et al. Cancer incidence in a population-based cohort of patients hospitalized with diabetes mellitus in Denmark. J. Natl. Cancer Inst. 1997; 89: 1360-5.

R15-4870 ENCePP. The ENCePP Code of Conduct for Scientific Independence and Transparency in the Conduct of Pharmacoepidemiological and Pharmacovigilance Studies. 21 November 2011, EMA/929209/2011. www.encepp.eu/code_of_conduct/documents/CodeofConduct_Rev2.pdf

R15-4871 Teppo L, Pukkala E and Lehtonen M. Data quality and quality control of a population-based cancer registry. Acta Oncologica. 1994; 33: 365-369.

R15-4873 NORDCAN Register [Internet]. Association of the Nordic Cancer Registries [cited 2015 June 18]. Available from: http://www-dep.iarc.fr/NORDCAN/english/frame.asp

R16-1231 Herrett et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015 44:827-36.

R16-2539 Murchie P, Campbell NC, Delaney EK, Dinant GJ, Hannaford PC et al. Comparing Diagnostic delay in Cancer: a cross-sectional study in three European Countries with primary care. led health care systems. Fam Pract. 2012 29: 69-78.

R17-0153 Boehringer Ingelheim International GmbH. Synjardy (empagliflozin, metformin hydrochloride) Summary of Product Characteristics. Available at: https://www.ema.europa.eu/en/documents/product-information/synjardy-epar-product-information_en.pdf. Accessed 21 August 2019

R19-1534 Wolf et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. Int J Epidemiol. 2019 Mar 11. pii: dyz034.

R19-2814 Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, Silverman MG, Zelniker TA, Kuder JF, Murphy SA, Bhatt DL, Leiter LA, McGuire DK, Wilding JPH, Ruff CT, Gause-Nilsson IAM, Fredriksson M, Johansson PA, Langkilde AM, Sabatine MS. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med 2019;380:347–357.

R19-2928 CPRD linked data [Internet, page last reviewed 8 August 2019]. Clinical Practice Research Datalink [cited 2019 Aug 21]. Available from: https://www.cprd.com/linked-data

12.2 UNPUBLISHSED REFERENCES

Not Applicable

13. ANNEXURE

ANNEX 1. LIST OF STAND-ALONE DOCUMENTS

Number	Document number	Date	Title
1	c27031418-01	08 June 2021	Post-authorisation safety study to assess the risk of urinary tract malignancies in relation to empagliflozin exposure in patients with type 2 diabetes: a multi-database European study – Fifth monitoring interim report

ANNEX 2. ENCEPP CHECKLIST FOR STUDY PROTOCOLS

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Doc.Ref. EMA/540136/2009

European Network of Centres for Pharmacoepidemiology and Pharmacovigilance

ENCePP Checklist for Study Protocols (Revision 4)

Adopted by the ENCePP Steering Group on 15/10/2018

The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) welcomes innovative designs and new methods of research. This Checklist has been developed by ENCePP to stimulate consideration of important principles when designing and writing a pharmacoepidemiological or pharmacovigilance study protocol. The Checklist is intended to promote the quality of such studies, not their uniformity. The user is also referred to the ENCePP Guide on Methodological Standards in Pharmacoepidemiology, which reviews and gives direct electronic access to guidance for research in pharmacoepidemiology and pharmacovigilance.

For each question of the Checklist, the investigator should indicate whether or not it has been addressed in the study protocol. If the answer is "Yes", the section number of the protocol where this issue has been discussed should be specified. It is possible that some questions do not apply to a particular study (for example, in the case of an innovative study design). In this case, the answer 'N/A' (Not Applicable) can be checked, and the "Comments" field included for each section should be used to explain why. The "Comments" field can also be used to elaborate on a "No" answer.

This Checklist should be included as an Annex by marketing authorisation holders when submitting the protocol of a non-interventional post-authorisation safety study (PASS) to a regulatory authority (see the Guidance on the format and content of the protocol of non-interventional post-authorisation safety studies). The Checklist is a supporting document and does not replace the format of the protocol for PASS presented in the Guidance and Module VIII of the Good pharmacovigilance practices (GVP).

Study title:
Post-authorisation safety study to assess the risk of urinary tract malignancies in relation to empagliflozin exposure in patients with type 2 diabetes: a multi-database European study

EU PAS Register® number: EUPAS16424
Study reference number (if applicable): c03856813-08

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	☒	☐	☐	5
1.1.2 End of data collection[2]	☒	☐	☐	5
1.1.3 Progress report(s)	☐	☐	☒	

[1] Date from which information on the first study is first recorded in the study dataset or, in the case of secondary use of data, the date from which data extraction starts.
[2] Date from which the analytical dataset is completely available.

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1.4 Interim report(s)	☒	☐	☐	5
1.1.5 Registration in the EU PAS Register®	☒	☐	☐	5
1.1.6 Final report of study results.	☒	☐	☐	5

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	
2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	6
2.1.2 The objective(s) of the study?	☒	☐	☐	7
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	8.2.1
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	8.1
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	8.1
3.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	8.1
3.4 Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☒	☐	☐	8.1
3.5 Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☒	☐	☐	10

Comments:

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.1 Is the source population described?	☒	☐	☐	8.2
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period	☒	☐	☐	8.2.2
4.2.2 Age and sex	☒	☐	☐	8.2.5
4.2.3 Country of origin	☒	☐	☐	8.2.1
4.2.4 Disease/indication	☒	☐	☐	8.2.2
4.2.5 Duration of follow-up	☒	☐	☐	8.2.2

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	8.2.5

Comments:

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.1 Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☒	☐	☐	8.3.1
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☒	☐	☐	8.3.1
5.3 Is exposure categorised according to time windows?	☒	☐	☐	8.3.1
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☒	☐	☐	8.3.1
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☒	☐	
5.6 Is (are) (an) appropriate comparator(s) identified?	☒	☐	☐	8.3.1

Comments:

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	8.3.2
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8.3.2
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☒	☐	☐	8.3.2
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	☒	☐	☐	8.3.3
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☒	☐	☐	8.9.1
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☒	☐	☐	8.9.1

Comments:

Section 8: Effect measure modification		**Yes**	**No**	**N/A**	**Section Number**
8.1	Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	☒	☐	☐	8.7.2

Comments:

Section 9: Data sources		**Yes**	**No**	**N/A**	**Section Number**
9.1	Does the protocol describe the data source(s) used in the study for the ascertainment of:				
	9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☒	☐	☐	8.4
	9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	8.4
	9.1.3 Covariates and other characteristics?	☒	☐	☐	8.4
9.2	Does the protocol describe the information available from the data source(s) on:				
	9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☒	☐	☐	8.4
	9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	8.4
	9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	8.4
9.3	Is a coding system described for:				
	9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	8.3
	9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	8.3
	9.3.3 Covariates and other characteristics?	☒	☐	☐	8.3
9.4	Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☒	☐	☐	8.3

Comments:

Section 10: Analysis plan		**Yes**	**No**	**N/A**	**Section Number**
10.1	Are the statistical methods and the reason for their choice described?	☒	☐	☐	8.7
10.2	Is study size and/or statistical precision estimated?	☒	☐	☐	8.7
10.3	Are descriptive analyses included?	☒	☐	☐	8.7
10.4	Are stratified analyses included?	☒	☐	☐	8.7
10.5	Does the plan describe methods for analytic control of confounding?	☒	☐	☐	8.7
10.6	Does the plan describe methods for analytic control of outcome misclassification?	☒	☐	☐	8.7
10.7	Does the plan describe methods for handling missing data?	☒	☐	☐	8.7
10.8	Are relevant sensitivity analyses described?	☒	☐	☐	8.7

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	8.6
11.2 Are methods of quality assurance described?	☒	☐	☐	8.8
11.3 Is there a system in place for independent review of study results?	☐	☒	☐	

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of:				
12.1.1 Selection bias?	☒	☐	☐	8.9
12.1.2 Information bias?	☒	☐	☐	8.9
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☒	☐	☐	8.9
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	8.4

Comments:

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	☐	☐	☐	9
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☒	☐	
13.3 Have data protection requirements been described?	☒	☐	☐	9

Comments:

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	4

Comments:

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	☒	☐	☐	11

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.2 Are plans described for disseminating study results externally, including publication?	☒	☐	☐	11

Comments:

Name of the main author of the protocol: [redacted]

Date: 27/January/2022

Signature: DocuSigned by [redacted] I am the author of this document January 28, 2022 | 1:33:11 AM JST [redacted]

ANNEX 3. READ CODES TO IDENTIFY TYPE 2 DIABETES

The codes will be updated for the data application.

Read code	Description
9OL..00	Diabetes monitoring admin
66A..00	Diabetic monitoring
66AS.00	Diabetic annual review
C10..00	Diabetes mellitus
9N1Q.00	Seen in diabetic clinic
C10F.00	Type 2 diabetes mellitus
9OL4.00	Diabetes monitoring 1st letter
66A2.00	Follow-up diabetic assessment
9NND.00	Under care of diabetic foot screener
66AP.00	Diabetes: practice programme
9OL1.00	Attends diabetes monitoring
66A4.00	Diabetic on oral treatment
68A7.00	Diabetic retinopathy screening
66AZ.00	Diabetic monitoring NOS
66AJ.00	Diabetic - poor control
C100112	Non-insulin dependent diabetes mellitus
66Ac.00	Diabetic peripheral neuropathy screening
9OL5.00	Diabetes monitoring 2nd letter
66A3.00	Diabetic on diet only
66Aq.00	Diabetic foot screen
C109.00	Non-insulin dependent diabetes mellitus
F420.00	Diabetic retinopathy
13B1.00	Diabetic diet
9OLA.00	Diabetes monitor. check done
7L19800	Subcutaneous injection of insulin
66AR.00	Diabetes management plan given
9h42.00	Excepted from diabetes quality indicators: Informed dissent
2G5E.00	O/E - Right diabetic foot at low risk
66A5.00	Diabetic on insulin

Read code	Description
2G5I.00	O/E - Left diabetic foot at low risk
9h41.00	Excepted from diabetes qual indicators: Patient unsuitable
F420000	Background diabetic retinopathy
9N4I.00	DNA - Did not attend diabetic clinic
66AI.00	Diabetic - good control
C109.12	Type 2 diabetes mellitus
66AQ.00	Diabetes: shared care programme
66AD.00	Fundoscopy - diabetic check
8BL2.00	Patient on maximal tolerated therapy for diabetes
1434.00	H/O: diabetes mellitus
9OL6.00	Diabetes monitoring 3rd letter
66Ai.00	Diabetic 6 month review
9N1v.00	Seen in diabetic eye clinic
8CA4100	Pt advised re diabetic diet
2BBP.00	O/E - right eye background diabetic retinopathy
2BBQ.00	O/E - left eye background diabetic retinopathy
8B3l.00	Diabetes medication review
66A9.00	Understands diet - diabetes
66Ab.00	Diabetic foot examination
42W..00	Hb. A1C - diabetic control
66AU.00	Diabetes care by hospital only
9NM0.00	Attending diabetes clinic
8H7r.00	Refer to diabetic foot screener
8I3X.00	Diabetic retinopathy screening refused
2G5F.00	O/E - Right diabetic foot at moderate risk
2G5J.00	O/E - Left diabetic foot at moderate risk
66A1.00	Initial diabetic assessment
F420400	Diabetic maculopathy
66AW.00	Diabetic foot risk assessment
9OLA.11	Diabetes monitored
13AB.00	Diabetic lipid lowering diet

Read code	Description
2BBM.00	O/E - diabetic maculopathy absent both eyes
ZC2C800	Dietary advice for diabetes mellitus
C100100	Diabetes mellitus, adult onset, no mention of complication
66Ao.00	Diabetes type 2 review
9OL8.00	Diabetes monitor.phone invite
9N4p.00	Did not attend diabetic retinopathy clinic
9N1i.00	Seen in diabetic foot clinic
14P3.00	H/O: insulin therapy
C101.00	Diabetes mellitus with ketoacidosis
66A8.00	Has seen dietician - diabetes
66AT.00	Annual diabetic blood test
C10FJ00	Insulin treated Type 2 diabetes mellitus
9OL3.00	Diabetes monitoring default
66AV.00	Diabetic on insulin and oral treatment
66A7.00	Frequency of hypo. attacks
9OL..11	Diabetes clinic administration
66AA.11	Injection sites - diabetic
66AH000	Conversion to insulin
8CS0.00	Diabetes care plan agreed
2G5G.00	O/E - Right diabetic foot at high risk
66AH.00	Diabetic treatment changed
2G5K.00	O/E - Left diabetic foot at high risk
9OLD.00	Diabetic patient unsuitable for digital retinal photography
F372.12	Diabetic neuropathy
F420600	Non-proliferative diabetic retinopathy
C106.12	Diabetes mellitus with neuropathy
9OL7.00	Diabetes monitor. verbal invite
C106.00	Diabetes mellitus with neurological manifestation
8H4F.00	Referral to diabetologist
8I3W.00	Diabetic foot examination declined
68A9.00	Diabetic retinopathy screening offered

Read code	Description
13AC.00	Diabetic weight reducing diet
F420100	Proliferative diabetic retinopathy
66AY.00	Diabetic diet - good compliance
C10FM00	Type 2 diabetes mellitus with persistent microalbuminuria
C109.13	Type II diabetes mellitus
C109.11	NIDDM - Non-insulin dependent diabetes mellitus
9OLZ.00	Diabetes monitoring admin.NOS
2G5A.00	O/E - Right diabetic foot at risk
2BBW.00	O/E - right eye diabetic maculopathy
2BBX.00	O/E - left eye diabetic maculopathy
2G5B.00	O/E - Left diabetic foot at risk
8HBG.00	Diabetic retinopathy 12 month review
8H7C.00	Refer, diabetic liaison nurse
C109J00	Insulin treated Type 2 diabetes mellitus
66Af.00	Patient diabetes education review
66A6.00	Last hypo. attack
ZLA2500	Seen by diabetic liaison nurse
8Hj0.00	Referral to diabetes structured education programme
7276.00	Pan retinal photocoagulation for diabetes
C10FL00	Type 2 diabetes mellitus with persistent proteinuria
8Hl1.00	Referral for diabetic retinopathy screening
68AB.00	Diabetic digital retinopathy screening offered
8Hj4.00	Referral to DESMOND diabetes structured education programme
F420200	Preproliferative diabetic retinopathy
C104.11	Diabetic nephropathy
8HTk.00	Referral to diabetic eye clinic
9h4..00	Exception reporting: diabetes quality indicators
C100.00	Diabetes mellitus with no mention of complication
2BBL.00	O/E - diabetic maculopathy present both eyes
66AM.00	Diabetic - follow-up default
9N2i.00	Seen by diabetic liaison nurse

Read code	Description
C10F.11	Type II diabetes mellitus
M271200	Mixed diabetic ulcer - foot
2BBR.00	O/E - right eye preproliferative diabetic retinopathy
66Ae.00	HbA1c target
2BBS.00	O/E - left eye preproliferative diabetic retinopathy
8H7f.00	Referral to diabetes nurse
42c..00	HbA1 - diabetic control
M271000	Ischaemic ulcer diabetic foot
679R.00	Patient offered diabetes structured education programme
F171100	Autonomic neuropathy due to diabetes
66Am.00	Insulin dose changed
ZL62500	Referral to diabetes nurse
C109700	Non-insulin dependent diabetes mellitus - poor control
F464000	Diabetic cataract
F420z00	Diabetic retinopathy NOS
C10FC00	Type 2 diabetes mellitus with nephropathy
C105.00	Diabetes mellitus with ophthalmic manifestation
8Hj5.00	Referral to XPERT diabetes structured education programme
2BBT.00	O/E - right eye proliferative diabetic retinopathy
C10F600	Type 2 diabetes mellitus with retinopathy
M271100	Neuropathic diabetic ulcer - foot
8Hl4.00	Referral to community diabetes specialist nurse
2BBV.00	O/E - left eye proliferative diabetic retinopathy
C104.00	Diabetes mellitus with renal manifestation
9N0m.00	Seen in diabetic nurse consultant clinic
9OLB.00	Attended diabetes structured education programme
9NN9.00	Under care of diabetes specialist nurse
66Aa.00	Diabetic diet - poor compliance
C107.00	Diabetes mellitus with peripheral circulatory disorder
F372.00	Polyneuropathy in diabetes
66AJz00	Diabetic - poor control NOS

Read code	Description
9N2d.00	Seen by diabetologist
8H2J.00	Admit diabetic emergency
2BBF.00	Retinal abnormality - diabetes related
9OL2.00	Refuses diabetes monitoring
66Ad.00	Hypoglycaemic attack requiring 3rd party assistance
8A13.00	Diabetic stabilisation
C10F700	Type 2 diabetes mellitus - poor control
9N0n.00	Seen in community diabetes specialist clinic
93C4.00	Patient consent given for addition to diabetic register
66AJ.11	Unstable diabetes
9M00.00	Informed consent for diabetes national audit
F420300	Advanced diabetic maculopathy
C10F900	Type 2 diabetes mellitus without complication
66AJ000	Chronic hyperglycaemia
66AK.00	Diabetic - cooperative patient
2G5C.00	Foot abnormality - diabetes related
66Ap.00	Insulin treatment initiated
9N0o.00	Seen in community diabetic specialist nurse clinic
9NN8.00	Under care of diabetologist
ZL62600	Referral to diabetic liaison nurse
C100z00	Diabetes mellitus NOS with no mention of complication
N030100	Diabetic Charcot arthropathy
F381311	Diabetic amyotrophy
TJ23A00	Adverse reaction to metformin hydrochloride
M037200	Cellulitis in diabetic foot
C10FN00	Type 2 diabetes mellitus with ketoacidosis
66AN.00	Date diabetic treatment start
8HHy.00	Referral to diabetic register
2G5H.00	O/E - Right diabetic foot - ulcerated
2G51000	Foot abnormality - diabetes related
2G5L.00	O/E - Left diabetic foot - ulcerated

Read code	Description
66Ak.00	Diabetic monitoring - lower risk albumin excretion
9OLM.00	Diabetes structured education programme declined
8CR2.00	Diabetes clinical management plan

Read code	Description
C10FK00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus
66A7100	Frequency of GP or paramedic treated hypoglycaemia
ZV65312	Dietary counselling in diabetes mellitus
C103.00	Diabetes mellitus with ketoacidotic coma
ZRB6.00	Diabetes wellbeing questionnaire
66AJ200	Loss of hypoglycaemic warning
9OLK.00	DESMOND diabetes structured education programme completed
C104z00	Diabetes mellitus with nephropathy NOS
F372100	Chronic painful diabetic neuropathy
F3y0.00	Diabetic mononeuropathy
C107.11	Diabetes mellitus with gangrene
F372.11	Diabetic polyneuropathy
66AL.00	Diabetic-uncooperative patient
9kL..00	Insulin initiation - enhanced services administration
9OLL.00	XPERT diabetes structured education programme completed
C109400	Non-insulin dependent diabetes mellitus with ulcer
C10D.00	Diabetes mellitus autosomal dominant type 2
M21yC00	Insulin lipohypertrophy
L180500	Pre-existing diabetes mellitus, insulin-dependent
C109900	Non-insulin-dependent diabetes mellitus without complication
66A7000	Frequency of hospital treated hypoglycaemia
8H4e.00	Referral to diabetes special interest general practitioner
C106z00	Diabetes mellitus NOS with neurological manifestation
C102.00	Diabetes mellitus with hyperosmolar coma
K01x100	Nephrotic syndrome in diabetes mellitus
66AJ300	Recurrent severe hypos
C101z00	Diabetes mellitus NOS with ketoacidosis
G73y000	Diabetic peripheral angiopathy
66Al.00	Diabetic monitoring - higher risk albumin excretion
ZRbH.00	Perceived control of insulin-dependent diabetes
N030000	Diabetic cheiroarthropathy

Read code	Description
8A12.00	Diabetic crisis monitoring
C105z00	Diabetes mellitus NOS with ophthalmic manifestation
9OLN.00	Diabetes monitor invitation by SMS (short message service)
8HBH.00	Diabetic retinopathy 6 month review
ZL22500	Under care of diabetic liaison nurse
9OLF.00	Diabetes structured education programme completed
F35z000	Diabetic mononeuritis NOS
8H3O.00	Non-urgent diabetic admission
C10FQ00	Type 2 diabetes mellitus with exudative maculopathy
C10G.00	Secondary pancreatic diabetes mellitus
C10F000	Type 2 diabetes mellitus with renal complications
C10F911	Type II diabetes mellitus without complication
C106100	Diabetes mellitus, adult onset, + neurological manifestation
C10FB00	Type 2 diabetes mellitus with polyneuropathy
C10F200	Type 2 diabetes mellitus with neurological complications
7L10000	Continuous subcutaneous infusion of insulin
66AG.00	Diabetic drug side effects
8HLE.00	Diabetology D.V. done
C10FH00	Type 2 diabetes mellitus with neuropathic arthropathy
F372200	Asymptomatic diabetic neuropathy
9OLG.00	Attended XPERT diabetes structured education programme
C109600	Non-insulin-dependent diabetes mellitus with retinopathy
U60231C	Adverse reaction to metformin hydrochloride
8Hg4.00	Discharged from care of diabetes specialist nurse
F420500	Advanced diabetic retinal disease
9OL9.00	Diabetes monitoring deleted
M21yC11	Insulin site lipohypertrophy
C10FE00	Type 2 diabetes mellitus with diabetic cataract
2BBk.00	O/E - right eye stable treated prolif diabetic retinopathy
F372000	Acute painful diabetic neuropathy
66AO.00	Date diabetic treatment stopp

Read code	Description
TJ23400	Adverse reaction to gliclazide
8I3k.00	Insulin therapy declined
C105100	Diabetes mellitus, adult onset, + ophthalmic manifestation
Cyu2.00	Diabetes mellitus
9OLJ.00	DAFNE diabetes structured education programme completed
C107.12	Diabetes with gangrene
C10zz00	Diabetes mellitus NOS with unspecified complication
44V3.00	Glucose tol. test diabetic
C10F400	Type 2 diabetes mellitus with ulcer
C10FJ11	Insulin treated Type II diabetes mellitus
42WZ.00	Hb. A1C - diabetic control NOS
ZLD7500	Discharge by diabetic liaison nurse
C107z00	Diabetes mellitus NOS with peripheral circulatory disorder
C10K.00	Type A insulin resistance
C10A100	Malnutrition-related diabetes mellitus with ketoacidosis
ZV6DA00	Admitted for commencement of insulin
C101000	Diabetes mellitus, juvenile type, with ketoacidosis
2BBl.00	O/E - left eye stable treated prolif diabetic retinopathy
9OLH.00	Attended DAFNE diabetes structured education programme
C106.13	Diabetes mellitus with polyneuropathy
R054200	Gangrene of toe in diabetic
C10y.00	Diabetes mellitus with other specified manifestation
C10F711	Type II diabetes mellitus - poor control
C10FL11	Type II diabetes mellitus with persistent proteinuria
U602300	Insul/oral hypoglyc drugs caus adverse eff therapeut use
C10z.00	Diabetes mellitus with unspecified complication
C109C00	Non-insulin dependent diabetes mellitus with nephropathy
C10F100	Type 2 diabetes mellitus with ophthalmic complications
679L000	Education in self-management of diabetes
C10z100	Diabetes mellitus, adult onset, + unspecified complication
C10FR00	Type 2 diabetes mellitus with gastroparesis

Read code	Description
8Hj3.00	Referral to DAFNE diabetes structured education programme
C10FA00	Type 2 diabetes mellitus with mononeuropathy
C10FD00	Type 2 diabetes mellitus with hypoglycaemic coma
2G5W.00	O/E - left chronic diabetic foot ulcer
F440700	Diabetic iritis
ZC2CA00	Dietary advice for type II diabetes
C10A.00	Malnutrition-related diabetes mellitus
C10N.00	Secondary diabetes mellitus
R054300	[D]Widespread diabetic foot gangrene
C104100	Diabetes mellitus, adult onset, with renal manifestation
C10N100	Cystic fibrosis related diabetes mellitus
F345000	Diabetic mononeuritis multiplex
C109711	Type II diabetes mellitus - poor control
C107200	Diabetes mellitus, adult with gangrene
C10F500	Type 2 diabetes mellitus with gangrene
2BBo.00	O/E - sight threatening diabetic retinopathy
C101100	Diabetes mellitus, adult onset, with ketoacidosis
C109712	Type 2 diabetes mellitus - poor control
C107100	Diabetes mellitus, adult, + peripheral circulatory disorder
8HVU.00	Private referral to diabetologist
C109J12	Insulin treated Type II diabetes mellitus
F381300	Myasthenic syndrome due to diabetic amyotrophy
ZV6DB00	Admitted for conversion to insulin
8HTi.00	Referral to multidisciplinary diabetic clinic
66Ae000	HbA1c target level - IFCC standardised
TJ23000	Adverse reaction to insulins
F420700	High risk proliferative diabetic retinopathy
C10F611	Type II diabetes mellitus with retinopathy
2G5V.00	O/E - right chronic diabetic foot ulcer
C109K00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus
C102100	Diabetes mellitus, adult onset, with hyperosmolar coma

Read code	Description
C10FF00	Type 2 diabetes mellitus with peripheral angiopathy
F420800	High-risk non-proliferative diabetic retinopathy
C107400	NIDDM with peripheral circulatory disorder
C10F300	Type 2 diabetes mellitus with multiple complications
8CP2.00	Transition of diabetes care options discussed
N030011	Diabetic cheiropathy
66At.00	Diabetic dietary review
C10M.00	Lipoatrophic diabetes mellitus
9M10.00	Informed dissent for diabetes national audit
C10FP00	Type 2 diabetes mellitus with ketoacidotic coma
C109611	Type II diabetes mellitus with retinopathy
C10FG00	Type 2 diabetes mellitus with arthropathy
C109J11	Insulin treated non-insulin dependent diabetes mellitus
C101y00	Other specified diabetes mellitus with ketoacidosis
C109411	Type II diabetes mellitus with ulcer
8I2S.00	Glitazones contraindicated
C109000	Non-insulin-dependent diabetes mellitus with renal comps
C109100	Non-insulin-dependent diabetes mellitus with ophthalm comps
C106y00	Other specified diabetes mellitus with neurological comps
TJ23.00	Adverse reaction to insulins and antidiabetic agents
9h43.00	Excepted from diabetes qual indicators: service unavailable
C106.11	Diabetic amyotrophy
TJ23z00	Adverse reaction to insulins and antidiabetic agents NOS
C104y00	Other specified diabetes mellitus with renal complications
C109612	Type 2 diabetes mellitus with retinopathy
C109212	Type 2 diabetes mellitus with neurological complications
C109200	Non-insulin-dependent diabetes mellitus with neuro comps
C106000	Diabetes mellitus, juvenile, + neurological manifestation
ZRB4.00	Diabetes clinic satisfaction questionnaire
C109412	Type 2 diabetes mellitus with ulcer
ZRB5.11	DTSQ - Diabetes treatment satisfaction questionnaire

Read code	Description
C109E00	Non-insulin depend diabetes mellitus with diabetic cataract
8I2P.00	Sulphonylureas contraindicated
C109B00	Non-insulin dependent diabetes mellitus with polyneuropathy
C109E11	Type II diabetes mellitus with diabetic cataract
8HKE.00	Diabetology D.V. requested
C10yz00	Diabetes mellitus NOS with other specified manifestation
C109C12	Type 2 diabetes mellitus with nephropathy
C109500	Non-insulin dependent diabetes mellitus with gangrene
C109H00	Non-insulin dependent d m with neuropathic arthropathy
C109D00	Non-insulin dependent diabetes mellitus with hypoglyca coma
C109011	Type II diabetes mellitus with renal complications
C109300	Non-insulin-dependent diabetes mellitus with multiple comps
U602312	Adverse reaction to insulins
C10zy00	Other specified diabetes mellitus with unspecified comps
U602311	Adverse reaction to insulins and antidiabetic agents
TJ23500	Adverse reaction to glipizide
66Ar.00	Insulin treatment stopped
C109H11	Type II diabetes mellitus with neuropathic arthropathy
889A.00	Diab mellit insulin-glucose infus acute myocardial infarct
C109H12	Type 2 diabetes mellitus with neuropathic arthropathy
9NiC.00	Did not attend DAFNE diabetes structured education programme
66AQ100	Declined consent for diabetes year of care programme
C10yy00	Other specified diabetes mellitus with other spec comps
C10F311	Type II diabetes mellitus with multiple complications
C109E12	Type 2 diabetes mellitus with diabetic cataract
C108y00	Other specified diabetes mellitus with multiple comps
TJ23300	Adverse reaction to glibenclamide
C109511	Type II diabetes mellitus with gangrene
C10y100	Diabetes mellitus, adult, + other specified manifestation
C10N000	Secondary diabetes mellitus without complication
66As.00	Diabetic on subcutaneous treatment

Read code	Description
66At100	Type II diabetic dietary review
C109012	Type 2 diabetes mellitus with renal complications
C109D12	Type 2 diabetes mellitus with hypoglycaemic coma
U602316	Adverse reaction to gliclazide
C102z00	Diabetes mellitus NOS with hyperosmolar coma
C109512	Type 2 diabetes mellitus with gangrene
C109F00	Non-insulin-dependent d m with peripheral angiopath
C109C11	Type II diabetes mellitus with nephropathy
C103z00	Diabetes mellitus NOS with ketoacidotic coma
9N1o.00	Seen in multidisciplinary diabetic clinic
C109G00	Non-insulin dependent diabetes mellitus with arthropathy
C105y00	Other specified diabetes mellitus with ophthalmic complicatn
C109F11	Type II diabetes mellitus with peripheral angiopathy
C10FM11	Type II diabetes mellitus with persistent microalbuminuria
C10F411	Type II diabetes mellitus with ulcer
66Au.00	Diabetic erectile dysfunction review
66Av.00	Diabetic assessment of erectile dysfunction
Cyu2000	Other specified diabetes mellitus
C109B11	Type II diabetes mellitus with polyneuropathy
C10F011	Type II diabetes mellitus with renal complications
C103y00	Other specified diabetes mellitus with coma
C109211	Type II diabetes mellitus with neurological complications
ZRB4.11	CSQ - Diabetes clinic satisfaction questionnaire
3883.00	Diabetes treatment satisfaction questionnaire
8IAs.00	Diabetic dietary review declined
2BBr.00	Impaired vision due to diabetic retinopathy
TJ23900	Adverse reaction to tolbutamide
C10FB11	Type II diabetes mellitus with polyneuropathy
C109D11	Type II diabetes mellitus with hypoglycaemic coma
C109111	Type II diabetes mellitus with ophthalmic complications
C109F12	Type 2 diabetes mellitus with peripheral angiopathy

Read code	Description
C10A000	Malnutrition-related diabetes mellitus with coma
C103100	Diabetes mellitus, adult onset, with ketoacidotic coma
C109A00	Non-insulin dependent diabetes mellitus with mononeuropathy
8HME.00	Listed for Diabetology admissn
C10FA11	Type II diabetes mellitus with mononeuropathy
U60231B	Adverse reaction to tolbutamide
U60231E	Adverse reaction to insulins and antidiabetic agents NOS
Cyu2300	Unspecified diabetes mellitus with renal complications
8HgC.00	Discharged from diabetes shared care programme
66AQ000	Unsuitable for diabetes year of care programme
C10F111	Type II diabetes mellitus with ophthalmic complications
TJ23B00	Adverse reaction to glucagon
C109G11	Type II diabetes mellitus with arthropathy
U602315	Adverse reaction to glibenclamide
C109G12	Type 2 diabetes mellitus with arthropathy
C109A11	Type II diabetes mellitus with mononeuropathy
C10K000	Type A insulin resistance without complication
U60231A	Adverse reaction to tolazamide
C108z00	Unspecified diabetes mellitus with multiple complications
C109112	Type 2 diabetes mellitus with ophthalmic complications
U602318	Adverse reaction to gliquidone
TJ23200	Adverse reaction to chlorpropamide
C10FE11	Type II diabetes mellitus with diabetic cataract
U602317	Adverse reaction to glipzide
C10G000	Secondary pancreatic diabetes mellitus without complication
ZRB6.11	DWBQ - Diabetes wellbeing questionnaire
C10F211	Type II diabetes mellitus with neurological complications
C10E512	Insulin-dependent diabetes mellitus with ulcer
C10FD11	Type II diabetes mellitus with hypoglycaemic coma
SL23100	Biguanide poisoning
Kyu0300	Glomerular disorders in diabetes mellitus

Read code	Description
C10A500	Malnutritn-relat diabetes melitus wth periph circul complctn
C10FC11	Type II diabetes mellitus with nephropathy
66Aw.00	Insulin dose
TJ23800	Adverse reaction to tolazamide

ANNEX 4. CODES TO IDENTIFY EXPOSURE VARIABLES

No.	**Variable**	**ATC codes[3]**	**Time-dependent (T) / fixed at cohort entry (F)**	**How the effect of variable is accounted for[1]**
1	AT / ITT: Empagliflozin	Empagliflozin A10BK03 (current), A10BX12 (previous) DPP-4 inhibitors ATC beginning with A10BH Fixed-dose combinations Will be identified from ATC class A10BD	T / F	E
2	AT / ITT: DPP-4 inhibitor		T / F	E
3	AT / ITT, FDCM[2]: empagliflozin		T / F	E
4	AT / ITT, FDCM[2]: DPP-4 inhibitor		T / F	E
5	Empagliflozin current use		T	E
6	Empagliflozin: cumulative duration of exposure		T	E
7	Empagliflozin: cumulative dosage of exposure		T	E
8	Empagliflozin: time since last dose		T	E
	Other OAD treatments			
9	Metformin	ATC beginning with A10BA	T	PS / A / VS, RS, TV, SG
10	Sulphonylureas	ATC beginning with A10BB	T	PS / A / VS, RS, TV

No.	**Variable**	**ATC codes[3]**	**Time-dependent (T) / fixed at cohort entry (F)**	**How the effect of variable is accounted for[1]**
11	Insulin	ATC beginning with A10A	T	PS / A / VS, RS, TV
12	Thiazolidinediones	ATC beginning with A10BG	T	PS / A / VS, RS, TV
13	Alpha glucosidase inhibitors	ATC beginning with A10BF	T	PS / A / VS, RS, TV
14	GLP-1 agonists	Will be identified from ATC classes A10BJ and A10BX	T	PS / A / VS, RS, TV
15	Other blood glucose lowering drugs, excluding insulins	ATC beginning with A10BX that is not included elsewhere		
16	Treatment complexity (mono, dual and triple combination therapy)		T	PS / A / VS, RS, TV
17	Switch or add-on		T	PS / A / VS, RS, TV
	Exclusion / Censoring			
18	DPP-4 (Linagliptin) and Empagliflozin	A10BD19	T	Censoring / Exclusion
19	Discontinuation of primary exposure		T	Censoring
20	Switch to other SGLT-2 inhibitor use	A10BK02, A10BX11, A10BK01, A10BX09, A10BK04, A10BD16, A10BD15, A10BD23	T	Censoring / Exclusion

No.	**Variable**	**ATC codes**[3]	**Time-dependent (T) / fixed at cohort entry (F)**	**How the effect of variable is accounted for**[1]
21	Use of fixed-dose combinations of SGLT-2 inhibitors with DPP-4 inhibitors	Identified from ATC class A10BD	T	Censoring / Exclusion

[1] E – Exposure, A – adjusted in the Cox model, VS – potential confounder and will be adjusted for using variable selection, PS – one of the PS matching variable at baseline, TV – time-varying confounder and adjusted for in MSM, RS – defines the risk set in NCC, SG – used as a subgroup in stratified analyses, / - and/or

2 FDCM – Fixed-dose combination with metformin

3 New drugs will be included in the variable definitions if they become available during the study period.

ANNEX 5. CODES TO IDENTIFY THE OUTCOME VARIABLES

The codes will be updated for the data applications.

No.	Variable	ICD-10 / ICD-O-3 codes	READ codes	Outcome / Censoring variable	Analyses[1]
1	Urinary Tract cancer (malignant and carcinoma in situ)	C64-C68 D09.0, D09.1	Byu9, B837, B83z, B4A, B49	Outcome	D, P, S, F , NCC, MSM Sens
2	Bladder cancer (malignant and carcinoma in situ)	C67, D09.0	B49, B837, selected BB4 codes	Outcome	D, P, S, F , Sens
3	Renal cancer (malignant)	C64, C65	B4A0, B4A1	Outcome	D, P, S, F , Sens
4	Other urinary tract cancer (non-renal, non-bladder cancers) – sensitivity (malignant)	C66,C68	Byu9	Sensitivity outcome	D, Sens
5	Urinary Tract cancers from GP data (CPRD only) – sensitivity		Byu9, B837, B83z, B4A, B49	Sensitivity outcome	Sens
6	Urinary tract neoplasms of uncertain or unknown behaviour - sensitivity	D41	B917, B91z, BA04	Sensitivity outcome	Sens
	Censoring variables				
7	Other non-UT cancers	C00-C99, D00-09 (excl. C64-68, D09.0, D09.1)	B (excl. Byu9, B837, B83z, B4A, B49, B49, B837, selected BB4 codes)	Censoring	
8	End of study period			Censoring	
9	End of GP practice coverage			Censoring	
10	Use of other SGLT-2 (since start of follow-up)			Censoring	
11	Use of fixed-dose combinations of			Censoring	

No.	Variable	ICD-10 / ICD-O-3 codes	READ codes	Outcome / Censoring variable	Analyses[1]
	SGLT-2 inhibitors with DPP-4 inhibitors				
12	Patient transfer			Censoring	
13	End stage renal disease / renal dialysis			Censoring	
14	All-cause mortality (excluding UT cancer mortality)			Censoring	
15	Discontinuation of primary exposure			Censoring for as-treated analysis	
16	UT cancer characteristics (based on staging, TNM codes, etc)			Sensitivity outcome / descriptive	D, Sens

1 D=Descriptive, P =Primary , S = secondary, F = further, Sens = sensitivity, NCC =Nested-case-control, MSM – marginal structural models

ANNEX 6. COVARIATES TO BE CONSIDERED FOR INCLUSION IN THE PROPENSITY SCORE MODEL.

No.	Variable
	Socio-demographic variables
1	Age
2	Sex
3	Socioeconomic status
4	Calendar year of index date
5	Duration of look-back period
6	Duration of treated diabetes at index date
	Diabetic complications
7	Diabetic retinopathy or diabetic maculopathy
8	Diabetic nephropathy
9	Diabetic neuropathy
10	Peripheral vascular disease
11	Diabetic lower limb severe complications
	Urinary tract related comorbidities
12	Kidney or genitourinary stones
13	Prior history of UTI or pyelonephritis
14	Liver disease
15	eGFR
16	Renal impairment
17	Prior ICU admission
18	Pancreatitis

19	BMI (when available)
20	Smoking (when available)
21	Alcohol use (when available)
22	HbA1c (when available)
23	Albuminuria testing
	Cardiovascular Diseases
24	Congestive heart failure
25	Blood pressure measurements
26	Hypertension
27	Stroke
28	Myocardial infarction
29	Autoimmune disease
30	COPD
31	Time since first diabetes diagnosis
	Other non-diabetes medications
32	Antihypertensives/diuretics
33	Non-steroidal anti-inflammatory drugs (NSAIDs)
34	Oral steroids
35	Statins, fibrates
36	Lipid modifying agents
37	Zoledronic acid
38	Antibiotics
39	Other drugs

BMI=body mass index; COPD=chronic obstructive pulmonary disease; eGFR=estimated glomerular filtration rate; HbA1c=glycated haemoglobin A1c; HF=heart failure; ICU=intensive care unit; MSM=marginal structural model; NCC=nested case-control; PS=propensity score; UTI=urinary tract infection.

SIGNATURE

PRINCIPAL INVESTIGATOR SIGNATURE

Study Title: **Post-authorisation safety study to assess the risk of urinary tract malignancies in relation to Empagliflozin exposure in patients with type 2 diabetes: a multi-database European study**

Trial Number: **1245.97**

Study protocol version: **8.0**

I herewith certify that I agree to content of Study protocol version 8.0 and to all documents referenced in the study protocol version 8.0.

Date: 02-Jun-2023

Name: [redacted]

Signature: DocuSigned by: [redacted] Signer Name: [redacted] Signing Reason: I approve this document Signing Time: 02-Jun-2023 | 12:05:57 AM PDT [redacted]

Affiliation: [redacted]