Boehringer Ingelheim

Protocol for non-interventional studies based on existing data

Document Number:	**c03856813-05**
BI Study Number:	1245.97
BI Investigational Product(s):	Jardiance® (empagliflozin)
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:	3.0
Date of last version of protocol:	19 February 2016
PASS:	Yes
EU PAS register number:	Study not yet registered. Will be registered following protocol regulatory endorsement and prior to data collection start.
Active substance:	A10BX12 empagliflozin
Medicinal product:	Jardiance
Product reference:	EMEA/H/C/002677
Procedure number:	EMEA/H/C/002677/MEA/004
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 other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor
Country(-ies) of study:	United Kingdom and Sweden
Authors:	[redacted]Pasi Korhonen, Fabian Hoti, Solomon Christopher, Rosa Juuti, Eero Mertano [redacted]

	[redacted]
Marketing authorisation holder(s):	Boehringer Ingelheim International GmbH [redacted]
MAH contact person:	[redacted]
EU-QPPV:	[redacted]
Signature of EU-QPPV:	The signature of the EU-QPPV is provided electronically
Date:	10 June 2016

1. TABLE OF CONTENTS

2. LIST OF ABBREVIATIONS

AT	As-treated analysis
ATC	Anatomical Therapeutic Chemical Classification System
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
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 hemoglobin A1c
HES	Hospital Episode Statistics
HF	Heart failure
HR	Hazard ratio
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

3. RESPONSIBLE PARTIES

Principal investigator
Pasi Korhonen
PhD, Adjunct Prof. Biostatistics,
EPID Research Oy, Finland

Participating institutions
EPID Research, Finland

Sponsor
Boehringer Ingelheim International GmbH

The study investigators at EPID Research Oy (EPID) 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.

4. ABSTRACT

Name of company:			
Boehringer Ingelheim			
Name of finished medicinal product: Jardiance			
Name of active ingredient: A10BX12 empagliflozin			
Protocol date:	**Study number:**	**Version/Revision:**	**Version/Revision date:**
16-Sep-2015	1245.97	3.0	10-Jun-2016
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 HCl 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 other SGLT-2 inhibitors and dipeptidyl peptidase-4 (DPP-4) inhibitors. Prior to the development of the protocol for this study, a feasibility assessment of three databases in Europe (CPRD and data with potential for linkage (HES, Cancer registry 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.		
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 other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor (2 separate comparison groups).		

	Primary objectives: (i) To estimate the adjusted hazard ratio and incidence rates of all urinary tract cancers (bladder, renal and other urinary tract cancers) in patients initiating empagliflozin compared to patients initiating other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor (ii) To estimate the adjusted hazard ratio and incidence rates of bladder cancer in patients initiating empagliflozin compared to patients initiating other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor (iii)To estimate the adjusted hazard ratio and incidence rates of renal cancer in patients initiating empagliflozin compared to patients initiating other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor **Secondary objective:** (i) To estimate the adjusted hazard ratio and incidence rates 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 adjusted hazard ratios and incidence rates of all urinary tract cancers with respect to empagliflozin initiation among sub-groups defined by age, gender, relevant comorbidities, concurrent use of metformin and other relevant treatments (ii) To estimate the adjusted hazard ratio and incidence rates of non-renal, non-bladder urinary tract cancers in patients initiating empagliflozin compared to patients initiating other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor Non-renal, non-bladder urinary tract cancers (referred to as other urinary tract cancers) includes ureter and urethra cancers
Study design:	This is an observational, 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 SGLT-2 inhibitors other than empagliflozin and 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, another SGLT-2 inhibitor, 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 channeling bias, individuals will be matched with similar treatment and clinical history at index date by deriving propensity scores. At least two sets of propensity scores, one for the comparison of empagliflozin versus other SGLT-2 inhibitors and one for the comparison of empagliflozin versus DPP-4 inhibitors, will be generated. Propensity-score-matched cohorts will be derived for each comparison from the study-base by matching patients who initiate empagliflozin with patients who initiate other SGLT-2 inhibitor and patients who initiate DPP-4 inhibitors at index date with a pre-fixed ratio (1:1 or 1:3 in a sensitivity analysis using Greedy matching methods). Thus an observational cohort is constructed to study the risk of urinary tract cancer malignancies among the three treatment groups. Depending on sample size, combination exposures with metformin will be accounted for through separate propensity scores, stratification, or adjustment.
Population:	The broad study population will include all patients from the Clinical Practice Research Database (CPRD-GOLD) in the United Kingdom (UK) and patients in the nationwide register in Sweden, who have purchased at least one prescription of empagliflozin, other SGLT-2 inhibitor drugs, or DPP-4 inhibitor drugs during the study period 2014-2019. Patients initiating SGLT-2 inhibitor or DPP-4 inhibitor fixed-dose combinations with metformin will be included. ***Index date:*** Date of first purchase/prescription of empagliflozin, other SGLT-2 inhibitor, or DPP-4 inhibitor during follow-up (2014-2019). Patients initiating SGLT-2 inhibitor or DPP-4 inhibitor fixed-dose combinations with metformin will 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 • 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-T2 diabetes • Use of a SGLT-2 inhibitor or a DPP-4 inhibitor 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 **Definition of exposure:** The primary definition of exposure for as treated (AT) analysis is whether an individual is exposed to one of the following three index drug groups during the study period starting from 6 months after the index date and ending at treatment discontinuation of index drug: • empagliflozin, • other SGLT-2 inhibitors (dapagliflozin or canagliflozin) or • DPP-4 inhibitor (sitagliptin, saxagliptin, linagliptin, vildagliptin, alogliptin) Patients initiating SGLT-2 inhibitor or DPP-4 inhibitor 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 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)

Data sources:	The following databases are country-specific data sources from which a linked master database for each country will be constructed. **UK:** CPRD GOLD, HES, cancer registry and ONS mortality sources. **Sweden:** The following nation-wide population based registers will be accessed: the National Patient Register, the Swedish Prescribed Drug Register, the Swedish Cancer Register, the Causes of Death Register and the National Diabetes Register. Initially, the country-level datasets will be analysed separately. Thereafter, a pooled dataset combining the patient level datasets from different countries will be analysed if data allows (homogeneity), in order 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 9.4.3.
Study size:	The study size will be driven by the uptake of empagliflozin for the treatment of T2D to improve glycaemic control in adults in the UK and Sweden. The study size required to detect a relative risk of 3 with 80% power ranges from 10,590 person-years of empagliflozin use for urinary tract malignancies to 34,220 person-years of empagliflozin use for renal cancer with incidence rates of about 80, 60 and 27 per 100,000 person-years, respectively (table below). The study size required to detect a relative risk of 2 with 80% power ranges from 29,330 person-years of empagliflozin use for urinary tract malignancies to 94,510 person-years of empagliflozin use for renal cancer with incidence rates as stated above (table below).

	Required Person-Years of Empagliflozin Use		
Level of Risk (HR) Detectable at 80% Power	Urinary Tract Malignancies	Bladder Cancer	Renal Cancer
2	29,330	43,180	94,510
3	10,590	15,480	34,220
Incidence Rates per 100,000 py	84.0	57.5	26.5

It is expected that by the end of 2019, approximately 18,000 empagliflozin-treated patients will accumulate in the CPRD and Sweden databases; however, the number of patients available for analysis may be lower.

The power to detect relative risks of 1.25, 1.5, 2 and 3 were determined based on the estimated number of individuals exposed to empagliflozin (approximately 18,000), a range of incidence rates of bladder and renal cancers in Nordic region (40, 60, 80 and 100 per 100 000 person years) with a weighted average of 2 years of follow-up time under 1:1 ratios of matching empagliflozin initiators with initiators of comparison drugs. The power calculations were based on Cox's proportional hazards (Cox's PH) model with type 1 error probability of 5% yielding power of 10%, 22%, 57% and 95% to detect relative risks of 1.25, 1.5, 2 and 3 respectively for an incidence of 40 per 100,000 person years and 17%, 47%, 92%, and 99%, respectively, for an incidence of 100 per 100,000 person years.

Based on the feasibility assessment, utilization of empagliflozin in the UK and Sweden is currently low. Accrual of empagliflozin users will be monitored in all data sources at 24 months after empagliflozin use starts being captured in the data sources (November 2016) and annually thereafter until 2019. 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 study period, expanding the study population, and/or changing the primary study design to a nested case-control study. All such changes will be approved via a protocol amendment.

Data analysis:	For the primary, secondary and further objectives, the main data analysis will be conducted in two stages: (i) construction of the

	propensity score (PS)-matched cohorts (for each comparison) by modelling the exposure to empagliflozin vs. SGLT-2 inhibitor and the exposure to empagliflozin vs. DPP-4 inhibitor and (ii) estimating the effect of exposure to empagliflozin on the urinary tract (bladder, renal and other) cancers using adjusted hazard ratio compared to those exposed to SGLT-2 inhibitor or DPP-4 inhibitor and their respective incidence rates. 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 as treated 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 another SGLT-2 inhibitor and metformin or 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 propensity scores, stratification, or adjustment. Incidence rates (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 adjusted for relevant variables using the Cox's proportional hazards model with the time-varying covariate approach. The adjusted hazard ratios and incidence rates will be presented along with 95% confidence intervals for the risk estimates. Sensitivity analyses will be performed using the nested-case-control (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 comparator groups. 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). Hazard ratios using the Cox's proportional hazard 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.
Milestones:	Jardiance (empagliflozin) was granted EU marketing authorisation in May 2014. Launch in the UK was in August 2014; launch in Sweden took place November 2014. Synjardy (empagliflozin/metformin HCl) was granted EU marketing

	authorisation in May 2015. Protocol version 1 was submitted to the EMA for review in September 2015 and version 2 in February 2016. Registration in the EU PAS register will occur two months following regulatory protocol endorsement. The study start will depend on the dates of protocol approval and market uptake and will take into account the 13-month lag time needed to access the data. The number of users in each treatment cohort will be monitored 24 months after empagliflozin use starts being captured in the data sources (November 2016) and annually thereafter until 2019. Annual interim reports of these monitoring data will be prepared in years 2017-2020 and will be submitted to EMA by the Sponsor within the earliest corresponding Periodic Safety Update Report (PSUR). 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, the use of other country data sources, and/or changing the primary study design to a nested case-control study will be approved via a protocol amendment. Taking into account the database lag, the analysis of data from May 2014 through December 2019 will be performed in March 2021 and the final report will be produced by June 2021.

5. AMENDMENTS AND UPDATES

None.

6. MILESTONES

Milestone	Planned Date
Protocol endorsed by the EMA	Q3 2016
Start of data collection	Expected during Q3 2016, depending on final protocol approval by the EMA
Interim report 24 months	Q1 2017 Will be based on data available 24 months after use of empagliflozin is first captured (November 2016)
Interim report 36 months	Q1 2018 Will be based on data available 36 months after use of empagliflozin is first captured (November 2017)
Interim report 48 months	Q1 2019 Will be based on data available 48 months after use of empagliflozin is first captured (November 2018)
Interim report 60 months	Q1 2020 Will be based on data available 60 months after use of empagliflozin is first captured (November 2019)
End of data collection	Q4 2019 Taking into account the database lags, the analysis of data through December 2019 will be performed in March 2021
Registration in the EU PAS register[1]	Two months following regulatory protocol endorsement
Final report of study results	Expected June 2021
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.
Note: Sponsor to submit to EMA the interim and final study reports within the earliest corresponding PSURs.
1. Website for the EU PAS Register: encepp.eu/encepp_studies/indexRegister.shtml.

7. RATIONALE AND BACKGROUND

Jardiance (empagliflozin), a highly potent and selective inhibitor of the sodium-glucose cotransporter 2 (SGLT-2), was approved in Europe in May 2014 for the treatment of type 2 diabetes mellitus (T2D) to improve glycaemic control in adults. Synjardy (empagliflozin/metformin HCl) 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 HCl is available in the following twice-daily doses:

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

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 HCl 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 SGLT-2 inhibitors and incident users of dipeptidyl peptidase-4 (DPP-4) inhibitors. Prior to the development of the protocol for this study, a feasibility assessment of three databases in Europe (the CPRD and data with potential for linkage (HES, Cancer registry 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.

Both comparator groups, other SGLT-2 and DPP-4 inhibitors, are an alternative treatment option to empagliflozin and are prescribed at a similar diabetes progression stage. Thus, these are considered suitable comparators (see Section 9.1 for further discussion of comparators).

7.1 EPIDEMIOLOGY OF URINARY TRACT MALIGNANCIES IN DIABETES PATIENTS

Based on data from the national cancer registries throughout the Nordic countries, the crude incidence rates for urinary tract malignancies (bladder and renal cancers) in the Nordic countries in 2013 was calculated as 43.0 per 100,000 person-years for the total population and 82.8 per 100,000 person-years for individuals over 40 years of age (Table 1) [R15-4873]. As the study population consists of type 2 diabetes 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 UK 2012

	Nordic Countries			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

Type 2 diabetes mellitus 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 diabetes [R10-6494, R12-2439, R12-0469]. A population-based cohort study on patients hospitalized with diabetes mellitus diagnoses in Denmark showed a higher risk of kidney cancers (with the standardized incidence ratios (SIRs) of 1.4 (95% 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 diabetes, 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 diabetes 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 incidence rates of bladder cancer in patients with diabetes range from 38 per 100,000

person-years in women from Denmark [R12-1334] to 140-150 per 100,000 person-years in men in Sweden and Denmark [R12-0368, R12-1334]. Meta-analyses of observational studies suggest that incidence of kidney cancer in patients with diabetes is approximately 40% higher than that in the general population [R13-1812]. The available crude incidence estimates of renal cancer in men with diabetes range between 40 and 50 per 100,000 PY in most studies.

Co-existence of hyperinsulinemia and hyperglycemia in type 2 diabetes 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), gender, chronic bladder irritation and infections, genetics and family history, chemotherapy and radiation therapy (R15-4866; R15-4856).

7.2 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 person-years in 1996 to 4.4 per 1,000 person-years in 2005. During the period 1996-2005, a change in oral glucose lower 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 person-years in 2007 to 0.8 per 1,000 person-years in 2009, at the same time that “other glucose-lowering drugs,” including DPP-4 inhibitors, increased from 0.2 per 1,000 person-years to 1.1 per 1,000 person-years [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 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 UK and Sweden

Drug category (ATC)	UK[1]	Sweden[2]
A10BA Biguanides	149,987	235,342
A10BB Sulphonylureas	64,006	35,890
A10BD Combinations	2,654	2,552
A10BF Alpha glucosidase inhibitors	424	703
A10BG Thiazolidinediones	9,849	1,535
A10BH DPP-4 inhibitors	30,918	13,430
A10BX01-07,10 Others, excl. insulins	9,088	15,151
A10BX09,11 SGLT-2 inhibitors (excl. empagliflozin)	5,840	832
A10BX12 Empagliflozin	108	373

[1] January 2014 – July 2015 (end date varies slightly by GP practice)
[2] 2014
[3] January – May 2015

7.3 EPIDEMIOLOGY OF DIABETES AND DIABETES TREATMENT PATTERNS IN SWEDEN

The total age-standardized prevalence of both pharmacologically and non-pharmacologically treated diabetes is 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).

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.

8. 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 other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor (two comparator groups).

Objectives:
Primary objectives:

(i) To estimate the adjusted hazard ratios and incidence rates of all urinary tract cancers (bladder, renal and other urinary tract cancers) in patients initiating empagliflozin compared to patients initiating other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor
(ii) To estimate the adjusted hazard ratios and incidence rates of bladder cancer in patients initiating empagliflozin compared to patients initiating other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor
(iii) To estimate the adjusted hazard ratios and incidence rates of renal cancer in patients initiating empagliflozin compared to patients initiating other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor

Secondary objective:

(i) To estimate the adjusted hazard ratio and incidence rates 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 adjusted hazard ratio and incidence rates of all urinary tract cancers with respect to empagliflozin initiation among sub-groups defined by age, gender, relevant comorbidities, concomitant use of metformin and other relevant treatments
(ii) To estimate the adjusted hazard ratio and incidence rates of non-renal, non-bladder urinary tract cancers in patients initiating empagliflozin compared to patients initiating other SGLT-2 inhibitors and to patients initiating a DPP-4 inhibitor

Non-renal, non-bladder urinary tract cancers (referred to as other urinary tract cancers) includes ureter and urethra cancers

9. RESEARCH METHODS

9.1 STUDY DESIGN

This is an observational, comparative, cohort safety study based on European healthcare databases and includes databases from the UK and Sweden**.**

The study will use an "incident users" design and compare new users of empagliflozin to two comparison groups, new users of SGLT-2 inhibitors other than empagliflozin and 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, another SGLT-2 inhibitor, 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 a SGLT-2 inhibitor or a DPP-4 inhibitor during the available look-back (pre-index) period [R13-1120, R14-4378].

Other SGLT-2 inhibitors have been selected as a comparator group as patients with T2D treated with SGLT-2 inhibitors other than empagliflozin are expected to constitute the most comparable patient group due to similar indications and target population. DPP-4 inhibitors have been selected as an additional 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 these three different drug groups may have inherent channeling bias wherein the treatment assignment is influenced by a range of factors such as disease progression, comorbidity, treatment history, age, and gender. Additionally, some of these variables lead to increased or decreased risk of cancer. To minimize channeling bias, individuals are matched with similar treatment and clinical history at index date by deriving propensity scores conditional on factors affecting the treatment and outcome [R13-1120, R14-4378]. Propensity-score-matched cohorts will be derived for each comparison from the study base by matching patients who initiate empagliflozin with patients who initiate other SGLT-2 inhibitors and patients who initiate DPP-4 inhibitors at index date with a pre-fixed ratio (1:1, or 1:3 in sensitivity analyses, using Greedy matching methods). Depending on sample size, combination exposures with metformin will be accounted for through separate propensity scores, 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 incidence rates and adjusted hazard ratios of multiple outcomes of interest among new users of empagliflozin compared with new users of other SGLT-2 inhibitors and 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 9.3.3]. 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 9.3.5]. In the context of data sources such as the CPRD and the Sweden national registers, the use of a cohort design has more advantages than limitations compared with the use of a nested case-control 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 three treatment groups. Several sensitivity analyses will be conducted including analyses using a nested-case-control (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. A pooled dataset combining the patient level datasets from different countries will be analysed if data allow (i.e. homogeneity is met), in order to gain power and produce as narrow confidence intervals as possible.

9.2 SETTING

9.2.1 Study populations

The source population will include all patients with T2D from the UK and Sweden. In the UK, the population will include patients with T2D from two of the CPRD datasets: the CPRD database of longitudinal medical records collected from UK primary care practices (CPRD-

GOLD). Sensitivity analyses may include a subset of the CPRD GOLD database with linkage to Hospital Episode Statistics (HES), and if available, cancer registry and mortality databases if linked data with sufficient sample size during the study period are available (see Section 9.4.1). In Sweden, patients with T2D in the nationwide registers will be included. All patients with T2D who have purchased at least one prescription of empagliflozin, other SGLT-2 inhibitor drugs, or DPP-4 inhibitor drugs during the study period 2014-2019 will be included in the study. Identification of diabetes will be based on medical/prescription records and all non-T2D patients will be excluded.

9.2.2 Study period

The study will include individuals who have purchased at least one prescription of empagliflozin, other SGLT-2 inhibitor drugs, or DPP-4 inhibitor drugs from 1st August 2014 through December 2019. The starting of the study period corresponds to the time of drug licensing (latter 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 2019. Based on the available patient numbers, a decision will be made regarding the length of the study period, the need to use other country data sources, and/or changing the primary study design to a nested case-control study. The maximum follow-up time for an individual will be 5 years.

9.2.3 Index date

The index date will be defined as the date of first purchase/prescription of empagliflozin, other SGLT-2 inhibitor, or DPP-4 inhibitor during the study period (2014-2019).

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

To characterise the empagliflozin, other SGLT-2 inhibitor, and 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, other SGLT-2 inhibitor, and DPP-4 inhibitor groups, 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 propensity score development, to control for possible differences in availability of information between the empagliflozin and comparator cohorts.

9.2.5 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 type 2 diabetes
    - Lists of codes that have been previously used to identify patients with T2D in the CPRD [P14-16383] are included in Annex 3. For Sweden 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).
- 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 type 1 diabetes or other specific non-type 2 diabetes
- Use of a SGLT-2 inhibitor or a DPP-4 inhibitor 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 end stage renal disease or receipt of renal dialysis recorded at any time prior to index date (i.e. during the available look-back time)

9.3 VARIABLES

The final list of variables and operational definitions will be presented in a separate statistical/epidemiological analysis plan (SEAP) to be developed prior to the start of data analysis.

9.3.1 Exposures

Empagliflozin (ATC code A10BX12; Annex 4) is the exposure of interest, free form or fixed-dose combination with metformin.

Two comparator exposure groups will be identified as follows. (See Section 9.1)

SGLT-2 inhibitors other than empagliflozin:

- Canagliflozin (ATC code A10BX11)
- Dapagliflozin (ATC code A10BX09)

- Canagliflozin and metformin hydrochloride fixed-dose combination (A10BD16)
- Dapagliflozin and metformin hydrochloride fixed-dose combination (A10BD15)

DPP-4 inhibitors:

- Alogliptin (ATC code A10BH04)
- Linagliptin (ATC code A10BH05)
- Saxagliptin (ATC code A10BH03)
- Sitagliptin (ATC code A10BH01)
- Vildagliptin (ATC code A10BH02)
- Alogliptin and metformin hydrochloride (A10BD13)
- Linagliptin and metformin hydrochloride (A10BD11)
- Saxagliptin and metformin hydrochloride (A10BD10)
- Sitagliptin and metformin hydrochloride (A10BD12)
- Vildagliptin and metformin hydrochloride (A10BD08)

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, other SGLT-2 inhibitor, 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 9.1)

If additional SGLT-2 inhibitor or DPP-4 inhibitor drugs are introduced in the UK and Sweden during the study period, they will also be considered to be members of the respective comparator group.

The primary definition of exposure, i.e. as-treated analysis (AT), is defined as an ongoing exposure to one of the following three index drugs or drug groups during the study period starting from 6 months after the index date (see Sections 9.2.3 and 9.3.5): empagliflozin, other SGLT-2 inhibitor, or DPP-4 inhibitor [Table 3].
As a typical prescription length is 1 month in UK and 3 months in Sweden, 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].

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 Empagliflozin (as-treated analysis)	Indicator of current use of empagliflozin, other SGLT-2 inhibitor, 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, other SGLT-2 inhibitor, 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) empaglifozin	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

Table 3 (cont'd) Exposure definitions for primary and secondary analyses

Exposure definition	Description	Time-dependent or Fixed at index date
Ever used metformin	Was metformin ever used – taking value 'yes' as soon as one prescription with metformin is purchased	Time-dependent
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

9.3.2 Outcomes

9.3.2.1 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 data and the national cancer registries in Sweden improving confidence in the accuracy and validity of these outcome diagnoses. [See Section 9.4]

Date of diagnosis of the first incidence of urinary tract cancers as specified in CPRD-GOLD or in the Sweden 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 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 using READ codes (Annex 5). A study evaluating the validity of cancer diagnoses in CPRD-GOLD 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 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, including utilization of the linked CPRD-GOLD, HES, and cancer registry data, will be performed to address some of the accuracy and validity issues in the primary outcomes [See Section 9.7.2].

9.3.2.2 Further Outcomes

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

9.3.3 Covariates

Variables potentially associated with urinary tract cancers, such as sociodemographic variables including age, sex, socioeconomic status, body mass index, smoking and alcohol consumption, concomitant medications, comorbidities, and duration of look-back period (see Section 7.1), 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.
Study participants will also be classified by indicator variables on the calendar time of cohort entry (by quarter) whether the index treatment (empagliflozin, other SGLT-2 inhibitor, or DPP-4 inhibitor) was added to existing medication (adding on), or if the index treatment was initiated as a replacement for another antihyperglycemia medication (switching to empagliflozin, other SGLT-2 inhibitor, 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 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 include specifying of variables for exact or propensity matching, risk set definition in nested-case-control (NCC) approach and/or covariates for deriving inverse probability of treatment weight (IPTW) in marginal structural models (MSM). Accounting for by analyses include stratification on and model adjustment for these covariates.
The examples of covariates in Table 4 describe the comorbidities, treatment history, and socio-demographic variables that will be included. 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.

Whereas propensity score (PS) matching is useful to account for channeling bias at index date, the same variable will be used for model adjustment at censoring or outcome if the variable is not balanced across treatment groups after PS matching.

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

Covariate	Exact matching	PS matching	Stratification	Risk set (NCC) / covariate in MSM	Model adjustment
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 antihyperglycemia 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
Prior history of UTI or pyelonephritis		Yes		Yes	Yes
Liver disease		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
Heart failure		Yes		Yes	Yes
Hypertension		Yes		Yes	Yes

Table 4 (cont'd) Baseline covariates to be included in primary and secondary analyses

Covariate	Exact matching	PS matching	Stratification	Risk set (NCC) / covariate in MSM	Model adjustment
Stroke		Yes		Yes	Yes
Myocardial infarction		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/diuretics					Yes
Non-steroidal anti-inflammatory drugs (NSAIDs)					Yes
Oral steroids					Yes
Statins, fibrates					Yes
Lipid modifying agents					Yes
Zoledronic acid					Yes
Antibiotics					Yes

9.3.4 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 9.2.5) 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 of the index drugs (empagliflozin, other SGLT-2 inhibitor, or DPP-4 inhibitor) plus a 3 months period following the switch, or as the stop date plus the a 3 months 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.

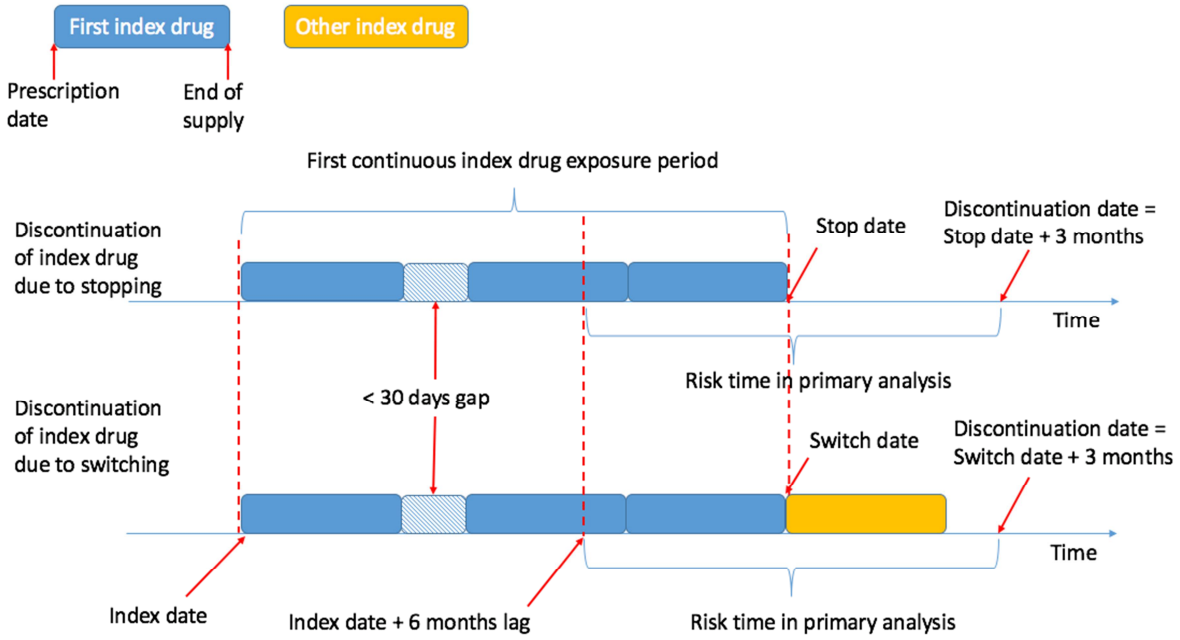

Figure 1 Discontinuation of index drug due to stopping of index drug or switching of index drug. In the primary analysis follow-up is censored at treatment discontinuation.

9.3.5 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 (Figure 1)

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, another SGLT-2 inhibitor, or a DPP-4 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.

9.4 DATA SOURCES

Prior to the development of the protocol for this study, a feasibility assessment of three databases in Europe (the CPRD and data with potential for linkage (HES, Cancer registry 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 CPRD and the national registries in Sweden

	UK				Sweden				
	CPRD	**HES**	**Cancer**	**Mortality**	**Patient**	**Prescription**	**Cancer**	**Cause of death**	**Diabetes**
Age	x	x	x	x	x	x	x	x	x
Sex	x	x	x	x	x	x	x	x	x
Medication	x	-	-		x[3]	x	-	-	x[5]
Defined Daily Dose	x	-	-	-	x	x	-	-	-
Purchase date	x[1]	-	-	-	x	x	-	-	-
Diagnoses	x	x	x	-	x[4]	-	x	-	x
Date of diagnosis	x	x	x[2]	-	x	-	x	-	-
Time of death	-	-	x[2]	x	-	-	x	x	-
Cause of death	-	-	x	x	-	-	x	x	-
Smoking	x	-	-	-	-	-	-	-	x
BMI	x	-	-	-	-	-	-	-	x
HbA1c	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

9.4.1 United Kingdom Clinical Practice Research Datalink (CPRD)

The UK CPRD-GOLD data contains the anonymised longitudinal medical records managed by GPs working the National Health Service (NHS) primary care setting. CPRD GOLD data cover approximately 8.8% of the UK population, including practices in England, Wales, Scotland and Northern Ireland. The data contains demographic information, diagnoses, prescriptions, tests and referrals. Data has been collected prospectively since 1987. Approximately 58% 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. 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 Cancer Registry (including histology, stage and grading) and 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.
Based on the feasibility assessment, the UK data sources evaluated (CPRD and data with potential for linkage - HES, Cancer registry and ONS mortality statistics) met the minimal data requirements (Table 5). The following CPRD data sources are considered: CPRD GOLD, HES, cancer registry and Office for National Statistics (ONS) mortality sources. In the UK, the full study data including follow-up period until the end of 2019, will be available during the first half of 2021. However, the linked data to the Cancer Register data is only available from 1985 to 2010 (as of June 2015). In the near future, the time period for this linked data is expected to be extended until December 2013, but the actual time of this extension is not yet known. It is also uncertain when the linked data for the study period of this particular study (2014-2019) will be available. Consequently, the primary analyses in CPRD will be based on the CPRD-GOLD data; sensitivity analyses may include linked HES and, if available, cancer registry data if linked data with sufficient sample size during the study period are available (Section 9.7.2.3.3).

9.4.2 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 and the National Diabetes Register. 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 2019 will be available during the first half of 2021.

9.4.3 Additional potential data sources

As described in Section 9.5, utilization of empagliflozin in the UK and Sweden is still low, therefore the study population may need to be expanded to include additional data sources. A potential additional data source in Europe includes the national registers in Finland, which have been evaluated in the feasibility assessment but lacked some potentially important confounder information. It is important for the choice of the appropriate data source to have the appropriate information on cancer and important risk factors included. As an additional option, a commercial data source in the US could be used.

Beside the evaluated data sources 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. In the following section, the proposed potential alternative data sources in Finland are described, as these data sources were evaluated as part of the feasibility assessment.

9.4.3.1 Finland national registries

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. 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, the nationwide prescription registers contain information on outpatient medication purchases in pharmacies. The register contains information on all purchased prescribed medicines within the reimbursement scheme. In addition, patients who are entitled for special refunds (65% or 100% reimbursement) of medicine expenses based on their chronic condition can be identified using the registry for reimbursed medications. For example, the special reimbursement category 103 is for diabetes. This database is used for identification of comorbidities. The national e-prescription register contains information on prescribed prescriptions irrespective of reimbursement status. The coverage of the e-prescription register is growing and depends on the prescribing unit.

Based on the feasibility assessment, the Finland registries were found to lack data on important potential confounders, specifically BMI, smoking, and laboratory measures such as HbA1c. Before data from the Finland registries could be used, an evaluation of these missing variables is needed to determine if they are confounders or effect modifiers for each of the comparisons. Data from Sweden and the UK would be used for this evaluation. In particular, for comparisons between empagliflozin and DPP-4 inhibitors, baseline characteristics of BMI, for example, may be different across the groups. BMI is an important risk factor or surrogate parameter for urinary tract cancers and empagliflozin initiators may potentially

have a higher BMI than DPP-4 inhibitor initiators given the modest weight loss that has been observed in some patients taking empagliflozin and other SGLT-2 inhibitors.

In addition, the ability to identify patients for the study through a disease cohort, rather than through prescription-only-based definitions, would be necessary before inclusion on the Finland data as a primary data source.

The following data sources would be included: the Finnish prescription register / the registry for reimbursed medications, the register of specialist medical care (HILMO), the register of primary medical care (AvoHILMO), the Finnish cancer register and the cause of death register. The study data would be extracted from the relevant registers using the general process as described above for Sweden. In Finland, the full study data including follow-up period until the end of 2019 would be available during the first half of 2021 only.

9.5 STUDY SIZE

The study size will be driven by the uptake of empagliflozin for the treatment of T2D to improve glycaemic control in adults in the UK and Sweden. The study size required to detect a relative risk of 3 ranges from 10,590 person-years of empagliflozin use for urinary tract malignancies to 34,220 person-years of empagliflozin use for renal cancer with incidence rates of 84.0, 57.5 and 26.5 per 100000 person-years, respectively (Table 6). The study size required to detect a relative risk of 2 with 80% power ranges from 29,330 person-years of empagliflozin use for urinary tract malignancies to 94,510 person-years of empagliflozin use for renal cancer with incidence rates as stated above (Table 6).

Table 6 Required sample sizes (in person-years) to estimate the effect sizes of 1.5, 2.0 and 3.0 for the three primary endpoints with their respective incidence rates with power of at least 80%

Primary endpoints	RR=1.5	RR=2	RR=3
All UT cancers (incidence 84.0 per 100 000)	96 140	29 330	10 590
Bladder cancer (incidence 57.5 per 100 000)	138 510	43 180	15 480
Renal cancer (incidence 26.5 per 100 000)	301 450	94 510	34 220

It is expected that by the end of 2019, approximately 18 000 empagliflozin-treated patients with age ≥18 years will accumulate in the UK and Sweden databases; however, the number of patients available for analysis may be lower.
The power to detect relative risks of 1.25, 1.5, 2 and 3 were determined based on the estimated number of individuals exposed to empagliflozin (approximately 18,000), a range of incidence rates of bladder and renal cancers in Nordic region (40, 60, 80 and 100 per 100 000

person years) with a weighted average of 2 years of follow-up time under matching ratios of empagliflozin initiators with initiators of comparison drugs of 1:1, 1:3, and 1:5 [Table 7]. The power calculations were based on Cox model with type 1 error probability of 5% yielding power of 10%, 22%, 57% and 95% to detect relative risks of 1.25, 1.5, 2 and 3, respectively, for an incidence of 40 per 100,000 person years and 17%, 47%, 92%, and 99%, respectively, for an incidence of 100 per 100,000 person years.

Table 7 Power of the study for different effect sizes and different matching ratios based on Cox model with type 1 error probability of 5% and 18,000 empagliflozin initiators with an average of 2 years of follow-up

Effect size (Relative risk / hazard ratio)	Incidence: 40 / 100000 person-years	Incidence: 60 / 100000 person-years	Incidence: 80 / 100000 person-years	Incidence: 100 / 100000 person-years
1:1 matching				
1.25	10%	12%	14%	17%
1.5	22%	31%	39%	47%
2.0	57%	75%	86%	92%
3.0	95%	99%	99%	99%
1:3 matching				
1.25	12%	16%	20%	24%
1.5	34%	47%	58%	68%
2.0	82%	94%	98%	99%
3.0	99%	99%	99%	100%
1:5 matching				
1.25	13%	18%	22%	27%
1.5	38%	52%	65%	74%
2.0	88%	97%	99%	99%
3.0	99%	99%	100%	100%

The follow-up time required to achieve 80% power with an estimated number of empagliflozin users of 18,000 was estimated for a range of incidence rates and relative risks. For an incidence rate of 80 per 100,000 person-years, 1.7 to 5.6 years of follow-up time is required to detect a relative risk of 2.0 and 1.5, respectively (Table 8).

Table 8 Required follow-up time (years) to achieve 80% power by levels of relative risk and incidence rates based on Cox model with type 1 error probability of 5% and 18,000 empagliflozin initiators

Incidence rates per 100,000 person-years	**RR = 1.5**	**RR = 2.0**
40	11.00	3.50
60	7.50	2.30
80	5.60	1.70
100	4.5	1.4

Contour plots of power for ranges of incidence rates, empagliflozin exposures, and risk estimates are provided in Figure 1.

Based on the feasibility assessment, utilization of empagliflozin in the UK and Sweden is currently still low. Accrual of empagliflozin users will be monitored in all data sources as described in Section 9.7.1.7. 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 November 2019. The numbers will be reported in annual interim reports in the 1st quarter of years 2017-2020 and the interim reports will be submitted to EMA by the Sponsor within the earliest corresponding PSUR. 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. 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 other country data sources, and/or changing the primary study design to a nested case-control study. All such decisions will be approved via a protocol amendment.

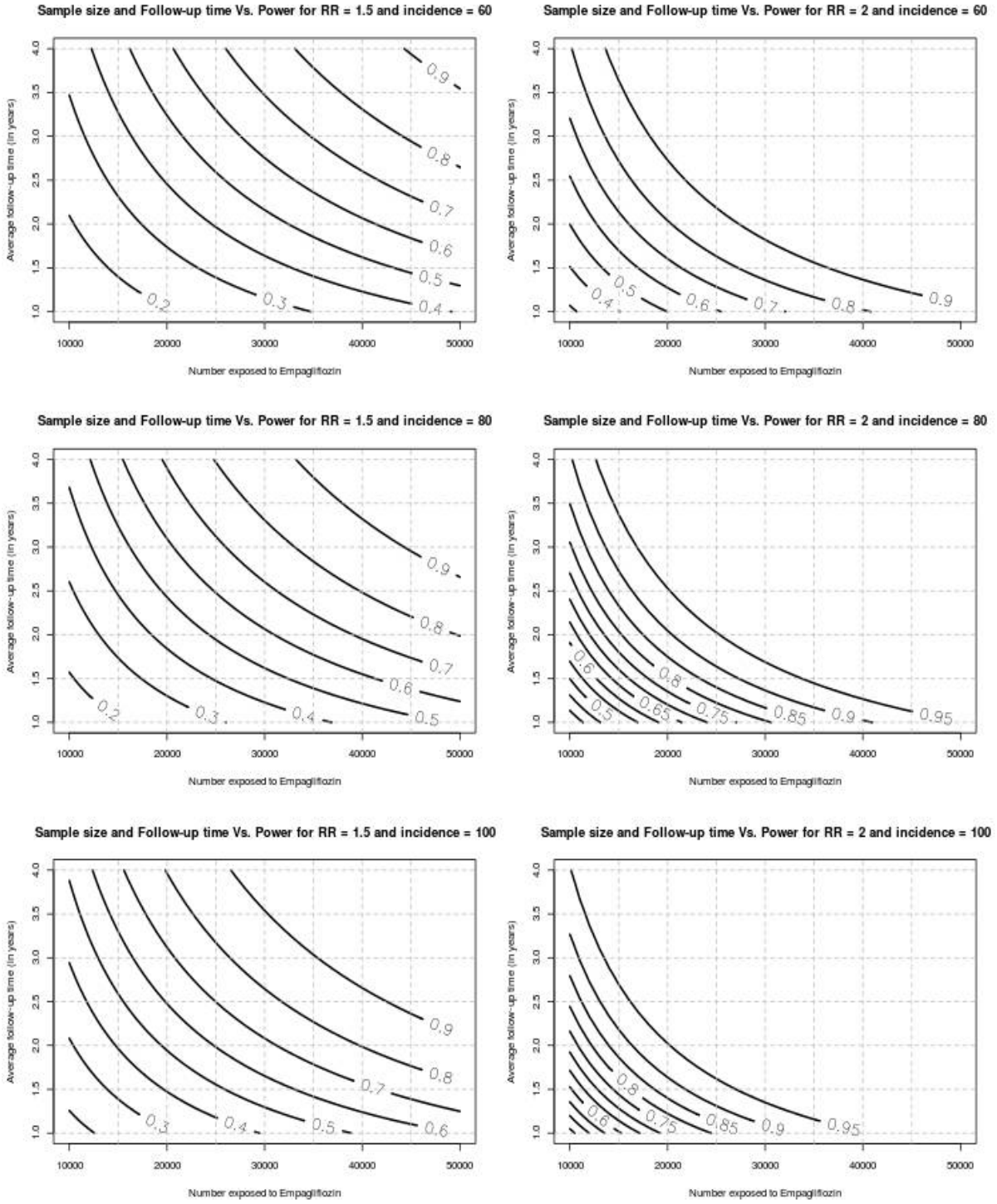

Figure 2 Contour plots showing power for a range of sample sizes, average follow-up times, levels of detectible risk, and incidence with a 1:1 matching ratio

9.6 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 statistical/epidemiological analysis plan (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.

9.7 DATA ANALYSIS

The final approach to data analysis will be presented in a separate statistical/epidemiological analysis plan (SEAP) to be developed prior to the start of data analysis. The SEAP will include a detailed algorithm for identifying T2DM 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.

9.7.1 Main analysis

For the primary, secondary and further objectives, the main data analysis will be conducted in two stages: (i) construction of the propensity score (PS)-matched cohort by modelling the exposure to empagliflozin vs. other SGLT-2 inhibitors and exposure to empagliflozin vs. DPP-4 inhibitor and (ii) estimating the effect of exposure to empagliflozin on the urinary tract (bladder, renal and other) cancers using adjusted hazard ratio and incidence rates compared to those exposed to other SGLT-2 inhibitor initiators or DPP-4 inhibitor. Additional sensitivity analyses will be performed to validate the robustness of alternative definitions of outcome, exposure, covariates and reduction of bias due to matching.

9.7.1.1 Matching at index date

For all individuals, propensity scores for each comparison (empagliflozin vs other SGLT-2 inhibitors; 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 propensity scores will be estimated using a binary logistic

regression model. At least two sets of propensity scores, one for the comparison of empagliflozin versus other SGLT-2 inhibitors and one for the comparison of empagliflozin versus DPP-4 inhibitors, will be generated. Depending on sample size, combination exposures with metformin will be accounted for through separate propensity scores, stratification, or adjustment.

The selection of variables to be included in the propensity score models (see Section 9.3.3) 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 propensity score 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 propensity scores [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 propensity score 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 propensity scores, stratification, or adjustment.

Matching will be performed separately for empagliflozin initiators with patients initiating each of the two comparators within each country database. The matching is described for other SGLT-2 inhibitor initiators and the same procedure will be repeated for DPP-4 inhibitor initiators. For each empagliflozin initiating individual *i* with a particular index date T_i^E, all individuals initiating other SGLT-2 inhibitors within three months before or after T_i^E will be chosen. Of these, one (or three in sensitivity analyses) individual(s) with the closest propensity score(s) to that of *i* within a predefined propensity score interval will be selected as comparators (Greedy matching method). The predefined propensity score interval will be calculated based on the standard deviation of the logit of PS (δ) multiplied by some coefficient k (say, 0.2δ, 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). Matching by time period, based on index dates, enables better control of time-varying confounders including changing prescription patterns for empagliflozin and potential confounding.

9.7.1.2 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 any other of the index drug (empagliflozin, other SGLT-2 inhibitor, or DPP-4 inhibitor) during follow-up or stop using the index drug. [see Section 9.3.4]

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

9.7.1.4 Comparative analyses

Relative risks (hazard ratios) for the defined outcomes will be estimated using Cox's PH model with time-varying covariates, if model pre-requisites are fulfilled. Hazard ratios will be presented along with 95% CIs.

The Cox's PH models will include the exposure and both comparator groups within the model. This will account for multiple testing and for retaining same variables in variable selection (see Section 9.7.1.6). The models are defined in terms of covariates adjusted for in addition to the primary exposure: (i) base-line 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's PH models will use datasets from one country at a time. The analyses will include datasets from UK and Sweden, separately.

Pooled analysis: Pooled analyses will utilise data from multiple countries in each model. Pooled analysis will be carried out when there is homogeneity of variable definitions and effect size of outcomes with respect to primary exposures among country-specific analyses.

9.7.1.5 Variable selection

All potential confounders, exact and PS matching variables will enter the outcome model only after being selected using pre-defined variable selection criteria to be detailed in the SEAP. Variable selection algorithm will be applied separately for country-specific and pooled analysis, if applicable.

9.7.1.6 Monitoring of accrual of empagliflozin users

Accrual of empagliflozin users, including accrual of metformin combination use with free or fixed dose combinations, will be monitored at 24 months after use of empagliflozin is first captured in the data sources (November 2016), and annually thereafter until 2019. Analyses will be conducted at 24 months after use of empagliflozin is first captured in the data sources (November 2016) and annually thereafter until 2019 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. Annual interim reports of these monitoring data will be prepared in years 2017-2020 and will be submitted to EMA by the Sponsor within the earliest corresponding Periodic Safety Update Report (PSUR). The available power to estimate the association between empagliflozin use and the outcomes will be examined. 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. 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 other country data sources, and/or changing the primary study design to a nested case-control study. A protocol amendment will be implemented to reflect these changes and submitted for EMA endorsement.

As a part of the interim reports the latest knowledge on the role of metformin in urinary cancers at time of the interim reports will be evaluated and reported.

9.7.1.7 Handling of missing data

A high frequency of missing values is not expected for most variables, with the possible exception of lifestyle variables. For the medical history conditions / comorbidities to be collected for inclusion in the propensity score, 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 (UK or Sweden), 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.

9.7.2 Further analysis

9.7.2.1 Stratified analysis

Depending on sample size, stratified analyses may be performed by estimating the hazard ratio 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). This analysis will be carried out using the pooled data (if appropriate, i.e. no heterogeneity between data sources) in order to have adequate power to estimate effect sizes within sub-groups.

The analyses will be stratified at least by the following subgroups.

- Concomitant metformin use (either as fixed-dose or free combination)

- Treatment complexity (monotherapy, dual combination therapy, triple combination therapy)

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

9.7.2.2 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 9.3.1 (current use, cumulative dosage, dosage per use and time since first use) will be utilized in secondary analyses.

9.7.2.3 Sensitivity Analyses

Additional sensitivity analyses may be defined in the SEAP.

9.7.2.3.1 Sensitivity analyses: study design

As sensitivity analyses to the two-stage modelling, two additional data analyses approaches will be implemented: nested-case-control (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 (UT) cancers) along with a matched sample of controls (individuals with no cancer event) among empagliflozin initiators and comparator groups. For each case, five controls will be sampled randomly from a corresponding risk set defined by relevant characteristics of the case. Hazard ratios from Cox's PH 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 channeling bias at index date, MSMs will be used to adjust for time-dependent confounding throughout follow-up by estimating inverse-probability of treatment weights (IPTW). Hazard ratios using the Cox's models will be presented along with 95% CIs.

9.7.2.3.2 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.

9.7.2.3.3 Sensitivity analyses: outcomes

Sensitivity outcomes include
- Non-renal, non-bladder urinary tract cancers
- Cancers of the urinary tract from the linked records in GPRD-GOLD, HES, and the cancer registry, if possible (if linked data with sufficient sample size during the study period is available)
- Neoplasms of uncertain or unknown behavior (non-malignant and not carcinomas in situ) of the urinary tract

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

In addition, 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).

9.7.2.3.4 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 9.3.4] In a sensitivity analysis this 3-month delay time will be extended up to 6 months.

9.8 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 European Medicines Agency 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.

9.9 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 expected number of empagliflozin users in the study is low, only 18 000, thus it is not sufficiently powered to detect weak effects, with <20% power for relative risk 1.25 and <50% power for relative risk of 1.5. Although, the study is sufficient to detect relative risks of 2.0 with 80% power, it is important to monitor the actual number of patients during the follow-up period. 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 other country data sources, and/or changing the primary study design to a nested case-control study. A protocol amendment will be implemented to reflect these changes and submitted for EMA endorsement.
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 to results.
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 urinary tract 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 propensity score for all unmatched cases after the first round of matching.
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 very 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).

9.9.1 BIAS

Potential source of difference in cancer incidence may arise due to channeling bias, detection bias and confounding. The channeling bias is addressed through both two-stage modeling approach wherein individuals are first matched on propensity scores 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 9.9, 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.

9.9.2 Generalizability

In the current study 2 population-based data sources (CPRD GOLD and Swedish 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. 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 T2DM. Therefore, the study results can be generalized to similar patients with T2DM in other geographic settings, including most industrialized countries.

10. PROTECTION OF HUMAN SUBJECTS

To ensure the full data protection of patients, all the research data in each country is anonymized. Approval from relevant Ethical/Research Review Boards will be required before conducting the study.

11. 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.

12. 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 reports to EMA 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.
Based on these results the principal and co-investigators will co-author scientific manuscript(s) of the results to be published. The publication strategy has been 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 study 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.

13. REFERENCES

13.1 PUBLISHED REFERENCES

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.
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.
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.
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.
R15-4864	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.
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.
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.
R12-1334	Carstensen B, Witte DR, Friis S. Cancer occurrence in Danish diabetic patients: duration and insulin effects. *Diabetologia* 2012; 55:948-958.
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.
R15-4862	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.
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.
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.
R13-5420	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.
R13-1395	ENCePP. ENCePP checklist for study protocols (revision 2). European Network of Centres for Pharmacoepidemiology and Pharmacovigilance; 14 January 2013. Available at: website: encepp.eu/standards_and_guidances/index.shtml. Accessed 7 February 2013.
R14-5282	ENCePP. Guide on methodological standards in pharmacoepidemiology (revision 3). EMA/95098/2010 Rev.3. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance; 2014. Available at: website: encepp.eu/standards_and_guidances/methodologicalGuide.shtml. Accessed 14 July 2014.
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
R15-4859	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.
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.
R11-5320	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.
R11-4318	ISPE, International Society for Pharmacoepidemiology. Guidelines for Good

	Pharmacoepidemiological Practices. Revision 2: April 2007. www.pharmacoepi.org/resources/guidelines_08027.cfm.
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.
R15-4857	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.
R15-4856	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.
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.
R12-0368	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.
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.
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.
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.
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.
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.
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.
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
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.
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.
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.
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.
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.
R10-6494	Vigneri P, Frasca F, Sciacca L, Pandini G, Vigneri R. Diabetes and cancer. *Endocrine-Related Cancer*, 2009; 16: 1103-23.
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.
R15-4868	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.
R16-2539	Murchie P1, Campbell NC, Delaney EK, Dinant GJ, Hannaford PC, Johansson L, Lee AJ, Rollano P, Spigt M. Fam Pract. 2012 Feb;29(1):69-78. doi: 10.1093/fampra/cmr044. Epub 2011 Aug 9. Comparing diagnostic delay in cancer: A cross sectional study in three European countries with primary care led health care systems.

13.2 UNPUBLISHED REFERENCES

N/A

ANNEX 1. LIST OF STAND-ALONE DOCUMENTS

N/A

ANNEX 2. ENCEPP CHECKLIST FOR STUDY PROTOCOLS

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

European Network of Centres for Pharmacoepidemiology and Pharmacovigilance

Doc.Ref. EMEA/540136/200

ENCePP Checklist for Study Protocols (Revision 2, amended)

Adopted by the ENCePP Steering Group on 14/01/2013

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 page number(s) 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). Note, the Checklist is a supporting document and does not replace the format of the protocol for PASS as recommended 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

Study reference number:
Study not yet registered in the EU PAS Register. Will be registered following protocol regulatory endorsement and prior to data collection start.

Section 1: Milestones	**Yes**	**No**	**N/A**	**Page Number(s)**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	☒	☐	☐	18
1.1.2 End of data collection[2]	☒	☐	☐	18
1.1.3 Study progress report(s)	☐	☒	☐	
1.1.4 Interim progress report(s)	☒	☐	☐	18
1.1.5 Registration in the EU PAS Register	☒	☐	☐	18
1.1.6 Final report of study results	☒	☐	☐	18

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Page Number(s)**
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)	☒	☐	☐	20
2.1.2 The objectives of the study?	☒	☐	☐	24
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	26
2.1.4 Which formal 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**	**Page Number(s)**
3.1 Is the study design described? (e.g. cohort, case-control, randomised controlled trial, new or alternative design)	☒	☐	☐	25
3.2 Does the protocol specify the primary and secondary (if applicable) endpoint(s) to be investigated?	☒	☐	☐	31

[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.

3.3 Does the protocol describe the measure(s) of effect? (e.g. relative risk, odds ratio, deaths per 1000 person-years, absolute risk, excess risk, incidence rate ratio, hazard ratio, number needed to harm (NNH) per year)	☒	☐	☐	43

Comments:

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Page Number(s)**
4.1 Is the source population described?	☒	☐	☐	35
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period?	☒	☐	☐	27
4.2.2 Age and sex?	☒	☐	☐	28
4.2.3 Country of origin?	☒	☐	☐	26
4.2.4 Disease/indication?	☒	☐	☐	26
4.2.5 Co-morbidity?	☒	☐	☐	28
4.2.6 Seasonality?	☐	☒	☐	
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	28

Comments:

Propensity score matching is used for sampling comparator groups and is explained in page 43.

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Page Number(s)**
5.1 Does the protocol describe how exposure is defined and measured? (e.g. operational details for defining and categorising exposure)	☒	☐	☐	29
5.2 Does the protocol discuss the validity of exposure measurement? (e.g. precision, accuracy, prospective ascertainment, exposure information recorded before the outcome occurred, use of validation sub-study)	☒	☐	☐	29
5.3 Is exposure classified according to time windows? (e.g. current user, former user, non-use)	☒	☐	☐	29
5.4 Is exposure classified based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☒	☐	
5.5 Does the protocol specify whether a dose-dependent or duration-dependent response is measured?	☒	☐	☐	29

Comments:

Section 6: Endpoint definition and measurement	**Yes**	**No**	**N/A**	**Page Number(s)**
6.1 Does the protocol describe how the endpoints are defined and measured?	☒	☐	☐	31
6.2 Does the protocol discuss the validity of endpoint measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, prospective or retrospective ascertainment, use of validation sub-study)	☒	☐	☐	31

Comments:

Section 7: Confounders and effect modifiers	**Yes**	**No**	**N/A**	**Page Number(s)**
7.1 Does the protocol address known confounders? (e.g. collection of data on known confounders, methods of controlling for known confounders)	☒	☐	☐	31; 43
7.2 Does the protocol address known effect modifiers? (e.g. collection of data on known effect modifiers, anticipated direction of effect)	☒	☐	☐	31; 43

Comments:

Section 8: Data sources	**Yes**	**No**	**N/A**	**Page Number(s)**
8.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
8.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview, etc.)	☒	☐	☐	35
8.1.2 Endpoints? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics, etc.)	☒	☐	☐	35
8.1.3 Covariates?	☒	☐	☐	35
8.2 Does the protocol describe the information available from the data source(s) on:				
8.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☒	☐	☐	35
8.2.2 Endpoints? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	35
8.2.3 Covariates? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, life style, etc.)	☒	☐	☐	35; Annex 6
8.3 Is a coding system described for:				
8.3.1 Diseases? (e.g. International Classification of Diseases (ICD)-10)	☒	☐	☐	31; 35; Annex 3
8.3.2 Endpoints? (e.g. Medical Dictionary for Regulatory Activities(MedDRA) for adverse events)	☒	☐	☐	31; 35; Annex 5
8.3.3 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	29; 35; Annex 4
8.4 Is the linkage method between data sources described? (e.g. based on a unique identifier or other)	☒	☐	☐	35

Comments:

Section 9: Study size and power	**Yes**	**No**	**N/A**	**Page Number(s)**
9.1 Is sample size and/or statistical power calculated?	☒	☐	☐	39

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Page Number(s)**
10.1 Does the plan include measurement of excess risks?	☒	☐	☐	45
10.2 Is the choice of statistical techniques described?	☒	☐	☐	43
10.3 Are descriptive analyses included?	☒	☐	☐	45
10.4 Are stratified analyses included?	☒	☐	☐	46
10.5 Does the plan describe the methods for adjusting for confounding?	☒	☐	☐	43
10.6 Does the plan describe methods addressing effect modification?	☒	☐	☐	43

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Page Number(s)**
11.1 Is information provided on the management of missing data?	☒	☐	☐	43
11.2 Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	43
11.3 Are methods of quality assurance described?	☒	☐	☐	48
11.4 Does the protocol describe possible quality issues related to the data source(s)?	☒	☐	☐	35
11.5 Is there a system in place for independent review of study results?	☐	☒	☐	

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Page Number(s)**
12.1 Does the protocol discuss:				
12.1.1 Selection biases?	☒	☐	☐	49
12.1.2 Information biases? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods)	☒	☐	☐	49
12.2 Does the protocol discuss study feasibility? (e.g. sample size, anticipated exposure, duration of follow-up in a cohort study, patient recruitment)	☒	☐	☐	39
12.3 Does the protocol address other limitations?	☒	☐	☐	48

Comments:

Section 13: Ethical issues	**Yes**	**No**	**N/A**	**Page Number(s)**
13.1 Have requirements of Ethics Committee/ Institutional Review Board approval been described?	☒	☐	☐	50
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☒	☐	
13.3 Have data protection requirements been described?	☒	☐	☐	43

Comments:

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Page Number(s)**
14.1 Does the protocol include a section to document future amendments and deviations?	☒	☐	☐	17

Comments:

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Page Number(s)**
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	☒	☐	☐	54
15.2 Are plans described for disseminating study results externally, including publication?	☒	☐	☐	54

Comments:

Name of the main author of the protocol: Pasi Korhonen, Fabian Hoti

Date: 15 September 2015

Signature: _____

ANNEX 3. READ CODES TO IDENTIFY TYPE 2 DIABETES

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

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
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
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
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
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
66Ak.00	Diabetic monitoring - lower risk albumin excretion
9OLM.00	Diabetes structured education programme declined
8CR2.00	Diabetes clinical management plan
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
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.
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
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
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
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
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
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
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	Time-dependent (T) / fixed at cohort entry (F)	How the effect of variable is accounted for *
1	AT / ITT: Empagliflozin	A10BX12	T / F	E
2	AT / ITT: DPP-4 inhibitor	ATC beginning with A10BH	T / F	E
3	AT / ITT: other SGLT-2 inhibitor	A10BX09, A10BX11	T / F	E
4	AT / ITT, FDCM**: empagliflozin	None yet	T / F	E
5	AT / ITT, FDCM**: DPP- 4 inhibitor	A10BD07, A10BD08, A10BD10, A10BD11, A10BD13	T / F	E
6	AT / ITT, FDCM**: other SGLT-2 inhibitors	A10BD15, A10BD16	T / F	E
7	Empagliflozin current use	A10BX12	T	E
8	Empagliflozin: cumulative duration of exposure	A10BX12	T	E
9	Empagliflozin: cumulative dosage of exposure	A10BX12	T	E
10	Empagliflozin: time since last dose	A10BX12	T	E
	Other OAD treatments			
11	Metformin	ATC beginning with A10BA	T	PS / A / VS, RS, TV, SG
12	Sulphonylureas	ATC beginning with A10BB	T	PS / A / VS, RS, TV
13	Insulin	ATC beginning with A10A	T	PS / A / VS, RS, TV
14	Thiazolidinediones	ATC beginning with A10BG	T	PS / A / VS, RS, TV
15	Alpha glucosidase inhibitors	ATC beginning with A10BF	T	PS / A / VS, RS, TV
16	GLP-1 agonists	A10BX02 – A10BX05, A10BX07, A10BX08, A10BX13, A10BX10	T	PS / A / VS, RS, TV

17	Treatment complexity (mono, dual and triple combination therapy)		T	PS / A / VS, RS, TV
18	Switch or add-on		T	PS / A / VS, RS, TV
	Exclusion / Censoring			
19	DPP-4 (Linagliptin) and Empagliflozin	A10BD19	T	Censoring / Exclusion
20	Discontinuation of primary exposure		T	Censoring

* 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

** FDCM – Fixed-dose combination with metformin

ANNEX 5. CODES TO IDENTIFY THE OUTCOME VARIABLES

No.	Variable	ICD-10 / ICD-O-3 codes	READ codes	Outcome / Censoring variable	Analyses *
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
6	Urinary Tract cancers from GP data (CPRD only) – sensitivity		Byu9, B837, B83z, B4A, B49	Sensitivity outcome	Sens
7	Urinary tract neoplasms of uncertain or unknown behavior - sensitivity	D41	B917, B91z, BA04	Sensitivity outcome	Sens
	Censoring variables				
8	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	
9	End of study period			Censoring	
10	End of GP practice coverage			Censoring	
11	Use of other SGLT-2 (since start of follow-up)			Censoring	
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			Sensitivity outcome /	D, Sens

	codes, etc)			descriptive	

* 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	Socio-economic 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	Prior ICU admission
16	Pancreatitis
17	BMI (when available)
18	Smoking (when available)
19	Alcohol use (when available)
20	HbA1c (when available)
21	Albuminuria testing
	Cardiovascular Diseases

22	Congestive heart failure
23	Hypertension
24	Stroke
25	Myocardial infarction
26	Autoimmune disease
27	COPD
28	Time since first diabetes diagnosis
	Other non-diabetes medications
29	Antihypertensives/diuretics
30	Non-steroidal anti-inflammatory drugs (NSAIDs)
31	Oral steroids
32	Statins, fibrates
33	Lipid modifying agents
34	Zoledronic acid
35	Antibiotics
36	Other drugs